

Vadim Ermolayev
Aleksander Spivakovsky
Mykola Nikitchenko
Athula Ginige
Heinrich C. Mayr
Dimitris Plexousakis
Grygoriy Zholtkevych
Oleksandr Burov
Vyacheslav Kharchenko
Vitaliy Kobets
(Eds.)



ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer

Proceedings of the 12th International Conference,
ICTERI 2016

Kyiv, Ukraine
June, 2016

Ermolayev, V., Spivakovsky, A., Nikitchenko, M., Ginige, A., Mayr, H. C., Plexousakis, D., Zholtkevych, G., Burov, O., Kharchenko, V., and Kobets, V. (Eds.): ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proc. 12th Int. Conf. ICTERI 2016, Kyiv, Ukraine, June 21-24, 2016, CEUR-WS.org, vol. 1614, online

Originally published online by CEUR Workshop Proceedings at CEUR-WS.org/Vol-1614.

This volume represents the proceedings of the 12th International Conference on ICT in Education, Research, and Industrial Applications, held in Kyiv, Ukraine, in June 2016. It comprises 62 contributed papers that were carefully peer reviewed (3-4 reviews per paper) and selected from 125 submissions. The volume opens with the abstracts of the keynote talks and invited tutorial. The rest of the collection is organized in 2 parts. Part I contains the contributions to the main ICTERI conference tracks, structured in four topical sections: (1) Advances in ICT Research; (2) Information Systems: Technology and Applications; (3) Academia/Industry ICT Cooperation; and (4) ICT in Education. Part II comprises the contributions of the three workshops co-located with ICTERI 2016, namely: 2nd International Workshop on Theory of Reliability and Markov Modeling for Information Technologies (TheRMIT 2016); 5th International Workshop on Information Technologies in Economic Research (ITER 2016); International Workshop on Professional Retraining and Life-Long Learning using ICT: Person-oriented Approach (3L-Person).

Copyright © 2016 for the individual papers by the papers' authors.
Copying permitted only for private and academic purposes. This volume is published and copyrighted by its editors.

Preface

It is our pleasure to present you the proceedings of ICTERI 2016, the twelfth edition of the International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications: Integration, Harmonization, and Knowledge Transfer, held at Kiev, Ukraine on June 21-24, 2016. This year's edition focused on research advances, business/academic applications of Information and Communication Technologies, design and deployment of ICT Infrastructures. Emphasis was also placed on the role of ICT in Education. These aspects of ICT research, development, technology transfer, and use in real world cases are vibrant for both the academic and industrial communities.

The ICTERI 2016 Call for Papers as well as the main conference itself were structured into four tracks reflecting these research fields.

The conference program was complemented by two invited keynote speeches, one invited tutorial, a Ph.D. mentors panel, industrial IT Talks sub-event, and the contributions to three co-located workshops:

- The 2nd International Workshop on Theory of Reliability and Markov Modeling (TheRMIT 2016) addressing long-standing research aspects of reliability modeling and assessment
- The 5th International Workshop on Information Technologies in Economic Research (ITER 2016) focused on cross-disciplinary issues in the use of Information Technology in economics and finance
- The International Workshop on Professional Retraining and Life-Long Learning using ICT: Person-oriented Approach (3L-Person 2016) presenting new uses of information technology for life-long learning

The rationale behind the Ph.D. Mentors Panel was to provide the opportunity to Ph.D. candidates to listen to and discuss the visionary ideas and promising topics for Ph.D. research offered by several renowned experts.

The tutorial by Prof. Oscar Corcho from the Universidad Politécnica de Madrid discussed the use of agreed vocabularies and appropriate data structures for supporting the sharing of open data in a smart city environment. The first keynote was delivered by Drs. John Davies and Sandra Stinčić Clarke from British Telecom on technical challenges for interoperability in the realm of the Internet of Things. In the second keynote Prof. Dr. Wolfgang Reisig of Humboldt University addressed the role of service orientation as a programming paradigm.

Overall ICTERI 2016 attracted a substantial number of submissions – a total of 122 comprising the main conference and workshops. Out of the 58 paper submissions

to the main conference we have accepted 27 high quality and most interesting papers to be presented at the main conference and published in our proceedings. The acceptance rate was therefore 46.5 percent. Our three workshops received overall 57 submissions, of which 35 were accepted by their organizers and also included in this volume.

The conference would not have been possible without the support of many people. First of all we would like to thank all the authors who submitted papers to ICTERI 2016 and thus demonstrated their interest in the research problems within our scope. We are very grateful to the members of our Program Committee for providing timely and thorough reviews and also for been cooperative in doing additional review work. We would like to thank the local organizers of the conference whose devotion and efficiency made this instance of ICTERI a very interesting and effective scientific forum. Finally a special acknowledgement is given to the support by our editorial assistant Victoria Kosa who invested a considerable effort in checking and proofing the final versions of our papers.

June, 2016

Vadim Ermolayev
Aleksander Spivakovskiy
Mykola Nikitchenko
Athula Ginige
Heinrich C. Mayr
Dimitris Plexousakis
Grygoriy Zholtkevych
Oleksandr Burov
Vyacheslav Kharchenko
Vitaliy Kobets

Committees

General Chair

Aleksander Spivakovsky, Verkhovna Rada of Ukraine, Kherson State University, Ukraine

Steering Committee

Vadim Ermolayev, Zaporizhzhya National University, Ukraine

Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria

Mykola Nikitchenko, Taras Shevchenko National University of Kyiv, Ukraine

Aleksander Spivakovsky, Verkhovna Rada of Ukraine, Kherson State University, Ukraine

Mikhail Zavileysky, DataArt, Russian Federation

Grygoriy Zholtkevych, V.N.Karazin Kharkiv National University, Ukraine

Program Chairs

Athula Ginige, Western Sydney University, Australia

Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria

Dimitris Plexousakis, Institute of Computer Science, FORTH, Greece

Workshop Chairs

Vadim Ermolayev, Zaporizhzhya National University, Ukraine

Grygoriy Zholtkevych, V.N.Karazin Kharkiv National University, Ukraine

Tutorial Chairs

Vadim Ermolayev, Zaporizhzhya National University, Ukraine

Grygoriy Zholtkevych, V.N.Karazin Kharkiv National University, Ukraine

IT Talks Chairs

Aleksander Spivakovsky, Verkhovna Rada of Ukraine, Kherson State University, Ukraine

Mikhail Zavileysky, DataArt, Russian Federation

Local Organization Chairs

Anatoly Anisimov (chair), Taras Shevchenko National University of Kyiv, Ukraine

Mykola Nikitchenko (chair), Taras Shevchenko National University of Kyiv, Ukraine

Volodymyr Shevchenko (vice-chair), Taras Shevchenko National University of Kyiv, Ukraine

Publicity Chair

Nataliya Kushnir, Kherson State University, Ukraine

Web Chair

Eugene Alferov, German Climate Computing Center, Germany

Program Committees

MAIN ICTERI 2016 Conference

Program Committee

Witold Abramowicz, Poznan University of Economics, Poland
Jan Aidemark, Linneaus University, Sweden
Eugene Alforov, German Climate Computing Center, Germany
Costin Badica, University of Craiova, Software Engineering Department, Romania
Nick Bassiliades, Aristotle University of Thessaloniki, Greece
Sotiris Batsakis, University of Huddersfield, UK
Antonis Bikakis, University College London, UK
Ruth Breu, Universität Innsbruck, Austria
Dmytro Buy, Taras Shevchenko National University of Kyiv, Ukraine
Lukas Chrpa, University of Huddersfield, UK
Michael Cochez, University of Jyväskylä, Finland
Jörg Desel, Fernuniversität in Hagen, Germany
Anatoliy Doroshenko, National University of Technology "Kyiv Polytechnic Institute", Ukraine
Vadim Ermolayev, Zaporizhzhya National University, Ukraine
David Esteban, TECHFORCE, Spain
Wolfgang Faber, University of Huddersfield, UK
Anna Fensel, Semantic Technology Institute (STI), University of Innsbruck, Austria
Giorgos Flouris, Institute of Computer Science, FORTH, Greece
Aditya Ghose, University of Wollongong, Australia
Brian Hainey, Glasgow Caledonian University, UK
Sungkook Han, Wonkwang University, The Republic of Korea
Ievgen Ivanov, Taras Shevchenko National University of Kyiv, Ukraine
Jan Jürjens, Fraunhofer Institute for Software and Systems Engineering, Germany
Dimitris Karagiannis, University of Vienna, Austria
Natalya Keberle, Zaporizhzhya National University, Ukraine
Vyacheslav Kharchenko, National Aerospace University "KhAI", Centre for National Aerospace University "KhAI", Centre for Safety Infrastructure-Oriented Research and Analysis, Ukraine
Vitaliy Kobets, Kherson State University, Ukraine
Oleksandr Kolgatin, Kharkiv National Pedagogical University named after G.S.Skovoroda, Ukraine
Christian Kop, Alpen-Adria-Universität Klagenfurt, Austria
Hennadiy Kravtsov, Kherson State University, Ukraine
Sergey Kryukov, Southern Federal University, Russian Federation
Vladimir Kukhareenko, National Technical University "Kharkiv Polytechnic Institute", Ukraine
Nataliya Kushnir, Kherson State University, Ukraine
Stephen W. Liddle, Brigham Young University, USA
Vira Liubchenko, Odessa National Polytechnic University, Ukraine
Mikhail Lvov, Kherson State University, Ukraine
Alexander Lyaletski, Taras Shevchenko National University of Kyiv, Ukraine
Hui Ma, Victoria University of Wellington, New Zealand
Frederic Mallet, Université de Nice-Sophia Antipolis, France
Wolf-Ekkehard Matzke, MINRES Technologies GmbH, Germany

Christine Natschläger, Software Competence Center Hagenberg GmbH, Austria
Mykola Nikitchenko, Taras Shevchenko National University of Kyiv, Ukraine
Yulia Nosenko, Institute of Information Technologies and Learning Tools of NAPS of Ukraine
Tope Omitola, University of Southampton, UK
Simon Parkinson, University of Huddersfield, UK
Vladimir Peschanenko, Kherson State University, Ukraine
Wolfgang Schreiner, RISC, Johannes Kepler University, Austria
Vladimir A. Shekhovtsov, Alpen-Adria-Universität Klagenfurt, Austria
Volodymyr G. Skobelev, Glushkov Institute of Cybernetic of NAS of Ukraine, Ukraine
Claudia Steinberger, Alpen-Adria-Universität Klagenfurt, Austria
Martin Strecker, IRIT, Paul Sabatier University, Toulouse, France
Julius Stuller, Institute of Computer Science, Academy of Sciences of the Czech Republic
Ilias Tachmazidis, University of Huddersfield, UK
Vagan Terziyan, University of Jyväskylä, Finland
Nikolay Tkachuk, National Technical University "Kharkiv Polytechnic Institute", Ukraine
Mauro Vallati, University of Huddersfield, UK
Leo Van Moergestel, Utrecht University of Applied Sciences, Netherlands
Paul Warren, Knowledge Media Institute, the Open University, UK
Maxim Vinnik, Kherson State University, Ukraine
Iryna Zaretska, V. N. Karazin Kharkiv National University, Ukraine
Grygoriy Zholtkevych, V. N. Karazin Kharkiv National University, Ukraine

TheRMIT 2016 Workshop

Workshop Chairs

Vyacheslav Kharchenko, National Aerospace University “KhAI”, Ukraine
Elena Zaitseva, Žilina University, Slovakia
Bogdan Volochiy, Lviv Polytechnic National University, Ukraine

Program Committee

Iosif Androulidakis, Ioannina University Network Operations Center, Greece
Nikolaos Bardis, Hellenic Army Academy, Greece
Mario Fusani, Institute of Information Science and Technologies, CNR, Italy
Anatoliy Gorbenko, National Aerospace University “KhAI”, Ukraine
Olexandr Gordieiev, University of Banking, National Bank of Ukraine
Yuriy Kondratenko, Black Sea State University named after Petro Mohyla, Ukraine
Vitaly Levashenko, Žilina University, Slovakia
Dmitriy Maevskiy, Odessa National Polytechnic University, Ukraine
Oleg Odarushchenko, Research and Production Company Radiy, Ukraine
Yuriy Ponochovnyj, Poltava National Technical National University, Ukraine
Inna Skarga-Bandurova, East Ukrainian National University, Ukraine
Vladimir Sklyar, Research and Production Company Radiy, Ukraine
Sergey Tyurin, Perm Research State Technical University, Russian Federation
Jüri Vain, Tallinn University of Technology, Estonia
Sergiy Vilkomir, East Carolina University, USA
Irina Yatskiv, Transport and Telecommunication Institute, Latvia
Vladimir Zaslavskiy, Taras Shevchenko National University of Kyiv, Ukraine

ITER 2016 Workshop

Workshop Chairs

Vitaliy Kobets, Kherson State University, Ukraine
Tetyana Payentko, National University of State Tax Service of Ukraine, Ukraine

Program Committee

Eugene Alforov, German Climate Computing Center, Germany
Tom Coupe, Kyiv School of Economics, Ukraine
Lina Dagiliene, School of Economics and Business, Lithuania

Fitim Deari, South East European University, Republic of Macedonia
Mohamed Fayad, San Jose State University, USA
Valentina Grigoryeva, Kherson State University, Ukraine
Dorota Jelonek, Czestochowa University of Technology, Poland
Vira Lyubchenko, Odessa National Polytechnic University, Ukraine
Dmitry Klyonov, Kherson State University, Ukraine
Ludmila Konstants, American University of Central Asia, Kyrgyz Republic
Olga Kotova, Kherson State University, Ukraine
Sergey Kryukov, Southern Federal University, Russian Federation
Sergei Makarenko, Kherson State University, Ukraine
Sergey Mazol, Academy of Public Administration, Minsk, Belarus
Carine Mesropyan, Southern Federal University, Russian Federation
Marin Neykov, University of National and World Economy (UNWE), Bulgaria
Nataliya Osipova, Kherson State University, Ukraine
Abdel-Badeeh Salem, Ain Shams University Cairo, Egypt
Viktor Selyutin, Southern Federal University, Russian Federation
Nina Solovyova, Kherson State University, Ukraine
Rasa Subačienė, Vilnius University, Lithuania
Elena Voronkova, State Fiscal Service National University, Ukraine
Alexander Weissbult, Kherson State University, Ukraine
Viktoriya Yatsenko, Kherson National Technical University, Ukraine
Rong Zhang, Nishinippon Institute of Technology, Japan
Grygoriy Zholtkevych, V.N. Karazin Kharkiv National University, Ukraine

3L-Person 2016 Workshop

Workshop Chairs

Oleksandr Burov, Institute of Information Technologies and Learning Tools, Ukraine
Svitlana Lytvynova, Institute of Information Technologies and Learning Tools, Ukraine

Program Committee

Valeriy Bykov, Institute of Information Technologies and Learning Tools, Ukraine
Natalia Derbentseva, Defence R&D, Canada
Olena Glazunova, National Agriculture University of Ukraine, Ukraine
Jose Orlando Gomes, Federal University of Rio de Janeiro, Brazil
Waldemar Karwowsky, University of Central Florida, USA
Tamara Koval, National Linguistic University, Ukraine, Ukraine
Hennadiy Kravtsov, Kherson State University, Ukraine
Volodymyr Kyseliov, Academy of Municipal Administration, Ukraine
Eugene Lavrov, Sumy State University, Ukraine
Mariya Shyshkina, Institute of Information Technologies and Learning Tools, Ukraine
Arvydas Martinkenas, Klaipeda University, Lithuania
Halyna Mygal, National Aerospace University named N.E.Zhukovskii, Ukraine.
Yuliya Nosenko, Institute of Information Technologies and Learning Tools, Ukraine
Olga Pinchuk, Institute of Information Technologies and Learning Tools, Ukraine
Eugene Popechitelev, St.-Petersburg State Electro-Technical University «LETI», Russian Federation
Oleg Spirin, Institute of Information Technologies and Learning Tools, Ukraine
Andrew Stryuk, Kryvyi Rih National University, Ukraine
Subhi Yamin Taisir, International Centre for Innovation in Education, Germany
Albert Zigler, University of Erlangen-Nuremberg, Germany

Additional Reviewers

Efstratios Kontopoulos, Aristotle University of Thessaloniki, Greece
Kalliopi Kravari, Aristotle University of Thessaloniki, Greece
Vitaly Levashenko, Žilina University, Slovakia
Olga Pinchuk, Institute of Information Technologies and Learning Tools, Ukraine
Nikolaos Tantouris, University of Vienna, Austria
Sergey Tyurin, Perm Research State Technical University, Russian Federation
Dimitris Vrakas, Aristotle University of Thessaloniki, Greece
Michael Walch, University of Vienna, Austria

Local Organizing Committee

Dmytro Buy, Taras Shevchenko National University of Kyiv, Ukraine
Ievgen Ivanov, Taras Shevchenko National University of Kyiv, Ukraine
Tatiana Rossada, Taras Shevchenko National University of Kyiv, Ukraine
Aleksey Senchenko, Taras Shevchenko National University of Kyiv, Ukraine
Mykola Ruban, Taras Shevchenko National University of Kyiv, Ukraine
Olena Shyshatska, Taras Shevchenko National University of Kyiv, Ukraine

ICTERI 2016 Sponsors

Platinum Sponsors



Oleksandr Spivakovsky's Educational Foundation (**OSEF**, <http://spivakovsky.fund/>) aims to support gifted young people, outstanding educators, and also those who wish to start up their own business. OSEF activity is focused on the support and further development of educational, scientific, cultural, social and intellectual spheres in the Kherson Region of Ukraine.



DataArt (<http://dataart.com/>) develops industry-defining applications, helping clients optimize time-to-market and minimize software development risks in mission-critical systems. Domain knowledge, offshore cost advantages, and efficiency – that's what makes DataArt a partner of choice for their global clients.

Silver Sponsors



Taras Shevchenko National University of Kyiv

(<http://www.knu.ua/en>) is a research university with a classic tradition and is a leading establishment of higher education in Ukraine. The university was founded in 1834 and has a long tradition of achievement and success both in teaching and in research. Presently, the university comprises 13 faculties and awards Junior Specialist's, Bachelor's, Specialist's and Master's degrees, Higher Qualification Post-graduate degrees and Doctoral degrees.



Logicify (<http://logicify.com/>) is an outsourcing company providing software development services. Company helps customers with issues and projects involving software. Logicify has been working in a variety of industries and fields, including telecom, video sharing, social media, insurance. It has several teams with specialized skills in different technologies that can relate to specific industries.

ICTERI 2016 Organizers



Ministry of Education and Science of Ukraine
<http://www.mon.gov.ua/>



Institute of Computer Science, FORTH, Greece
<https://www.ics.forth.gr/>



Western Sydney University, Australia
<http://www.westernsydney.edu.au/>



Alpen-Adria-Universität Klagenfurt, Austria
<http://www.uni-klu.ac.at/>



Kherson State University, Ukraine
<http://www.kspu.edu/>



Taras Shevchenko National University of Kyiv, Ukraine
<http://www.univ.kiev.ua/en/>



V.N. Karazin Kharkiv National University, Ukraine
<http://www.univer.kharkov.ua/en>



Zaporizhzhya National University, Ukraine
<http://www.znu.edu.ua/en/>



Institute of Information Technologies and Learning Tools
of the National Academy of Pedagogical Sciences
of Ukraine
<http://iitlt.gov.ua/en/>



DataArt Solutions Inc., Russian Federation
<http://dataart.com/>

Table of Contents

Invited Contributions

The Internet of Things – Technical Challenges for Interoperability	1
<i>John Davies and Sandra Stincic Clarke</i>	
Service Orientation as a Paradigm of Computing	2
<i>Wolfgang Reisig</i>	
Tutorial: Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing	3
<i>Oscar Corcho</i>	

ICTERI 2016: Track 1 - Advances in ICT Research

Complexity Class of Semantics-related Tasks of Text Processing	4
<i>Oleg Bisikalo and Ilona Bogach</i>	
A Complete Axiomatization for Reduced Clock Constraint Specification Language	13
<i>Bogdan Chornomaz, Kirill Rukkas and Kseniia Troino</i>	
Evidential Paradigm as Formal Knowledge Presentation and Processing . .	25
<i>Alexander Lyaletski, Alexandre Lyaletsky and Andrei Paskevich</i>	
A Human Communication Network Model	33
<i>Oksana Pichugina and Babak Farzad</i>	
A Model-Based Framework for Adaptive Resource Management in Mobile Augmented Reality System	41
<i>Mykola Tkachuk, Oleksii Vekshyn and Rustam Gamzayev</i>	
Descriptive Models of System Dynamics	57
<i>Grygoriy Zholtkevych, Konstantin Nosov, Yuri Bespalov, Larisa Rak, Elena Vysotskaya, Yulia Balkova and Vadim Kolomiychenko</i>	

ICTERI 2016: Track 2 - Information Systems: Technology and Applications

Information Support of the Computer-aided Fixture Design System	73
<i>Vitalii Ivanov, Svitlana Vashchenko and Yiming Rong</i>	
Information Technology of Predicting the Characteristics and Evaluating the Success of Software Projects Implementation	87
<i>Andriy Krasiy and Tetiana Hovorushchenko</i>	

Design and Development of Information System of Scientific Activity Indicators	103
<i>Aleksandr Spivakovsky, Maksym Vinnyk, Yulia Tarasich and Maksym Poltoratskiy</i>	
Adaptation of Legacy Fortran Applications to Cloud Computing	111
<i>Eugene Tulika, Anatoliy Doroshenko and Kostiantyn Zhreb</i>	
ICTERI 2016: Track 3 - Academia/Industry ICT Cooperation	
Application of the Method for Concurrent Programs Properties Proof to Real-World Industrial Software Systems	119
<i>Taras Panchenko</i>	
Safety-critical Certification of FPGA-based Platform against Requirements of U.S. Nuclear Regulatory Commission (NRC): Industrial Case Study	129
<i>Vladimir Sklyar</i>	
ICTERI 2016: Track 4 - ICT in Education	
Simulation Model for Computerized Testing of Learning Success in Quality Management Systems	137
<i>Alexander Alexeyev, Nataliya Konovalova, Kateryna Lozova and Elena Korol</i>	
Method of Evaluation of Electronic Educational Resource Quality for Conformity Assessment (Certification)	148
<i>Dmytro Fedasyuk, Nataliya Hots, Mykola Mykyychuk, Viktor Kuts and Yuliia Dzikovska</i>	
Hybrid Cloud-Oriented Educational Environment for Training Future IT Specialists	157
<i>Olena Glazunova and Tetyana Voloshyna</i>	
Information, Communication, and Modeling Technologies in Prosthetic Leg and Robotics Research at Cleveland State University	168
<i>Yuriy Kondratenko, Gholamreza Khademi, Vahid Azimi, Donald Ebeigbe, Mohamed Abdelhady, Seyed Abolfazl Fakoorian, Taylor Barto, Arash Roshanineshat, Igor Atamanyuk and Dan Simon</i>	
University Curricula Modification Based on Advancements in Information and Communication Technologies	184
<i>Yuriy Kondratenko, Dan Simon and Igor Atamanyuk</i>	
The Experience of the Master Classes as a Means of Formation of Readiness of Teachers to Implement Innovation	200
<i>Nataliya Kushnir, Nataliya Osipova, Nataliya Valko and Olena Litvinenko</i>	

Systems of Computer Mathematics for Educational Purposes as the Means of Reception of Procedural Knowledge	215
<i>Michael Lvov, Ludmila Shishko, Irina Chernenko and Evgen Kozlovskiy</i>	
Educator's e-Portfolio in the Modern University	231
<i>Nataliia Morze and Liliia Varchenko-Trotsenko</i>	
The Hybrid Cloud-based Service Model of Learning Resources Access and its Evaluation	241
<i>Mariya Shyshkina</i>	
Monitoring of Efficiency of Feedback Systems Use on the Base of Kherson State University	257
<i>Evgeniya Spivakovska, Maksym Vinnyk, Yulia Tarasich and Olesya Kuchma</i>	
Tips and pitfalls for blended learning: Redesigning a CS curriculum using IT	273
<i>Leo van Moergestel, Ander de Keijzer and Esther van der Stappen</i>	
Application of 3D Printing and 3D Scanning Technologies in Educational Activities	284
<i>Kseniia Yashyna, Oleksandr Krupnyk and Oksana Karpenko</i>	
 2nd International Workshop on Theory of Reliability and Markov Modeling for Information Technologies (TheRMIT)	
Calculation Methods of the Prognostication of the Computer Systems State under Different Level of Information Uncertainty	292
<i>Igor Atamanyuk, Yuriy Kondratenko and Vyacheslav Shebanin</i>	
Classification and Research of the Reactor Protection Instrumentation and Control System Functional Safety markov models in a Normal Operation Mode	308
<i>Yevgeniy Bulba, Yuriy Ponochozny, Vladimir Sklyar and Aleksandr Ivasiuk</i>	
Use of Natural LUT Redundancy to Improve Trustworthiness of FPGA Design	322
<i>Alex Drozd, Miroslav Drozd and Mykola Kuznietsov</i>	
Ontological Approach to the Assessment of Information Sufficiency for Software Quality Determination	332
<i>Tetiana Hovorushchenko and Oksana Pomorova</i>	
Semi-Markov Availability Models for an Infrastructure as a Service Cloud with Multiple Pools	349
<i>Oleg Ivanchenko and Vyacheslav Kharchenko</i>	

Application of Structure Function in System Reliability Analysis based on Uncertain Data	361
<i>Vitaly Levashenko, Elena Zaitseva and Miroslav Kvassay</i>	
Metamorphic Viruses Detection Technique Based on the the Modified Emulators.....	375
<i>Oksana Pomorova, Oleg Savenko, Sergii Lysenko and Andrii Nicheporuk</i>	
Mobile Post-Emergency Monitoring System for Nuclear Power Plants	384
<i>Anatoliy Sachenko, Volodymyr Kochan, Vyacheslav Kharchenko, Hubert Roth, Vasyl Yatskiv, Mykhaylo Chernyshov, Pavlo Bykovyy, Oleksiy Roshchupkin, Vasyl Koval and Herman Fesenko</i>	
Availability Model of Critical NPP I&C Systems Considering Software Reliability Indices	399
<i>Bogdan Volochiy, Vitaliy Yakovyna and Oleksandr Mulyak</i>	
Analysis of Methods for Providing Availability and Accessibility of Cloud Services.....	414
<i>Max Yanovsky, Olga Yanovskaya and Vyacheslav Kharchenko</i>	
Complexity-based Prediction of Faults Number for Software Modules Ranking Before Testing: Technique and Case Study	427
<i>Svitlana Yaremchuk and Vyacheslav Kharchenko</i>	
Importance Analysis of k-out-of-n Multi-State Systems based on Direct Partial Logic Derivatives	441
<i>Elena Zaitseva, Miroslav Kvassay, Vitaly Levashenko and Jozef Kostolny</i>	
5th International Workshop on Information Technologies in Economic Research (ITER)	
Forecasting Economic Indices of Agricultural Enterprises Based on Vector Polynomial Canonical Expansion of Random Sequences.....	458
<i>Igor P. Atamanyuk, Yuriy P. Kondratenko and Natalia N. Sirenko</i>	
Machine Learning Methods and "Real-Time" Economics	469
<i>Vladimir Dorozhinsky, Arsen Khadikov and Nina Zholtkevych</i>	
Government Financial Accountability and Transparency in the Digital World	477
<i>Lidiia Hladchenko</i>	
Using an Evolutionary Algorithm to Improve Investment Strategies for Industries in an Economic System.....	485
<i>Vitaliy Kobets and Maksym Poltoratskiy</i>	

Nonlinear Dynamic Model of a Microeconomic System with Different Reciprocity and Expectations Types of Firms: Stability and Bifurcations .	502
<i>Vitaliy Kobets and Alexander Weissblut</i>	
Adjusting Business Processes by the Means of an Autoregressive Model Using BPMN 2.0.....	518
<i>Vitaliy Kobets and Valeria Yatsenko</i>	
Intelligent Cyber Defense System	534
<i>Myroslav Komar, Anatoliy Sachenko, Sergei Bezobrazov and Vladimir Golovko</i>	
Fractal Analysis of Currency Market: Hurst Index as an Indicator of Abnormal Events	550
<i>Olena Liashenko and Tetyana Kravets</i>	
Using Hybrid Algorithms Based on GMDH-Type Neural Networks for Solving Economic Problems	558
<i>Olha Moroz</i>	
Modeling Input Financial Flows of Insurance Companies as a Component of Financial Strategy	566
<i>Tetiana Paientko and Tatsiana Verezubova</i>	
Simulation of the Impact of Social Media on Promoting Education Services	575
<i>Iryna Ushakova</i>	
A Company-Specific Business Process "Accounting of Cash on the Bank Accounts" in the Global Electronic Payment System	591
<i>Viktoria Yatsenko</i>	
International Workshop on Professional Retraining and Life-Long Learning using ICT: Person-oriented Approach (3L-Person)	
Life-Long Learning: Individual Abilities versus Environment and Means..	608
<i>Oleksandr Burov</i>	
The Use of Cloud Services for Learning Foreign Language (English)	620
<i>Hennadiy Kravtsov and Olga Gnedkova</i>	
Models and Tools for Information Support of Test Development Process in Learning Management Systems	632
<i>Olena Kuzminska and Mariia Mazorchuk</i>	
A Review of Teaching Methods for Life-Long Learning.....	640
<i>Vira Liubchenko</i>	
Professional Development of Teachers Using Cloud Services During Non-formal Education	648
<i>Svitlana Lytvynova and Oksana Melnyk</i>	

Collaboration between Research Institutions and University Sector Using Cloud-based Environment	656
<i>Yuliya Nosenko, Mariya Shyshkina and Vasyl Oleksiuk</i>	
Mobile Learning Technologies for Learning English	672
<i>Nataliya Osipova, Olga Gnedkova and Denis Ushakov</i>	
Perspective Analysis of the Use of Electronic Social Networks in a Learning Environment	680
<i>Olga Pinchuck</i>	
The Introduction of the Competence-based Approach in Educational Process of Training of Skippers	687
<i>Tatyana Zaytseva</i>	

The Internet of Things – Technical Challenges for Interoperability

Keynote Talk

John Davies ¹ and Sandra Stinčić Clarke ¹

¹ BT Group plc., United Kingdom
(john.nj.davies, sandra.stincic)@bt.com

Abstract. In this talk, we look at the various players in the Internet of Things ecosystem and explain the critical role of an information broker in an open ecosystem. We look at the challenge of data interoperability in the IoT context and describe the HyperCat standard, a specification for representing and exposing Internet of Things data catalogues to improve data discoverability and interoperability. The central idea is to enable distributed data repositories (data hubs) to be used jointly by applications through making it possible to query their catalogues in a uniform machine-readable format. This enables the creation of “knowledge graphs” of available datasets across multiple hubs that applications can exploit and query to identify and access the data they need, whatever the data hub in which they are held. This is achieved through employing the same principles on which linked data and the web are built: data accessible through standard web protocols and formats (HTTPS, JSON, REST); the identification of resources through URIs; and the establishment of common, shared semantics for the descriptors of datasets. We exemplify by way of several live demonstrations of IoT applications which use data from HyperCat-enabled data hubs.

Keywords. Internet of Things, Information Broker, Data Interoperability, Distributed Data Repository, Data Discovery, Data Cataloguing

Key Terms. Interoperability, Integration, ICT Infrastructure, Information Technology, KnowledgeRepresentation

Service Orientation as a Paradigm of Computing

Keynote Talk

Wolfgang Reisig¹

¹ Humboldt-Universität zu Berlin
Institut für Informatik
Unter den Linden 6 Berlin, 10099 Germany
reisig@informatik.hu-berlin.de

Abstract. Typical University curricula teach computable functions over symbol chains as the central topic of Theoretical Computer Science. This is motivated by the idea that the essentials of any kind of computation processes can be abstracted to computable functions. We challenge this idea, motivated by the observation that modern ICT embedding systems are reactive, communicating, and non-terminating. Neither do they compute functions, nor do they always process symbol chains. An obvious example are service-oriented architectures with their “always on” principle. An adequate, unique theory for such (and many other) architectures is missing; instead we witness a lot of different approaches and modeling techniques.

This talk surveys published proposals that attack the above problem. We try to discriminate fundamental notions and concepts such as refinement and composition, concurrency, and locality, as they prevail in different modeling techniques for service-oriented and other architectures. Those ideas lead to a fresh look at some aspects of education, research, and Industrial Applications of ICT.

Keywords. Theory of Informatics, paradigms of informatics, modeling techniques, Service-oriented Computing, associative composition

Key Terms. ICT Infrastructure, Webservice, Process

Building Agreed Vocabularies and Data Structures for Successful Open City Data Sharing

Invited Tutorial

Oscar Corcho ¹

¹ Facultad de Informática, Campus de Montegancedo,
Universidad Politécnica de Madrid
Boadilla Del Monte, 28660 Spain
ocorcho@fi.upm.es

Abstract. Existing EU directives and national laws on the reuse of Public Sector Information have motivated the creation of a large set of open data portals across Europe (following a global worldwide trend towards openness in the publication of open data). This is also the case in Spain, where the national government and national agencies, as well as regional governments and institutions, and local city councils, have started publishing actively open data in their corresponding portals. However, this wealth of data has been generated in a bottom up fashion, what means that the selection of the datasets to be published is done by those publishers independently, and that the formats and data structures in which such datasets are available are generally very heterogeneous. This generates problems for those companies and individuals who want to reuse such data across several institutions.

This tutorial will present the work that has been done in the context of one of the Spanish normalization working groups towards the definition of an open data maturity model for cities and for the selection of a set of ten datasets to be published by cities, and which is now being extended to handle many more datasets. The goal of this tutorial is to show how this process could also be applied in other countries and by other cities in different parts of the world.

Keywords. Smart city, open data, public sector information reuse, PSI reuse, vocabularies.

Key Terms. KnowledgeEngineeringProcess.

Complexity Class of Semantics-related Tasks of Text Processing

Oleg Bisikalo¹, Ilona Bogach²

Vinnitsia National Technical University, 95 Khmelnytske shose, Vinnitsia 21021, Ukraine

¹ obisikalo@gmail.com

² ilona.bogach@gmail.com

Abstract. Formal features of semantics-related tasks of text processing are reviewed; NP-complete procedural complexity of the class is substantiated. To diminish procedural complexity, the rationale behind applied formal linguistic knowledge is demonstrated, based on the analogy of the knapsack problem and automated text abstracting. Versatile approach for text processing is proposed, considering relations between entities; informational estimates are obtained and recommendations set forward.

Keywords. procedural complexity, NP-completeness, semantics-related tasks, text processing, informational estimate.

Key Terms. Text processing, computational linguistic, synthesis of natural language information.

1 Introduction

Some of the most complicated tasks in computational linguistics are those associated with semantics parsing and synthesis of natural language information. According to the authors, taking into account certain common formal features, such tasks are to be separately classified as semantics-related tasks of text processing, viz. text annotation and abstracting, searching key words, dialogue support etc. Sharper focus on this class and respective scientific research papers are driven by escalating demand for linguistic Internet technologies throughout the world.

Semantics-related tasks of text annotation and abstracting can save time for the experts, provided there is a proper quality of solutions. Summary is a coherent text, concisely depicting core topic as well as objectives, methods and findings of the research or insight, unlike annotation, which is a brief description of the content and general information on the topic. While the main purpose of annotation is to draw attention the text, the summary, containing just 10-23% of the text, allows users to arrive at conclusions as accurately as from the text, having spent twice as less time [1].

The main challenges of the above-mentioned tasks are caused by polysemy of natural language, the topical issue in computational linguistics. It is important to obtain the assessment of complexity of different approaches to solutions of semantics-related tasks, from direct enumeration to such heuristic methods, which enable people to understand the sense of new text information quickly. This will identify the rationale and efficiency of additional procedures of linguistic text analysis, singling out the so-called stop-words, attracting expertise etc. Thus, crucial task is to identify formal properties, including informational and general assessment of the complexity of the class of semantics-related tasks of text processing.

The objective of the research is to assess procedural class complexity of semantics-related tasks and determine efficient approaches to solutions.

2 Analysis of subject domains

To formally include semantics-related tasks in separate class, the general notion of class complexity should be considered. In the theory of algorithms complexity class is a set of computational tasks, approximately similar in terms of computing complexity. Otherwise, complexity class is a predicate set (function, having a word at the entry and coming back with a result 0 or 1), which is used for computing approximately similar number of resources [2].

There is a category of “the most complicated” for each category of tasks. It means that any task from the class goes down to such one, and the task belongs to the class. Such tasks are called complete for the class. NP-complete tasks are the most common.

Usually complexity class is determined by predicate sets with certain properties. Common determination of the class is as follows: complexity class X is called predicate set $P(x)$, computed by Turing machines, using resource computation $O(f(n))$, wherein n is the length of word x .

In most cases computation time is selected for the resource (number of tact in Turing machine) or operation area. Languages, which are identified by predicates from any class (i.e. sets of words, for which predicate turns 1), are also called those that belong to the same class.

Class P (Engl. *Polynomial*) is a set of tasks, providing relatively quick algorithms of solutions. Class P is included in broader classes of algorithms complexity.

Examples of P class are integral addition, division, matrix multiplication, determination of graph connectivity, ranging of sets from n numbers.

Non-deterministic polynomial task is a set of recognition problems, where solutions can be promptly checked at Turing machine, providing certain additional data solutions certificates. Equivalently *NP class* can be identified as a class, which contains tasks, admitting polynomial time of solution at non-deterministic Turing machine. There are examples of tasks, which are currently either classified or not classified as P , but belonging to NP :

- Tasks with Boolean formulae – find out with the *Boolean formula*, if there is a set of input variable, which turns 1. Certificate is such a set.
- Tasks on complete subgraphs – according to graph data, find out, whether it contains complete subgraphs of specified size. Certificate is a number of vertex, making complete subgraph.

– Find out the availability of Hamilton cycle in graph. Certificate is a sequence of vertex, making Hamilton cycle.

NP-complete problems are the most complicated among class *NP*. If anyone could cope with any of them for polynomial time, all tasks of *NP* class would be solved for polynomial time. Some examples of *NP*-complete problems are travelling- salesman task problem, Steiner problem, independent set problem, games Sapper, Tetris, Knapsack problems etc. For the time being, all those problems require exponential algorithms of solution.

To assess complexity of semantics-related tasks of text processing, proposed to be included, significant specific criteria of the results on understanding text information should be taken into account. Therefore, let's consider the issue of polysemy of the words in natural language from the formal view on word meanings. Thesauri usually provide all possible meanings of each word form with respective lexeme sign, which combines a certain set of words. The same spelling of words, belonging to different word forms, is a driver of escalation of scope of searching in the process of determination of proper meaning (polysemantic) of the word in each sentence of the text. Formally for r_i lexeme signs in i -sentence of the selected text the general scope

of search equals to all possible options of meanings $(k)^{r_i}$, with the only one correct according to the author (k – average polysemy coefficient of certain language).

Linguistic research substantiated the following hypothesis: the higher the level of analyticity in the language, the more frequently the same lexeme sign is used for different functions, and the larger is average polysemy coefficient. For example, Spanish language is more analytical than German, its polysemy coefficient makes the value of 6,9 of per lexeme, and for German – being less analytical language, polysemy coefficient is 5,6 per lexeme [3]. Average polysemy coefficient considerably varies for different parts of speech for most synthetic Slavic languages. For example, for nouns – 4,32 meanings per lexeme, for adjectives: 5– for specific and 3,5 for abstract ones; as for Russian language, average polysemy coefficient makes 3,1 meaning per lexeme [4]. Thus, it can be inferred that the lower limit of general scope of V search for the text is no less than

$$V \geq \sum_{i=1}^m 3^{r_i}, \quad (1)$$

where, m is a number of sentences in the text.

Apart from the degree of language analyticity, character and subject domain of the text can affect average polysemy coefficient. The latter is reduced by terminological steadiness of certain subject domain and austere (scientific) writing style, and increase by a number of adverbs, metaphors, elements of the so-called Aesopian language etc. Anyway, it is clear that problems of text understanding are formally referred to *NP*-complete complexity due to a step function (1). Moreover, it is not difficult for people to understand familiar language, including unknown text, which testifies for natural mechanisms of effective selection of the most proper combinations of meanings of all lexemes, contrary to complete search of all possible meanings.

We also considered common approaches for semantic analysis of text information, which differentiate the notions of lexical functions and semantic ones. In terms of semantics of the separate sentence linguists revealed 40÷60 (depending on the

language) of lexical functions, which mostly connect separate pairs of words or collocations. Accurate differentiation of all possible cases means the following: complexity by number of pairs at least from r_i to 2 with the coefficient of 40, i.e.

$$V' \geq 40 \cdot \sum_{i=1}^m \frac{r_i!}{2!(r_i - 2)!}.$$

The next step of formal adumbration of the sentence content is the notion of semantic relation (scheme), e.g. in [5] aggregation of 21 relations in 6 types, assigned by 9 triadic (quadruple)- predicates. Complexity of such approach is proportional to a number of allocations from r_i to 3 with the coefficient of 9, viz.

$$V'' \geq 9 \cdot \sum_{i=1}^m \frac{r_i!}{(r_i - 3)!}.$$

It should be noted that nearly all people have never heard of lexical functions and semantic relations, or have never thought of them, but it has not prevented them from understanding their language.

Thus, the rationale behind separating the class of semantics-related tasks is as follows: on the one hand, it is characterized by *NP*-complete complexity, wherein $V'' \geq V' \geq V$, and, on the other hand, there is an objective existence of natural algorithms of thinking that enable to solve the tasks of the class efficiently.

3 Automated abstracting as an example of semantics-related tasks of text processing

Preliminary analysis is the ground to ascertain that certain semantics-related problems are not only classified as *NP*-complete by procedural complexity, but are also similar to them by the formulation. Let us consider the afore-mentioned Knapsack problem as a proof, demonstrating convenient analogy for comparison and estimation of procedural complexity of automated text abstracting tasks [6]. Generally, tasks can be formed as follows: we need to select a certain number of objects from assigned set of objects with properties value and weight so that we obtain a maximal aggregate value along with the limit for the aggregate weight.

Without taking into account additional information by parameter analogues “value” and “weight” in Knapsack problems, it is obvious that there are parameters “importance” and “size” of fragments in automated abstracting tasks. Thus, in general case, abstracting is to result in a minimal scope of text, provided that it contains the most important phrases (sentences), whereat the text is supposed to keep the essence, and the last additional requirement makes the tasks of automated abstracting even more complicated. We assume that by the analogy described above, the task of automated abstracting is related to *NP*-complete problems.

As we know, classifying certain computation problems as *NP*-complete brings finding approximate algorithms [7] to focus of the scientists, since the unavailability of polynomial solutions makes the scientific paper futile. The problem of combinatory optimization of knapsack packing is a classical example of unsatisfactory time for solution by precise methods of full enumeration (for the sake of increasing necessary memory), dynamic programming or branches and limits. It shifts the focus to obtaining approximate results by greedy algorithm, genetic algorithms or other methods of discreet optimization. Unlike its analogy, approaches in solutions of

automated abstracting tasks have been historically construed as approximate methods [8], which considered additional linguistic information depending on the specifics of the task, e.g. TRM – Text Relation Map.

Classical *TRM* method takes into account weighted word vectors, corresponding to fragments (sentences) of the selected document, wherein graph is used as a formal model of semantic relations between structural units of text. Graph vertexes are text fragments, edges connecting the vertexes with a high level of approximation (semantic relation). Identifying key text fragments (vertex of the graph) for abstracting is based on criteria of a number of semantic relations of some fragments with others (ribs, coming out from vertex of the graph). It is proposed to combine TRM method with statistics methods TFIDF and *TLTF* in different options to additionally identify the weight of separate words of the document [9].

Estimation of procedural complexity of traditional *TRM method*. Wherein n is a number of words in the text, and m is a number of fragments (e.g., sentences). Generally thinking, we assume that there is an equal number of words in each sentence making $n' = n/m \approx r_i$. Then one operation of finding scalar outcome of two vectors with dimension n' (for 2-x sentences) requires computation

$$k_1 = 2^{\frac{n}{m}} - 1. \quad (2)$$

Since general number of fragments is m , the number of operation of scalar outcome of its vectors equals

$$k_2 = \sum_{j=1}^m j. \quad (3)$$

Sum of terms of arithmetic progression (3):

$$k_2 = \sum_{j=1}^m j = \frac{m(m+1)}{2}. \quad (4)$$

On the assumption of (2) and (4), general number of computation for identification of measures of semantic similarity of text fragments by TRM method equals to

$$K_2 = k_1 k_2 = (2^{\frac{n}{m}} - 1) \frac{m(m+1)}{2}. \quad (5)$$

Thus [10], limiting the estimate of procedural complexity by TRM method $O(nm)$ does not exceed the complexity of 2-classed polynom for the number of words in the text n and proves the efficiency of applying procedures of linguistic analysis. Though, we should admit that the best results of automated abstracting cede in authorship or expert options.

So, the following is effective for typical semantics-related tasks of automated abstracting: a) consideration of relevant linguistic properties and parameters of separate words, text or selection of texts; 6) identifying the most informative metrics for estimation of abstracting quality taking into account peculiarities of the text.

4 Informational estimate of the approach to text abstracting with relations between entities taken into account

Information flow analysis can be deemed as an alternative method for estimating complexity of semantics-related tasks. As opposed to procedural complexity, which is identified as a general estimate without specification, and considering peculiarities of the method of solution, informational estimate is procedure-oriented. Therefore, we propose to review informational estimate of the universal approach to text processing with relations between its entities (lexemes) taken into account.

Key feature of semantics-related tasks is deemed to be in determination and processing of content entities of the text. From informational view, understanding the sense of the sentence by a person is accompanied by recognizing separate words of the sentence and relations between pairs of the words with respective construction of the relations tree [11]. It should be recognized that in general using calculus of probability and in particular the notion of entropy in building NLP system goes back to the works of academician Markov regarding mathematical analysis of literary texts [12] and Claude Shannon regarding information value of English alphabetical symbols [13]. Though, such works focus on determination of the probability of correct string of symbols on the level of one word or several consecutive words.

Thus, famous work [14] covers maximum-likelihood approach for automatically constructing maximum entropy models and describes how to implement this approach efficiently, using as examples several problems in natural language processing. Partial results for constructing context-dependent models are obtained, viz. for segmentation of sentences and optimization of other parameters of machine translation. The following is proposed in the work [15]: multilayer neural network architecture that can handle a number of NLP tasks with both speed and accuracy by entropy-based criteria. Proposed are unified neural network architecture and learning algorithm that can be applied to various natural language processing tasks including: part-of-speech tagging, chunking, named entity recognition, and semantic role labeling. Obtained results allow automating the processes of useful markers to the text, though they do not take into account the level of general understanding of the text, which can be achieved by people.

The difference of our approach is that we consider all understanding processes to be carried out by comparative analysis and attracting information from the general linguistic base of knowledge of the subject. If each of those stages is accompanied by increasing of information, the following will be hypothetical in the universal approach:

- Ration of general understanding of text T can vary from minimal to maximal depending on the scope and other parameters of general linguistic base of individual's knowledge;
- Quality of determination of contents entities is proportional to the level of general understanding of text, to be confirmed by formal properties.

Informational estimate of this information is as follows:

1. Hereby we determine the entropy scope of one word in the text, for the case when appearance of the word is an independent and accidental event x with l of the following possible states

$$H(x) = - \sum_{j=1}^l p(x_j) \cdot \log p(x_j),$$

And maximal average estimate is made for the equally likely case

$$H_w = \log_2 l \text{ [Bit]}.$$

A number of different words (lexemes) of the text T can be considered variable l , and it is obvious that $l \leq n$.

2. Let's determine maximal estimate of entropy scope of all words of the sentence, provided that appearance of next word with n' words of this sentence does not depend on the previous one

$$H(x) = -n' \cdot \sum_{j=1}^l p(x_j) \cdot \log p(x_j),$$

or for equally likely case

$$H_{sw} = n' \cdot \log_2 l \text{ [Bit]}. \quad (6)$$

3. Let's determine entropy scope of paired association, provided that the words of the sentence in the form of a certain set $X = \{x_1, \dots, x_{n'}\}$ are known and recognized by the individual. For the pair to appear independently as an accidental event, potential number of pairs y can be $n' \times (n' - 1) = (n')^2 - n'$, where sentence with n' words makes parsing tree from n' pairs, taking into account bilateral relation of subject-predicate. On the other hand, key diagonal of such matrix is excluded, since the word in the sentence cannot be connected with itself. Thus

$$H(y) = - \sum_{i=1}^{(n')^2 - n'} p(y_i | X) \cdot \log p(y_i | X),$$

Accordingly, we get the following for equally likely cases

$$H_p \approx \log_2 (n')^2 = 2 \cdot \log_2 n' \text{ [Bit]}.$$

4. Entropy scope of all pairs of separate sentences can be determined by combinational properties of tree formation from n' pairs, which are selected from $n' \times (n' - 1)$ possible. In case of the most rigid condition of independent combination of words into n' pairs we have the following

$$H(y) = -n' \cdot \sum_{i=1}^{(n')^2 - n'} p(y_i | X) \cdot \log p(y_i | X),$$

And for equally likely case

$$H_{sp} \approx n' \cdot \log_2 (n')^2 = 2n' \cdot \log_2 n' \text{ [Bit]}. \quad (7)$$

As a result of increasing basic scope of word entropy into sentences (6) by additional entropy of its pairs (7) we get maximal general entropy of one sentence of the text

$$H_{sent} = H_{sw} + H_{sp} = n' \cdot (\log_2 l + 2 \cdot \log_2 n') \text{ [Bit]}. \quad (8)$$

Thus, application of proposed universal approach to processing the texts with m sentences increases general linguistic knowledge base of the individual by $m \cdot n' \cdot (\log_2 l + 2 \cdot \log_2 n') \text{ [Bit]}$.

Analysis of phrase (8) demonstrates that in case of inconsiderable fluctuation of a number of significant words n' in a sentence, l – a number of recognized word forms (lexemes) of the text- remains a key parameter of general linguistic knowledge of the individual. It can be also inferred that there is a common estimate range $O(m \cdot n' \cdot l)$ of procedural complexity of the universal approach, which is commensurate to procedural complexity of TRM method for typical semantics-related tasks of automated abstracting.

To diminish procedural complexity of the proposed approach, which appears to be promising for solving a number of semantics-related tasks [11], frequency characteristics of vocabulary stock of natural language should be considered. Since considerable word forms (lexemes) of the sentence carry the most comprehensive information, direct exclusion of the so-called stop-words in the process of text parsing and identification of co-references of the pronoun considerably decreases value l . Thus, according to [16, 17] for synthetic Russian language specific weight in the corpus of such parts of speech as parenthesis, pronouns (for nouns, adjectives and adverbs), prepositions, conjunctions, particles make 38,1%. The situation is different for analytical languages like English, viz. according to [18], specific weight of nouns, verbs, adjectives and adverbs from most frequently used words makes 96,4%. On the other hand, exclusion of loss- making relations for relatively small number of prepositions of English sentence enables to considerably decrease the value n' . All those processes are provided by modern parsers.

5 Conclusion

It has been substantiated that semantics-related tasks should be identified as separate class, characterized by NP -complete complexity along with algorithms of natural reasoning, which provide efficient solutions of tasks of this class. The analogy between knapsack problem and automated abstracting has demonstrated that using linguistic knowledge is traditionally included in text processing algorithms and enables to decrease procedural complexity to polynomial.

Based on the obtained informational estimate of universal approach in text processing with relations between entities taken into account, maximal general entropy of one sentence of the text has been determined. Since complexity of proposed method is polynomial, and technological possibilities of modern parsers provide respective procedures of linguistic analysis of text, processing of the latter with m sentences under that approach can practically increase general linguistic knowledge base of the individual by $m \cdot n' \cdot (\log_2 l + 2 \cdot \log_2 n')$ [Bit].

Promising direction for development of the research is to determine how acquired knowledge about relations between relevant entities of the text affect the accuracy of likelihood estimation $p(y_i | X)$.

References

1. U.Hahn, I.Mani: The Challenges of Automatic Summarization. IEEE Computer Society. Vol. 33, № 11, 29-36 (2000)
2. T.H.Cormen, C.E.Leiserson, R.L.Rivest, C.Stein: Introduction to Algorithms (2nd ed.), MIT Press and McGraw-Hill. Chapter 34: NP-Completeness, 966–1021 (2001)
3. Contrastive analysis of substantive polysemy : Based on German and Spanish languages, theme of the thesis on spec. 10.02.20, assistant professor of philology Tatiana Anatoliyevna Yakovleva, Majored in: comparative and historical, typological and comparative linguistic (2001)
4. L.V.Nikolayeva: The notion of polysemy in nominative systems of terms. Culture of the People K of Black Sea Region, № 110, T. 2, 65-67 (2007)
5. Major types of semantic relations between subject domain terms [E-resource] / Available: <http://cyberleninka.ru/article/n/osnovnye-tipy-semanticheskikh-otnosheniy-mezhdu-terminami-predmetnoy-oblasti>
6. K. Riedhammer, D. Gillick, B. Favre, and D. Hakkani-Tur: Packing the Meeting Summarization Knapsack. In Proc. Annual Conference of the Int'l Speech Communication Association (INTERSPEECH), 2434– 2437 (2007)
7. Identification of complexity classes [E-resource] / Available: <http://www.cs.berkeley.edu/~luca/cs172/karp.pdf>
8. N.B.Shakhovska: Automated text abstracting. National University Bulletin: Lviv polytechnics, № 743: Informational systems and networks, 210–218 (2012)
9. O.V.Kanisheva: Using maps of text ratios for automated abstracting. National University Bulletin Lviv Polytechnics, № 770: Informational systems and networks, 108–122 (2013)
10. O.V.Bisikalo: Automated annotation of texts based on linguistic images. Cybernetic management and informational, № 1, 46–51 (2014)
11. O.V.Bisikalo: Identification of pertinent parameters of text based on the relations between lexical units. National Technical University Bulletin «XIII», Series: Mechanical and technological systems and complexes, X: HTY «XIII», № 21 (1130), 83–89 (2015)
12. Andrei Andreyevich Markov. [E-resource] / Available: <http://www-gap.dcs.st-and.ac.uk/~history/Biographies/Markov.html>
13. C. E. Shannon. Prediction and entropy of printed english. Bell Systems Technical Journal, 30:50–64, 1951.
14. A.Berger, S.Della Pietra, V.Della Pietra: A Maximum Entropy Approach to Natural Language Processing [E-resource] / Available: <http://www.cs.columbia.edu/~jebara/6772/papers/maxent.pdf>
15. Ronan Collobert, Jason Weston, Léon Bottou - Natural Language Processing (almost) from Scratch [E-resource] / Available: <http://arxiv.org/pdf/1103.0398.pdf>
16. Russian National Corpus [E-resource] / Available: <http://www.ruscorpora.ru/corpora-stat.html>
17. New frequency dictionary of Russian language [E-resource] / Available: http://dict.ruslang.ru/freq.php?act=show&dic=freq_pos&title=%C4%E0%ED%ED%FB%E5%20%EE%20%F7%E0%F1%F2%EE%F2%ED%EE%F1%F2%E8%20%F7%E0%F1%F2%E5%F0%E5%F7%ED%FB%F5%20%EA%EB%E0%F1%F1%EE%E2%20%28%ED%E0%20%EC%E0%F2%E5%F0%E8%E0%EB%E5%20%EF%EE%E4%EA%EE%F0%EF%F3%F1%E0%20%F1%EE%20%F1%ED%FF%F2%EE%E9%20%E3%F0%E0%EC%EC%E0%F2%E8%F7%E5%F1%EA%EE%E9%20%EE%EC%EE%ED%E8%EC%E8%E5%E9%29
18. Adam Kilgariff. BNC database and word frequency lists [E-resource] / Available: <http://www.kilgariff.co.uk/bnc-readme.html>

A Complete Axiomatization for Reduced Clock Constraint Specification Language

Bogdan Chornomaz, Kirill Rukkas, and Kseniia Troino

Department of Computer Science,
Kharkiv National University

Abstract. Clock Constraint Specification Language, or CCSL, is a domain-specific language designed to model distributed real-time systems in terms of logical time, that is of sequences of events. Typical application of CCSL is to serve as a specification language for verification of specified systems.

In this paper we provide a sound and complete axiomatic for propositional logic over large fragment of CCSL which we call reduced CCSL, or RCCSL. This axiomatics appears to be rather simple, thus enabling effective verification of RCCSL specifications.

Keywords: time model, verification model, propositional logic, completeness model, specification language

Key terms: Computation, FormalMethod, SpecificationProcess, MathematicalModel, DistributedSystem

1 Introduction

Models dealing with discrete logical time rather than with real-valued “physical time” are well known to computer science, one classical example being Lamport’s algorithm for distributed clock synchronization [2]. In this paper we study model called Clock Constraint Specification Language (CCSL), proposed by F. Mallet in his dissertation [3]. Initially developed as UML profile for MARTE, CCSL later become domain-specific language on its own.

Constraints developed with CCSL allow some obvious logical reasoning, however the natural question arises: to which extent can this reasoning be carried out. Considerable efforts involving reasoning about CCSL constraints are put from the standpoint of formal verification of MARTE models with CCSL constraints. Usually this verification is carried out by transforming model into some framework which provides a model-checking ability, for example UPAAL [5] or Fiacre [1].

We take a different approach, having in mind a rather theoretical goal. We restrict ourselves with a fragment of CCSL called *Reduced CCSL* (RCCSL), for which we provide sound and complete system of axioms for propositional logic over it. The question of constructing such axiomatics for CCSL itself remains open.

As we see it, this result can be interesting in two ways. First, complete and sound axiomatic gives way to effective verification of constraint system. On the other hand, lots of effort are put into extending CCSL by endowing it with clock compositions or by introducing delays, see for example [4]. These attempts reveal a demand for deeper understanding of which elementary dependencies between clock can exist and which expressive power do they bring about. The language of logic which we utilize here can be exceptionally well suited for answering such questions.

The structure of the paper is as follows: In Section 2 we introduce basic terminology from CCSL and from logic background. In Section 3 we introduce axiomatics which, as we argue later, is a complete and sound axiomatics for modeling time systems with RCCSL. In Section 4 we take first preliminary steps towards the proof of completeness. Section 5 contains the most essential part of the paper, providing the central part of the proof of completeness. Section 6 concludes the paper.

2 CCSL and RCCSL

We start with a short introduction to CCSL terminology following [6]. We define *time structure* as a tuple $(\mathcal{I}, \leq, \mathcal{C}, \pi)$ where \mathcal{I} is at most countable set of *instants*, \mathcal{C} is a finite set of clocks, \leq is a preorder on \mathcal{I} such that $(x]$ is finite for all $x \in \mathcal{I}$, where $(x]$ is a principal preorder-ideal of x , that is

$$(x] = \{y \mid y \leq x\};$$

finally, π is a function $\pi: \mathcal{I} \mapsto \mathcal{C}$, mapping each instant into corresponding clock, such that for every $c \in \mathcal{C}$ each preimage $\mathcal{I}_c = \pi^{-1}(c)$ is linearly ordered and nonempty. We denote an equivalence relation corresponding to \leq by \doteq , that is, $x \doteq y$ if $x \leq y$ and $y \leq x$.

Thus defined, clock systems models a situation when we are given a set of clocks, each producing signals for which we only know their relative order of appearance. Signals in different clocks may be incomparable, or, on the contrary, may happen simultaneously. Proposition 1 in [6] can clarify this parallel.

Proposition 1 *For a time structure $(\mathcal{I}, \leq, \mathcal{C}, \pi)$, each \mathcal{I}_c is well-ordered with ordinal type at most ω , where ω is first finite ordinal.*

In order to visualize time system we use a modified version of Hasse diagram, see Figure 1 below. Vertical lines depicts instants of corresponding clocks, slanted lines between clocks depict covering relation of \leq , finally instants equivalent with respect to \leq are connected by horizontal lines.

For an instant i of \mathcal{I} we define *height* $h(i)$ of i as the length of the largest increasing chain in \mathcal{I} , ending in i minus one. This is a standard order-theoretic definition which can be reformulated in an inductive way as follows

- $h(i) = 0$, whenever i is minimal in \mathcal{I} ;

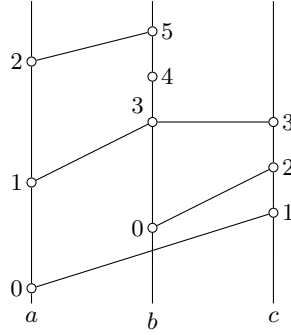


Fig. 1. Example of Hasse diagram depicting a time structure with clocks a , b and c .

$$- h(i) = \max \{h(j) \mid j < i\} + 1.$$

Heights of instants are also depicted on Figure 1 below

We say that time system is linear if \leq is a linear quasi-order, that is \leq becomes a linear order after factorization by corresponding equivalence relation. When depicting a linear time system, we will omit slanted and horizontal lines on Hasse diagram, the order between instants in this case is represented solely by their relative height.

We define *run* as a time system with the set \mathcal{I} defined as a custom subset of $\mathcal{C} \times \mathbb{N}$, with $\pi(a, k) = a$ and $(a, k) \leq (b, l)$ if and only if $k \leq l$, for all (a, k) and (b, l) in $\mathcal{C} \times \mathbb{N}$. In a run every set \mathcal{I}_c can be treated as a finite or infinite sequence of natural numbers. Trivially, every run is a linear time system. On the other hand, every time system can be considered a run, as stated in the Proposition 2 below, we refer to [6] for proof.

Proposition 2 For a time system $T = (\mathcal{I}, \leq, \mathcal{C}, \pi)$ define a run $L(I) = (\mathcal{I}', \mathcal{C})$ where \mathcal{I}' is defined as

$$\mathcal{I}' = \{(\pi(x), h(x)) \mid x \in \mathcal{I}\}.$$

Then $L(I)$ is a linear time system, and if I is linear, then $L(I) \cong I$.

Let us fix a potentially infinite set of clocks \mathcal{C} and a set \mathcal{S}^* of binary relation symbols

$$\mathcal{S}^* = \{\equiv, \prec, \preceq, \subseteq, \#\},$$

called *coincidence*, *precedence*, *cause*, *subclocking* and *exclusion* correspondingly. Now we introduce CCSL as a propositional language over a set \mathcal{T} of *terms*, where each term is defined as a triple xRy , x and y are clocks and R is a relational symbol from \mathcal{S}^* . Thus, $\mathcal{T} = \mathcal{C} \times \mathcal{S} \times \mathcal{C}$.

Thus, examples of terms are: $a \equiv b$, $a \# a$ or $b \prec d$; and CCSL formulas are: $a \equiv b$, $\neg(a \# b) \wedge a \prec c$, $\neg(\neg(a \# b) \wedge (b \preceq c \vee \neg(a \equiv b)))$.

Reduced CCSL, or RCCSL, is defined in a similar way, by excluding precedence from the set of possible relational symbols. That is, we fix

$$\mathcal{S} = \{\equiv, \preceq, \subseteq, \#\}.$$

In the example above, $\neg(a\#b) \wedge a \prec c$ is not an RCCSL formula, but $\neg(a\#b) \wedge a \preceq c$ is.

CCSL terms can be interpreted on time systems with the set of clocks \mathcal{C} as follows:

- $a \equiv b \Leftrightarrow$ for any $x \in \mathcal{I}_a$ there is $y \in \mathcal{I}_b$ with $x \doteq y$ and vice versa;
- $a \prec b \Leftrightarrow$ there is a strict extensive $h: \mathcal{C}_a \rightarrow \mathcal{C}_b$, that is, $x < h(x)$, for all $x \in \mathcal{C}_a$;
- $a \preceq b \Leftrightarrow$ there is an extensive $h: \mathcal{C}_a \rightarrow \mathcal{C}_b$, that is, $x \leq h(x)$, for all $x \in \mathcal{C}_a$;
- $a \subseteq b \Leftrightarrow$ for any $x \in \mathcal{I}_a$ there is $y \in \mathcal{I}_b$ with $x \doteq y$;
- $a \# b \Leftrightarrow x \neq y$ for all $x \in \mathcal{C}_a, y \in \mathcal{C}_b$.

Figure 2 illustrates the interpretation of CCSL terms on time systems.

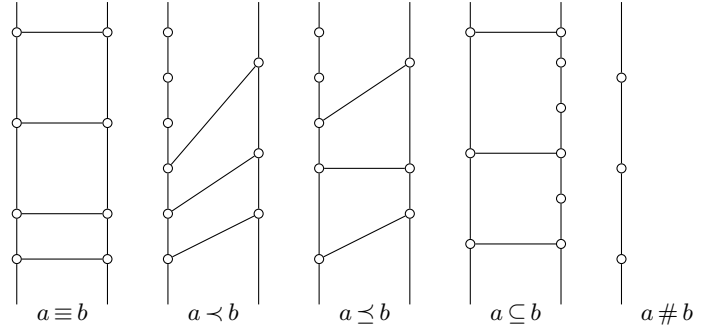


Fig. 2. Cause and subclocking relation on clocks a and b

After we interpret all CCSL terms, the interpretation of CCSL formulas on time systems is straightforward. For example, time system on Figure 1 satisfies formula $(a \prec b) \wedge (\neg(b\#c) \vee \neg(a\#c))$.

3 Axiomatics

Notice, that not all CCSL formulas, satisfiable as propositional formulas, can be satisfied on time system. For example, a formula $\neg(a \equiv a)$ is clearly satisfiable if we put $(a \equiv a) = \text{False}$. On the other hand, this formula can hold on no time structure. In fact, the following formulas, which we call axioms, hold on any time structure.

Axiomatics A1 (\mathbf{A}_0)

1. \equiv is an equivalence relation, which is congruent with respect to every other relation in \mathcal{S} , i.e.

$$\forall * \in \mathcal{S} \forall a, b, a', b' \in \mathcal{C}, a \equiv a', b \equiv b' : a * b \Leftrightarrow a' * b';$$

2. \preceq and \subseteq are quasiorders (i.e. reflexive and transitive) sharing associated equivalence relation \equiv ;
3. $a \subseteq b \Rightarrow b \preceq a$;
4. $\#$ is irreflexive and symmetrical;
5. $a \subseteq b, b \# c \Rightarrow a \# c$.

We say that a CCSL formula is *valid* if it holds under any interpretation on time structures. We say that an axiomatics is *sound* if all its axioms are valid formulas. Similarly, we call axiomatics *complete*, if any valid formula can be inferred from it. Throughout the paper, we consider all propositional axioms and propositional inference rules over CCSL terms to hold.

We denote Axiomatics A1 by \mathbf{A}_0 and refer to [6] for its soundness. In fact, in the following two sections we will show that this axiom set is also complete. Each of the axioms in \mathbf{A}_0 is not a singular propositional axiom, but rather a set of axioms, described in generally used terminology.

Define *relation structure* as a pair (\mathcal{C}, R) , where R is a subset in \mathcal{T} , which we treat as a valuation on a set of CCSL terms on \mathcal{C} . Usually we will write relation structure simply as R . For each relation symbol $*$ we define its corresponding relation in a relation structure:

$$*_R = \{(a, b) \mid a, b \in \mathcal{C}; (a, *, b) \in R\}$$

For a set of propositional formulas F over T we write $R \models F$ iff all formulas in F hold under truth assignment R , and say that R *comply* with F . Given a time structure T we define $R(T)$ as a valuation of terms given by their interpretation on T . We say that time structure T complies with F , denoted $T \models F$, if $R(T)$ does.

Using completeness of propositional logic we infer following general fact, which is essential for our proof of completeness

Proposition 3 \mathbf{A} is complete iff there is a model for a relation structure R whenever R complies with \mathbf{A} .

Proof. (\Rightarrow) : Let R comply with \mathbf{A} but does not have a model. Let F_R be a propositional formula which holds only for R . As there is no model for R , $\neg F_R$ is a valid formula and thus $\mathbf{A} \vdash \neg F_R$. By propositional inference rules this is equivalent to the formula $\neg \mathbf{A} \vee \neg F_R$ being propositionally valid, but it does not hold on R , a contradiction.

(\Leftarrow) : Let F be a valid formula not inferred from \mathbf{A} . Then there is a propositional structure R such that R complies with \mathbf{A} but not with F . By assumption, there is a time structure T with $R(T) = R$. But then F does not hold on T , which means F is not valid, a contradiction. ■

4 Completeness: preliminary reduction

Our first goal is to eliminate relations \equiv and $\#$. We say that relational structure R is *clarified* if \equiv is an equivalence relation. We say that time structure is clarified if its relational structure is.

For a relational structure (\mathcal{C}_0, R_0) we define its *factorization* as a relational structure (\mathcal{C}_e, R_e) , denoted $(\mathcal{C}_e, R_e) = (\mathcal{C}_0, R_0)/\equiv_0$, such that:

- \mathcal{C}_e is a set of equivalence classes of \mathcal{C}_0 by \equiv_0 ;
- $R_e = \{([a]_{\equiv_0}, *, [b]_{\equiv_0}) \mid a, b \in \mathcal{C}_0, * \in \mathcal{S}; (a, *, b) \in R_0\}$.

The fact that \equiv_0 is a congruence guarantees that the definition of R_e is consistent. Obviously, R_e is clarified for any R_0 . Let us now define simplified axiom system \mathbf{A}_e , which defines axiomatics for clarified time systems.

Axiomatics A2 (\mathbf{A}_e)

1. \equiv is an equity;
2. \preceq and \subseteq are partial orders;
- 3 - 5. same as in \mathbf{A}_0

To justify passing from \mathbf{A}_0 to \mathbf{A}_e let us prove the following two easy lemmas:

Lemma 1 *If a relational model R_A complies with \mathbf{A}_0 then R_A/\equiv_A complies with \mathbf{A}_e .*

Proof. Obvious. ■

Lemma 2 *Given a relational model R_A , if there is model for R_A/\equiv_A then there is model for R_A .*

Proof. Let \mathcal{C} be clocks of R_A/\equiv_A and let T' be a model for R_A/\equiv_A . Then clocks from \mathcal{C} are equivalence classes of clocks from \mathcal{C}_A . Define time system T with clocks \mathcal{C}_A such that $\mathcal{I}_a^T = \mathcal{I}_{[a]}^{T'}$. Now it is a trivial fact to check that T is a model for R_A . ■

As our next step we relax restrictions on $\#$ relation. From \mathbf{A}_e we can easily deduce that

$$\exists c: c \subseteq a, c \subseteq b \Rightarrow \neg a \# b.$$

Indeed

$$a \# b, c \subseteq a, c \subseteq b \Rightarrow c \# b, c \subseteq b \Rightarrow c \# c,$$

a contradiction.

It would be convenient for us if this implication would work the other way as well, i.e. if $\exists c: c \subseteq a, c \subseteq b \Leftrightarrow \neg a \# b$. However it is easy to construct a counterexample to this statement, on the other hand, it is always possible to "extend" the temporal structure by adding a clock (or several), so that it become true, see Figure 3 below:

We want to make the same trick, but with relation structures rather than with temporal structures.

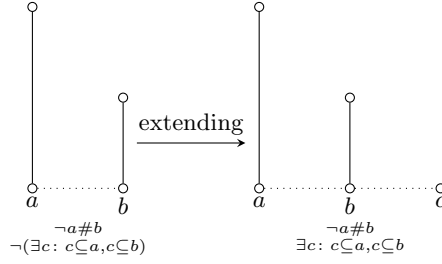


Fig. 3. Extension of a temporal structure

For a relation structure (\mathcal{C}_A, R_A) and a set of clocks $\mathcal{C}_B \subseteq \mathcal{C}_A$ by *restriction* of relation structure R_A to \mathcal{C}_B , denoted $R_A|_{\mathcal{C}_B}$, we understand a relation structure (\mathcal{C}_B, R_B) , where:

$$R_B = R_A \cap \mathcal{C}_B \times \mathcal{S} \times \mathcal{C}_B.$$

Next, by *extension* of a relation structure R_A we understand a relation structure R_B such that $R_A = R_B|_{\mathcal{C}_A}$. We say that relation structure is *subclock-closed*, iff for it holds

$$\exists c: c \subseteq a, c \subseteq b \Leftrightarrow \neg a \# b \quad (\mathfrak{S})$$

The time structure is *subclock-closed* iff its relation structure is subscklock-closed. Theorem 1 allows us to consider only subscklock-closed relation structures:

Theorem 1 *For each relation structure satisfying \mathbf{A}_e there is a subscklock-closed extension satisfying \mathbf{A}_e .*

Proof. Take a non subscklock-closed relation structure R_A satisfying \mathbf{A}_e . Let \triangleleft be some strict linear order on the set \mathcal{C}_A . Define a set R as:

$$R = \{c_{ab} \mid a, b \in \mathcal{C}_A, a \triangleleft b; \neg a \# b, \neg(\exists c: c \subseteq a, c \subseteq b)\}.$$

Define the set of clocks \mathcal{C}_B of our to-be-constructed system as:

$$\mathcal{C}_B = \mathcal{C}_A \cup R.$$

Now we need to define relations R_B in three cases: for pair of old clocks, for pair of new clocks and for a pair of an old and a new clock. In the first case we simply put $R_B|_{\mathcal{C}_A} = R_A$, which automatically assures that R_B is an extension of R_A .

In case of two elements from R we put:

$$\begin{aligned} c_{ab} \preceq c_{de} &\Leftrightarrow c_{ab} = c_{de}; \\ c_{ab} \subseteq c_{de} &\Leftrightarrow c_{ab} = c_{de}; \\ c_{ab} \# c_{de} &\Leftrightarrow c_{ab} \neq c_{de}. \end{aligned}$$

Finally, when elements are from different sets, put:

$$\forall a, b, c: c_{ab} \not\preceq d, d \not\subseteq c_{ab}$$

and

$$\begin{aligned} d \preceq c_{ab} &\Leftrightarrow d \preceq a \text{ or } d \preceq b; \\ c_{ab} \subseteq d &\Leftrightarrow a \subseteq d \text{ or } b \subseteq d; \\ c_{ab} \# d &\Leftrightarrow d \# c_{ab} \Leftrightarrow \neg c_{ab} \subseteq d \Leftrightarrow a \not\subseteq d \text{ and } b \not\subseteq d \end{aligned}$$

Generally, for a pair of clocks a, b from \mathcal{C}_A such that $\neg a \# b$, $\neg(\exists c \in \mathcal{C}_A: c \subseteq a, c \subseteq b)$ by $\overline{c_{ab}}$ we understand element c_{ab} in case when $a \triangleleft b$ and element c_{ba} in case when $b \triangleleft a$.

Observe, that R_B is subclock-closed, indeed:

$$\begin{aligned} a, b \in \mathcal{C}_A, \neg a \# b, \neg(\exists c \in \mathcal{C}_A: c \subseteq a, c \subseteq b) &\Rightarrow \exists c = \overline{c_{ab}} \in \mathcal{C}_B: c \subseteq a, c \subseteq b \\ a \in \mathcal{C}_A, c_{de} \in R, \neg a \# c_{de} &\Rightarrow c_{de} \subseteq a; \\ c_{ab}, c_{de} \in R, \neg c_{ab} \# c_{de} &\Rightarrow c_{ab} = c_{de}. \end{aligned}$$

So what is left to check is that R_B satisfy \mathbf{A}_e , let us do it.

1. \preceq and \subseteq are partial orders:

Check the transitivity of \preceq : if $f \preceq e \preceq a$, $a \neq e$, $e \neq f$ then, as all elements in R are not larger than any element of \mathcal{C}_A , we have two possibilities: either $a, e, f \in \mathcal{C}_A$, in which case the transitivity is trivial, or $a = c_{b,d} \in R$, $e, f \in \mathcal{C}_A$, but then:

$$e \preceq c_{b,d} \Leftrightarrow e \preceq b \text{ or } e \preceq d \Rightarrow f \preceq b \text{ or } f \preceq d \Leftrightarrow f \preceq c_{b,d}.$$

The reflexivity and the fact that associated equivalence relation is an equity are trivial. The proof for \subseteq is similar.

2. $a \subseteq b \Rightarrow b \preceq a$: obvious.
3. $\#$ is irreflexive and symmetrical: obvious.
4. $a \subseteq b, b \# c \Rightarrow a \# c$:

follows from the fact that R_B is subclock-closed and that \subseteq is a partial order, indeed let $\neg a \# c$ then $\exists d: d \subseteq a, d \subseteq c$. But then $d \subseteq a \subseteq b$ and so $\neg b \# c$, a contradiction.

■

Now, if we fix axiom system \mathbf{A}_F , which is a proper subset of \mathbf{A}_e ,

Axiomatics A3 (\mathbf{A}_F)

1. \preceq and \subseteq are partial orders;
2. $a \subseteq b \Rightarrow b \preceq a$;

then Theorem 1 together with Lemmas 1 and 2 yield the following corollary

Corollary 1 *If every subclock-closed relational structure compliant with \mathbf{A}_F can be realized by clarified subclock-closed time system, then every relational structure compliant with \mathbf{A}_0 can be realized by some time system.*

Proof. Let R be a relational structure compliant with \mathbf{A}_0 . Then by Lemma 1 $R_E = R/\equiv$ is compliant with \mathbf{A}_e . Take a subclock-closed extension R_F of R_E , which exists by Theorem 1. Now R_F satisfy \mathbf{A}_e and thus \mathbf{A}_F and by the hypothesis of the corollary there is subclock-closed time system T_F such that $R_F = R(T_F)$.

Notice that while $R(T_F)$ contains only interpretations of formulas with \preceq and \subseteq , by \mathfrak{S} it can be extended in a straightforward fashion to formulas with $\#$ and \equiv . Thus, T_F restricted to \mathcal{C}_F is a model for R_F . The claim of the corollary now follows by Lemma 2. ■

5 Completeness: modelling \subseteq and \preceq

Theorem 2 *Take a set of clocks \mathcal{C} and a pair of partial orders \preceq and \subseteq on it, such that $a \subseteq b \Rightarrow b \preceq a$. Then there is a subclock-closed time structure T over the same clocks such that $\subseteq_T = \subseteq$ and $\preceq_T = \preceq$.*

Proof. Let $n = |\mathcal{C}|$. Fix some linear order \triangleleft on \mathcal{C} and denote by c_i the i -th clock in \mathcal{C} relative to this order. Fix p linear orders $\pi_1 \dots \pi_n$ on \mathcal{C} so that

1. each π_i is an extension of \preceq ;
2. $\bigcap_{i=1 \dots p} \pi_i = \preceq$

Clearly, such orders can be found and p can always be chosen so that $p \leq n$.

Next, define function $f: \mathcal{C} \rightarrow \mathbb{N}^+$ as:

$$f(x) = \begin{cases} 1 & \forall y \neq x: y \not\preceq x \\ \sum_{y \preceq x, y \neq x} f(y) & \text{otherwise} \end{cases}$$

It is clear that, although this definition is "recursive", the recursion is only seeming: for bottom elements with regards to \preceq , i.e. for elements of height 1, the sum is empty, and so f equals 1; for elements of height 2 f is defined via elements of height 1, etc., see Figure 4 below.

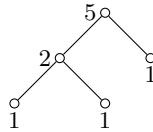


Fig. 4. Function f recursively defined for a partial order \preceq

Let $F = \sum_x f(x)$. Choose $l_1, \dots, l_p \in \mathbb{N}$ as:

$$\begin{aligned} l_1 &= 1; \\ l_i &= F * (l_1 + \dots + l_{i-1}) + 1. \end{aligned}$$

Or, using a direct formula:

$$l_i = (F + 1)^{i-1}.$$

Now, define \mathcal{I} to be the chain with $F * (l_1 + \dots + l_p)$ elements:

$$\mathcal{I} = \{(i, c, j) \mid i = 1 \dots p; c \in \mathcal{C}; j = 1 \dots l_i * f(c)\}$$

with order given by

$$(i, c, j) \leq (q, d, r) \Leftrightarrow \begin{cases} i < q \\ i = q, c >_{\pi_i} d \\ i = q, c = d, j \leq r \end{cases}$$

So this order is "almost" lexicographic, except that the second letter is each time ordered differently, depending of the first one.

Define clock c_T in T as:

$$c_T = \{(i, d, j) \in \mathcal{I} \mid d \subseteq c\}.$$

We claim that thus defined time structure T satisfies the requirements of the theorem.

From the definition of c_T it is obvious that T is subclock-closed, and that it satisfies $\subseteq_T = \subseteq$. The nontrivial part is to show that $\preceq_T = \preceq$, which we will do now by separately showing that $\preceq_T \subseteq \preceq$ and $\preceq \subseteq \preceq_T$.

1. $\preceq_T \subseteq \preceq$:

Let $a, b \in \mathcal{C}, a \preccurlyeq b$. Then for each i we have $a \leq_{\pi_i} b$. Define the function $h : a_T \rightarrow b_T$ as:

$$h(i, x, j) = (i, b, g_a(i, x, j))$$

where

$$g_a(i, x, j) = \left| \{(i, x', j') \in a_T \mid (i, x', j') \leq (i, x, j)\} \right|.$$

Observe, that to assure that h is correctly defined, we must check that $g_a(i, x, j) \leq l_i * f(b)$, but indeed:

$$\begin{aligned} g_a(i, x, j) &= \left| \{(i, x', j') \in a_T \mid (i, x', j') \leq (i, x, j)\} \right| \\ &\leq \left| \{(i, x', j') \in a_T\} \right| \\ &= \sum_{x' \subseteq a} f(x') * l_i \leq \sum_{x' \preccurlyeq a} f(x') * l_i \\ &\leq \sum_{x' \preccurlyeq b, x' \neq b} f(x') * l_i = l_i * f(b). \end{aligned}$$

So h is defined correctly, it is obviously strictly increasing and from $(i, x, j) \in a_T$ follows $x \preccurlyeq a \preccurlyeq b$, and $\forall w: (i, x, j) > (i, b, w)$ yields $h(i, x, j) \leq (i, x, j)$.

2. $\preceq \subseteq \preceq_T$:

Take $a, b \in \mathcal{C}$, $a \not\preceq b$, and take k so that

$$a \not\preceq_{\pi_k} b \Leftrightarrow b \leq_{\pi_k} a.$$

If $a \preceq_T b$ then there is an increasing function $h: a_T \rightarrow b_T$ such that $\forall w: h(w) \leq w$. Observe that $f(a) * l_k$ elements represented as (k, a, u) do not belong to b_T , from which we conclude:

$$\begin{aligned} f(a) * l_k &\leq \left| \{(i, x, j) \in b_T \mid (i, x, j) < (k, a, f(a) * l_k)\} \right| \\ &= \left| \{(i, x, j) \in b_T \mid (i, x, j) < (k, a, 1)\} \right| \\ &\leq \left| \{(k, x, j) \in b_T \mid (k, x, j) < (k, a, 1)\} \right| \\ &\quad + \left| \{(i, x, j) \in b_T \mid i \leq k-1\} \right| \\ &\leq \left| \{(k, x, j) \in b_T \mid a \leq_{\pi_k} x \leq_{\pi_k} b\} \right| + \left| \{(i, x, j) \in \mathcal{I} \mid i \leq k-1\} \right| \\ &= 0 + F * (l_1 + \dots + l_{k-1}) = l_k - 1, \end{aligned}$$

a contradiction.

■

Corollary 2 *Every subclock-closed relational structure compliant with \mathbf{A}_F can be realized by clarified subclock-closed time system*

Combining Corollaries 1, 2 and Proposition 3 we obtain

Theorem 3 *Axiom system \mathbf{A}_0 is complete and sound axiom system with time systems as its models.*

As a spin-off, let us notice that time system constructed in Theorem ?? is a run and is finite. These properties are also preserved by extension in Theorem 1 and by factorization in Lemma 1. Thus, we have the following propositions.

Statement 1 *Axiom system \mathbf{A}_0 is complete and sound axiom system with runs as its models.*

Statement 2 *Axiom system \mathbf{A}_0 has finite model property.*

6 Conclusion and future work

We had constructed sound and complete axiomatics for propositional logic over RCCSL, with completeness being the nontrivial part of this construction. We hope that the proposed model-theoretical approach would help to define canonical clock constraints that would be essential from theoretical perspective. As

an example, this paper shows that when arguing about clock constraints, coincidence could be easily removed from consideration, which is obvious. What is not so obvious is that arguing about exclusion could be replaced by arguing about subclocking.

Natural question that arises from this perspective is to extend our result to wider fragments of logic over CCSL, we formulate it as a series of problems.

Problem 1 *What is the complete and sound system of axioms for propositional logic over CCSL.*

Problem 2 *What is the complete and sound system of axioms for propositional logic over CCSL with clock compositions.*

Problem 3 *What is the complete and sound system of axioms for propositional logic over CCSL with delays.*

Our preliminary research shows that axiom system for complete CCSL might be much more complicated than the one for RCCSL. On the other hand, augmenting CCSL with composition might quite naturally fall into our approach of extending time systems.

References

1. Dhaussy, P., Menad, M.: A transformation approach for multiform time requirements. *Software Engineering and Formal Methods*, volume 8137 of *Lecture Notes in Computer Science*, 1630, (2013) *Commun. ACM*, 21(7):558565, 1978.
2. Lamport, L.: Time, clocks, and the ordering of events in a distributed system. *Commun. ACM*, 21(7):558565, 1978.
3. Mallet, F.: *Logical Time @ Work for the Modeling and Analysis of Embedded Systems*, Habilitation thesis. LAMBERT Academic Publishing (2011).
4. Mallet, F., Millo J.-V., Romenska, Y.: State-based representation of CCSL operators. *Research Report RR-8334*, INRIA (2013).
5. Mallet, F., Petterson, P., Seceleanu, C., Suryadevara, J.: Verifying MARTE/CCSL mode behaviors using UPPAAL. *Software Engineering and Formal Methods*, volume 8137 of *Lecture Notes in Computer Science*, 115, (2013)
6. Mallet, F., Zaretska, I., Zholtkevych, G., Zholtkevych, G.: Clocks Model for Specification and Analysis of Timing in Real-Time Embedded Systems. *ICTERI*:475-489 (2013)

Evidential Paradigm as Formal Knowledge Presentation and Processing

Alexander Lyaletski, Alexandre Lyaletsky, and Andrei Paskevich

Taras Shevchenko National University of Kyiv, Ukraine
Paris-Sud University, Orsay, France
forlav@mail.ru foraal@mail.ru andrei@lri.fr

Abstract. Investigations on the design and creation of high-performance information systems based on a number of paradigms supporting a human activity in formalized text processing were started in the beginning of 1960s – the time of the appearance of computers of such a high performance that programming of complex intelligent processes became possible. The so-called evidential paradigm is among them and it can be considered as a certain way of the integration of all reasonable paradigms intended for the development of languages for presenting formalized text in the form most appropriate for a user, formalization and evolutional development of a computer-made proof step, information environment having an influence on a current evidence of a computer-made proof step, and interactive man-assistant search of a proof. This paper contains a short description of the evidential paradigm and its implementation in the form of systems for presenting and processing formal knowledge.

Keywords. paradigm, Evidence Algorithm, formalized text, formal language, automated reasoning, automated theorem-proving, classical logic, intuitionistic logic, modal logic, number computation, symbolic transformation, deduction, induction, proof search, sequent calculus

Key Terms. MachineIntelligence

1 Introduction

The term “artificial intelligence” cannot be reduced only to the creation of devices or their components that completely or partially simulate the physical activity of humans; it also touches questions relating to the ability of a human to do reasoning in the framework of formalized languages. Investigations in this direction as well as in the direction of design and creation of high-performance intelligent information systems, now called automated reasoning systems and computer algebra systems started in the beginning of 1960s – the time of the appearance of computing machines of such a high speed, information capacity, and flexibility, that the programming of complex intelligent processes became possible. As a result, several different paradigms of the intelligent processing of computer formalized knowledge have appeared.

Efficient processing of formalized knowledge presupposes carrying out deep investigations in the fields of automated reasoning and construction of linguistic tools intended for supporting daily mental activity of a human, which presupposes an auspicious combination of theory and practice.

Attempts to find a reasonable balance between theory and practice have lead to creation and development of a number of certain paradigms reflected in intelligent systems such as automated theorem-proving and computer algebra systems.

In Ukraine, such investigations were initiated by Academician V.M.Glushkov in the end of 1960s, who announced the Evidence Algorithm (EA) programme in [1]¹ intended for making research on automated theorem proving and symbolic computations simultaneously in languages for presenting formalized texts in the form most appropriate for a user, formalization and evolutionary development of a computer-made proof step, information environment having an influence on a current evidence of a computer-made proof step, and interactive man-assistent search of a proof. In [4], the modern vision of the EA programme was called the *evidential paradigm*.

The implementation of Glushkov's approach in the framework of the evidential paradigm has found its partial reflection in the form of the so-called *SAD* system (System for Automated Deduction) intended for the presentation and processing of (formalized) mathematical texts in English and accessible at <http://nevidal.org/>. This paper is devoted to a short description of the peculiarities and features of the evidential paradigm and SAD.

2 Some remarks on formalized knowledge processing

Let us consider some of the issues, with which computer science is constantly dealing in solving problems associated with the formalized knowledge processing based on the following paradigms: numerical, analytical (symbolic), deductive, inductive and integrative.

The *numerical paradigm* reflects tools and methods for finding approximate or exact solutions of problems of pure or applied mathematics. It is based on constructing finite sequences of numerical calculations and actions over finite sets of numbers.

The *symbolic (analytical) paradigm* is based on the ability of computers to perform complex symbolic transformations, do numerical calculations, plot functions, construct mathematical models of certain processes, and so on. It is focused on the construction and use of computer algebra systems and appeared as an alternative to the numerical paradigm in the middle of 1960s, when computers with high enough performance were constructed.

The *deductive paradigm* reflects the declarative way of the representation and processing of computer knowledge. It is based on the fact that existing knowledge is represented in the form of certain formal texts or their pieces (usually axioms,

¹ A detailed enough description of EA and the investigations connected with it can be found in [2] and [3].

definitions, propositions, theorems, etc.) and that an additional knowledge is produced by means of applying certain inference (reasoning) rules for deducing new facts. The systems for the representation and processing of knowledge based on this paradigm are called automated reasoning systems, most of which exists in the form of automated theorem-proving systems, as the deductive approach is most relevant and efficient for tasks requiring reasonings “from general to particular” used in many application areas.

The *inductive paradigm* is based: in the natural sciences – on the transition from individual observations to general conclusions, while in mathematics – on the induction method used for the “checking” of a selected property for partially or fully ordered objects.

The *integration paradigm* “brings together” all of the above-mentioned paradigms. It can be divided into two types:

- *integration at the design phase*, when the ability of hierarchical building-in new components and subsystems into a designed system and connecting it to available computer services is provided during the design time of the system and
- *integration at the operation stage*, i.e. combining existing computer services into one (new) system (a special interest to the development of such services has been caused by a wide use of the Internet for the data exchange and transition of a necessary information, which, in its turn, led to the creation of relevant data exchange standards).

As to the integration at the operation stage, there can be mentioned Open Mechanized Reasoning Systems [5] and OpenMath [6] projects, in the framework of which certain specification and communication languages for solving tasks of theorem proving and symbolic computation were constructed.

The most famous representatives of the approach to the integration at the design stage are the QED [7] project as well as the Mizar [8] and Theorema [9] systems.

As for the Evidence Algorithm, it can be considered as the first representative of the integration paradigm at the design stage, which gave a reason for calling it by the evidential paradigm.

3 Peculiarities of evidential paradigm

The evidential paradigm is developed for mainly solving such problems of formalized text processing as proving a theorem under consideration or verifying a given proof of a theorem.

According to the evidential paradigm, the scheme of the processing of formalized (non necessary mathematical) texts is as follows (a more detailed description can be found in [2]):

Text prepared by a user for proving/verifying a theorem =====>
(using a parser)
A first-order self-contained text =====> *A computer-made proof/verification*
(using a prover)
=====> *Text in a form comprehensible for a human*
(using an editor)

Thus, the following problems arise in the framework of the evidential paradigm: creation of formal mathematical languages, construction of deductive methods and tools, creation and usage of information environment, and development of interactive modes.

Let us briefly describe each of them and some of the ways used for solving them. Naturally, the main attention will be paid to the linguistic and deductive tools satisfying the evidential paradigm requirements. In this connection, the other tools are only slightly contoured.

3.1 Linguistic tools

The following requirements were formulated for a language that should be constructed in the framework of the evidential paradigm.

Its syntax and semantics should be formalized. It should allow writing the axioms of a theory under consideration as well as auxiliary propositions, lemmas, theorems, and their proofs to ensure the self-containedness of a text and wording of definitions. The language thesaurus should be extensible and separated from the language grammar. In addition, it should be close to the languages of mathematical publications in order to provide convenience and comfort to a user in creating and processing a text in interactive mode. Besides, it should give the possibility to write formulas of the first-order language. Moreover, it should allow to formulate tasks that are not directly related to the deduction search.

The first sketch of such a language appeared in 1970 [10]. Its final (Russian) version called TL (Theory of Language) was published in [11].

In 2000, a new, improved, English-language version of TL called ForTheL (FORmal THEory Language) was constructed [12]. The main objective of the construction of ForTheL (and TL) was to provide an initial (mathematical) environment for solving a task under consideration as well as for the further development of evidential (logical) engines and strengthening of their capabilities.

Like any usual mathematical text, a ForTheL text consists of definitions, assumptions, affirmations, theorems, proofs, etc. The syntax of a ForTheL sentence follows the rules of the English grammar. Sentences are built of units: statements, predicates, notions (that denote classes of objects) and terms (that denote individual entities). Units are composed of syntactical primitives: nouns which form notions (e.g. “subset of”) or terms (“closure of”), verbs and adjectives which form predicates (“belongs to”, “compact”), symbolic primitives that use a concise symbolic notation for predicates and functions and allow to consider usual quantifier-free first-order formulas as ForTheL statements. Of course, just a little fragment of English is formalized in the syntax of ForTheL.

There are three kinds of sentences in the ForTheL language: assumptions, selections, and affirmations. Assumptions serve to declare variables or to provide some hypotheses for the subsequent text. For example, the following sentences are typical assumptions: “Let S be a finite set.”, “Assume that m is greater than n .”. Selections state the existence of representatives of notions and can be used to declare variables, too. Here follows an example of a selection: “Take an even prime number X .”. Finally, affirmations are simply statements: “If p

divides $n - p$ then p divides n ". The semantics of a sentence is determined by a series of transformations converting a ForTheL statement to the corresponding first-order formula for processing it by the deductive tools of SAD.

ForTheL sections are: sentences, sentences with proofs, cases, and top-level sections: axioms, definitions, signature extensions, lemmas, and theorems. A top-level section is a sequence of assumptions concluded by an affirmation. Proofs attached to affirmations and selections are simply sequences of low-level sections.

3.2 Deductive tools

From the very beginning of its appearance, EA has paid great attention to developing machine proof search methods suitable for the various fields of mathematics and taking into account (informal) human reasoning techniques.

Deductive tools should satisfy the following requirements: (1) the syntactical form of an original task under consideration should be preserved; (2) deduction should be carried out within the signature of an initial theory (i.e. without applying skolemization); (3) proof search should be goal-driven; (4) equality (equations) handling should be separated from a deductive processes.

As a result, the sequent approach was selected as basic for the construction of logical engines in the framework of the evidential paradigm. This permitted to satisfy (1) in a good enough form. Besides:

- for satisfying (2), a technique for the optimization of quantifier handling was proposed on the basis of an original notion of an admissible substitution,
- for satisfying (3), proof search was proposed in the form "driving" the process of an auxiliary goal generation by taking a formula (goal) under consideration into account,
- for satisfying (4), special methods for equality processing and equation solving were developed (allowing the use of algebra systems and problem solvers if necessary).

Investigations in this direction began in 1963, when the problem of automated theorem proving in group theory was formulated by V. Glushkov. As a result, the procedure admitted its interpretation as a specific, sound and complete, sequent calculus later called the AGS (Auxiliary Goals Search) calculus was proposed in [13]. It used Kanger's approach to quantifier handling and, in this connection, it was less efficient than the usual resolution-type methods.

Attempts to overcome this disadvantage of AGS led to the construction of an original sequent calculus [14, 15] on the basis of a new notion of an admissible substitution distinguished from that was used before its appearance. It was implemented in the (Russian) SAD system and its implementation demonstrated the usefulness of deductive approach to constructing such calculi.

Since then, these investigations were stopped until 1998, when the paper's authors started participating in the Intas project 96-0760 "Rewriting Techniques and Efficient Theorem Proving" (1998-2000). This project gave an impulse for modifying the calculus from [15] in several directions, one of which concerns the case of classical logic (see, for example, [16]) and the others – non-classical ones. As a result, a special research was conducted and there was obtained a wide

spectrum of interesting results on efficient enough inference search in classical and intuitionistic logics as well as in their modal extensions. Note that this research for classical logic was used in implementing the logical engine of the English SAD system.

Note that the (new) notion of admissibility is not enough for the construction of sound calculi in the intuitionistic case. This situation can be corrected by using the notion of compatibility firstly used in [17] for the construction of the sound (and complete) tableau calculus with free variables for intuitionistic logic and introduced in a number of calculi for modal extensions of classical and intuitionistic logics [18].

3.3 Information environment and interactive modes

We distinguish *three types* of information environment. The first is a *proof environment* intended for keeping all the data necessary for solving a task under consideration. By now, it presents a certain self-contained ForTheL-text prepared with the help of a user. All mathematical facts accumulated during previous sessions and needed for solving a task form the second type of environment called *internal environment*. The whole information that can be received via Internet by means of using mathematical services existent in Internet give the third type, the so-called *external environment*.

Interactive modes can serve for interfacing a computer assistant (in our case – the SAD system) with both a user and computer mathematical services existent in the external information environment. They can be designed and implemented only after fixing the form of internal and proof environments.

4 SAD system

The earlier-mentioned English SAD system satisfies well enough the requirements of the evidential paradigm and now it can be considered as its three-level implementation that is intended for theorem proving and text verification [19].

At the *first level*, its parser module analyzes an input mathematical text, its structure, and its logical content, encoded in the ForTheL statements. After this, the translation of the text into its internal presentation is made.

The result of translation gives a list of goal statements to be deduced from their predecessors. Note that there exists a module for processing first-order formulas presenting in a “dialect” of ForTheL, which can also be used if convenient.

At the *second level*, the goal statements are generated one-by-one and subsequently processed by the so-called foreground reasoner of SAD for reducing them to a number of subtasks for a prover splitting a goal under consideration to several simpler subgoals or generating an alternative subgoal. The module becomes redundant when SAD solves a problem for automated theorem proving.

Proof search tasks are resolved by a background prover at the *third level*. The SAD native prover is based on a special goal-driven sequent calculus for the classical first-order logic with equality. Note that it exploits the above-mentioned

notion of admissible substitution, which permits to preserve the initial signature of a task so that special accumulated equations can be sent to a specialized solver, e.g. an external computer algebra system (CAS). Additionally note that the SAD system was implemented in such a way that it can use some other external first-order provers, such as Otter [20], SPASS [21], or Vampire [22].

As a final step, the SAD outputs the result of its session.

Finally, note that by now, a number of experiments has been made in the SAD. They are related to: inference search in first-order sequent logic (a user may give his own problems or refer to a problem from the known TPTP problem library), theorem proving in the context of a self-contained ForTheL-text, and verification of self-contained ForTheL-texts.

The most interesting examples on verification are:

- Ramsey's finite and infinite theorems;
- some properties of finite groups;
- theorem that the square root of a prime number is irrational;
- Cauchy-Bouniakowsky-Schwarz inequality from mathematical analysis;
- Chinese remainder theorem and Bezout's identity in terms of abstract rings;
- Tarski's fixed point theorem;
- theorem on the stability of the refinement relation over a set of program specification.

5 Conclusion

The above-given material shows that the evidential paradigm is well in line with the modern trends in the processing of formalized (not obligatory mathematical) texts and that the SAD system looks fruitful from the point of view of implementing the existing principles of the construction of computer mathematical services being defined as information systems that are able to carry out numerical calculations and/or symbolic transformation and/or deductive constructions. The paper's authors hope that the ideas presented in the paper and used in the SAD system, will attract the attention of researchers in the field of artificial intelligence and will be useful in developing computer services intended for formal knowledge presentation and processing.

References

1. V. M. Glushkov. Some problems in automata theory and artificial intelligence. *Cybernetics and System Analysis*, Vol. 6, No. 2, Springer, New York, 1970, P. 17-27.
2. A. Lyaletski, M. Morokhovets, and A. Paskevich. Kyiv School of Automated Theorem Proving: a Historical Chronicle. In book: *"Logic in Central and Eastern Europe: History, Science, and Discourse"*, University Press of America, USA, P. 431-469, 2012.
3. A. A. Letichevsky, A. V. Lyaletski, and M. K. Morokhovets. Glushkov's Evidence Algorithm. *Cybernetics and System Analysis*, Vol. 49, No. 4, Springer, New York, P. 3-16, 2013.

4. A. Lyaletski and M. Morokhovets. Evidential paradigm: a current state. *Programme of the International Conference "Mathematical Challenges of the 21st Century"*. University of California, Los Angeles, USA, P. 48, 2000.
5. A. Armando, P. Bertoli, A. Coglio, F. Giunchiglia, J. Meseguer, S. Ranise, C. Talcott. Open Mechanized Reasoning Systems: A preliminary report. *Proceedings of the Workshop on Integration of Deduction Systems (CADE 15)*, 1998.
6. OpenMath.org: <http://www.openmath.org/>
7. The QED Manifesto: <http://www.cs.ru.nl/freek/qed/qed.html>
8. Mizar Home Page: <http://mizar.uwb.edu.pl/>
9. Theorema: <https://www.risc.jku.at/research/theorema/>
10. V. M. Glushkov, Yu. V. Kapitonova, A. A. Letichevskii, K. P. Vershinin, and N. P. Malevanyi. Construction of a practical formal language for mathematical theories. *Cybernetics and System Analysis*, Vol. 8, No. 5, Springer, 1972, P. 730-739.
11. V. M. Glushkov, K. P. Vershinin, Yu. V. Kapitonova, A. A. Letichevsky, N. P. Malevanyi, and V. F. Kostyrko. On a formal language for representation of mathematical texts. *Automated Proof Search in Mathematics*, GIC, AS of UkrSSR, Kiev, 1974, P. 3-36 (in Russian).
12. K. Vershinin and A. Paskevich. ForTheL – the language of formal theories. *International Journal of Information Theories and Applications*, 7(3), 2000, P. 120-126.
13. F. V. Anufriyev. Analgorithm of theorem proving in logical calculi. *Theory of Automata*, GIC, AS of UkrSSR, Kiev, 1969, P. 3-26 (in Russian).
14. A. I. Degtyarev and A. V. Lyaletski. Deductive tools of the Evidence Algorithm. *Abstracts of the All-union Conference "Methods of Mathematical Logic in Artificial Intelligence and Systematic Programming"*, Vol. I, Palanga, Lithuania, P. 89-90, 1980 (in Russian).
15. A. I. Degtyarev and A. V. Lyaletski. Logical inference in the system for automated deduction, SAD. *Mathematical Foundations of Artificial Intelligence Systems*, GIC, AS of UkrSSR, Kiev, P. 3-11, 1981 (in Russian).
16. A. Degtyarev, A. Lyaletski, and M. Morokhovets. Evidence Algorithm and sequent logical inference search. *Lecture Notes in Computer Science*, 1705, Springer-Verlag, P. 44-61, 1999.
17. B. Konev and A. Lyaletski. Tableau proof search with admissible substitutions. *Proceedings of the International Workshop on First-order Theorem Proving*, Koblenz, Germany, P. 35-50, 2005.
18. A. Lyaletski. Mathematical Text Processing in EA-style: a Sequent Aspect. *Journal of Formalized Reasoning (Special Issue: Twenty Years of the QED Manifesto)*, vol. 9, No. 1, P. 235-264, 2016.
19. K. Verchinine, A. Lyaletski, and A. Paskevich. System for Automated Deduction (SAD): a tool for proof verification. *Lecture Notes in Computer Science*, 4603, Springer, P. 398-403, 2007.
20. Otter prover: <https://www.cs.unm.edu/mccune/otter/>
21. SPASS Homepage: <http://www.spass-prover.org/>
22. Vampire's Home Page: <http://www.vprover.org/>

A Human Communication Network Model

Oksana Pichugina¹, Babak Farzad²

¹ Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
pichugina.os@mail.ru

² Brock University, St. Catharines, Canada
bfarzad@brocku.ca

Abstract. A number of attributed formation models based on Erdos-Renyi and Barabasi-Albert random graph models are presented. One of them is a Human Communication Network (HCN) model based on time restrictions on face-to-face communication. Construction of this weighted network requires a few numerical parameters and allows to transform any unweighted node attributed network into weighted. This transformation helps solving numerous problems in Network Analysis such as community detection, network topology inference, etc. Understanding nature of human communication networks allows to solve many practical problems starting with fast spreading any information and innovation through the networks and ending with detecting key people, collaboration with whom helps achieving different goals.

Keywords. Social Networks, Community Detection, Attributed Networks, Node Partition, Random Graphs

Key Terms. Network Decoration, Community Detection, Random Graphs

1 Introduction

Network Analysis is an area of research that has been studied intensively lately. Researches investigate structural characteristics of different networks, network formation models, and many other related questions. Among a variety of networks, social networks, which reflect a diversity of people relationships, are a priority [5], [7]. Study of social networks is important since it helps understanding how our world is organized, what place each of us takes in it, how this situation affects us and how the knowledge can be used to achieve our goals. Social networks are characterized by heterogeneity of nodes and edges, sparsity, high average clustering coefficient, small average shortest path length and power-law degree distribution, existing observable and tightly bound groups of elements called communities [5]. Most of these properties are united in "small-world networks" and "scale-free networks" concepts [1],[4]. Many attempts have been made to construct social networks, but still no satisfactory solution to simulate all the listed properties is found [1],[4],[5]. It is also important to find efficient ways of these community detection (CD) [6],[8]. We believe that the key

in qualitative CD in social networks is in using the heterogeneity (making them multi-layer ones) and study the issue of such networks formation.

2 Definitions and Notations

Definition 1. [6] *A social network is a hybrid graph, which is represented in the form:*

$$G = (V, E, \Lambda, \Lambda'), \quad (1)$$

where V is the set of nodes (the social network's users), E is the set of edges (these users relationships), Λ and Λ' contain an information about attributes related to each node $v \in V$ and each edge $\{u, v\} \in E$, respectively.

The network represented in the form (1) is an *attributed network* if $\Lambda \cup \Lambda' \neq \emptyset$. So, any attributed network representing individuals' relationships is social.

Let $\Lambda \in \mathbb{R}^{n \times K}$, $\Lambda' \in \mathbb{R}^{m \times K'}$ hence V, E are of size $|V| = n$, $|E| = m$. Initially, we consider an unweighted *node-attributed network* $G = \{V, E, \Lambda\}$, $K \geq 1$. Then we assign weights to its edges (*decorate* the edges) and come to consideration of a weighted network $G^w = \{V, E, \Lambda, \Lambda'\}$ with Λ' being a matrix-column of edge weights ($K' = 1$). G^w is a *node-edge-attributed* and used then for CD.

Introduce some notations: $G^{[\cdot]} = \{V, E^{[\cdot]}, \Lambda\}$ - is an unweighted node-attributed network with an adjacency matrix $B^{[\cdot]} = (b_{ij}^{[\cdot]}) \in \mathbb{R}^{n \times n}$. After decoration $E^{[\cdot]}$ by weights, the new weighted network is denoted by $G^{w[\cdot]} = \{V, E^{[\cdot]}, \Lambda, \Lambda'\}$ and its weighted adjacency matrix (WAM) - by $A^{[\cdot]} = (a_{ij}^{[\cdot]}) \in \mathbb{R}^{n \times n}$.

The *node degree* $d_i^{[\cdot]}$ of a node $v_i \in V$ in $G^{[\cdot]}$ is the number of its incident edges: $d_i^{[\cdot]} = |N_i^{[\cdot]}|$ where $N_i^{[\cdot]} = \{u \in V : u \leftrightarrow v_i, \{u, v_i\} \in E^{[\cdot]}\}$. The *node strength* $s_i^{[\cdot]}$ of $v_i \in V$ is a sum of weights of its incident edges in $G^{[\cdot]}$. In terms of adjacency and weighted adjacency matrices, these values are: $d_i^{[\cdot]} = \sum_j b_{ij}^{[\cdot]}$, $s_i^{[\cdot]} = \sum_j a_{ij}^{[\cdot]}$.

A *network cover* is a division of the network nodes $\mathcal{C} = \{C_l\}$ satisfying $\bigcup_{l=1}^L C_l = V$. If in the division the node clusters $C_l, l \in J_L = \{1, \dots, L\}$, are pairwise disjoint, then it is called a *network partition*.

Assume that the nodes are decorated by K discrete attributes $\{AT^k\}$, $\Lambda = (at_i^k)$ where at_i^k is the value of AT^k for a node v_i , and there are L_k different values of AT^k . Let $AC_l^k \in V$ be a set of nodes with l -th value of AT^k . We call it a *node attribute cluster* (AC) and denote a G -partition into ACs related to different values of AT^k by $\mathcal{AC}^k = \{AC_l^k\}_{l \in J_{L_k}} (n_l^k = |AC_l^k|)$. Let $G^{[\cdot]k}$ be a $G^{[\cdot]}$ -subnetwork related to AT^k . A *sum of unweighted networks* $\{G^k\}_k$ of the same node set V is an unweighted network $G = \{V, \bigcup_k E^k, \Lambda\}$. A *linear combination of weighted networks* $\{G^{wk}\}_k$ of a node set V is a weighted network $G^w = \{V, \bigcup_k E^k, \Lambda, \Lambda'\}$ with a WAM $A = \sum_k \alpha_k A^k$ where $\{\alpha_k\} \subset \mathbb{R}$ are coefficients of this linear combination. Let $\omega(G^{[\cdot]})$ be denoted a *weight* of a network $G^{[\cdot]}$ ($\omega(G^{[\cdot]}) = \sum_{ij} a_{ij}^{[\cdot]}$). A network of a weight one is a *normalized network*.

The networks' linear combination is a *weighted network sum* if

$$G^w = \sum_k W^k \cdot G^{wk} \text{ where } \{W^k\} > \mathbf{0}, \sum_k W^k = 1. \quad (2)$$

3 Motivation

Let us consider a social network. Suppose that, in addition to basic information about the node and edge sets, there is available some extra information about the nodes and edges features (social semantic networks are highly helpful here [2]). These additional characteristics are called attributes and the procedure of their complementing is decoration of the network [3], [5] resulted in creation of an attributed network [8]. Applying CD on the network we, typically, get communities closely related to one node attribute and this dominant attribute does not allow us to observe communities in other layers related to the rest, less important, node attributes. For instance, in the humankind network the dominant attribute would be belongingness to families. If we are interested in study communities, say, in work place, then the family division is an obstacle on this way. However, if it is possible to transform the network into weighted, moreover, to assign edge weights to each layer subnetworks of the multi-layer network, the problem of multi-layer CD (MLCD) can be solved. For that we just detect the dominant attribute and extract the corresponding subnetwork from consideration repeating then the procedure on the remaining network.

The crucial part of the approach is constructing edge sets of the one-layer subnetworks and distributing weights within them. The first one is a problem of the attributed network formation considered in Sect.4.1 (the edge inference problem [5]), the second one is the edge attribute inference problem [5]. The last one we solve for a social network of people face-to-face communication in Sec.4.2.

4 Human Communication Network Models

At this section we touch formation of attributed networks. We are wondering how an attributed network (1) is formed if the information about nodes V and their attributes A is known. In other words, we review formation of an edge set E and its attributes A' and refer to them as Problems 1 and 2, respectively.

4.1 Attributed Network Formation

We consider a number of ways to solve Problem 1. For convenience, we interpret the presented network formation models in terms of communication of people spending a time together during common activities/interests (AIs). Here nodes are people and their AIs are the nodes' attributes.

Model 1 - an association network model. An association network G^a [5] is an example of an attributed network where links exist between any nodes with common attributes. It can be interpreted as a network of virtual contacts of people with common interests where supporting such contacts does not need anything.

The auxiliary network G^{wk} corresponds to each activity/interest (AI) AT^k ; G^w is representable as a weighted network sum (2) of K networks, which are collections of complete graphs: $G^{wk} = \bigcup_{l \in J_{L_k}} K_{n_l^k}$. Thus the network G^a is a cover of

K overlapping V -partitions by a disjoint union of complete graphs.

Model 2 - an attributed networks model based on Erdos-Renyi Model.

Suppose that for existing an edge a similarity of node attributes is necessary, but not enough because of randomness. Similar to Model 1, we represent the network G^w by (2). Edges in G^{wk} are created randomly with probability p_l^k between two nodes v_i, v_j sharing the l -th value of the attribute AT^k . Hence G^k is a node partition by Erdos-Renyi Random Graphs (ERRGs) [4]: $G^{wk} = \bigcup_{l \in J_{L_k}} ERRG(p_l^k, n_l^k)$

and the resulting network G^w is an overlapping of K partitions by ERRGs.

In terms of human communication, Model 2 simulates a real situation where a group of people is formed simultaneously. Contacts of each user occur randomly without analysing any prior information due to its inaccessibility. The communication can be established on a regular basis only if these people actually have common interests. Different type of contacts are formed independently.

Model 3 - an attributed networks model based on Barabasi-Albert Model.

In comparison with Model 2, here we review a situation where a group of people is formed gradually. First of all, group members aspire to contacts with popular and authoritative colleagues in each area of expertise. First, these contacts are formed for the most important AIs, then for the less significant. A chance to clarify common interests is higher if the contact already exists.

As before, G^w is a weighted network sum (2). $\{G^{wk}\}$ are formed consecutively by k in accordance with decreasing priorities of node attributes. For each k an edge set E^k is formed between nodes with the same value of AT^k consecutively by i with probabilities depending on degrees of all preceding nodes $\{d_{i'}^k\}_{i' < i}$ and parameters p^k, p'^k ($p^k \leq p'^k$) for new and previously established contacts.

There are many ways of a generalisation to attributed networks of Barabasi-Albert Preferential Attachment Model [1]. For instance, each auxiliary network G^k is formed as follows: disjoint subsets of nodes of different ACs are connected by preferential attachment and then the isolated subnetworks are connected forming the whole node partition \mathcal{AC}^k . These all partitions are united into a cover with respect to node attributes priorities, A , and pre-assigned order of the nodes arising. The network layers are dependent regardless we consider the case $p^k = p'^k, \forall k$ (Model 3.1) or another one ($\exists k : p^k < p'^k$). Model 3.1 simulates a node partition by Barabasi-Albert Graphs (BAGs). Each G^{wk} can be represented in a manner of Models 1, 2: $G^{wk} = \bigcup_{l \in J_{L_k}} BAG(n_l^k, \alpha_l^k)$ where α_l^k is the power of preferential attachment in AC_l^k .

4.2 The Human Communication Model

The models presented in Sect. 4.1 - Model 2 and Model 3 - are able to simulate networks of real, face-to-face contacts implying requirements to spend time for keeping in touch. Suppose an edge set E was formed according to Models 2 or 3. To finish the network G^w formation, Problem 2 has to be solved and the matrix A' be formed. Here we present a way to distribute edge weights according to assumptions typical, in our opinion, for real people interaction. We will refer to the obtained network as a Human Communication Network (HCN).

The HCN model assumptions. *Condition 1.* People AIs had already formed;
Condition 2. Connections between people are possible if they have common AIs;
Condition 3. Each person distributes uniformly the time t^k allotted for supporting a contact related to the AI AT^k between friends of this interest;
Condition 4. For everyone possibility of the communication is restricted by time T . If for a person the time is not enough for supporting his/her contacts, then the time allotted for supporting a contact related to the AT^k and $AT^{k'}$ is distributed proportionally to t^k and $t^{k'}$, respectively;
Condition 5. If two persons with the same interest are ready to devote time to each other, then, if necessary, they come to a compromise following certain rules.
 Formalise Conditions 1-5 in terms of the WAM A . We rewrite (2) in the form:

$$G^w = \sum_k G^{w'k}, \text{ where } G^{w'k} = W^k \cdot G^{wk}. \quad (3)$$

In addition to G^w satisfying Conditions 1-5, we introduce networks G^{w*} , $G^{w'*}$ satisfying Conditions 1-3 and 1-4, respectively.
 Similarly to the (3), G^{w*} and $G^{w'*}$ are representable as networks sums: $G^{w*} = \sum_k G^{*k}$, $G^{w'*} = \sum_k G'^{*k}$ where G^{*k} , G'^{*k} are subnetworks of G^{w*} and $G^{w'*}$ related to AT^k . Respectively, the following holds for the corresponding WAMs:

$$A = \sum_k A'^k, A^* = \sum_k A^{*k}, A'^* = \sum_k A'^{*k}. \quad (4)$$

Let a set of v_i, v_j common attribute values be found as follows: $\mathcal{E}_{ij} = \{k : at_i^k = at_j^k\} \subseteq J_K$. Then $N_i^{[.]^k} = \{v_j \in N_i^{[.]} : k \in \mathcal{E}_{ij}\}$ is a set of v_i -neighbours with the same AT^k -value as v_i in $G^{[.]}$. We expand the notations of the node degree and strength from the set $N_i^{[.]}$ into the sets $\{N_i^{[.]^k}\}_k$: a) $d_i^{[.]^k} = |N_i^{[.]^k}|$ is the node attribute AT^k -degree of $v_i \in V$ in $G^{[.]}$; b) $s_i^{[.]^k} = \sum_{v_j \in N_i^{[.]^k}} a_{ij}^{[.]^k}$ - is the AT^k -strength of v_i in $G^{[.]}$. Respectively, the node strength in $G^{[.]}$ is $s_i^{[.]} = \sum_k s_i^{[.]^k}$.

1. We start with assigning edge weights in G^{w*} :
 - (a) Condition 1 says that the network G^{w*} is decorated by discrete attributes AT^k and the matrix A is known;
 - (b) Condition 2 means that the links are formed only by similarity of the node's attributes, hence if $i, j : \mathcal{E}_{ij} = \emptyset \Rightarrow \{v_i, v_j\} \notin E$.
 - (c) Condition 3 allows to determine the ratio of A^{*k} -elements: if i, j, j', k, k' such that $at_i^k = at_j^k$, $at_i^{k'} = at_{j'}^{k'}$, then

$$\frac{at_{ij}^{*k}}{a_{ij'}^{*k'}} = \frac{t^k}{t^{k'}}. \quad (5)$$

Since there is no restrictions on the communication time in G^{w*} , it implies that all of the contacts are supported at the appropriate level. So, the weights in G^{w*} , G^{w**} can be assigned with respect to the maximal needed time t^k :

$$\forall i, j : k \in \mathcal{E}_{ij} \ a_{ij}^{*k} = t^k; \ a_{ij}^* = \sum_{k \in \mathcal{E}_{ij}} t^k. \quad (6)$$

Notice that A^* is symmetric thus G^{w*} is undirected. The communication time of each person depends on the number of the contacts of each type therefore the node strengths in G^{w*k} , G^{w*} are defined as follows: $s_i^{*k} = d_i^k t^k$, $s_i^* = \sum_{k,j} a_{ij}^{*k} = \sum_k s_i^{*k} = \sum_k d_i^k t^k$. In terms of the HCN model, the values s_i^* , s_i^{*k} can be interpreted as the time that a person i could devote for the communication overall and for the particular AI, correspondingly.

2. Moving on to the network $G^{w'*}$, we add Condition 4 - the time restriction - to the network G^{w*} . This condition determines how much time a person i is ready to spend for supporting each AI-contact depending on his/her priorities and the number of these contacts. It can be expressed as the restriction on node strengths by T -value: $\forall i s_i^{*k} \leq T$. If it holds, then the above restriction holds and $G^{w'*} = G^{w*}$, otherwise the weights a_{ij}^{*k} are scaled to meet the time restriction:

$$a_{ij}'^{*k} = \nu_i^* a_{ij}^{*k} \quad (7)$$

where the scaling parameter ν_i^* depends on the node i strength: $\nu_i^* = \min\left(1, \frac{T}{s_i^*}\right)$. Substitution (7) into (5) yields: $\frac{a_{ij}'^{*k}}{a_{ij}'^{*k'}} = \frac{t^k \cdot \nu_i^*}{t^{k'} \cdot \nu_i^*} = \frac{t^k}{t^{k'}}$. It means that each person distributes his/her own time independently from each other and guided common priorities \bar{W} accumulated in $\bar{t} = (t^k)$: $\bar{W} = \bar{t}/|t|$. Find the WAM of $G^{w'*}$ by (4),(7):

$$a_{ij}'^* = \sum_k a_{ij}'^{*k} = \nu_i^* \sum_k a_{ij}^{*k} = \nu_i^* \cdot a_{ij}^*. \quad (8)$$

The weights $a_{ij}'^*$, $a_{ji}'^*$ determine how much time a person i is ready to devote for communication with a person j and vice versa. It is clear that normally these are different values, $a_{ij}'^* \neq a_{ji}'^*$. Thus the network $G^{w'*}$ is directed that does not display a face-to-face communication.

3. To describe the real situation, we consider constructing the final network G^w from $G^{w'*}$. By adding Condition 5, the abstract directed network $G^{w'*}$ is transformed into the undirected G^w with weights equal to time that both persons - i and j - actually devote to each other. The weights are obtained as a result of a compromise between these persons who are ready to spend together not the same time.

Let persons i and j have a real contact ($\mathcal{E}_{ij} \neq \{\emptyset\}$) and are looking for a compromise ($a_{ij}'^* \neq a_{ji}'^*$). The result of their common decision can be expressed as function of these weights $a_{ij} = f(a_{ij}'^*, a_{ji}'^*)$. The function $f(\cdot)$ can be chosen in different way. For instance, we choose a simple averaging: $a_{ij} = \frac{1}{2}(a_{ij}'^* + a_{ji}'^*)$. Then, by (8) and due to a symmetry of A^* , we have:

$$a_{ij} = 0.5(\nu_i^* a_{ij}^* + \nu_j^* a_{ji}^*) = 0.5 \cdot a_{ij}^* (\nu_i^* + \nu_j^*). \quad (9)$$

Distribution of weights within $\{G^{*k}\}$ is obtained from (4), (6), (9): $a_{ij} = \sum_k a_{ij}'^k = \frac{\nu_i^* + \nu_j^*}{2} \sum_k a_{ij}^{*k} = \frac{\nu_i^* + \nu_j^*}{2} \sum_{k \in \mathcal{E}_{ij}} t^k = \frac{\nu_i^* + \nu_j^*}{2} \sum_k t^k b_{ij}^k$ wherefrom

$$a_{ij}'^k = 0.5(\nu_i^* + \nu_j^*) t^k b_{ij}^k, \quad a_{ij}^k = a_{ij}'^k / W^k = 0.5|\bar{t}|(\nu_i^* + \nu_j^*) b_{ij}^k. \quad (10)$$

4.3 Human Communication Network Simulation

Example 1 - Model 2 simulation. First, we demonstrate a solution of Problem 1 for Model 2 (see Sect. 4.1). Parameters of a simulated node-attributed network G are: the order $n = 60$, the number of node attributes $K = 3$, the nodes are divided randomly into $\{L_k\}_k = \{5, 4, 6\}$ attribute clusters of the same sizes: $(n_l^k) = (12^5, 15^4, 10^6)$. The result of the simulation with parameters $(p_l^k) = (0.3^5, 0.3^4, 0.5^6)$ is shown in Figure 1.

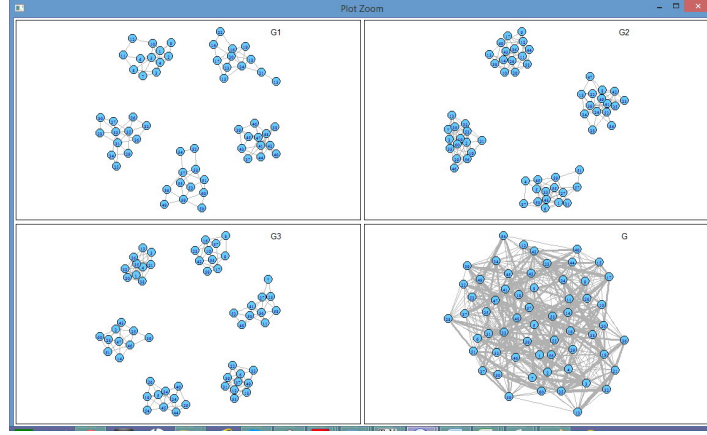


Fig. 1. Model 2 - the weighted network G^w and its subnetworks $G^1 - G^3$

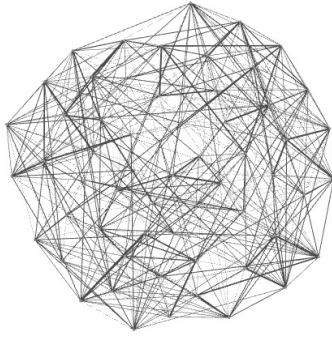


Fig. 2. The HCN G^{wI}

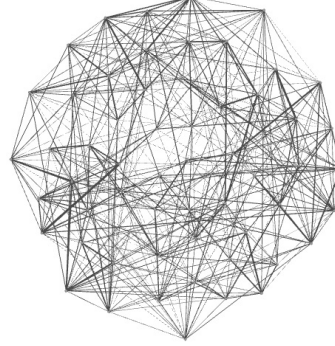


Fig. 3. The HCN G^{wII}

Example 2 - HCN Model 2 simulation. We took the unweighted network G from Example 1 and converted it into the HCN-Model 2 network (see Sect. 4.2) decorating edges by weights according to (9), (10). Two values of the time resource $\bar{T} = (T^I, T^{II})$ and the vector $\bar{t} = (t^1, t^2, t^3) = (4, 3, 2)$ are used. The vector of priorities of AIs is $\bar{W} = \frac{(4, 3, 2)}{|(4, 3, 2)|} = (0.45, 0.33, 0.22)$. We constructed two networks G^{wI}, G^{wII} corresponding to T^I, T^{II} . The time restrictions are

chosen in the following way: a) in the network G^{wI} for majority, 80%, of people the time T^I is sufficient to support their contacts completely; b) for the network G^{wII} the situation is opposite - most, 80%, of people should distribute their time resource T^{II} . For the simulated in Example 1 network these parameters are $\bar{T} = (56, 40)$. In Figures 2-3 we can see the resulted HCNs and observe that edge weights in G^{wI} are more heterogeneous than the ones in G^{wII} . Most likely, the reason is in absence in G^{wI} , in most cases, of necessity to redistribute the time resource. After normalizing G^w , the weights of the subnetworks $\{G^{wk}\}$ are $(\omega(G^{wk})) = (0.772, 1.330, 0.962)$, hence they are all not normalised and G^{w2} is the "haviest".

Results of community detection. CD on G does not show community structure in the network whilst CD on G^w quite accurately yields the partition \mathcal{AC}^2 into ACs related to AT^2 , namely, in 80% of cases two ACs of \mathcal{AC}^2 were detected correct, rest two - with one error each in G^{wI} .

5 Conclusions and Future Work

The presented Human Communication Network (HCN) model demonstrates an approach to reconstructing missing network information about edges and edge weights based on node attributes and assumptions on nature of interaction in the networks. To the edge inference problem we apply an extension of Erdos-Renyi and Barabasi-Albert random graph models to multi-layer node attributed networks. There is shown that, in spite of interconnection of HCN layers, CD is running better in these networks decorated by weights.

The results we are planning to expand to other kinds of networks and use for designing new MLCD algorithms and solving node attribute inference problems.

References

1. Albert, R., Barabasi, A.-L.: Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47-97 (2002)
2. Breslin, J., Passant, A., Decker, S.: The Social Semantic Web, 2010 edition. ed. Springer, Heidelberg; New York (2009)
3. Brning, J., Geiler, V.A., Lobanov, I.S.: Spectral Properties of Schrodinger Operators on Decorated Graphs. Mathematical Notes. 77, 858-861 (2005)
4. Erds, P., Rnyi, A: On random graphs, I. Publicationes Mathematicae (Debrecen). 6, 290-297 (1959)
5. Kolaczyk, E.D.: Statistical Analysis of Network Data, Springer Series in Statistics. Springer New York, New York (2009)
6. Lin, Z., Zheng, X., Xin, N., Chen, D.: CK-LPA: Efficient community detection algorithm based on label propagation with community kernel. Physica A: Statistical Mechanics and its Applications. 416, 386-399 (2014)
7. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, 1 edition. ed. Cambridge University Press, Cambridge; New York (1994)
8. Zhou, Y., Cheng, H., Yu, J.X.: Clustering Large Attributed Graphs: An Efficient Incremental Approach. In: 2010 IEEE 10th International Conference on Data Mining (ICDM), pp. 689-698 (2010)

A Model-Based Framework for Adaptive Resource Management in Mobile Augmented Reality System

Mykola Tkachuk¹, Oleksii Vekshyn¹, Rustam Gamzayev¹

¹ National Technical University “Kharkiv Polytechnic Institute”,
Frunze str., 21, Kharkov, Ukraine
tka@kpi.kharkov.ua, alexeyvekshin@gmail.com,
rustam.gamzayev@gmail.com

Abstract. A 3-level modeling framework to adaptive resource management in mobile augmented reality systems (MARS) is proposed, which is based on comprehensive data structuring and analyzing of their specific hard- and software features. At the conceptual modeling level an ontological specification of MARS resources is constructed, at the logical modeling level a case-based reasoning algorithm is elaborated, and as a physical model the reference software architecture is designed. This approach was successfully tested to solve the task to adaptive management of image resolution on mobile device (MD) according to changes of computational loading that finally enabled better video stream quality in MARS.

Keywords. model, adaptation, resource management, augmented reality, case-based reasoning.

Key Terms. ModelBasedSoftwareDevelopmentMethodology, Model, SoftwareSystem

1 Introduction

Nowadays, mobile information systems become more and more popular. One of the most complex and dynamically grown type of these systems are mobile augmented reality systems (MARS) [1]. Such systems require more hardware resources than standard mobile applications (e.g. social network clients, instant messengers etc.), and this fact leads to supporting problems of different devices such as mobile phones and tablets. One of the possible solutions is an execution of complex business logic on the server side, where computational capabilities are higher than on the mobile client side; but on the other hand this, could lead to problems with application response time and energy efficiency because of more intensive usage of wireless networking technologies. One of the overriding trends in software development is using of complex systems construction principles, especially cybernetic adaptive control schemes for software components control including appropriate decision making models and quality evaluation metrics [2]. These approaches are also useful in case of mobile infor-

mation systems development, in particular for MARS. Such systems development requires effective utilization of restricted mobile device (MD) resources and on the other hand implementation of complex real-time algorithms.

In this paper we propose an adaptive model-based framework for resource management in MARS. Additionally a prototype implementation of adaptive MARS is presented and series of experiments with proposed framework and MARS are provided. These experiments highlight positive effect of using adaptive approaches in MARS development.

This paper is structured in the following way: Section 2 depicts briefly some modern trends in this research domain, with respect to some adaptation issues; Section 3 provides main concept of adaptive model-based framework for resource management in MARS; Section 4 presents some ARS domain modeling issues like ontological specifications; in Section 5 the algorithmic model for adaptive resources management and metrics are given; in Section 6 presents proof of concept, introduces software prototype and experimental results. Finally, a short outlook on the result achieved and some future work is presented.

2 Related Work in the Domain of Adaptive Augmented Reality System Development

An Augmented reality (AR) is a representational form for a real physical environment, which is extended by adding of computer-generated data [3]. AR registers physical objects in three dimensions and combines them with the virtual ones. Apart from virtual reality, AR combines models of objects from real world with additional information, but virtual reality completely replaces the real world with the virtual one.

Currently AR supports several data sources: two-dimensional markers; data received from GPS-modules (Global Positioning System) [4] and from build-in gyroscopes. Additionally ARS could use some modern technologies like images recognition without any markers and GPS data [5].

With respect to supported data source, it is possible to construct some MARS classification, which is presented in Fig. 1. There are four “top-level” types of MARS; each of these types has some features and requirements to MD performance. Marker-based MARS operates with special markers, which stores some data and links to additional information. Non-marker based MARS are more complex types of these systems, it based on image recognition algorithms and requires more resources to find and recognize free-form objects in an input image. Geolocation MARS uses build-in GPS sensors in order to get information about real environment and augment it with some virtual data. Finally, infrared-sensors based MARS operates with some infrared sensors, which is able to detect object and moves in real environment, subsequently this class of MARS is very useful in entertainment and simulators. In scope of this research we are focused on Marker-based MARS, because this class of MARS has the lowest complexity, easy to implement and does not require any additional equipment apart from MD.

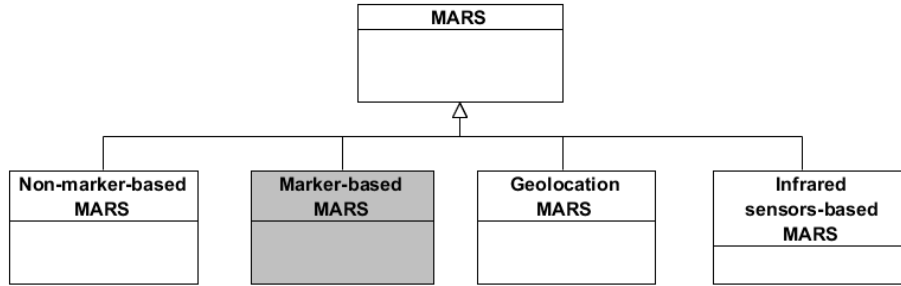


Fig. 1. Classification of MARS types

In terms of software development there are some modern software frameworks to develop MARS: Metaio Mobile SDK (Software Development Kit) [6], D’Fusion Mobile [7] and Qualcomm [8].

Nowadays, software adaptation is one of the common trends in modern software engineering (see, e.g., in [9]), and especially in mobile application development. There are several approaches to adaptation in mobile systems, some of them are represented in projects like Q-CAD (QoS and Context Aware Discovery), MADAM (Mobility and Adap-tation Enabling Middleware), IST-MUSIC (Self-Adapting Applications for Mobile Users in Ubiquitous Computing Environments) [10], CloudRidAR (A Cloud-based Architecture for Mobile Augmented Reality) [11] and researches like Elastic Application Model [12] and resource management in mobile cloud computing [13].

Q-CAD is a resource discovery framework which enables mobile applications to discover and to select resources best satisfied the user’s needs. MADAM and ITS-MUSIC frameworks provide model-driven development approach enabling to assemble applications through a recursive composition process. In this case, variability is achieved by plugging into the same component type different component’s implementation with similar functional behavior [10]. In [14], a new approach to the composition of mismatching components in context-aware systems is introduced. CloudRidAR is a cloud-based framework to MARS development which provides development facilities to construct MARS using all advantages of cloud computing and code offloading, but on the other hand this framework forces developer to use quite complex design solutions (e.g. could computing, workflows, etc) [11].

Elastic Application Model [12] based on code offloading, but flexible application architecture and models are built only on the server side called “weblets” and MD hosts only simple client application, which is connected to these weblets. Resource management in mobile cloud computing [13] addressed to cloud structure, code offloading and energy efficiency but not to runtime application adaptation.

To sum up related work investigation we can conclude that none of existing approached doesn’t provide complete model based framework to adaptive resource management in MARS, but take into account particular aspects of this problem.

In the Fig. 2 and in the Fig. 3, the interface of marker-based MARS is presented. This MARS detects source object in the input video stream recognize it, obtain 3D graphical model of this gear and finally augment source image with this model.

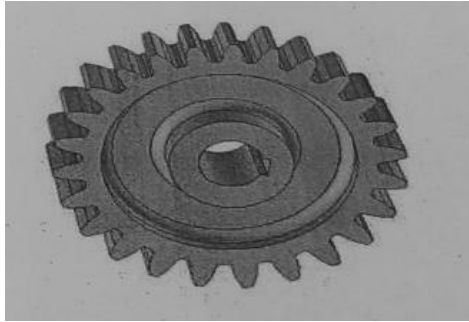


Fig. 2. Not augmented image

Fig. 2 demonstrates interface of the marker-based MARS with not augmented image. In current state MARS is ready to analyze source image and search for markers.

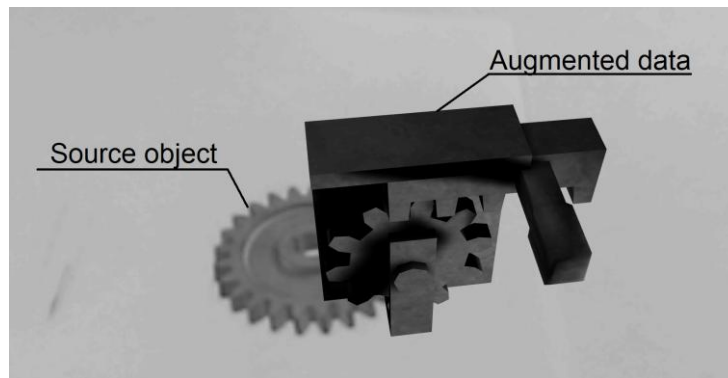


Fig. 3. Example image, augmented by marker-based MARS

In the Fig. 3 result image is shown, MARS finds marker in the source image (in this case – gear), obtains related data and augments source image with this data. In this case, result of augmentation is the gear's 3D model which is presented over sources gear image.

Marker-based MARS [15] uses different marker recognition approaches, such AS QR-codes [16] or barcodes [17]. One of the typical examples of such MARS is QR Droid application [18].

3 Adaptive Model-based Framework for Resource Management

Taking into account the results of the provided analysis and based on the understanding of modern trends in the domain of adaptive MARS-development (see Section 2), we can conclude that it is necessary to elaborate a complex model-based framework for adaptive resource management in MARS. This assumption is completely corresponded with such well-proved and recognized approaches in modern software development as model-driven development (MDD) and model-driven architecture (MDA) [19]. Last time these issues are already discussed intensively in a lot of publications about resource management in distributed real-time systems [20], in SOA- and cloud-centered applications [13, 21, 23], but there is a lack on such work in the domain of MARS development. That is why we propose to construct such model-based framework in the following way:

- to elaborate a domain model specify to describe all hard- and soft-ware resources to be analyzed and considered in any adaptive procedure in MARS, and such a vision services as a conceptual level in the proposed framework;
- to propose some algorithmic approaches to manage these resources in adaptive mode, and this vision about the resources in MARS can be considered as a logical modeling level in our framework;
- to develop a reference software architecture to implement a logical model of MARS with appropriate components and interfaces, and such architecting has to be recognized as physical modeling level in this framework.

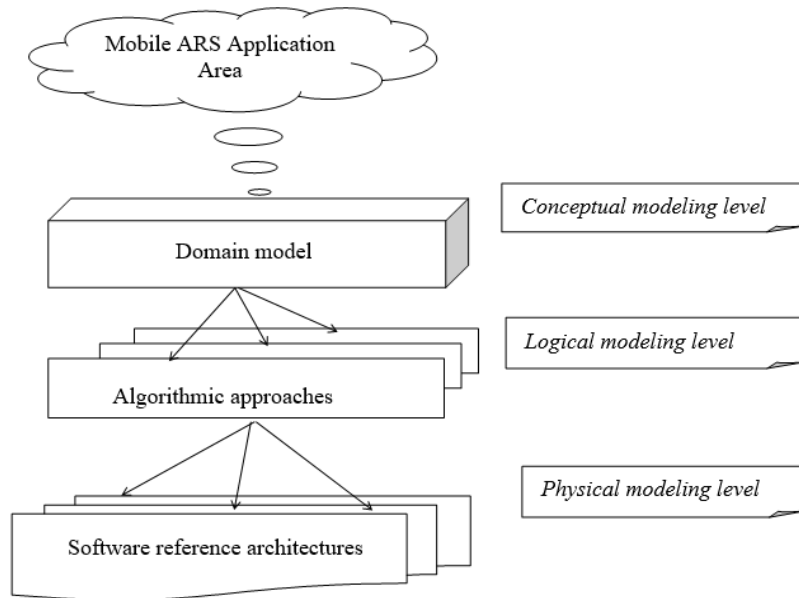


Fig. 4. 3-level modeling framework for MARS adaptation

It is to note that according to this model-based vision about resource management in MARS for the one and the same domain model a lot of different algorithmic approaches can be elaborated, and for any such an approach several reference software architectures might be implemented.

The interacting between these abstraction levels which build the proposed model-based framework is shown in Fig. 4.

In this way, it is possible to elaborate a lot of knowledge-oriented and reusable algorithm-centered solutions and software components, which support the adaptive approach to resource management in MARS. More detailed these issues are considered below.

4 Ontological Specifications for MARS Domain Modeling

Taking into account the 3-level modeling framework for adaptive resource management in MARS proposed in previous Section 3, it is needed to elaborate the domain model for this purpose. This model should represent all relevant hard- and software capabilities of MD which can be considered as adaptable parameters in an appropriate algorithmic modeling approach.

Ontology models are widely used to represent relationships between concepts in some application domain, and they can be applied for different purposes in software engineering, e.g.:

1. for information sharing between human and machines in Semantic Web applications [23];
2. for natural language processing, and knowledge engineering,
3. in software product line engineering [24], etc.

For example, [25] represent the domain model including 4 ontologies like User, Service, Environment and Device. It describes general relationships that occur during ARS development. In [26] ontology is used to build platform for educational institutions that could be used for rapid ARS development. In [27] the ontological specifications are used to connect together the knowledge about the users, environment, and user aims in the given application domain (museum).

Most of models mentioned above describe different MARS components taking into account some static resource allocation, and they do not represent system features needed for adaptive resource management. To close this gap in our approach the new MARS model was created using the OWL notation, and the OWLGrEd tool is used for this purpose [28], which provides the UML-like graphical editor. In Fig. 5, the proposed ontological domain model for MARS is presented.

There are following main entities included in our domain model

1. *Augmented Reality Application*: is an application that provides a technology to analyze elements in the real physical environment, and to extend them with additional virtual objects or with additional information.

2. *ACU (Adaptive Control Unit)*: this is control management element that is responsible to adapt system to the mobile device. It could use different techniques to optimize application, e.g. to change display resolution.
3. *Mobile Device*: this is a small device (iPhone, PDA, notebook, tablet etc.), with restricted amount of system resources, which does not have permanent access to the power sources, and which uses wireless communication technologies. For resource adaptations, the following device parts (model entities) could be used: *Screen* – by changing screen resolution performance of the device could be changed; *RAM* – depending on the allocated memory different application modes might be used; *Battery* - usually a charged level is used as a main parameter to switch device into power save mode; modern *CPU* could change frequency and switch its operation mode.
4. *AR Object Recognizer*: is a module to analyze and to extend extracted physical objects with additional or virtual information.

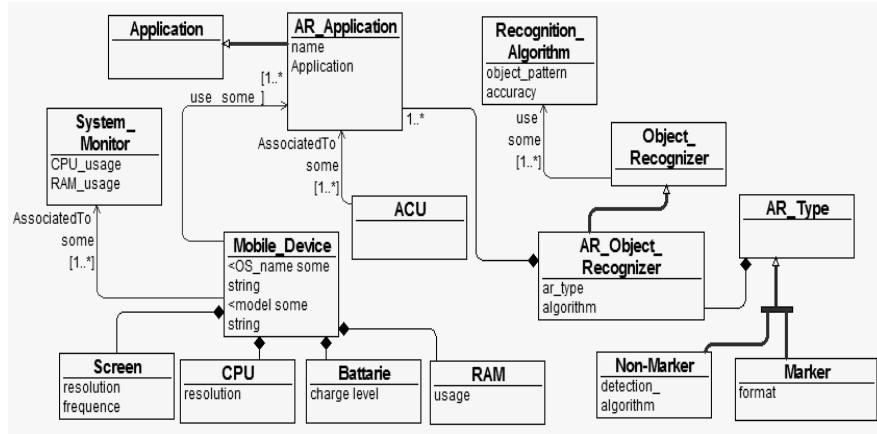


Fig. 5 Ontological domain model for resource adaptation in MARS

Below this domain model is used to elaborate an algorithmic approach to adaptive resource management in MARS.

5 Algorithmic Model of Adaptive Resource Management in MARS

According to the proposed adaptive model-based framework (see Section 3) at its logical level an algorithmic model to resources management has be constructed with respect to specific hard- and software characteristics of MD which is used to get client applications running within a given MARS.

5.1 Formal Definition

In order to formalize the proposed framework to find adaptive solutions for resource management we can use an algorithmic modeling approach [29], and the appropriate algorithmic model AM can be defined as the following tuple

$$AM = \langle Workflow(Methods), InfoBase, Metrics \rangle, \quad (1)$$

where $Workflow(Methods)$ are some algorithms which implement the given methods ($Methods$), $InfoBase$ is an information base to be used for these methods, and $Metrics$ is a collection of metrics to assess a quality of adaptation process in mobile ARS. The choice of a set of adaptation methods in (1) depend on specific features of ARS's resources, which should be managed, and one of such possible way will be considered in the next subsection. A set of metrics in (1) also has to be defined taking into account the appropriate hard- and software properties of MD which is used in a target ARS (see e.g. in [10, 15]).

5.2 Case-based reasoning (CBR) within an algorithmic model

Taking into account a complex and weak-formalized character of ARS – functioning, namely: parallel and multi-threaded calculation processes, turbulence loading on MD, permanent changes on number of users etc., it is reasonable to use so-called soft calculation methods [30]: neuronal net technologies, fuzzy-logic methods, generic algorithms, case-based reasoning (CBR) and some others. In particular, exactly CBR-methods can be considered as an effective way to develop decision-making procedures for management of complex software system (see e.g. in [31-33]). According to this statement the collection of Methods in formula (1) can be specified in the following form

$$Methods = (NNM, kNNM, kwNNM), \quad (2)$$

where NNM is a Nearest Neighbor Method, kNNM is a k-Nearest Neighbors Method, and kwNNM is k-weighted Nearest Neighbors Method [31].

The main idea of all CBR-methods is that any new problem occurred in some application domain can be resolved using already existing solution for the similar situation (called precedent or case). The several CBR differ each other in a search algorithm to find an appropriate precedent in the given case - database. For this purpose, it is also important to elaborate an adequate description for the precedent's representation, which reflects all relevant issues of ARS functionality.

5.3 Information base for CBR-method

Corresponding to formula (1), InfoBase is an information base which is used to apply the CBR-methods defined in (2). It includes a set of precedents, and any such precedent can be defined in the following way

$$c = (\vec{p}, \vec{s}), \quad (3)$$

where \vec{p} is a vector of parameters to characterize a given problem situation, \vec{s} is a vector of parameters to represent an appropriate solution for this problem.

Taking into account the hardware - and software issues of MD which is included in ARS, vector \vec{p} can be given as:

$$\vec{p} = (CPU, RAM, BAT, RES, FPS), \quad (4)$$

where CPU is a current level of processor loading (in %), RAM is a current level of RAM usage, BAT is a current level of battery charging; RES is a number of possible screen resolution modes in MD, FPS is a measure of a screen refresh rate.

Vector \vec{s} in formula (2) can be represented as the tuple

$$\vec{s} = (Width, Height), \quad (5)$$

where Width and Height are respectively a width and a height of a video frame on MD.

5.4 Metrics for ARS resource estimation

In [15] is mentioned that a performance of a ARS-client application is depend on its screen resolution, and accordingly to this reason a number of frame per second (FPS) can be used as one of the metrics from its set Metrics defined in formula (1). This factor is depend on some parameters: on power of MD processor, on size of its RAM, on screen resolution of MD, and on screen resolution of video-camera.

Therefore, a collection of metrics *Metrics* in (1) has the following definition

$$\vec{s} = (T, R), \quad (6)$$

where T is a metric to estimate a number of frame per second, P is a metric to measure a MD total productivity (R).

A value of metric T can be calculated using the standard function Count(), namely:

$$T = Count(FPS) \quad (7)$$

A value of metric R (named below as a productivity index) defines a MD performance ratio, which is dimensionless parameter and it can be defined as following

$$R = w_c \frac{CPU_{cur}}{CPU_{total}} + w_r \frac{RAM_{cur}}{RAM_{total}} + w_b \frac{BAT_{cur}}{BAT_{total}} \quad (8)$$

where CPU_{cur} is a current MD processor loading ratio (in %); RAM_{cur} is a current RAM usage ratio (in Kb); BAT_{cur} is a current battery charging ratio (in Ah); CPU_{total}, RAM_{total}, BAT_{total} are respectively the nominal values of the given parameters; w_c , w_r , w_b are some weighting coefficients for these parameters, and the following

condition must be fulfilled $w_c + w_r + w_b = 1$. In this work we have defined the following value ranges for the index R : it is critical if $R \geq 0.95$; it is high if $0.6 \leq R < 0.95$; it is normal if $0.25 \leq R < 0.6$; and it is low if $0 \leq R \leq 0.25$.

Therefore, the metrics defined in formula (6) – (8) allow us to estimate the computational resources of an appropriate MD which is used in a target ARS with respect to our final goal: to provide an adaptive recourses management in this ARS.

6 Prove of Concept: Software Prototype and Experimental Results

6.1 Software prototype design and implementation

In order to prove efficiency of the proposed approach the MARS prototype with integrated ACU has been developed. The main purpose of developed MARS is to recognize marker on a cinema poster, search information about this cinema in an appropriate database and augment source video stream with this additional information at real-time mode. Such MARS with integrated ACU which analyzes environmental parameters and adopts frame size with respect to MD current state and resources utilization rate. In the Fig. 6 is presented adaptive MARS functioning algorithm in form of UML activity diagram.

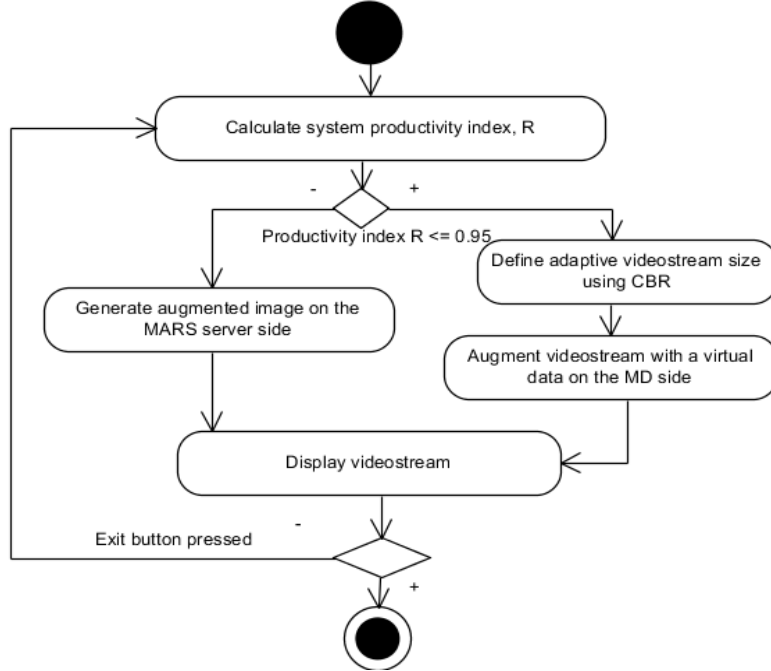


Fig. 6. Adaptive MARS functioning algorithm

Initial activity in this algorithm is – calculation of a productivity index. If this index is less than 0.95 ($R \leq 0.95$) it is possible to augment data on the MD side, so the next steps are to define adaptive video stream size, using CBR method, and augment source video stream with a virtual data on a MD side, and finally – display augmented video stream to user. If index $R > 0.95$ it is not possible to augment data on the MD side and in this case it is necessary to use external resources to augment image from input video stream (e.g. server in client-server MARS) and show result to user. This activity could be interrupted by user's event (pressing exit button).

Presented algorithm takes into account some code offloading possibility (in case of quite high computation load on a MD), subsequently it is useful to select three-tier software architecture to implement adaptive MARS prototype. Such architecture allows us to implement recognition component on the server side in order to offload some logic from MD in case of critical computational load. Another plus point of this architecture is a possibility to deploy centralized precedent database on the server side and distribute this precedents among different MDs.

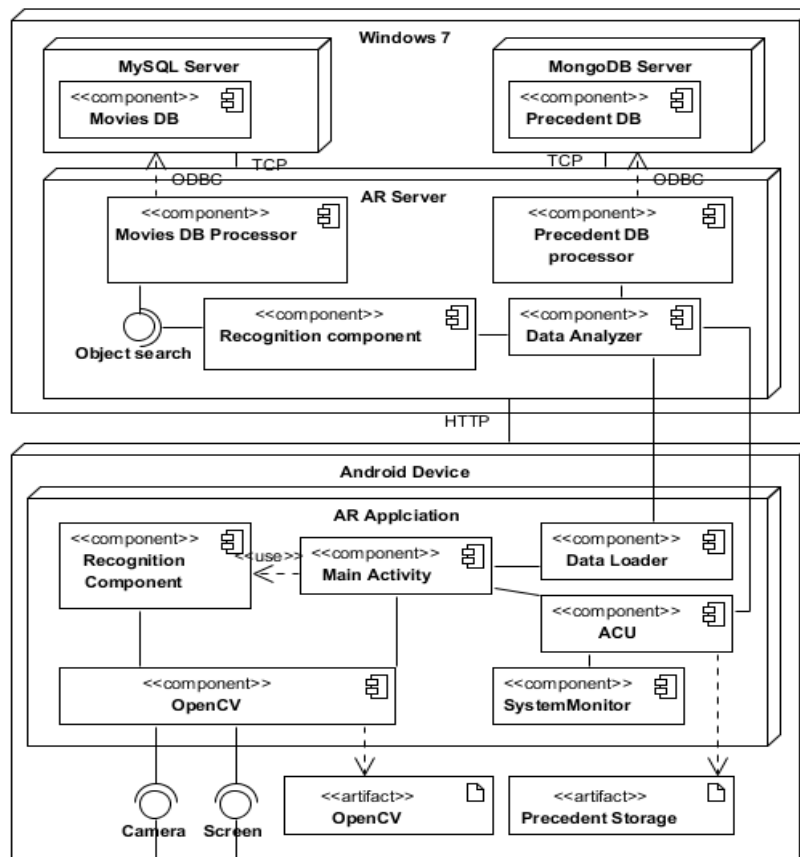


Fig. 7. Adaptive MARS functioning algorithm

In the Fig. 7 presented three-tier component software architecture of MARS in form of UML deployment diagram. This architecture provides few crucial advantages such as: high scalability, data processing security, lower resource requirements for clients MD.

Client MARS application implemented with Android platform [34], using embedded Berkeley DB [35] and OpenCV library [36]. To develop server-side application PHP programming language [37], Apache web-server [38], MySQL [39] and MongoDB [40] have been selected.

ACU on a MD node (Android Device) contains the following components: Precedent-Storage – local precedent DB, SystemMonitor – the component which provides data regarding current state of MD and ACU – the component, which implements CBR methods on the mobile client side. On the server node into adaptation process are involved the following components: Precedent DB Processor is the accessor component for centralized Precedents DB, and Data Analyzer is the component that implements movie's additional data search by input images hash-code.

With respect to the proposed architectural solution, three databases have been implemented: 1) local embedded precedents DB (PrecedentStorage), this DB is used by ACU; 2) the remote centralized DB (Precedent DB) to store all precedents, all local DBs are synchronized with this one; 3) movies DB (Movies DB), this DB stores information about movies, which are handled in MARS prototype. Conceptual data model of the Precedent DB is presented in the Fig. 8 as the UML class diagram. This data model takes into account the following entities: Device, Precedent, Param, ListPrecedent, Platform, TypeParameter. Therefore, it allows handling data, required by CBR method in our domain.

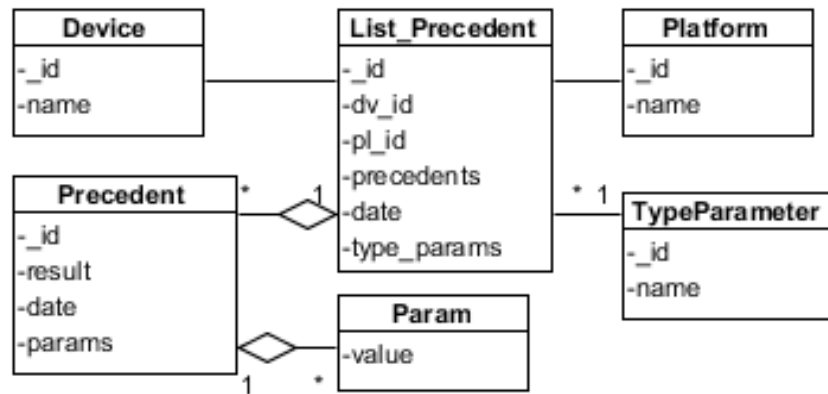


Fig. 8. Conceptual data model of the Precedent DB

To develop this database we have selected non-relational (NoSQL) database management system MongoDB [40]. This DBMS provides high-speed data processing and stores the appropriate information in object-oriented form.

6.2 Estimation results and their analysis

In order to estimate an efficiency of the implemented MARS prototype the experiments have been performed using the following scheme: 1) selection of mobile devices for experiments; 2) precedents DB generation; 3) run MARS prototype with different operating modes for ACU (with enabled and disabled ACU).

To test MARS prototype two types of MD were selected, the detailed characteristics of these devices are presented in the Table 1.

Table 1. Technical characteristics of testing MD

Device name	Processor	RAM	Maximal resolution
Nexus 7	Qualcomm APQ8064 (1.5 GHz)	2 GB RAM	1920x1080
Fly IQ4416	MT6572 (1.3 GHz)	512 MB RAM	800x600

In the Fig. 9 presented example of tuple, which describe particular precedents included in the precedents DB.

Nº	Result	CPU	RAM	BATTERY	RESOLUTION	FPS
1	640x480	0.7	0.7	0.4	800x600	13
2	800x600	0.2	0.5	0.9	640x480	17
3	1024x768	0.3	0.8	0.5	1920x1080	2

Fig. 9. Example of precedents tuple in precedents DB

Two series of experiments have been provided with these mobile devices. The first experiment has been provided with disabled ACU (i.e. without adaptation), and the second one with enabled ACU. During these experiments the image resolution of MD screen in case of different values of productivity index (see equation 8) had been measured. In this experiment we take into account two intervals from normal and high ranges of the productivity index: $0.4 \leq R < 0.6$ and $0.6 < R \leq 0.8$. The results of these experiments are presented in Table 2 and Table 3 respectively.

Table 2. Experimental results in case of disabled ACU

Mobile device	Resolution	T (FPS) for $0.4 \leq R < 0.6$	T (FPS) for $0.6 < R \leq 0.8$
Nexus 7	640x480	15	13
Nexus 7	800x600	14	10
Fly IQ4416	800x600	10	9
Fly IQ4416	480x320	15	14

The data from the Table 2 show that in case of the fixed image resolution on MD, and if the value of productivity index R is increased: from the values range $[0.4; 0.6]$ to the range $[0.6; 0.8]$, then the $T(FPS)$ metric value is decreased, namely, these values are placed in interval $[9, 14]$. In other words, the maximum value's difference $T(FPS)$ is about 35.7% apart from difference 33,3% for case of $[10, 15]$ ($0.4 \leq R < 0.6$). The reason of such trend in this experiment is the disabled mode of ACU.

Table 3. Experimental results in case of enabled ACU

Mobile device	Resolution	T (FPS)	Resolution	T (FPS)
		$0.4 \leq R < 0.6$		$0.6 < R \leq 0.8$
Nexus 7	1024x768	11	800x600	13
Nexus 7	800x600	14	640x480	16
Fly IQ4416	640x480	13	640x480	14
Fly IQ4416	800x600	9	640x480	13

Table 3 represents experiment results with enabled ACU. ACU component monitors computational load on a mobile device and in case of its increasing correct image resolution; such correction leads to stabilization of $T(FPS)$ metric: this metric changes only in 18.75%. Additionally, in this experiment value of the maximal difference for $T(FPS)$ in case of $0.4 \leq R < 0.6$ is 37.5%. That is why we can make conclusion that the proposed adaptive resource management approach enables better video stream quality for mobile device in MARS.

7 Conclusions and Future Work

In this paper we have presented the model-based framework to adaptive resource management in mobile augmented reality systems (MARS), which is based on the 3-level data structuring and analyzing of their specific hard- and software features. At the conceptual modeling level the appropriate ontological specification of MARS resources was constructed, at the logical modeling level a case-based reasoning approach is utilized, and as a physical model to implement an adaptive resource management in MARS the 3-level reference software architecture is elaborated. This approach was successfully applied in order to solve the task to adaptive management of screen image resolution on mobile device according to changes of its computational loading that finally enabled better video stream quality in MARS.

In future we are going to extend a collection of decision search methodologies in order to improve an adaptation process and compare its efficiency with case-based reasoning approach implementation. Besides that is it supposed to develop a more sophisticated adaptive MARS domain model with wider amount of input and output parameters, which should enable a more configure options in the proposed model-based adaptive resource management framework.

8 References

1. Lopez, H., Navarro, A., Relano, J.: An Analysis of Augmented Reality Systems. In: Proceedings of the 2010 Fifth International Multi-conference on Computing in the Global Information Technology (ICCGI '10), pp. 245--250 (2010)
2. Joao W. C., Scott D. M., Kai-Yuan C., Aditya P. M.: Software Cybernetics. Encyclopedia of Computer Science and Engineering. WILEY (2008).
3. Furth, B.: Handbook of Augmented Reality. Springer-Verlag, New York (2011)
4. Official U.S. Government information about the Global Positioning System (GPS) and related topics, <http://www.gps.gov>
5. MacWilliams, A.: Decentralized Adaptive Architecture for Ubiquitous Augmented Reality Systems. PhD thesis, Institut fur Informatik der Technischen Universitat Munchen (2005)
6. Official Web-site of Metaio Mobile SDK, <http://www.metaio.com/software/mobile-sdk/>
7. Official Web-site of D'Fusion Mobile project, <http://www.t-immersion.com/products/dfusion-suite/dfusion-mobile>
8. Official Web-site of Qualcomm AR SDK, <http://www.qualcomm.com/solutions/augmented-reality>
9. Kell, S.: A Survey of Practical Software Adaptation Techniques. In: J. of Universal Computer Science, Vol. 14, pp. 2110--2157 (2008)
10. Kakousis, K., Paspallis, N., Papadopoulos, G. A.: A survey of software adaptation in mobile and ubiquitous computing. In: J. of Enterprise Information Systems, Vol. 4, pp. 355--389 (2010)
11. Huang, Z., Li, W., Hui, P., Peylo, C.: CloudRidAR: A Cloud-based Architecture for Mobile Augmented Reality. In: Proceedings of MARS '14, pp. 29-34 (2014)
12. Zhang, X., Jeong, S., Kunjithapatham A., Gibbs, S.: Towards an elastic application model for augmenting computing capabilities of mobile platforms. In: Proceedings of 3rd International ICST Conference on Mobile Wireless Middleware, Operating Systems, and Applications (Mobile-Ware), pp. 161--174 (2010)
13. Ionescu, A.: Resource Management in Mobile Cloud Computing. Informatica Economica, Vol. 19, pp. 55--66 (2015)
14. Camara, J., Salaun, G., Canal, C.: On run-time behavioral adaptation in context-aware systems. In Proceedings of M-ADAPT'07 at ECOOP '07. pp. 26--34 (2007)
15. Vekshyn O., Tkachuk M. Algorithmic Software Adaptation Approach in Mobile Augmented Reality Systems. In: Proceedings of 7-th International Conference on Software Engineering Advances, Lisbon, Portugal, pp. 40--43 (2012)
16. Official Web-site of QRCode, <http://www.qrcode.com>
17. Official Web-site of Barcodes & ID Key Standards, <http://www.gs1.org/gsm/kc/barcodes>
18. Official Web-site of QR Droid, <https://play.google.com/store/apps/details?id=la.droid.qr&hl=ru>
19. Sommerville, I. Software Engineering. Addison Wesley (2011)
20. Tian He et.al. Feedback Control-based Dynamic Resource Management in Distributed Real-time Systems. In: J. of Systems and Software, vol. 80, pp. 997--1004 (2007)
21. Ghanbari, H. Model-Based Dynamic Resource Management for Service Oriented Clouds. PhD thesis, York University Toronto, Ontario (2014)
22. Sun, Y. White, J., Eade, S.: A Model-Based System to Automate Cloud Resource Allocation and Optimization. In: J. of Model-Driven Engineering Languages and Systems, Vol. 8767, p 34 (2014)
23. Fensel, D., Kerrigan, M., and Zaremba M.: Implementing Semantic Web Services: The SESA Framework. Springer-Verlag Berlin Heidelberg(2008)

24. Tenório T., Dermeval D., Bittencourt I.: On the use of ontology for dynamic reconfiguring software product line products. In: ICSEA 2014, Proceedings of the 9th International Conference on Software Engineering Advances, pp. 545--550 (2014)
25. Hervas, R., Bravo, J., Fontecha, J., Villarreal, V.: Achieving Adaptive Augmented Reality through Ontological Context-awareness Applied to AAL Scenarios. In: J. of Universal Computer Science, vol. 19 (9), pp. 1334--1349 (2013)
26. Chuan-Jun S., Tzu-Ning Y., Yu-Ming Y.: Ontology-based Mobile Augmented Reality for Personalized U-Campus. In: Proceedings APIEMS 2012, pp.2037--2046 (2012)
27. Hatala, M., Wakkary, R.: Ontology-Based User Modeling in an Augmented Audio Reality System for Museums. In: J. of User Modeling and User-Adapted Interaction, Vol. 5(3--4), pp. 339--380 (2005)
28. Bārzdīņš, J., Bārzdīņš, G., Čerāns, K., Liepiņš, R., Sproģis, A.: OWLGrEd: a UML Style Graphical Notation and Editor for OWL 2. In: Proceedings of 7th International Workshop "OWL: Experience and Directions", (2010)
29. Ramesh, K., Karunanidhi, P. Literature Survey on Algorithmic and Non-Algorithmic Models for Software Development Effort Estimation. In: Int. J. of Engineering and Computer Science, Vol. 2 (Issue 3), pp. 623--632 (2013)
30. Aliev A., Soft Computing and its Applications. World Scientific (2001)
31. Maiden, N., Sutcliffe, A. Case-based Reasoning in Software Engineering. In: IEE Colloquium on Case-Based Reasoning, Digest No.036, pp. 1-3, (1993)
32. Limthanmaphon, B., Zhang, Y. Web Service Composition with Case-Based Reasoning. In: Proceedings of 14th Australian Database Conference (ADC2003), Vol. 17, (2004)
33. Tkachuk, M., Polkovnikov, S., Bronin, S. Adaptive Control Framework for Software Components: Case-based Reasoning Approach. In: Proceedings of 6th International Workshop on Software Cybernetics (IWSC – 2009), Seattle, USA, pp. 47--56 (2009).
34. Official Web-site of Android platform, <https://www.android.com/>
35. Official Web-site of Oracle Berkeley DB, <http://www.oracle.com/technetwork/database/database-technologies/berkeleydb>
36. Official Web-site of OpenCV project, <http://www.opencv.org>
37. Official Web-site of PHP language, <http://www.php.net>
38. Official Web-site of Apache project, <http://www.apache.org>
39. Official Web-site of MySQL project, <http://www.mysql.com>
40. Official Web-site of MongoDB , <http://www.mongodb.org>

Descriptive Models of System Dynamics

Zholtkevych G. N.¹, Nosov K. V.¹, Bespalov Yu. G.¹, Rak L. I.¹,
Vysotskaya E. V.², Balkova Y. B.¹, and Kolomiychenko V. K.²

¹ V. N. Karazin Kharkiv Nat. Univ.,
4, Svobody Sqr., Kharkiv, 61077, Ukraine
g.zholtkevych@gmail.com, k.nosov@yahoo.com

² Kharkiv Nat. Univ. of Radio-Electronics
Science ave., 14, Kharkiv, 61166, Ukraine

Abstract. Nowadays in the course of investigation researchers are facing large arrays and datasets required fast processing, analysis and drawing adequate conclusions. Data mining, statistical methods and big data analytics provide an impressive arsenal of tools allowing scientists to solve these tasks. However, investigators often require techniques that enable with use of relatively simple and cheap measurements of easily accessible parameters to build useful and meaningful concepts.

In our paper two classes of dynamical models aimed at revealing between-component relationships in natural systems with feedback are presented. The idea of both models follows from the frameworks of theoretical biology and ecology regarding pairwise interactions between the parts of the system as background of system behavior. Both deterministic and stochastic cases are considered, that allow us to determine the direction of pairwise relationships in the deterministic case and the direction and strength of relationships in the stochastic one.

Keywords: dynamical systems, deterministic models, stochastic models, Markov chains

Keyterms: Model, MathematicalModel, MathematicalModeling, ComputerSimulation

1 Introduction

The world of scientific research has been immersing into an extraordinary information explosion over past decades, accompanied by the rapid growth in the use of Internet and the number of connected computers worldwide. We see a rate of increase in data growth that is faster than at any period throughout history. Enterprise application and machine-generated data continue to grow exponentially, challenging experts and researchers to develop new innovative techniques to evaluate hardware and software technologies and to develop new methods of big data investigation [1].

Heterogeneity, scale, timeliness, complexity, and privacy problems with big data impede at all phases of obtaining value from data. The problems begin during data acquisition, when the amount of data requires us to make decisions, currently in an *ad hoc* manner, about importance and interpretability of data. Besides, much data today are not natively in structured format, have gaps and incomplete. Hence, data analysis, organization, retrieval, and modeling are foundational challenges. Finally, presentation of the results and its interpretation by non-technical domain experts is crucial to extracting actionable knowledge.

Our study is devoted to the well-known problem of revealing conditions of stability in natural systems providing long and steady development and existence of systems. Today there is a large amount of online big data collections comprising datasets taken from different branches of biology, health sciences, ecology etc. As examples we can mention Data Centre of International Council for the Exploration of the Sea (includes hundreds of thousands marine biology related datasets), which were used in the current investigation.

The problem of homeostasis and stability in the living organisms community or natural systems (biological or ecological) is closely related to the problem of dynamic stability. The practical aspect of this problem is connected to the disturbance in stability of systems, that is often accompanied, for example, by outbreaks in number or biomass of species.

The study of stability in communities or natural systems is closely connected to investigation of relationships that determine the dynamic characters of a system, i. e. relationships between systems parameters having influence on the system dynamics.

For decades systemic methods, for example, based on the Shannon index of diversity, have been used for studying the relationships between the structure and stability of a system. Generalizing that and other approaches, Margalef [2] states that “the ecologist sees in any measure of diversity an expression of the possibilities of constructing feedback systems, or any sort of links, in a given assemblage of species”. Similar ideas were therefore presented in studying the structure of correlation pleiads, in using cluster analysis and other statistical techniques to establish such relationships for investigating similar problems.

Despite different approaches to revealing between-component relationships, in biology and ecology there is a general approach in studying such relationships, based on the following pairwise relationships: $(+, +)$, $(-, -)$, $(-, +)$, $(-, 0)$, $(+, 0)$, $(0, 0)$. For the multi-component systems, this set of relationships exhausts all possible pairwise inter-component

relationships categorized by the type of effect and have been studied at length, particularly, in biology and ecology [3–5]. Therefore, in the current paper the analysis of the relationships structure is based on the idea of regarding the objects (i. e., living organisms in a community, species etc.) as components of a system between which the pairwise relationships mentioned above are possible. This allows us to present the structure of relationships in an explicit form of relationships between the components of a natural system.

It should be noted, that mentioned relationships not always can be revealed with the help of statistical methods. For example, correlation analysis is initially used for estimation of a relationship between two or more variables, but it covers only statistical relation and cannot reveal a cause-effect relationship [6].

There are statistical methods (structural relation modeling, analysis of path and adjacent techniques), which are devoted to revealing between-component relationships (and other tasks as latent variables' analysis) and can be used for causality analysis [7, 8]. But these methods express the relationships in a system in the terms of regression coefficients and not in the form of paired relationships. Besides, interpretation of results of this analysis is occasionally difficult (e.g. studying relationship between a feedback system and homeostasis in a community).

2 Theory

Here we present two dynamical models developed for revealing between-component relationships on the base of observation data obtained from a real natural system.

First model has deterministic dynamic, finite number of states and discrete time. As it is described in [9] at length, here we describe the model in brief. The second model is stochastic and will be describe in more detailed. Both models have a common background, so we begin with its description and later will go to specific properties of each models.

We assume that a natural system to be modelled comprises N components, which can be denoted by A_1, A_2, \dots, A_N . It is assumed that the component take integer values $1, 2, \dots, K$, i. e. K value for each component. The value 1 means a minimum amount of a component, the value K means maximum, i. e. the component value varies from 1 to K .

The system develops in discrete time and the moments of time are to be denoted $t = 0, 1, \dots$. So, the value of the component A_i at the moment of time $t = 0, 1, \dots$ are numbers $A_i(0), A_i(1), \dots$

The next properties of a system are different for deterministic and stochastic cases, so we shall describe them separately.

2.1 Deterministic model revealing the direction of between-component relationships

We begin with deterministic case discussed, as mentioned, in [9] and was named the Discrete model of dynamical systems with feedback. For the deterministic system its state at the moment $t + 1$ is fully and definitively determined by the state at the moment t .

If the system at the moment t is in the state $(A_1(0), A_2(0), \dots, A_N(0))$, all the following states can be written as the trajectory, where each column is a state at corresponding moment:

$$\begin{pmatrix} A_1(0) & A_1(1) & A_1(2) & \dots \\ A_2(0) & A_2(1) & A_2(2) & \dots \\ \vdots & \vdots & \vdots & \dots \\ A_N(0) & A_N(1) & A_N(2) & \dots \end{pmatrix}. \quad (1)$$

In the theory of dynamical systems [10], such a system is called a free dynamical system with discrete time. The system has only finite number of states, so there exists a positive integer \mathcal{T} , which can be called a period of the trajectory, for which the conditions of periodicity hold $A_i(s) = A_i(s + \mathcal{T})$ for enough large s .

Taking into account the periodicity, it is possible to extract the following minor form (1)

$$\begin{pmatrix} A_1(s) & A_1(s+1) & \dots & A_1(s+\mathcal{T}-1) \\ A_2(s) & A_2(s+1) & \dots & A_2(s+\mathcal{T}-1) \\ \vdots & \vdots & \ddots & \vdots \\ A_N(s) & A_N(s+1) & \dots & A_N(s+\mathcal{T}-1) \end{pmatrix} \quad (2)$$

presenting full description of the dynamics of the system.

Now we introduce the concept of relationships between components. Let $\Omega = \{-, 0, +\}$. A relationship between specified components A_i and A_j is determined as an entry from the set $\Omega \times \Omega$ and denoted by $\Lambda(A_i, A_j) = (\omega_1, \omega_2)$, where $\omega_1 \in \Omega$, $\omega_2 \in \Omega$. If $\Lambda(A_i, A_j) = (\omega_1, \omega_2)$, this means that:

- if $\omega_1 = \{-\}$, then large values of the A_j will lower the value of the A_i .
- if $\omega_1 = \{0\}$, then the A_j doesn't influence the value of the component A_i .

- if $\omega_1 = \{+\}$, then large values of the A_j will raise the value of the A_i .

The relationship Λ is antisymmetric in the following sense: $\Lambda(A_i, A_j) = (\omega_1, \omega_2)$ implies $\Lambda(A_j, A_i) = (\omega_2, \omega_1)$.

Assume that all the relationships $\Lambda(A_j, A_i)$ between all pairs (A_j, A_i) of components A_1, A_2, \dots, A_N are given. For each A_j and each $(s, u) \in \Omega \times \Omega$ it is possible to find the set of components, with which A_j has the relationship (s, u)

$$L_j(s, u) = \{A_i | \Lambda(A_j, A_i) = (s, u)\}.$$

Let $\varkappa = \{1, 2, \dots, K\}$ is to be the set of the states of an individual component and $N_j(s, u)$ is the number of components in the set $L_j(s, u)$, $j = 1, 2, \dots, N$, $(s, u) \in \Omega \times \Omega$. A transition from the state at t to the state at $t + 1$ is described by N transition functions F_j , each of which defines the mapping

$$\varkappa^{N_j(+,+) + N_j(+,0) + N_j(+,-) + N_j(-,+) + N_j(-,0) + N_j(-,-)} \mapsto \varkappa.$$

Two types of relationships, intrinsic to natural systems. For more detailed description of the dynamics of a natural system, one needs to specify the explicit form of the transitional mapping. We introduced two approaches based on the concepts of biological interactions: weight functions' approach and approach based on principles of Justus von Liebig's law.

Define the following functions on the set \varkappa : $\text{Inc}(A) = \min\{K, A + 1\}$, $\text{Dec}(A) = \max\{1, A - 1\}$.

The system dynamics based on weight functions' approach. First we define the type of dynamics, which takes into account the weighted sum of all $A_j(t)$ (inclusive $A_i(t)$) for calculating the value of the component A_i at the moment $t + 1$.

As we defined above, for each j ($j = 1, 2, \dots, N$) and each pair $(s, u) \in \Omega \times \Omega$ there exists the set $L_j(s, u)$ with $N_j(s, u)$ entries. Assume that the function $\varphi_{j,1}^{(s,u)}(\cdot), \varphi_{j,2}^{(s,u)}(\cdot), \dots, \varphi_{j,N_j(s,u)}^{(s,u)}(\cdot)$ are to be the functions of interactions of those components, with which the A_j has relationships (s, u) .

The functions are defined on the discrete set \varkappa and have the following properties: (i) $\varphi_{j,k}^{(+,+)}(\cdot), \varphi_{j,k}^{(+,0)}(\cdot), \varphi_{j,k}^{(+,-)}(\cdot)$ are increasing functions; (ii) $\varphi_{j,k}^{(-,+)}(\cdot), \varphi_{j,k}^{(-,0)}(\cdot), \varphi_{j,k}^{(-,-)}(\cdot)$ are decreasing functions; (iii) $\varphi_{j,k}^{(s,u)}(1) = 0$ for any $(s, u) \in \Omega \times \Omega$.

We also introduce the numbers $\delta_j > 0$ ($j = 1, 2, \dots, N$) which can be called thresholds of sensitivity.

For the system's state at the moment of time t the following value is calculated

$$\begin{aligned} d_j = & \sum_{A_k \in L_j(+,+)} \varphi_{j,k}^{(+,+)}(A_k(t)) + \sum_{A_k \in L_j(+,0)} \varphi_{j,k}^{(+,0)}(A_k(t)) + \\ & \sum_{A_k \in L_j(+,-)} \varphi_{j,k}^{(+,-)}(A_k(t)) + \sum_{A_k \in L_j(-,+)} \varphi_{j,k}^{(-,+)}(A_k(t)) + \\ & \sum_{A_k \in L_j(-,0)} \varphi_{j,k}^{(-,0)}(A_k(t)) + \sum_{A_k \in L_j(-,-)} \varphi_{j,k}^{(-,-)}(A_k(t)). \end{aligned} \quad (3)$$

The value of the component A_j is being changed according to the value d_j by the following rules

1. if $d_j \geq \delta_j$, then $A_j(t+1) = \text{Inc}(A_j(t))$;
2. if $d_j \leq -\delta_j$, then $A_j(t+1) = \text{Dec}(A_j(t))$;
3. if $-\delta_j < d_j < \delta_j$, then $A_j(t+1) = A_j(t)$.

Now, the meaning of introduced transition functions can be explained in clear way. For example, the functions $\varphi_{j,k}^{(-,+)}(\cdot)$ ($k = 1, 2, \dots, N_j(-,+)$) reflects the influence upon the component A_j of components in the set $L_j(-,+)$, which are related with A_j by relationship $(-,+)$. The greater the influence (i. e. the greater values of $A_i(t)$ from the set $L_j(-,+)$), the lower the values of d_j .

The dynamics based on the Liebig's law of the minimum. Next approach is based on principles of Justus von Liebig's law (Liebig's law of the minimum) and essentially differs from the first approach, which is basically additive.

Omitting the details, enough to say, that according to this approach, transition from the state $(A_1(t), A_2(t), \dots, A_n(t))$ is defined by relations of the state with two matrices C and C^* playing the role of a threshold.

The system identification with use of the observation data. While dealing with real data, we often don't observe the data in dynamics. Often real data come unordered in time in contrast to data used for time series modeling. So we don't observe any dynamism described by the trajectory (1) or the minor (2).

Usually, the result of observation is represented by a table of cases:

$$\tilde{M} = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1B} \\ C_{21} & C_{22} & \dots & C_{2B} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N1} & C_{N2} & \dots & C_{NB} \end{pmatrix}, \quad (4)$$

where columns correspond to cases and rows correspond to components (N components and B cases). We emphasize unordered character of the data above, i. e. there is no time order between the cases in the table \tilde{M} .

Here we describe a principle allowing to reveal the system relationships of above mentioned type on the basis of the observation table \tilde{M} .

This algorithm determines inter- and intra-component relationships, which are as close as possible to relationships, which form matrix (2) in some sense.

Assume that relationships structure is given. In that case for initial state $(A_1(0), A_2(0), \dots, A_N(0))$ and for given sets $L_1(u, s), L_2(u, s), \dots, L_N(u, s), u \in \Omega, s \in \Omega$ the minor (2) can be calculated. Let P is to be the correlation matrix (Pearson or Spearman) between the rows of the minor (2) with entries $r_{i,j}$. Also, for the table \tilde{M} , the correlation matrix \tilde{P} (with entries $\rho_{i,j}$) of its rows can be calculated.

Introduce the measure of distance between the matrices P and \tilde{P}

$$D(P, \tilde{P}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N (r_{ij} - \rho_{ij})^2. \quad (5)$$

We state the task of minimization $D(P, \tilde{P})$ by all possible vectors of initial states $(A_1(0), A_2(0), \dots, A_N(0))$ and all allowable sets $L_j(s, u), s \in \Omega, u \in \Omega$ for all j

$$D(P, \tilde{P}) \mapsto \min \begin{array}{l} \text{(by all initial states \&} \\ \text{by all allowable sets } L_j(s, u)). \end{array} \quad (6)$$

The stated task means the search for such relationships between components, that the minor (2) is to be as close as possible to the table of observations regarding the measure (5).

The following theorem proved in [9] shows that this task is well-grounded in probabilistic sense.

Theorem 1. *If the table of observations \tilde{M} is obtained from the minor (2) by equiprobable choice of columns, then the Pearson correlation matrix of the observations table \tilde{P} converges to the correlation matrix of minor P (in probability)*

$$\lim_{B \rightarrow \infty} \rho_{ij} = r_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, N.$$

The same result takes place for the Spearman correlation matrix as well.

2.2 Additive stochastic model of between-component relationships

Our second model is also described by a set of components A_1, A_2, \dots, A_N taking discrete values $1, 2, \dots, K$.

But, in contrast to the first one, the second model introduces into consideration not only direction of relationships (in fact, for the first model we considered three direction — negative, neutral, and positive), but also a strength of relationships. Hence, relationships in this case can be recognized besides directions by the strength.

The structure of relationships between components A_1, A_2, \dots, A_N is described by the following relationships matrix

$$\mathcal{M} = \begin{pmatrix} m_{1,1} & m_{1,2} & \dots & m_{1,N} \\ m_{2,1} & m_{2,2} & \dots & m_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ m_{N,1} & m_{N,2} & \dots & m_{N,N} \end{pmatrix}.$$

Any entry $m_{i,j}$ reflects the strength and direction of influence of the component A_j upon the component A_i . The direction of influence is expressed by the sign of the value $m_{i,j}$ (may be $-, 0, +$) and the strength — by modulus of $m_{i,j}$ and varies from 0 to 1. So, $-1 \leq m_{i,j} \leq 1$ for each i, j . The influence of the component A_i on A_j is expressed by $m_{j,i}$. It is easy to see, that the relationship between the components A_i on A_j is described by the pair $(m_{i,j}, m_{j,i})$, which is close by implication to the relationship (ω_1, ω_2) introduced for the first model.

Now describe the dynamics of transition from the state of the system at the moment t to the state at the next moment $t + 1$. As for the weight functions' approach, we assume, that a set of functions $\psi_{i,j}(\cdot)$, ($i, j = 1, 2, \dots, N$) reflecting relationships between all pairs of components, including inner relationships, are given. The functions $\psi_{i,j}(\cdot)$ have the following properties: (i) $\psi_{i,j}(\cdot)$ are defined on the set \mathcal{X} ; (ii) $\psi_{i,j}(1) > 0$; (iii) $\psi_{i,j}(\cdot)$ are increasing functions on \mathcal{X} .

Also assume that a positive number δ playing the role of threshold, is given. Let the system is to be in the state $(A_1(t), A_2(t), \dots, A_N(t))$. For each pair of indices define the random variable $\xi_{i,j}$ as follows

$$\xi_{i,j} = \begin{cases} \text{sign}(m_{i,j}) \cdot \psi_{i,j}(A_j(t)) & \text{with probability } |m_{i,j}|, \\ 0 & \text{with probability } 1 - |m_{i,j}|. \end{cases}$$

Then we calculate the set of N random variables $d_i = \sum_{j=1}^N \xi_{i,j}$, $i = 1, 2, \dots, N$.

Using the set (d_1, d_2, \dots, d_N) , it's possible to calculate the set of probabilities (p_i^-, p_i^0, p_i^+) for each i according to the rule

$$p_i^- = P(d_i \geq \delta), p_i^0 = P(-\delta < d_i < \delta), p_i^+ = P(d_i \leq -\delta)$$

for each i from 1 to N .

This definition implies the equality $p_i^- + p_i^0 + p_i^+ = 1$. The transition from the state at the moment t to the next state at $t + 1$ is defied by following rule

$$A_i(t+1) = \begin{cases} \text{Dec}(A_i(t)) & \text{with probability } p_i^-, \\ A_i(t) & \text{with probability } p_i^0, \\ \text{Inc}(A_i(t)) & \text{with probability } p_i^+. \end{cases}$$

That is, at the moment $t + 1$ the value of A_i can increase by 1, remain the same or decrease by 1 with probabilities p_i^-, p_i^0, p_i^+ correspondingly. Applying this rule for each i , the probabilities of transition from any appropriate state $(A_1(t), A_2(t), \dots, A_N(t))$ can be calculated.

It can be shown, that if each row of the matrix \mathcal{M} include both negative and positive entries, we obtain the Markov chain with K^N states $A_1(t), A_2(t), \dots, A_N(t)$ ($A_i \in \mathcal{X}$). Besides, this chain is regular, so there a unique steady-state stochastic vector \mathbf{w} .

Now the reasons that state behind this model, can be explained. We assume, that a natural system is described by this model, and the probability of staying the system in states converges to the entries of the vector \mathbf{w} . Using the states A_1, A_2, \dots, A_N and the components of the steady-state vector \mathbf{w} we can calculate a weighted Pearson correlation matrix [11] between the components. Denote such the matrix by $R_{\mathbf{w}}$.

We suppose, that the true dynamics of our natural system is not visible, i. e. we cannot observe time series of states, but can record a state of the system at random moments of time. These observations are collected in the observation table \tilde{M} having N variables and B cases (after B observations) in similar way as for analogous table (4) of the first model. Let the Pearson correlation matrix between rows of (4) is denoted by \tilde{R} .

Theorem 2. If the observation table \tilde{M} is obtained according to the way described above, we have

$$\tilde{R} \rightarrow R_{\mathbf{w}} \text{ in probability when } B \rightarrow \infty.$$

This means component-wise convergence.

Proof. Omitted for short.

Introduce the measure of proximity for the matrices R and \tilde{R}

$$D(R_{\mathbf{w}}, \tilde{R}) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N (\tilde{R}_{i,j} - [R_{\mathbf{w}}]_{i,j})^2. \quad (7)$$

The result proved in the theorem 2 means that the sample observation matrix consistently represents a true dynamics, not observed straightforwardly. This result works as a base for identifications of entries of the relationships matrix \mathcal{M} . Therefore we can try to calculate transition probabilities of the Markov chain, that provide the best approximation of a true correlation matrix by a sample matrix in the sense of the measure (7). So, \mathcal{M} is obtained by resolving the following optimization task

$$D(R_{\mathbf{w}}, \tilde{R}) \mapsto \min \text{ by entries } m_{i,j}.$$

In fact, we find the relationships matrix \mathcal{M} , which makes the modelled correlation matrix as close as possible to the observe correlation matrix.

3 Case Studies

We present here three examples from different areas, where our models were applied.

First example concerns analysis of system factors affecting activity of social networks users, playing an important role in modern culture [12]. The structure of relationships between components of the system for two states of the Internet-forum on fantasy literature were calculated and compared. This comparison aimed to reveal system aspects of forum visiting in two periods. One state can be regarded as “low-performance”, other as “high-performance” according to number of written fanfictions (also abbreviated as fan fics, fanfics) of visitors at the site dedicated to the cycle of novels of Joanne Rowling about Harry Potter (snapetales.com). The period of first half of December 2010 is regarded as “high-performance”, the second period of the first half of December 2014 is called “low-performance”. For these two periods a statistically significant difference according to Student t -test ($p < 0.05$) in average number of visits per day was also detected.

The fanfictions were divided into 4 categories according to their length: fanfictions of small, large, and medium size; the last, fourth category includes fanfictions not related to the novels about Harry Potter.

The following values were taken as the components of the system reflecting the authors activity

- the number of fanfictions *of small size* per day related to the cycle of novels about Harry Potter (denoted by **MIN**);
- the number of similar fanfictions *of large size* per day (**MAX**);
- the number of similar fanfictions *of medium size* denoted by (**MID**);
- the number of fanfictions denoted by not related to the cycle of novels about Harry Potter, based on another literary works (**OTHER**).

For the “high-performance” and “low-performance” periods, the structure of relationships were built. We identified the models using the Pearson correlation matrix and the approach on the base of von Liebig law, with $K = 3$ levels of components values. The structure of relationships for both period is presented in Figs. 1 and 2. The notation on the graphs corresponds to the models and is quiet understandable: the components are presented as rectangular connected by ovals presenting considered relationships. For example, $\Lambda(\text{MIN}, \text{MID}) = (+, -)$, that is clearly shown on the graph.

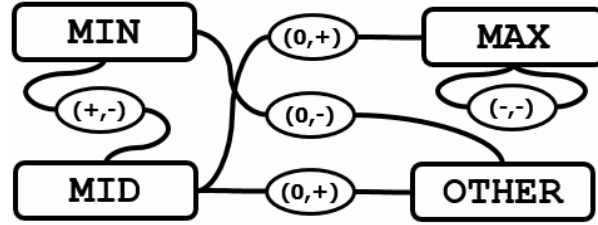


Fig. 1. The structure of relationships for “high-performance” period. Rounded rectangulars present the components of the system, the ovals include relationships between the components.

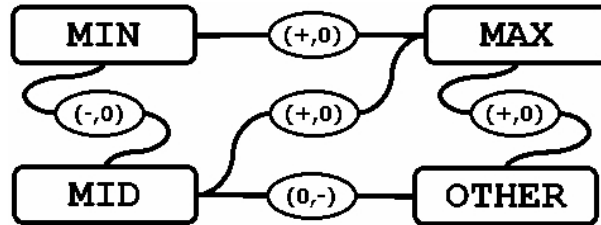


Fig. 2. The structure of relationships for “low-performance” period.

Comparing the graphs in Fig. 1 and Fig. 2 shows a system-forming role of the component **MID** for the “high-performance” period, in which **MID** positively affected other three components. This affect disappeared in the “low-performance” period together with loss a stabilizing mechanism through the relationship $(+, -)$ between **MID** and **MIN** supporting a dynamic equilibrium of the system.

These results are consistent with empirically established ideas about significant positive role of fanfictions of medium size (**MID**) in a functioning of social networks of this category and their close relation to short-sized fanfictions (**MIN**) representing a reaction of the most dynamic part of users. Differences in role of **OTHER** correspond to significance of “offtopic” as an index of deterioration in work of dedicated web-sites.

Our next example concerns the system relations of anthropometric parameters of adolescents suffering diseases of cardiovascular system [13].

Anthropometry is important in school medical, in particular, for determining the factors of predisposition of adolescents to cardiovascular disorders. At the same time, among other drawbacks of currently used anthropometric methods they often refer to insufficient use of systematic approach, among other things, in description of regularities in formation of body’s proportions in the individual development of adolescents.

Here we present a demo of application of DMDS for this purpose, calculated on the material of adolescents anthropometry with arterial hypertension and other forms of cardiovascular disorders. Body compositions related to overweight plays an important role in development of arterial hypertension. Taking that into account, the models for four following components were built: hip circumference, waist circumference, chest circumference, and shoulder breadth divided by height of a subject. The Spearman correlation and Liebig’s approach with $K = 3$ levels of components were used in modeling.

Comparison of these graphs has revealed a different role of such anthropometric parameters as the hip circumference for two group of adolescents under investigation. In the group with disorders different from arterial hypertension high values of hip circumference increase other three components. Simultaneously, shoulder breadth negatively affects hip circumference, that should form a proportion of male’s future body perceiving by subconscious as harmonious on the base on evolutionary history and recognized as such by modern physiology and medicine — the proportions of male “triangle” directed beneath by edge. The structure of relationships in the group with hypertension prevents the formation of such a standard and associates with the accumulation of a depot fat in

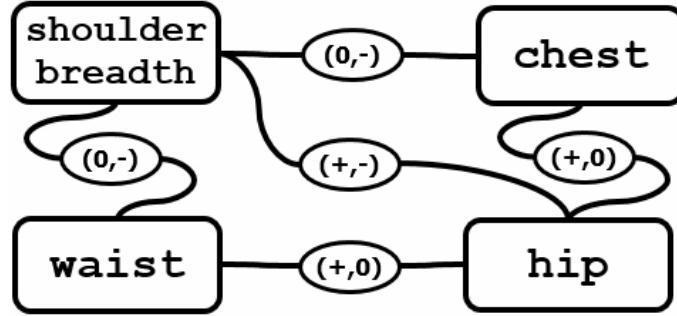


Fig. 3. The structure of relationships for adolescents without arterial hypertension.

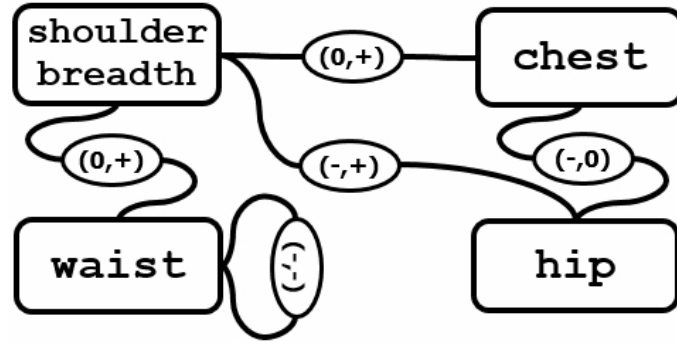


Fig. 4. The structure of relationships for adolescents with arterial hypertension.

certain parts of a human body: relatively high values of the hip circumference negatively affects shoulders breadth and chest circumference, not directly affecting waist circumference, on which shoulders breadth positively influences.

These results, regarded by authors as preliminary, do not contradict known facts about the impact of anthropometric parameters on the risk of development of hypertension in adolescents groups.

Our last example was taken from industrial fishery of Atlantic cod (*Gadus morhua*) at North Sea. The fishery of the cod plays important role in the economy of several countries and provokes considerable interest to use of mathematical models in industrial ichthyology describing large fluctuations of catching [14] (well-known example of this kind is collapse of the Atlantic northwest cod fishery in 1992).

As the demo the additive stochastic model of relationships structure between dimensional parameters of cod populations was considered. The average fish body length (\mathbf{L}), the difference between Upper Length Bound

and Lower Length Bound (\mathbf{vL}), the average stomachs weight (\mathbf{M}), and the average weight of preys of cod (\mathbf{dM}) were taken as components of the model. Additive stochastic models were built according to data of International Council for the Exploration of the Sea for two years (1984 and 1989) preceding to rapid changes of CPUE (the catch per unit effort). We used the model with $K = 4$ levels of components values. In the matrix

Table 1. Relationships matrices for 1984 and 1989

1984					1989				
	\mathbf{L}	\mathbf{vL}	\mathbf{M}	\mathbf{dM}		\mathbf{L}	\mathbf{vL}	\mathbf{M}	\mathbf{dM}
\mathbf{L}	-0.936	-0.379	0.893	0.581	\mathbf{L}	-0.843	0.137	0.953	0.059
\mathbf{vL}	0.325	-0.737	-0.405	0.519	\mathbf{vL}	0.963	-0.721	-0.997	0.494
\mathbf{M}	-0.184	-0.868	0.969	-0.081	\mathbf{M}	-0.882	-0.054	0.788	0.224
\mathbf{dM}	0.016	-0.941	0.999	-0.028	\mathbf{dM}	0.091	0.066	0.941	-0.995

corresponding to 1984, which precedes significant (till 1990) decrease of catching, there are large (above 0.85) negative effects of high values of \mathbf{vL} on \mathbf{M} and \mathbf{dM} . That is, increasing the diversity of dimensional characteristics of the cod population, that improves the consumption possibilities of forage reserve by the cod, leads to exhaustion of food resources (reducing the number of available preys) and deterioration of preys quality (reducing the average size of forage organisms), and results in deterioration of food supply of the cod, that lowers the values of \mathbf{M} and \mathbf{dM} .

In the matrix corresponding to 1989, which precedes sharp increase of CPUE, recorded a year later, in 1991, there exist extremely small (below 0.07) negative effects of high values of \mathbf{vL} on \mathbf{M} and \mathbf{dM} . In this case, the increasing diversity of sizes, that enhances abilities of consumption of forage reserve, does not lead to exhaustion and deterioration of the latter. This result of modeling explains differences described above in the dynamics of catching in accordance with modern concepts of industrial ichthyology. We also note the difference in the value of positive influence of \mathbf{L} on \mathbf{vL} : 0.325 and 0.963 for 1984 and 1989 correspondingly.

Presented results bring hope for the possibility of developing methods for a forecast of cod catching with use of the stochastic models of this class, built on the base of actual material on size structure of the population.

4 Conclusion

In the paper we followed the established framework in model development, appropriated for natural sciences. Typical approach in development, among others, comprises the data selection, specification of assumptions and simplifications, selection of a mathematical modeling framework, estimation of parameter values, model diagnostics, model validation, model refinements and model application. It's clear, that all these stages of building mathematical models for biological systems are too complicated, but the most difficult task among them is the model parameters estimation for identifying structure in the underlying biological networks.

The models presented in the paper are created for description of biological and ecological systems, based on pairwise relationships characterized by the direction (positive, negative, or neutral) for both models and by the strength varied from 0 to 1 in the stochastic model only.

The task of parameter estimation is a true challenging problem for both models and requires development of special algorithms of numerical optimization. For example, if the system has N components and the number of levels is to be assumed K , for the first deterministic model the number of initial states is equal to K^N and the number of possible relationships' structures is equal to 3^{N^2} . For solving the stated optimization problem (6), one should built the minor (2) with use of an initial state and a relationships' structure, calculate correlation matrix P and calculate the distance (5). So, the exhaustive search of both initial states and relationships' structures jointly gives us $K^N 3^{N^2}$ variants, that is a huge number for even moderate N and K .

The case studies presented in the paper, considered by the authors as preliminary and illustrating, offer the prospects of applications of proposed models.

The results of modeling of system aspects in anthropometry of adolescents present the approaches to use of this simple and cheap method for identifying the risk groups of the progress of arterial hypertension. These approaches may be applied in school medicine an, if necessary, in extreme situations for mass screening as well.

The investigation of system factors of functioning of web-site dedicated to fiction about characters from original works about Harry Potter, due to use of components of the system, that are invariant to the content of the web-site, may have a broader meaning in analysis of the social networks performance.

The model of the cod population as a whole does not contradict known facts on the role of fish size and state of a forage reserve in the population dynamics. At the same time, these results reveal some promises and can be used in the development of approximate methods for prediction of populations of commercial fish with use of relatively simple and inexpensive methods of data acquisition, including even the commercial reports concerning the assortment of fish products.

References

1. Chen, M., Mao, S., Liu, Y.: Big data: A survey. *Mobile Networks and Applications*. Vol. 19, 171–209 (2014)
2. Margalef R.: *Perspectives in ecological theory*. Chicago University Press (1968)
3. Abrams, P. A.: On classifying interactions between populations. *Oecologia*. 2, vol. 73, 272–281 (1987)
4. Lidicker, W. Z. Jr.: A Clarification of Interactions in Ecological Systems. *Bio-Science*. 8, vol. 29, 475–477 (1979)
5. Wootton, J. T., Emmerson, M.: Measurement of Interaction Strength in Nature. *Annual Review of Ecology, Evolution, and Systematics*. Vol. 36, 419–444 (2005)
6. Shipley B.: *Cause and Correlation in Biology*. Cambridge University Press (2004)
7. Pearl, J.: *Causality: models, reasoning, and inference*. Cambridge University Press (2009)
8. Spirtes, P., Glymour, C. N., Scheines, R.: *Causation, Prediction, and Search*. MIT Press (2000)
9. Zholtkevych, G. N., Bespalov, G. Y., Nosov, K. V., Abhishek, M.: Discrete Modeling of Dynamics of Zooplankton Community at the Different Stages of an Anthropogeneous Eutrophication. *Acta Biotheoretica*. 4, vol. 61, 449–465 (2013)
10. Kalman, R. E., Falb, P. L., Arbib, M. A.: *Topics in mathematical system theory*. McGraw-Hill (1969)
11. Blalock, H. M.: *Social statistics*. McGraw-Hill New York (1972)
12. Boyd, D., Ellison, N.: Social Network Sites: Definition, History, and Scholarsh. *J. Computer-Mediated Comm.* 1, vol. 13, 210–230 (2007)
13. Cole, T. J., Bellizzi, M. C., Flegal, K. M., Dietz, W. H.: Establishing a standard definition for child overweight and obesity worldwide: International Survey. *BMJ Clinical Research*. Vol. 320, 1–6 (2000)
14. ICES: Benchmark Workshop on Arctic Stocks (WKARCT), 26–30 January 2015, ICES Headquarters, Denmark. Report of the ICES (2015)

Information Support of the Computer-aided Fixture Design System

Vitalii Ivanov¹, Svitlana Vashchenko¹, and Yiming (Kevin) Rong²

¹ Sumy State University

ivanov@tmvi.sumdu.edu.ua, sveta@opm.sumdu.edu.ua

² South University of Science and Technology of China

rongym@sustc.edu.cn

Abstract. The article contains the rationale for the development of the computer-aided fixture design system (CAFD). The place of the CAFD among other automated systems is considered and the data flows are found between them. The structure of the CAFD system is suggested, which is distinguished by the module of optimization and allows to select the optimal fixture for the given production conditions. The purpose and place of each module in the structure of the CAFD system is grounded. The database of the CAFD system is a collection of libraries connected by logical connections. The database includes 15 libraries of engineering and design, general engineering, regulatory, reference, and optimization and calculation character. The classifications for the supporting, locating and clamping elements, locating charts and the clamping charts of the workpieces are developed.

Keywords: Computer-aided fixture design, production planning, optimal fixture, module of database management

KeyTerms: Development, SoftwareEngineeringProcess, Integration

1 Introduction

In modern manufacturing engineering a major challenge is the contradiction between the need to reduce the time required to design and manufacture the products and the increasing complexity of product design. Over the past 15 years the product range has increased more than 2 times, constantly increasing their complexity, the demands for the accuracy and product quality are increasing [1,2]. Today's market requires more varieties of products, and consequently the equipment and processes should be more flexible to meet the needs of the market and reduce the amount of time to appear on the market. This necessitates the development and implementation of the innovative production solutions for the implementation of the processes which should be aimed at the intensification and automation of production. Integrated CAD / CAE / CAPP / CAM technologies provide a solution to problems of design, analysis and optimization, development of the technology of manufacturing products, production process automation and inspection. This can increase the productivity and quality of the developed production solutions.

One of the promising areas of the intensification of the manufacturing production planning is the development and implementation of the computer-aided fixture design (CAFD) technologies which allow, in the automatic mode, to design fixtures, evaluate their effectiveness and develop the necessary design and production documentation. These systems can be integrated with CAD / CAE / CAPP / CAM systems, and thus perform a full cycle of design, analysis, synthesis and manufacture of fixtures.

Fixtures are an integral part of a closed loop technological system "machine tool – fixture – workpiece – cutting tool" and are designed for accurate locating and reliable clamping of the workpieces during the machining on the metal-cutting machine tools. Fixtures affect the production of the competitive manufacturing engineering products, as evidenced by the following data: fixtures constitute 70–80% of the total tooling [3]; the proportion of fixtures is 10–20% of the total manufacturing system cost [4]; 80–90% of the time required for the production planning correspond to the design and manufacturing of fixtures [3]; 40% of rejected parts are due to dimensioning errors that are attributed to poor fixture design [5]; 70% of the new fixtures are a modification of the existing [6].

In today's manufacturing engineering, which is characterized by the instability of the range and volume of manufacturing capacity, a rational choice of the fixtures is actual, which put forward the following demands: flexibility, sufficient for machining of the parts within the machine tool specification; assurance of the specified accuracy of the machining; mechanized or automated changeovers when switching to a machining of the parts of different size; high stiffness of the parts and assembly units, which can take considerable cutting forces and ensure maximum use of the equipment power; tool availability to handle the maximum number of surfaces in one setup; high unification of the parts and assembly units, providing cost reduction of fixtures; high functional and production reliability of a fixture and its elements; economy [7]. The basic requirements for the creation of the fixtures are reduced to six groups, including the physical requirements, tolerance, constraining requirements, affordability, collision prevention, usability [8].

Therefore the main objective of this article is to present the conceptual structure of the CAFD system and to develop the information support for the entire fixture design process including analysis, synthesis, optimization, and verification procedures in order to ensure the effectiveness of production planning.

2 Related Work

Today people have many years of experience of developing and implementing of the CAFD systems which considerably accelerated and improved the process of fixture design, allowing the designers, at the design stage, comprehensively analyze the properties of the future fixture before its manufacture.

According to their purpose the CAFD are divided into systems designed for dedicated [9], modular [6], [10,11], and adjustable fixtures [9], [12]. CAFD systems for dedicated fixtures are meant for detail design of all fixture functional elements with the following assembling to meet the requirements of the specific sizes of the

workpieces. CAFD systems for modular fixtures are meant for fixture design process by means of assembling of fixtures from the set of prefabricated functional elements. The use of adjustable fixtures supposes locating and clamping the workpieces in specific range of sizes. Therefore such CAFD systems provide manufacturing engineers with design and assembly tools for fixture design from original and standardized fixture elements. A thorough review of the existing CAT and CAFD systems is considered in works [4], [8,9], [13,14,15,16].

According to the level of automation the CAFD systems are traditionally classified into interactive (I-CAFD), semi-automated (Semi-AFD), and automated (AFCD) [12]. The analysis of the existing CAFD systems showed that the typical structure [16] consists of four modules which provide a step-by-step solution of setup planning, fixture planning, fixture unit design, and verification of the fixture (Fig. 1). Verification is the essential stage in fixture design allows evaluation the designed fixture to draft proposal and production conditions.

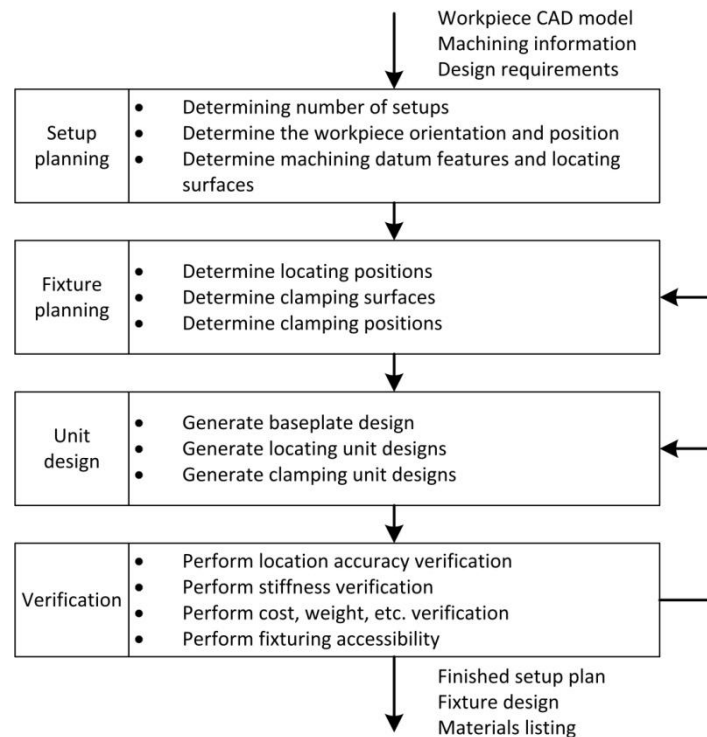


Fig. 1. The typical structure of the CAFD system (adapted from [16])

Today there is the problem of choosing the optimal fixtures. A variety of fixtures which are used for the machining of similar parts, on the one hand, provides many variants, on the other – significantly complicates the task of determining the optimal fixture for specific production conditions. Existing CAFD systems doesn't allow the execution of optimization procedures according to production conditions.

Taking into account that the fixtures are complex technical systems that differ according to different parameters (accuracy, flexibility, cost, reliability, performance, steel intensity, etc.) the effective choice of the best among the set of competitive variants is advisable to carry out by means of multicriteria optimization. Considering the complexity of the analysis and calculation of all the parameters necessary for the design of the fixtures, it is advisable to automate the process.

3 The Structure of the CAFD

3.1 The Place of the CAFD Among Other Automated systems

CAFD is one of the elements of computer integrated manufacturing, along with CAD / CAE / CAPP / CAM systems. This approach allows you to implement the fixture life cycle from design to its manufacturing, actualizing the geometric modeling and engineering analysis, process planning and automation of the manufacturing. The analysis of the data flows (Fig. 2) confirms that the CAFD structure must be built taking into account the integration of these systems.

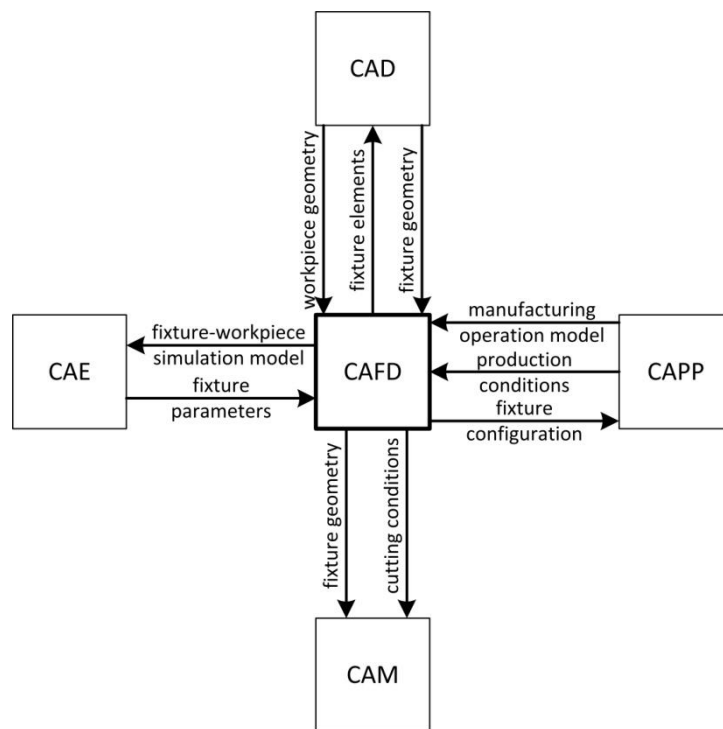


Fig. 2. The place of the CAFD among other automated systems

3.2 CAFD Structure

3.2.1 General Information

The suggested structure of CAFD system (Fig. 3) is the five modules that operate according to the developed algorithm in stages. The information support of the designing stages is provided by the database. The implementation of the mathematical models of all the necessary calculations is performed in the CAFD working modules which include module of the database management. The provision of the visualization of the results of the CAFD design involves integration with the three-dimensional solid design package Autodesk Inventor that allows working with solid state 3D models of functional elements of the fixtures which are contained in the relevant libraries.

3.2.2 Module of Input Data

Input data when designing are: the design parameters of a workpiece (geometry, overall dimensions, weight, material, workpiece stiffness, etc.); technological parameters (type of manufacturing operation, metal-cutting equipment, cutting tools, etc.); manufacturing parameters (type of manufacturing, batch size, etc.); other factors (economic, ergonomic, aesthetic, etc.).

3.2.3 Module of the Manufacturing Analysis of the Workpiece

The analysis of a 3D model allows determining the functional surfaces of the workpiece, which can be: work; locating; clamping. The work surfaces (one or more) – are surfaces that require machining on a given manufacturing operation. Locating surfaces (usually a combination of surfaces) – are surfaces that determine the position of the workpiece in three-dimensional space, constraining the workpiece a certain number of degrees of freedom (maximum six – three movements, three rotations). The clamping surfaces (one or more) – are designed for applying certain clamping force to ensure the permanent position of the workpiece during the process of forming surfaces.

The analysis of the above mentioned surfaces allows to determine from 3D model of the part: geometry (shape) of the surfaces; spatial relationship of the surfaces relatively to other surfaces; dimensional characteristics of the surfaces (length, width, height, diameter). From the drawing (or other accompanying documentation) it is necessary to get information: tolerance range for each of the surfaces; surface roughness; material; hardness of the workpiece material; type of heat treatment; other technical requirements.

Having information about the machining surfaces, the plan of the machining is formed. At this stage we distinguish: a schematic diagram of the surfaces machining; cutting conditions (cutting depth, feed, cutting speed); cutting force; required power of the machine tool.

The set of locating surfaces implement a theoretical locating chart with the definition of the points of contact of the workpiece with the locating elements of the fixtures.

The analysis of surfaces for clamps allows to select one of the typical clamping chart, as well as the points of contact of the workpiece with the clamping elements.

Based on the plan of machining, locating and clamping charts of the workpiece the setup and machining plan is formed, indicating the values and the direction of the cutting and clamping forces in a particular coordinate system.

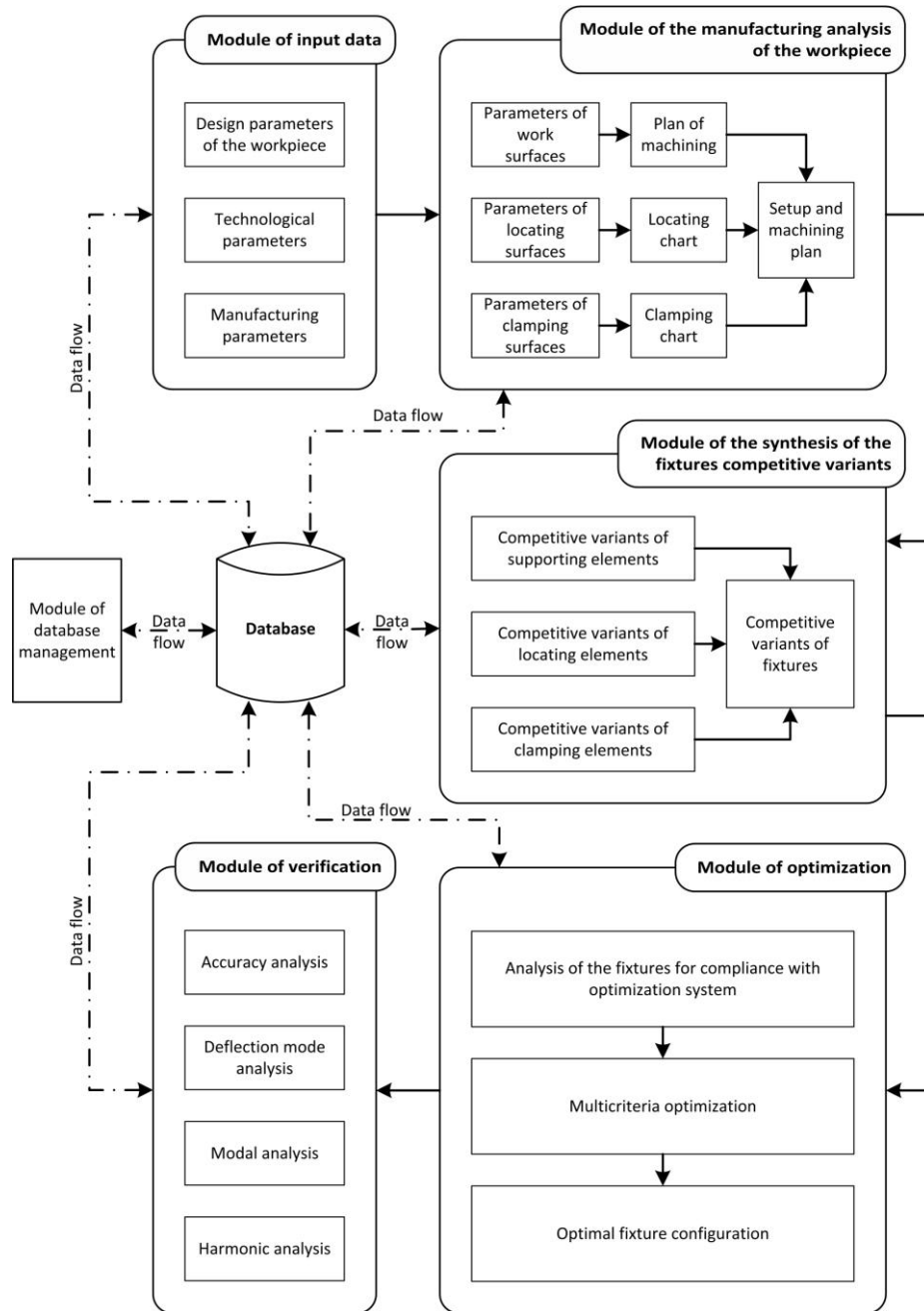


Fig. 3. The proposed CAFD system structure

3.2.4 Module of the Synthesis of the Fixtures Competitive Variants of the Fixtures

Based on the developed rules, including the selected locating chart and a clamping chart of the workpiece the locating and clamping elements are selected which correspond to the specified parameters. According to the overall dimensions of the workpiece, the selected locating and clamping charts the choice of the supporting elements is carried out. This allows to form a set of competing options of the supporting, locating and clamping elements. Then on their basis and taking into account the previously created fixtures contained in the library of the fixtures, the competitive variants of the fixtures are formed according to the developed algorithm [17].

3.2.5 Module of Optimization

The implementation of the module of optimization allows to choose the optimal fixture for the specific manufacturing conditions on the basis of the multicriteria optimization. The first step in this module is the choice of the fixtures that satisfy the system of technical constraints, based on a mathematical model and also reference data. For the fixtures that remained due to the imposition of technical constraints, the calculation of the parameters according to each criterion of optimality is performed. Then a multicriteria optimization according to the method of successive concessions is carried out [18]. As a result of the optimization an optimal fixture is found with identified criteria evaluations.

3.2.6 Module of Verification

The engineering analysis is a very important step because it allows you to control the basic parameters of the fixture. At this stage, the possibility to perform calculations on the accuracy of the workpiece machining, the research of the system "fixture – workpiece" on the deflected mode analysis, modal analysis, harmonic analysis is provided.

3.2.7 Output Data

The results of the CAFD system are: 3D-model of the fixture and specification of the elements, that are part of it; the list of the optimality criteria and calculated criteria evaluations for the selected fixture; the results of the engineering calculations; the information about the conditions of production.

4 Database

4.1 General Information

An important condition for the effective functioning of any CAFD system is the database that provides information support of the design process of the fixture, the accumulation and the storage of reference, design and technological and methodological information. The developed database is a collection of individual libraries-tables, in which the information necessary for the performance of the design is systematized according to the purpose. The detail of a conceptual database scheme is shown in Fig. 4. Between the respective tables and CAFD modules of the fixture the connections are identified (Fig. 5) that implements data flows which ensure the design process.

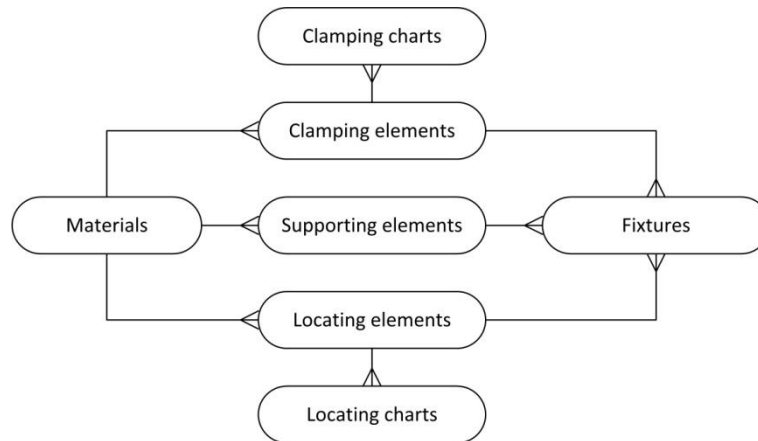


Fig. 4. Conceptual scheme of the data presentation (fragment)

4.2 Libraries

In general, the system provides 15 libraries used for full information support of the design process of the fixture (Fig. 5).

The library of the design parameters of the parts contains information about structural elements of the parts, their variety, standard sizes and so on.

The library of the technological information includes the list of data about the types of workpieces and the rational conditions of their use, types of the heat treatment and the typical processes of its implementation, methods of machining, classification of the manufacturing operations and steps, the typical structures of the manufacturing operations, the recommended parameters of the size accuracy and the quality of the surfaces.

The library of the manufacturing parameters of the machining contains information about the type of manufacturing, manufacturing capacity, batch size, level of flexibility, and the level of machining efficiency.

The library of the metal-cutting equipment contains information about the equipment for the drilling-milling-boring manufacturing operations. All the elements of the given library are systematized according to various parameters: the level of automation; the type of the layout of a machine tool; manufacturing capabilities; workspace overall dimensions; the type and a set of a tool magazine; the frequency of the spindle rotation; the power of the force actuator and so on. The library contains information about 1075 machine tools of vertical and horizontal layouts of more than 30 world manufacturers.

The library of the cutting tools contains technical information and manufacturers' recommendations about the cutting mode for drills, core drills, reamers, boring heads, spotfacers, chamfer bits, taps and also face-milling, end-milling, keyway-milling, plain-milling and disk cutters that are used for drilling-milling-boring manufacturing operations.

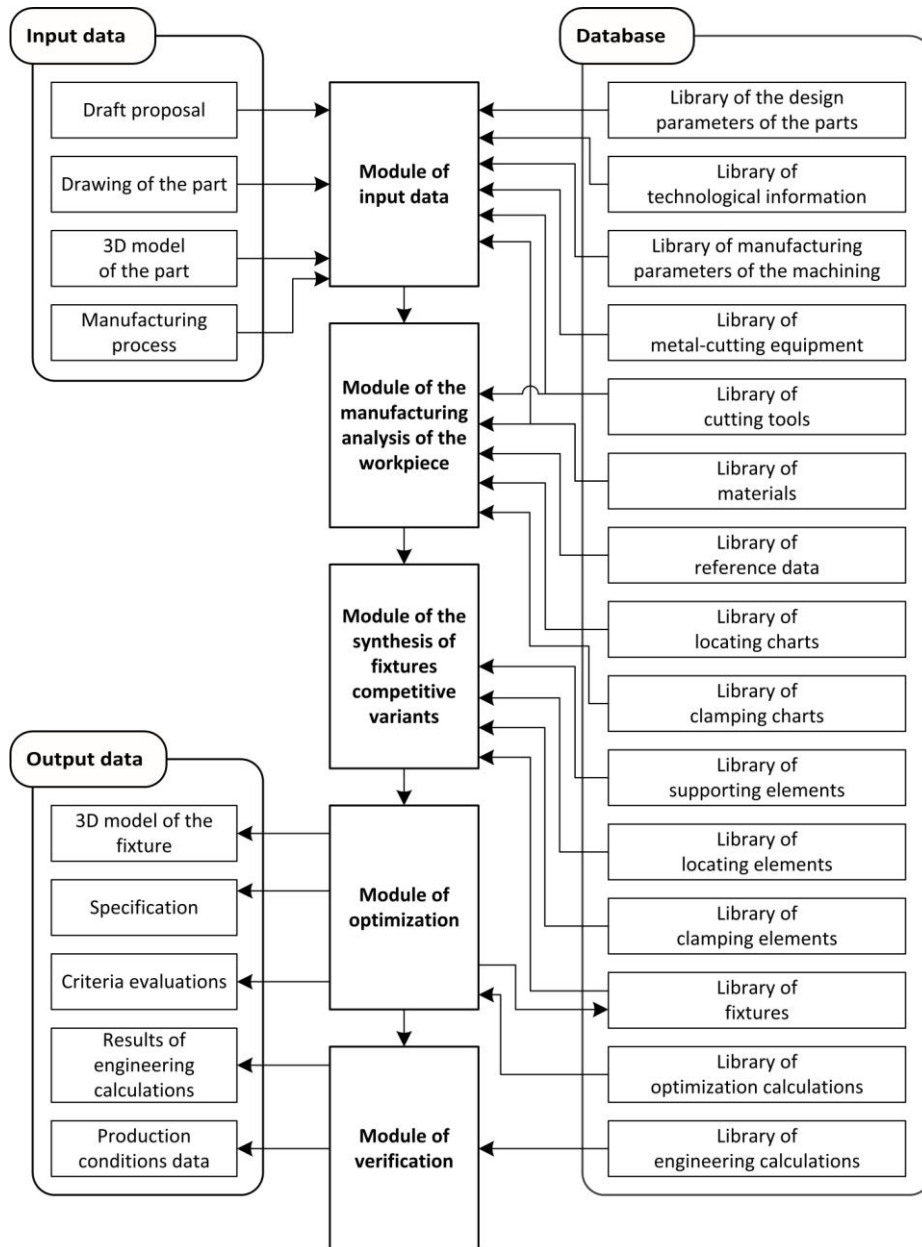


Fig. 5. The scheme of the interaction of the libraries of the database with the modules of the computer-aided fixtures design system

The library of the materials contains information about the physical, mechanical and technological properties, chemical composition of the manufacturable materials. In order to maximize the coverage of a large number of possible options for the produc-

tion of the machining workpieces, the library contains data on ferrous metals and alloys; nonferrous metals and alloys; non-metals.

The library of reference data includes information about tolerances and fits, quality class, tolerances of the form and position, processing errors, roughness of the surfaces, information on the cutting modes of the materials, standard time for machining of the surfaces, standard time for the fixture assembling.

The library of the locating charts of the workpieces contains classification of the typical locating charts (Fig. 6) of the prismatic parts, rotational parts (shafts, shaft-collars, flanges etc.), flat parts, and complex parts (levers, connecting rods, brackets, cantilevers, etc.). According to the library information it is possible to provide a defined locating of 90% of the workpieces that are machined on the machine tools of the drilling-milling-boring group.

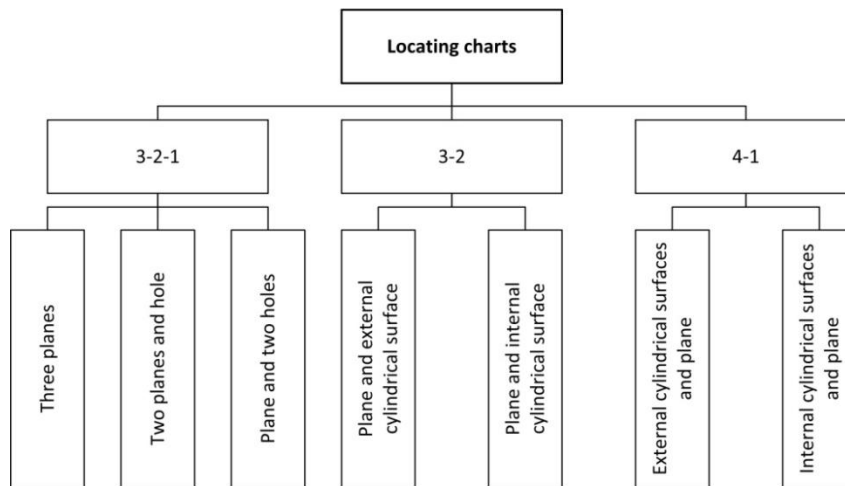


Fig. 6. Classification of the locating charts of the workpieces

The library of the clamping charts contains information about the most rational ways of workpieces clamping, based on the principles of the equilibrium position of the workpiece, and also take into account the sequence of the clamping of the workpieces in the charts with many clamping elements.

The library of the supporting elements includes a wide range of supporting elements (plates, cubes, angular plates, tombstones, etc.) that are the basis for creating of the fixtures.

The library of the locating elements contains a variety of locating elements (support plates, supports, V-blocks, locating pins), intended to implement theoretical locating charts.

The library of the clamping elements has a large selection of clamping elements (swing clamps, cam clamps, lever clamps, clamping straps, etc.) to ensure reliable clamping of the workpiece in the fixture during the cutting process.

The libraries of the supporting, locating and clamping elements provide basic technical characteristics of each element, including code, dimensions, mounting dimen-

sions, weight, material, size of the work surfaces, adjustment range, clamping force, etc., and links to related 3D models. Also the practical advice is given on the effective scope for different production conditions.

For the convenience of the identification of the functional elements in the proposed CAFD system a coding system is developed. These codes are unique for each elements. Such codes help the designer, looking only at the code, to make conclusions about the element. According to the developed classifier it is advisable to code all the functional elements with the help of the code from letters and numbers (Fig. 7), where the first letter indicates the purpose of the functional element (supporting – S; locating – L; clamping – C), the second sign – sort of the element, the third – type, the fourth – standard size (Fig. 8).

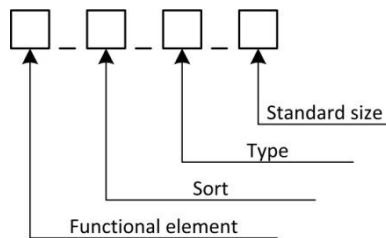


Fig. 7. Structural formula of the functional elements of the fixtures

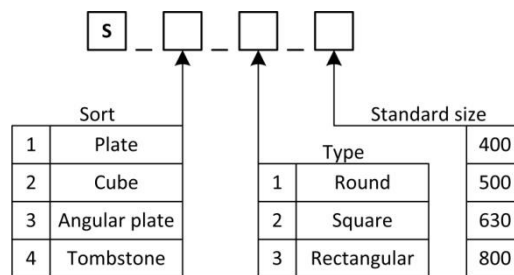


Fig. 8. Detailing of the structural formula according to the example of the supporting elements

The library of the fixtures is designed for the storage of the information about finished structural solutions which can be implemented with the help of the libraries of the supporting, locating and clamping elements for the typical parts of manufacturing engineering. In fact, the library is an archive of the finished project solutions obtained in CAFD before, with the definition of the object of machining, production conditions and technical specifications of the fixture.

The library of the optimization calculations is developed for the information support when choosing the optimal fixture for the given production conditions and provides a list of optimality criteria with the objective functions and technical constraints.

The library of engineering calculations contains information about reference data, calculation models, calculation templates that help to investigate the fixtures.

4.3 Working With the Libraries

The main functions of the developed database are the accumulation, updating, storage and provision of information to implement the design stages. A physical model of the proposed database is implemented with the help of the database management system MySQL. Working with database within the frames of the proposed CAFD is conducted both at the program level (provision with the information on the needs from the calculation modules), and at the level of user interface (editing data in the tables). The latter is implemented using a separate module of the database management, which is a part of the program complex of the CAFD system (Fig. 9).

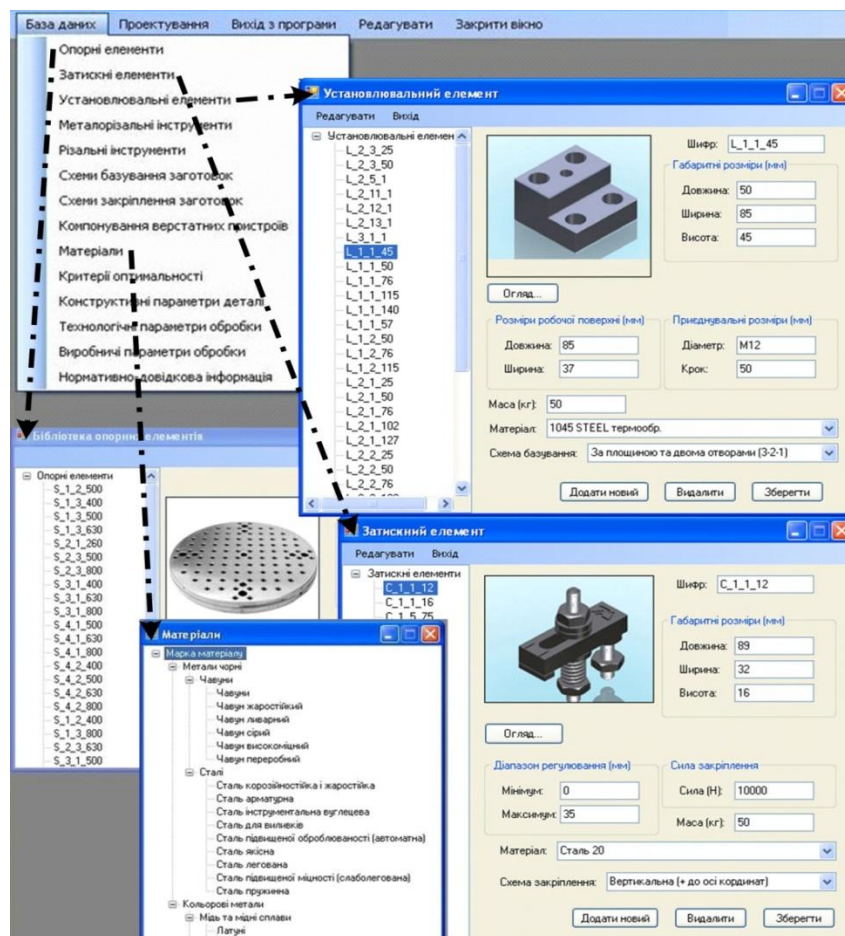


Fig. 9. The module of database management

The main module of the entire complex is made according to the MDI-interface technology that provides the ability to open several documents. The call of commands that allow the user to work with the information is through the main menu. The windows

for the work with the individual libraries are implemented in the form of the subordinated forms that are created within the container – the main form. The user interface provides access to the database online.

Each command of the submenu provides access to the screen forms, which provide the users of navigation means in the database, view and editing of the available information.

When designing the user interface of the screen forms of the control module of the database, standard components for Windows are used, that provides an intuitive interface.

5 Conclusions

According to the results of the study the necessity of developing CAFD with optimization procedures is grounded, the use of which will shorten the term for the production planning.

The structure of the developed CAFD includes not only conventional design stages, which are implemented as individual modules. The key difference of the proposed structure is the module of optimization, the implementation of which will generate the most advantageous variants for the fixtures.

The information support of the design process is performed by the developed database, the model of which covers the whole range of relevant information: of production, general engineering, regulatory, reference, and optimization and calculation character. According to the results of the analysis the connections of the calculation modules are established with the tables of the database.

The further work is directed to replenish the information storage of the databases, rank and improve data flows and logical relationships between them in the functioning of CAFD, the development of algorithms and methods of the automated synthesis of the fixtures and their multicriteria optimization that would ensure effective interaction with the computer-aided systems of the design and production orientation.

References

1. The Future of Manufacturing: Bringing the real and virtual world together, <http://w3.siemens.com/topics/global/en/industry/future-of-manufacturing/Documents/feature-infografik/all/en/index.html?stc=wwcg102138#/conclusions/469>.
2. Bi, Z.M., Lang, S.Y.T., Verner, M., Orban, P.: Development of Reconfigurable Machines. *Int J. Adv Manuf Techol*, 39, 1227–1251 (2008).
3. Karpus, V.E., Ivanov, V.O., Kotliar, O.V., Minenko, D.O., Ivanova, M.S.: Intensification of Manufacturing Processes. Sumy State University, Sumy (2012) [in Ukrainian].
4. Bi, Z.M., Zhang, W.J.: Flexible Fixture Design and Automation: Review, Issues and Future Directions. *Int. J. Prod Res*, 39, 2867–2894 (2001).
5. Nixon, F.: *Managing to Achieve Quality and Reliability*. McGraw Hill, Maidenhead (1971).
6. Rong, Y., Zhu, Y.: *Computer-Aided Fixture Design*. Marcel Dekker, New York (1999).

7. Karpus, V.E., Ivanov, V.A.: Universal-Composite Adjustable Machine-Tool Attachments. *J. Rus Eng Res*, 28, 1077–1083 (2008).
8. Boyle, I., Rong, Y., Brown, D.: A Review and Analysis of Current Computer-Aided Fixture Design Approaches. *Int. J. Robotics and Computer-Integrated Manufacturing*, 27, 1–12 (2011).
9. Rong, Y., Huang, S.H., Hou, Z.: *Advanced Computer-aided Fixture Design*. Elsevier Academic Press, New York (2005).
10. Nee, A.Y.C., Senthil Kumar, A., Tao, Z.J.: *An Advance Treatise on Fixture Design and Planning*. World Scientific, Singapore (2004).
11. Rong, Y.: Chapter 4. Computer-Aided Modular Fixture Design. In: Leondes, C.T. (ed.) *Computer Aided and Integrated Manufacturing Systems. Vol. 4: Computer-Aided Design / Computer-Aided Manufacturing (CAD/CAM)*, pp. 103–170. World Scientific, Singapore (2003).
12. Nee, A.Y.C., Whybrew, K., Senthil Kumar, A.: *Advanced Fixture Design for FMS*. Springer-Verlag, London (1995).
13. Hargrove, S.K., Kusiak, A.: Computer-aided Fixture Design: A Review. *Int. J. Prod Res*, 32, 733–753 (1994).
14. Zholtkevych, G.: *Computer-aided Tooling: Theory and Practice*. Tehnica, Kyiv (1998) [in Ukrainian].
15. Kang, X., Peng, Q.: Recent Research on Computer-Aided Fixture Planning. *J. Recent Patents on Mechanical Engineering*, 2, 8–18 (2009).
16. Wang, H., Rong, Y., Li, H., Price, S.: Computer-Aided Fixture Design: Recent Research and Trends. *J. Computer-Aided Design*, 42 (2010), 1085–1094.
17. Karpus, V.E., Ivanov, V.A.: Choice of Optimal Construction of Modular Reusable Fixtures. *J. Rus Eng Res*, 32, 213–219 (2012).
18. Karpus, V.E., Kotliar, O.V., Ivanov, V.O. Optimization of Manufacturing Processes of Rotational Parts, NTMT, Kharkiv (2012) [in Ukrainian].

Information Technology of Predicting the Characteristics and Evaluating the Success of Software Projects Implementation

Andriy Krsiy¹, Tetiana Hovorushchenko²

¹ Khmeltsky National University, Khmeltsky, Ukraine
andriy-krsiy@yandex.ua

² Khmeltsky National University, Khmeltsky, Ukraine
tat_yana@ukr.net

Abstract. The aim of this research is the development of the information technology (model, method and tools) of predicting the characteristics and evaluating the success of the software projects implementation based on the analysis of the software requirements specification (ITPCES). ITPCES structure was first time proposed. One of the non-realized components of ITPCES is the intelligent system of predicting the characteristics and evaluating the success of the software projects implementation (SPCES). This research is devoted to design of SPCES and experiments with it. SPCES gives the conclusion about probably category of success of the software project implementation by analyzing the software requirements specification (at the early stages of the life cycle). The SPCES conclusions provide to the customer the ability of the comparing the proposed software projects and provide to the customer the data for the grounded and informed choice of the most successful software project.

Keywords: software requirements specification (SRS), SRS indicators, software project characteristics, success of project implementation, the category of success of project implementation.

Key Terms: Model-Based Software System Development, Software Component, Software System, Specification Process.

1 Introduction

In recent years, the software industry has reached a level of evolution where the development of software systems is user-oriented [1]. At the present definition of software quality [2], if the goals of the project don't meet the needs of users, the software will not be qualitative and successful, even if the modern technologies and the most qualified developers were involved to its development. But until now the development of successful and high-quality software products don't become the norm - statistics [3] says that in 2012 only 39% of software projects are successful, but 43%

of software projects are challenged and 18% of software projects are failed, i.e. 61% software projects aren't successful and qualitative - Fig. 1.

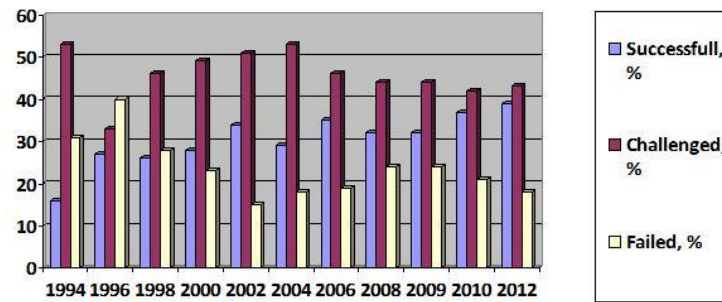


Fig. 1. Statistics of success of software projects implementation in 1994-2012

In [4-6] the fact, that almost all the causes of software incidents and accidents are latent in the software requirements specification (SRS), is confirmed. The vast majority of software accidents arises from false requirements but not from coding bugs. Software versions, written by the different developers for the same requirements, contain the number of the common bugs associated with errors or inaccuracies of requirements (SRS) [5]. Paper [7] says, that the main causes of the failure of software projects are the misconceptions of project managers on real deadline and budget for providing the user functional requirements. So paper [7] again confirmed that the most of the software problems are associated with SRS. Then the quality and success of the software project implementation depend on the SRS, resulting in the need to deepen the analysis of specifications.

Then *the actual task* is the ability to evaluate the potential success of software project implementation based on the software project characteristics (project cost, duration, complexity, usability, cross-platform, quality), the predicted values of which can be obtained by analyzing the SRS indicators. *The success of software project implementation* is timely execution of software project within the allocated budget and with realization of all necessary features and functionality.

The analysis of SRS structure [8, 9] showed, that the SRS requirements provide the set of indicators, on the basis of which the customer and the developer can get the predicted quantitative values of the characteristics of software projects. For establishment of the dependence between the basic software project characteristics and SRS indicators the software requirements specifications and finished applications (realized by these SRS) were analyzed. For this analysis six types of software projects (Web-applications, mobile applications, e-learning applications, applications for statistics and accounting, automated systems, information systems) were considered. For each type of software projects 30-50 tasks of different complexity were studied. For each task 1-3 SRS (proposed by the various developers) and 1-3 finished applications (written by the analyzed specifications) were selected. For this the course projects in discipline "Technology of software systems design", the diploma papers, the projects of students scientific group «SOFTWARE» of Khmelnytsky National

University (15% of all projects; moreover only student projects, that were devoted to solving the real-world tasks and were successfully applied in different industries, was considered), the SRS and applications of the software companies of Khmel'nitsky ("Avivi", "Smile", LLC «STU Electronics») were studied. Thus, we selected 200 tasks, for which 410 SRS and appropriately 410 applications were developed (for various industries, i.e. software projects of different types were selected) and we analyzed them: what SRS indicators differed in the selected specifications, what characteristics of finished applications changed depending on it, and what values had the SRS indicators in these SRS. The conducted analysis of finished SRS and applications led to the conclusion about dependence the basic software project characteristics on the SRS indicators for all types of software projects [10].

The analysis of the methods and tools for determination of the software projects characteristics [11, 12] led to the conclusion that they are focused on ready code, but not on existing SRS that is unusable at the early stages of the software projects life cycle. The research of the methods and tools of the SRS analysis [13-15] showed that they are aimed at monitoring the implementation of requirements and don't determine the predicted values of software projects characteristics. Thus, the existing methods and tools of SRS analysis and software project characteristics determination are not acceptable for the quantitative evaluation of the software project characteristics based on only requirements analysis and for evaluating the success of the software projects implementation.

The *task of this research* is the development of the information technology (model, method and tools) of predicting the characteristics and evaluating the success of software projects implementation based on analysis of the SRS.

2 Information Technology of Predicting the Characteristics and Evaluating the Success of Software Projects Implementation (ITPCES)

Considering the definition of information technology [16, 17], the structure of the information technology of predicting the characteristics and evaluating the success of software projects implementation is represented on Fig. 2:

Information Technology of Predicting the Characteristics and Evaluating the Success of Software Projects Implementation (ITPCES)				
Object – the success of software projects implementation	Goal – predicting the characteristics and evaluating the success of software projects implementation	Model –neuronet model of predicting the software projects characteristics based on the SRS analysis	Method – method of evaluating the success of software projects implementation based on analysis of SRS (ME SSPI)	Tools – intelligent system of predicting the characteristics and evaluating the success of software projects implementation (SPCES)

Fig. 2. The structure of information technology of predicting the characteristics and evaluating the success of software projects implementation

Fig. 2 shows that the basis of ITPCES is the developed neuronet model of predicting the software projects characteristics based on the SRS analysis and method of evaluating the success of software projects implementation based on analysis of SRS (MESSPI) and also intelligent system of predicting the characteristics and evaluating the success of software projects implementation, which should be design.

The neuronet model of predicting the software projects characteristics based on the SRS analysis was developed for evaluation the software projects characteristics based on the processing of SRS indicators [11]. The basis of this model is the artificial neural network (ANN), which performs the approximation of SRS indicators and provides the predicted relative evaluation of the characteristics of the software, which will be developed by the analyzed specification. The input data for ANN are three sets of indicators: the set of indicators of section 1 of the SRS $R1=\{Tv, Qv, Sa, Qcs, Sc\}$, where Tv – predicted realization time, Qv – quantity of performers, Sa – predicted quantity of users, Qcs – quantity of software components, Sc – predicted size (LOC); the set of indicators of section 2 of the SRS $R2=\{Cos, Cdb, Cc, Cdt, Cud, Sud\}$, where Cos – cost of used operating systems, Cdb – cost of used databases, Cc – cost of used compilers, Cdt – cost of development tools, Cud – quantity of user documentation pages, Sud – cost of user documentation; the set of indicators of section 1 of the SRS $R3=\{Qfr, Cfr, Qa, Cb, Cui, Qmi, Cmi, Qai, Cai, Qci, Cci, Qnfr, Cnfr\}$, where Qfr – quantity of functional requirements, Cfr – cost of functional requirements, Qa – quantity of algorithms, Cb – average predicted cost of bug, Cui – cost of user interfaces, Qmi – quantity of intermodule interfaces, Cmi – cost of intermodule interfaces, Qai – quantity of hardware interfaces, Cai – cost of hardware interfaces, Qci – quantity of communication interfaces, Cci – cost of communication interfaces, $Qnfr$ – quantity of non-functional requirements, $Cnfr$ – cost of non-functional requirements. The result of ANN functioning is the set of the predicted relative evaluation of the software project characteristics $SCH=\{Cs, Dsp, Cx, Cp, Ub, Qs\}$, where Cs – software project cost, Dsp – duration, Cx – complexity, Cp – cross-platform, Ub – usability, Qs – quality [11]. These characteristics provide the comprehensively analysis of the possible success of software projects implementation – in terms not only quality of developed software products (quality, cross-platform, usability), but quality of software projects management (cost, duration, complexity). ANN was realized in Matlab, was trained with training sample of 6030 vectors by different training methods and was tested with testing sample of 610 vectors [10]. The analysis of charts of the ANN training and testing led to the conclusion that the ANN was trained with high accuracy and precision. In [10] the analysis of ANN training results (by different training functions with different performance functions) was also conducted. The performance function $msereg$ and the training functions OSS , SCG , $RPROP$ were selected on the basis of the following criteria: training performance, training time, and number of epochs.

The structure chart of the ANN layers in Simulink is shown on Fig. 3.

The developed *method of evaluating the success of software project implementation based on analysis of SRS (MESSPI)* consists of the next stages [18]:

1. neuronet prediction of characteristics of software project based on the analysis of specification (the basis of which is the neuronet model of predicting the software projects characteristics based on the SRS analysis [10, 11]). The result of this stage is the set of the predicted relative evaluations of the software project characteristics $SCH=\{Cs, Dsp, Cx, Ub, Cp, Qs\}$, $Cs \in [0..1]$, $Dsp \in [0..1]$, $Cx \in [0..1]$, $Ub \in [0..1]$, $Cp \in [0..1]$, $Qs \in [0..1]$, where 0 – insufficient data for prediction of the characteristics (in this case MESSPI does not work), 0.08 – characteristic negative affects on the success of software project implementation (high cost, duration, complexity, low usability, cross-platform, quality), 1 – characteristic positive impacts on the success of software project implementation (low cost, duration, complexity, high usability, cross-platform, quality);
2. interpretation of the received relative values of the software project characteristics – criteria for this interpretation is the integrative indicator of software project (Fig. 4):

$$Iip_{Sp}=0.5*0.866*(Cs*Cx+Cx*Dsp+ Dsp*Ub+ Ub*Cp+Cp*Qs+ Qs*Cs) \quad (1)$$

3. evaluation of the degree of success of the software project implementation on the basis of the integrative indicator:

$$P_{Iip}=Iip_{Sp}/Iip_{max}=Iip_{Sp}/2.598=0.385*Iip_{Sp} \quad (2)$$

4. testing of the stability and acceptability of compensations of software project characteristics: the indicator Ace_{Sp} of stability and acceptability of compensatory effects of the characteristics has the value “True”, if the hexagon (Fig. 4) is convex (if the sum of the angles of hexagon is 720° and sines of angles have the same signs).

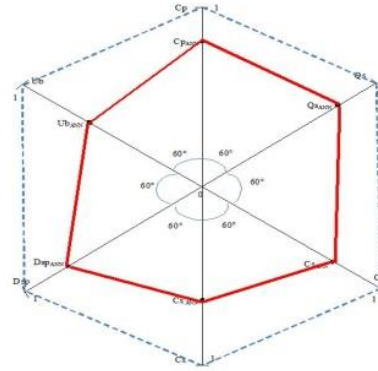
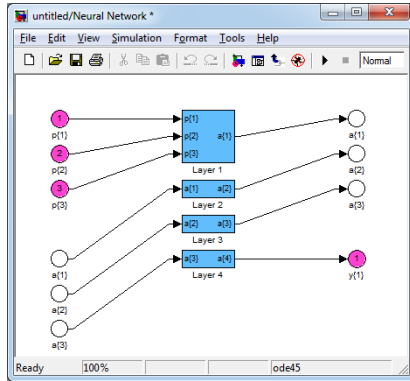


Fig. 3. The structure chart of ANN layers **Fig. 4.** The graphical representation of Iip_{Sp} and Iip_{max}

Thus, the input data for MESSPI is the set of SRS indicators, and the result of the method is the evaluation of the degree of success of the software project implementation [18], which provides to perform the reasonable choice of SRS for the further implementation of the project.

For the completion of ITPCES we need to develop *the intelligent system of predicting the characteristics and evaluating the success of software projects implementation* based on the developed method MESSPI [18].

3 Intelligent System of Predicting the Characteristics and Evaluating the Success of Software Projects Implementation

The input of the intelligent system of predicting the characteristics and evaluating the success of software projects implementation (SPCES) are the selected in [10] 24 SRS indicators, and the result of its work are: the relative values of the software project characteristics, the conclusion about stability and acceptability of compensatory effects of the software project characteristics, the integrative indicator of software project (graphical representation and value), the value of the degree of success of the software project implementation and the conclusion about category of success of software project implementation (the successful, the challenged or the failed project is expected).

The structure (algorithmic-focused vision with elements of architectural solutions) of the intelligent system of predicting the characteristics and evaluating the success of software projects implementation is represented on Fig. 5.

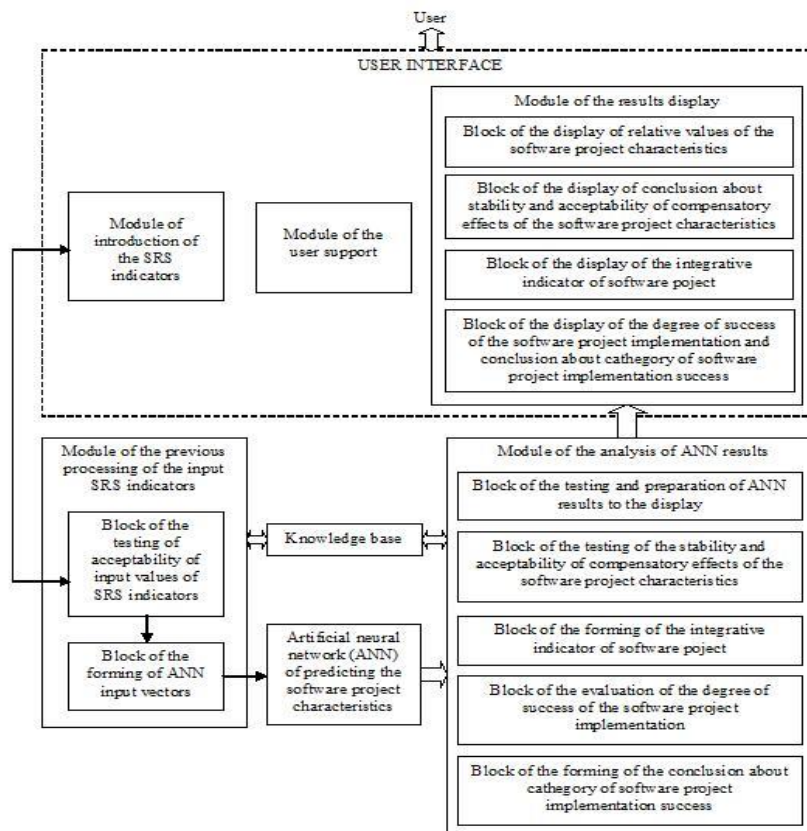


Fig. 5. The structure (algorithmic-focused vision with elements of architectural solutions) of intelligent system of predicting the characteristics and evaluating the success of software projects implementation (SPCES)

SPCES consists of the next components:

1. *module of introduction of the SRS analysis* – is the part of the user interface; reads the user information about the quantitative values of 24 SRS indicators, which are necessary for prediction of the software project characteristics;
2. *module of the user support* – is the part of the user interface; provides to the user the information about the structure of the software requirements specification, about the SRS indicators (which are required for prediction of the software projects characteristics), about the valid (for the system) ranges of the SRS indicators values (defined in [10] based on the analysis of the above-described 410 SRS), about the process of forming the results of the system (SPCES) functioning;
3. *module of the previous processing of the input SRS indicators* – tests the acceptability of the input values of the SRS indicators under the rules of the knowledge base; forms the input vector for ANN: ANN has 5 inputs x' , 6 inputs x'' and 13 inputs x ; on the inputs x' the indicators of the section 1 of the SRS are submitted, on the inputs x'' - the indicators of the section 2 of the SRS, on the inputs x – the indicators of the section 3 of the SRS under the rules of the knowledge base;
4. *knowledge base* – consists of the data section and rules section; in the data section accumulates the values of the SRS indicators and results of the system SPCES functioning; the rules section contains: rules for the testing of acceptability of input values of SRS indicators, rules for the forming of ANN input vectors, rules for the testing and preparation of ANN results to the display, rules for the testing of the stability and acceptability of compensatory effects of the software project characteristics, rules for the forming of the conclusion about category of success of the software project implementation;
5. *artificial neural network (ANN)* of predicting the software project characteristics – detailed described in paragraph 2 and in the papers [10, 11];
6. *module of the analysis of ANN results* – tests and prepares of ANN results to the display, calculates the value of indicator Ace_{sp} of stability and acceptability of compensatory effects of the characteristics according to the 4-th stage of the method MESSPI [18] and forms the conclusion about stability and acceptability of compensatory effects of the software project characteristics, forms the graphical representation and calculates the value of the integrative indicator Iip_{sp} of the software project according to the 2-nd stage of the method MESSPI [18], evaluates the degree P_{ip} of success of software project implementation according to the 3-rd stage of the method MESSPI [18] and forms the conclusion about category of software project implementation success (conclusions are formed using the rules from the knowledge base);
7. *module of the results display* – is the part of the user interface; provides to the user the relative values of the software projects characteristics, the conclusion about stability and acceptability of compensatory effects of the software project characteristics, the graphical representation and the value of the integrative indicator of the software project, the degree of success of software project implementation and the conclusion about category of software project implementation success (successful, challenged or failed project is expected).

The rules for the testing of acceptability of input values of SRS indicators (are substantiated by the valid (for the SPCES) ranges of the SRS indicators values, that were defined in [10] based on the analysis of the above-described 410 SRS) have the form:

1. if $T_v \in [1..24]$ (months), then flag=true, else flag=false;
2. if $Q_v \in [1..10]$ (persons), then flag=true, else flag=false;
3. if $S_a \in [1..1000]$ (persons), then flag=true, else flag=false;
4. if $Q_{cs} \in [1..50]$ (components), then flag=true, else flag=false;
5. if $S_c \in [50..50000]$ (lines of code), then flag=true, else flag=false;
6. if $C_{os} \in [0..1250]$ (USD), then flag=true, else flag=false;
7. if $C_{db} \in [0..1250]$ (USD), then flag=true, else flag=false;
8. if $C_c \in [0..1250]$ (USD), then flag=true, else flag=false;
9. if $C_{dt} \in [0..1250]$ (USD), then flag=true, else flag=false;
10. if $C_{ud} \in [1..50]$ (pages), then flag=true, else flag=false;
11. if $S_{ud} \in [50..2500]$ (USD), then flag=true, else flag=false;
12. if $Q_{fr} \in [5..300]$ (requirements), then flag=true, else flag=false;
13. if $C_{fr} \in [50..4750]$ (USD), then flag=true, else flag=false;
14. if $Q_a \in [1..500]$ (algorithms), then flag=true, else flag=false;
15. if $C_b \in [10..960]$ (USD), then flag=true, else flag=false;
16. if $C_{ui} \in [50..3000]$ (USD), then flag=true, else flag=false;
17. if $Q_{mi} \in [50..2450]$ (interfaces), then flag=true, else flag=false;
18. if $C_{mi} \in [25..2500]$ (USD), then flag=true, else flag=false;
19. if $Q_{ai} \in [5..100]$ (interfaces), then flag=true, else flag=false;
20. if $C_{ai} \in [25..1500]$ (USD), then flag=true, else flag=false;
21. if $Q_{ci} \in [5..125]$ (interfaces), then flag=true, else flag=false;
22. if $C_{ci} \in [25..1750]$ (USD), then flag=true, else flag=false;
23. if $Q_{nfr} \in [1..9]$ (requirements), then flag=true, else flag=false;
24. if $C_{nfr} \in [50..4000]$ (USD), then flag=true, else flag=false;
25. if flag=true, then the input values of the SRS indicators are acceptable, else if flag=false the input values of the SRS indicators are not acceptable, in this case the method MESSPI and the system SPCES cannot be used for this SRS and project.

The rules for the forming of ANN input vectors (are substantiated by the quantities of the elements of the above-described sets R1-R3) have the form:

1. on the input x'_i ($i=1..5$) the value of i-th element of set R1 of the indicators of the section 1 of the SRS is submitted;
2. on the input x''_k ($k=1..6$) the value of k-th element of set R2 of the indicators of the section 2 of the SRS is submitted;
3. on the input x_j ($j=1..13$) the value of j-th element of set R3 of the indicators of the section 3 of the SRS is submitted;
4. if the user doesn't enter the value of indicator, then corresponding input of ANN is -1.

The rules for the testing and preparation of ANN results to the display (are substantiated by the above-described approach to ANN training) have the form:

1. if $C_s=0$ or $D_{sp}=0$ or $C_x=0$ or $C_p=0$ or $U_b=0$ or $Q_s=0$, then insufficient data for prediction of the software project characteristics, in this case the method MESSPI and the system SPCEs cannot be used for this SRS and project;
2. output $y_1 - C_s$ – the relative value of the software project cost, output $y_2 - D_{sp}$ – the relative value of the software project duration, output $y_3 - C_x$ – the relative value of the software project complexity, output $y_4 - U_b$ – the relative value of the software project usability, output $y_5 - C_p$ – the relative value of the software project cross-platform, output $y_6 - Q_s$ – the relative value of the software project quality.

The rules for the testing of the stability and acceptability of compensatory effects of the software project characteristics (are substantiated by the above-described method MESSPI) have the form:

1. if the hexagon (Fig. 4) is convex (if the sum of the angles of hexagon is 720° and sines of angles have the same signs), then the indicator Ace_{sp} of stability and acceptability of compensatory effects of the characteristics has the value "True";
2. if the hexagon (Fig. 4) isn't convex (if the sum of the angles of hexagon isn't 720° or sines of angles have the different signs), then the indicator Ace_{sp} of stability and acceptability of compensatory effects of the characteristics has the value "False";
3. if $Ace_{sp}=\text{True}$, then the characteristics are stable, the compensations of characteristics are acceptable, the method MESSPI and the system SPCEs are suitable for this software project and this SRS, else if $Ace_{sp}=\text{False}$, then the characteristics are unstable, the compensations of characteristics are unacceptable, the method MESSPI and the system SPCEs are not suitable for this software project and this SRS.

The degree of success of the software project implementation, which is defined under the 3-rd stage of method MESSPI, is uninformative to the developers and to the customers through the complexity and ambiguity of interpretation of its value in the predicting the category of the success of the software project. For the facilitation of the interpretation of the value of the degree of success of the software project implementation we define thresholds values of this degree, which provide the conclusion about the category of the project success. For establishment of these thresholds values (for creation of the rules for the forming of the conclusion about category of success of the software project implementation) we have analyzed 410 above-described SRS, for which the degrees P_{lip} of success of the software project implementation were determined according to method MESSPI [18], and 410 finished applications, for which the categories of the success are known. In general, based on the proposed definition of the success of software project implementation and current reports [3], there are three categories of success of the software projects implementation: successful (are projects, that delivered on time, on budget and have required features and functions), challenged (are projects, that late, over budget,

and/or with less than the required features and functions), failed (are projects, that cancelled prior to competition or delivered and never used). The results of this analysis are shown in Table 1.

Table 1. Predicted relative values of characteristics, calculated integrative indicators and degree of success of four software projects implementation

	Web-applications	Mobile applications	E-learning	Applications for statistics, accounting	Automated systems	Information systems
Failed software projects	$P_{lip} \leq 0.17$	$P_{lip} \leq 0.19$	$P_{lip} \leq 0.15$	$P_{lip} \leq 0.16$	$P_{lip} \leq 0.17$	$P_{lip} \leq 0.18$
Challenged software projects	$0.17 < P_{lip} \leq 0.62$	$0.19 < P_{lip} \leq 0.60$	$0.15 < P_{lip} \leq 0.61$	$0.16 < P_{lip} \leq 0.62$	$0.17 < P_{lip} \leq 0.61$	$0.18 < P_{lip} \leq 0.59$
Successful software projects	$P_{lip} > 0.62$	$P_{lip} > 0.60$	$P_{lip} > 0.61$	$P_{lip} > 0.62$	$P_{lip} > 0.61$	$P_{lip} > 0.59$

The rules for the forming of the conclusion about category of success of the software project implementation (considering the empirical estimates from Table 1, which in general correspond to the statistical evaluations [3] from Fig. 1) have the form:

1. if the value of the degree of success of the software project implementation $P_{lip} \leq 0.19$, then software project is predictably failed;
2. if the value of the degree of success of the software project implementation $0.19 < P_{lip} \leq 0.62$, then software project is predictably challenged;
3. if the value of the degree of success of the software project implementation $P_{lip} > 0.62$, then software project is predictably successful.

4 Experiments with SPCEs

The input data for the SPCEs are the SRS indicators for five software projects, that were developed by the different groups of developers for the solution of the one task – the development of the automated system for large-format photo print – to the order by LLC «Deymos», Khmelnytsky (Table 2).

Table 2. The values of the indicators of five SRS, that developed by the different groups of developers for the solution of the one task

№ pr.	The set of indicators of section1 of SRS	The set of indicators of section2 of SRS	The set of indicators of section3 of SRS	
1	Tv=6, Qv=3, Sa=220, Qcs=13, Sc=10900	Cos=260, Cdb=324, Cc=216, Cdt=270, Cud=11, Sud=690	Qfr=83, Cfr=1075, Qa=108, Cb=216, Cui=705, Qmi=680, Cmi=563,	Qai=29, Cai=415, Qci=30, Cci=400, Qnfr=3, Cnfr=910
2	Tv=13, Qv=6, Sa=495, Qcs=25, Sc=26090	Cos=624, Cdb=640, Cc=648, Cdt=639, Cud=25, Sud=1329	Qfr=160, Cfr=2510, Qa=258, Cb=505, Cui=1687, Qmi=1310, Cmi=1315,	Qai=53, Cai=805, Qci=65, Cci=925, Qnfr=5, Cnfr=2110
3	Tv=18, Qv=8, Sa=770, Qcs=37, Sc=36940	Cos=249, Cdb=300, Cc=219, Cdt=283, Cud=10, Sud=650	Qfr=149, Cfr=2530, Qa=247, Cb=499, Cui=1683, Qmi=1302, Cmi=1319,	Qai=52, Cai=811, Qci=67, Cci=918, Qnfr=5, Cnfr=2100
4	Tv=24, Qv=10, Sa=1000, Qcs=50, Sc=50000	Cos=0, Cdb – not defined, Cc – not defined, Cdt – not defined, Cud – not defined, Sud – not defined	Qfr=300, Cfr=4750, Qa=500, Cb=960, Cui=3000, Qmi=2450, Cmi=2500,	Qai=53, Cai=805, Qci=65, Cci=925, Qnfr=5, Cnfr=2110
5	Tv – not defined, Qv=1, Sa – not defined, Qcs – not defined, Sc – not defined	Cos=620, Cdb=641, Cc=645, Cdt=653, Cud=27, Sud=1326	Qfr=167, Cfr=2498, Qa=262, Cb=509, Cui=1691, Qmi=1313, Cmi=1310,	Qai=56, Cai=798, Qci=62, Cci=929, Qnfr=5, Cnfr=2125

After the introduction of the SRS indicators values the module of the previous processing of the input SRS indicators saves this data in the data section of the knowledge base and tests the acceptability of the input values of the SRS indicators

under the rules of the rules section of the knowledge base - if input values are not valid, then the system gives the message to the user: "The input values of SRS indicators are unacceptable, the system of predicting the characteristics and evaluating the success of software projects implementation cannot be used for such SRS". For the proposed software projects the valid values of the SRS indicators were entered, so the ANN input vector is formed for each software project. Block of the forming of ANN input vectors generates the vectors for the appropriate ANN inputs - Table 3. The ANN of predicting the software project characteristics of the project processes the input vector and gives the results (the predicted relative evaluations of the software project characteristics), that also is shown in Table 3.

Table 3. ANN input and output vectors for five software projects

№ pr.	Input 1 (x')	Input 2 (x'')	Input 3 (x)	Output (Y)
1	[6;3;220;13; 10900]	[260;324;216; 270;11;690]	[83;1075;108; 216;705;680; 563;29;415;30; 400;3;910]	[0.789;0.782; 0.792;0.79; 0.795;0.792]
2	[13;6;495;25; 26090]	[624;640;648; 639;25;1329]	[160;2510;258; 505;1687;1310; 1315;53;805;65; 925;5;2110]	[0.518;0.521; 0.521;0.53; 0.53;0.518]
3	[18;8;770;37;369 40]	[249;300;219; 283;10;650]	[149;2530;247; 499;1683;1302; 1319;52;811;67; 918;5;2100]	[0.539;0.537; 0.542;0.533; 0.54;0.542]
4	[24;10;1000;50; 50000]	[0;-1;-1;-1;-1]	[300;4750;500; 960;3000;2450; 2500;53;805;65; 925;5;2110]	[0.389;0.082; 0.39; 0.097; 0.093;0.389]
5	[-1;1;-1;-1;-1]	[624;648;648; 648;25;1329]	[167;2498;262; 509;1691;1313; 1310;56;798;62; 929;5;2125]	[0.68;0.522; 0.681;0.52; 0.52;0.68]

Block of the testing and preparation of ANN results to the display tests the ANN results – if the value of even one ANN output is 0, the system gives the message to the user: "The data for predicting the software project characteristics are insufficient, so the system of predicting the characteristics and evaluating the success of software projects implementation cannot be used for such SRS". For the proposed software projects the input data were sufficient, so the ANN results were prepared to the display according to the above rules.

Block of the testing of the stability and acceptability of compensatory effects of the software project characteristics calculates the indicator Ace_{sp} of stability and

acceptability of compensatory effects of the characteristics (Table 4) and forms the conclusion about stability and acceptability of compensatory effects of the software project characteristics. If $Ace_{sp} = \text{False}$, the user gets the message: "The software projects characteristics are unstable, the compensations of characteristics are unacceptable, so the system of predicting the characteristics and evaluating the success of software projects implementation is not suitable for this project and for this SRS". If $Ace_{sp} = \text{True}$, then the system calculates the integrative indicator of software project, the degree of success of the software project implementation and forms the conclusion about category of software project implementation success. For the proposed software projects №1, №2, №3, №5 the characteristics are stable, the compensations of characteristics are acceptable, so the obtained predicted relative values of the characteristics are processed according to the method MESSPI. For the proposed software project №4 the system gives the message to the user: "The software projects characteristics are unstable, the compensations of characteristics are unacceptable, so the system of predicting the characteristics and evaluating the success of software projects implementation is not suitable for this project and for this SRS".

Block of the forming of the integrative indicator of software project forms the graphical representation and calculates the value of integrative indicator Iip_{sp} of software project (Table 4).

Block of the evaluation of the degree of success of the software project implementation estimates the value of the degree of success P_{lip} of the software project implementation (Table 4).

Block of the forming of the conclusion about category of software project implementation success uses the rules of the knowledge base and forms the conclusion about category of software project implementation success (Table 4).

Table 4. The values of the indicator Ace_{sp} of stability and acceptability of compensatory effects of the characteristics, integrative indicator Iip_{sp} , the degree of success P_{lip} of the software project implementation and the conclusion about category of software project implementation success for five above software projects

№ pr.	Ace_{sp}	Iip_{sp}	P_{lip}	Conclusion about category of software project implementation success
1	True	1,621	0,624	Successful project
2	True	0,711	0,274	Challenged project
3	True	0.754	0.290	Challenged project
4	False	-	-	The software projects characteristics are unstable, the compensations of characteristics are unacceptable, so this system is not suitable for this project and for this SRS
5	True	0.942	0.363	Challenged project

Let's analyze of the results: the category of software project implementation success was defined for the projects №1, №2, №3, №5, for which the predicted values of characteristics are stable, their compensations are acceptable. The software project №1 has the best characteristics, it predictably belongs to the category of successful projects. The software projects №2, №3, №5 have the worst characteristics and predictable are classified as challenged projects. For the project №4 the system SPCES cannot determine the category of implementation success because the predicted characteristics are unstable and their compensations are unacceptable. So the conclusion of the intelligent system of predicting the characteristics and evaluating the success of software projects implementation recommends to LLC "Deymos" to order the implementation of software project №1 (to the development of the automated system for large-format photo print) that will be successful with the greatest probability.

Nowadays the developer and customer select the software project based on only own intuition and the cost and duration that predicted in the SRS. But SRS developers cannot always correctly predict the oriented cost and duration of software project during development of the SRS. Predicted (in the SRS) values of cost and duration for the four examined alternative software projects (for which SPCES determines the category of implementation success) are represented in Table 5.

Table 5. Predicted (in the SRS) values of cost and duration for the four software projects

Characteristics of software project	Values for Project №1	Values for Project №2	Values for Project №3	Values for Project №5
Predicted cost, specified in the SRS	11875 USD	11125 USD	10625 USD	10812 USD
Predicted duration, specified in the SRS	6 months	13 months	18 months	Not defined

The values of the characteristics of software projects from Table 5 show, that all four software projects have the different duration but the same cost that predicted in the SRS. But the results of Table 3 show, that projects have significantly different relative values of all characteristics, including the cost, which were calculated taking into account all significant SRS indicators. Thus, the relative cost ranges from 0.518 (for Project №2) to 0.789 (for Project №1). So, if we evaluate the cost, taking into account all significant SRS indicators, then it values are not the same for the four examined projects. As for the value of project duration, this value isn't defined for project №5, for example, then software project in this case will be evaluated solely on the basis of its cost value. Therefore, the customer and developer can make the wrong conclusion about choice of project on the basis of solely cost and time that predicted in the SRS. In addition, such conclusion is difficult in the real conditions. For example, according to Table 5, the lowest cost has the software project №3, and the lowest duration has the software project №1 (but the value is unknown for the software project №5), i.e. the customer and the developer must make the choice of software project in this case on the basis of one criterion - or by cost, or by the duration.

In addition, the success of software projects implementation depends not only on the cost and duration, but also on the functionalities of developed software, i.e. on the rest of the main characteristics of the software project - complexity, usability, cross-platform and quality, which aren't defined in the SRS explicitly in the quantitative form. In addition, Table 4 shows that the examined software projects have the different category of software project implementation success. Therefore, the values of main characteristics, provided by ANN and the conclusions of SPCES about the category of software project implementation success will help to make the right choice and to implement the software project which will be successful with the greatest probability (among from four examined software projects is Project №1). But if the developer and the customer made the choice of the software project on the basis of the only duration, they probably would choose the project №3, which really has a low degree of success of the implementation and with the high probability will be challenged software project (wrong choice).

5 Conclusions

In the article the structure of information technology of predicting the characteristics and evaluating the success of software projects implementation (ITPCES) is first time proposed. The basic components of ITPCES are the previously developed by the author the neuronet model of predicting the software projects characteristics and the method of evaluating the success of software projects implementation based on analysis of SRS and also (yet not developed) the intelligent system of predicting the characteristics and evaluating the success of software projects implementation, to the designing of which this research is dedicated.

The structure of the intelligent system of predicting the characteristics and evaluating the success of software projects implementation (SPCES) are proposed. SPCES consists of the next components: module of introduction of the SRS analysis; module of the user support; module of the previous processing of the input SRS indicators; knowledge base; artificial neural network (ANN); module of the analysis of ANN results; module of the results display. This system gives the conclusion about the probably category of success of the software project implementation based on analysis of the SRS (at the early stages of the life cycle).

The practical significance of the proposed information technology ITPCES is this fact, that system's conclusions about the category of the success of software project implementation provide to the customers the comparison of the proposed software projects and the data for the reasoned and informed choice of the most successful software project (not just on the basis of the project cost and duration, as is currently).

The authors' following perspective for future researches are: 1) development of DEF0-block diagram and UML component/deployment diagrams for the SPCES; 2) realization of the intelligent system of predicting the characteristics and evaluating the success of software projects implementation for prediction of characteristics and evaluation of success of software project implementation based on analysis of the

SRS; 3) realization of the information technology of predicting the characteristics and evaluating the success of software projects implementation.

References

1. ISO 10002:2014 Quality management – Customer satisfaction – Guidelines for complaints handling in organizations (2014)
2. Bourque, P., Fairley, R.: Guide to the software engineering body of knowledge (SWEBOK): Version 3.0. A project of the IEEE Computer Society (2014)
3. The Standish Group International: CHAOS Manifesto – Think big, act small. Technical report, CHAOS Knowledge Center (2013)
4. Levenson, N.G.: Systemic factors in software-related spacecraft accidents. In: AIAA Space Conference and Exposition, pp.1-11 (2001)
5. Levenson, N.G.: Software challenges in achieving space safety. *Journal of the British Interplanetary Society*. 62, 265-272 (2009)
6. Ishimatsu, T., Levenson, N., Thomas, J., Fleming, C., Katahira, M., Miyamoto, Y., Ujiie, R.: Hazard analysis of complex spacecraft using systems-theoretic process analysis. *Journal of Spacecraft and Rockets*. 51, 509-522 (2014)
7. Yourdon, E.: *Death March: The complete software developer's guide to surviving "Mission impossible" projects*: 2nd edition. Prentice Hall (2003)
8. IEEE 1031-2011 – IEEE Guide for the Functional Specification (2011)
9. IEEE 29148-2011 - Systems and software engineering - Life cycle processes - Requirements engineering (2011)
10. Krasiy, A.: Research of artificial neural network's component of method of characteristics prediction and evaluating the success of software projects implementation. *Transactions of Khmelnytsky National University*. 1, 24-31 (2016) [in Ukrainian]
11. Krasiy, A.: Modelling of process of prediction of software characteristics based on the analysis of specifications. *Computer-Integrated Technologies: Education, Science, Industry*. 66-76 (2014) [in Ukrainian]
12. Fenton, N.: *Software metrics: A rigorous approach* (3rd edition). CRC Press (2014)
13. Chen, A., Beatty, J.: *Visual models for software requirements*. MS Press, Washington (2012)
14. Fatwanto, A.: Software requirements specification analysis using natural language processing technique In: *International Conference on Quality in Research*, pp.105-110 (2013)
15. Rehman, T., Khan, M.N.A., Riaz, N.: Analysis of requirement engineering processes, tools/techniques and methodologies. *I.J. Information Technology and Computer Science*. 40-48 (2013)
16. Leon, A., Leon, M.: *Fundamentals of Information Technology*: 2nd Edition. Vikas Publishing (2009)
17. Korneev, I., Ksandopulo, G., Mashurtsev, V.: *Information Technologies: Tutorial*. Publishing house "Prospekt" (2009) [in Russian]
18. Hovorushchenko, T., Krasiy, A.: Method of Evaluating the Success of Software Project Implementation Based on Analysis of Specification Using Neuronet Information Technologies. *Proceedings of the 11th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer*. CEUR-WS. 1356, 100-107 (2015)

Design and Development of Information System of Scientific Activity Indicators

Aleksandr Spivakovsky, Maksym Vinnyk, Yulia Tarasich and Maksym Poltoratskiy

Kherson State University, 27, 40 rokiv Zhovtnya St., 73000 Kherson, Ukraine

{Spivakovsky, Vinnik, YuTarasich}@kspu.edu,
max1993poltorackii@gmail.com

Abstract. The article provides a brief overview of the most popular information systems of evaluation of scientific activity of scientists. The vision of functional capabilities of processing of the system scientometric indicators of the scientific team, the organization and its business units on the basis of scientific profiles of existing scientometric and bibliometric systems are described. The example of the implemented solutions with the authors description of its components, basic algorithms and used technologies is presented.

Keywords: scientific activity, information systems, scientometric systems, bibliometric systems, scientometric indicators.

Key Terms: ICTInfrastructure, ICTComponent, InformationTechnology, WebService

1 Introduction

Scientific information is a special kind of information that affects the development of any and all sectors of modern society. Analysis of scientific information can be divided into polysyllabic such as information on the research teams, scientific collections, scientist, scientific works and more. Elementary but the objective component, in our opinion, is the scientific activities of the scientist. Today there are many information systems that attempt to create methods and technologies of processing and saving information on the activities of scientists.

The most outstanding services with rapidly growing impact are Google Scholar, Scopus, Orcid, Academia.edu, Research Gate, Mendeley, arXiv.org, cs2n, Epernicus, Myexperiment, Network.nature, Science community.

These services contribute to satisfying the needs of the scientific community. In fact, this positively influences scientific and technical progress and creates a new paradigm of scientific research. A big number of the recently created scientometric services allow assessing the relevance of the research results by a scientist. Having these measurements at hand opens up new opportunities and prospects. In this article we consider the existing information systems for the processing of scientific activities

(section 2), describe your own vision and capabilities to design and develop our system (section 3), as well as the basic methods and technologies (section 4) used for its implementation.

2 Related works

After analyzing the information systems that run on the activities of scientists, scientific groups, publishers, etc..., we offer to look for the most interesting projects.

Bibliometrics of Ukrainian Science. The pilot project of information-analytical system "Bibliometrics of Ukrainian Science", is implemented by the Department of bibliometric and scientometrics of information and analytical support of Vernadsky National Library [1].

The system "Bibliometrics of Ukrainian Science" is representation of information of Ukraine scientists' profiles who provided information about their publication in the Internet; national component of the project Ranking of Scientists (Cybermetrics Lab).

Information resources of systems are formed by processing: created by scientists on the platform of Google Scholar bibliometric profiles containing information of their publication activity results, bibliometric indicators of Scopus, Web of Science, Ranking Web of Research Centers. Updating of information on value of Hirsch index in bibliometric profiles of scientists is executing on monthly, the value of other indicators is updated quarterly (Hirsch index of scholar is h , if he has h publications, each of which is cited at least h times) [1].

Scopus. Scopus is a single the world's largest abstract database, which indexes more than 17 000 items of scientific, technical and medical journals about 4,000 international publishers [2].

Scopus system is designed to maintain efficient workflow of researchers, helping them to: find new articles from the area of their specialization; find information about the author; analyze the publication activity in the subject area; track citation; view the h -index; identify the most cited articles and authors; assess the relevance of the study.

Scopus enables researchers to combine their articles under a single profile [2].

Google Scholar. Google Scholar is freely accessible search system, which indexed the full text of the scientific publications all formats and disciplines.

Google Scholar executes not only informational, but scientometric function. From the list of results on a hyperlink Search Cited by we can obtain the information how many and what documents are linked on the publication in data base Google Scholar. The number in Cited by reflects the degree of authoritativeness and publicity of publication [3].

Web of Science. Web of Science – International established database of Scientific Citation, it is presented by company Thomson Reuters. Web of Science gives possibility to search among 12 000 magazines and 148 000 materials of conferences in the field of natural, social, human sciences and arts, which allows to obtain the most relevant information for your questions. In addition to search, Web of Science establishes a reference link between the specific research using the cited materials and thematic

links between articles established reputable re-searchers working in this field. It is the most extensive database of abstracts. It is available by subscription [4].

Russian Science Citation Index (RSCI). RSCI is a national information-analytical system, accumulating more than 2 million publications of Russian authors, as well as information about the citation of these publications from more than 3,000 Russian magazines. It is designed not only for the operational support of research to date reference and bibliographic information, but is also a powerful tool to carry out evaluation of the impact and effectiveness of research organizations, scientists, the level of scientific journals, etc. [4].

Earlier research team of Kherson State University (KSU), which included the authors of the article, took part in a number of international and national projects whose aim was the development and implementation of scientific and management processes of analytical information systems and services [10].

In addition, this article is a continuation of the previous work of the authors [5] which addressed the issue of openness of scientific activities of Ukrainian scientists, as well as the construction of an open scientific training system, one of the main elements of which are the scientometric information processing system.

The authors also conducted a study of the technical component of the implementation of feedback services in the KSU [6], as well as the formation of the ICT infrastructure at higher education institution [7, 8].

3 Vision of the system. Criteria

Analysis of information systems described in Section 2 (Table 1) confirms once again the need to implement a system that would allow build the consolidated ratings of scientists, scientific groups and organizations in the automatic mode.

Why consolidated? A significant part of scientometric databases and systems, which are presented in the scientific world are closed, and accordingly assess only the academic publications that are indexed by them, while the rest of the scientific work in this assessment are not included. For example, Scopus indexes, indicators of the other part of scientometric databases are not always accounted for as tangible.

In addition, for the analysis of the scientific activities of scientists' group, or a specific organization, it should be carried out manually. The only option of its partial automation is now rating the organization's profile in Google Scholar (which makes the system "Bibliometrics of Ukrainian Science"). But what should do if this profile is not created? Or if not all scientists working in the organization or are part of the research team, and their articles are incorporated in the profile?

Thus, the main task of building our system is the realization of the possibility of automatic processing of scientometric and bibliometric indicators of scientific groups and organizations on the basis of analysis of scientific profiles of known scientometric databases and systems, including automatic search and its analysis.

Table 1. Compare features considered information systems

Information Systems / criteria	Scopus	Google Scholar	RSCI	"Bibliometrics of Ukrainian Science"	Our System
Scientist profile	+	+	+	-	+
Scientific institution Profile	-	+	-	-	+
Profile of structural units of scientific institutions	-	+	-	-	+
Construction ratings	-	-	-	+	+
Scientometric and bibliometric indicators	+	+	+	+	+
Personal notifications and reports	-	-	-	-	+
The openness of the system	+	+	+	+	+
Possibility of automatic comparison of the scientific work of several scientists, organizations, groups, etc.	-	-	-	-	+

3.1 Concept of solutions

Goals and objectives.

Estimated system should have the following features: parsing pages of scientists in scientometric systems and databases; processing and display scientometric indicators of the author, the scientific team, the organization; formation of a library of publications of scientists and the ability to sort by the specified criteria: the author's name; the name of the organization of the university; the department, etc.; statistical processing of the obtained information; the ability to compare the quality performance of universities, research groups, etc.; the possibility of multi-threaded data processing; presence feedback.

Allocation objects of system: parser; scientometric system and the database works with resource; data store; Web- site of resource; reporting.

Parser and data warehouse include the following attributes: author's name; link to profile scientometric databases; number of publications; scientometric indicators (Hirsch index, citation index, I-code, etc.); publications; links to the publications; publication description.

Web-site of resource. Select the following attributes: rating of universities in Ukraine; rating of Ukrainian scientists; rating of university scientists; rating of the university departments; rating of the university faculties; profiles of the scientists of the university.

Reporting includes the following attributes: scientometric indicators; graphics.

The interaction of key system components shown in Fig.1.

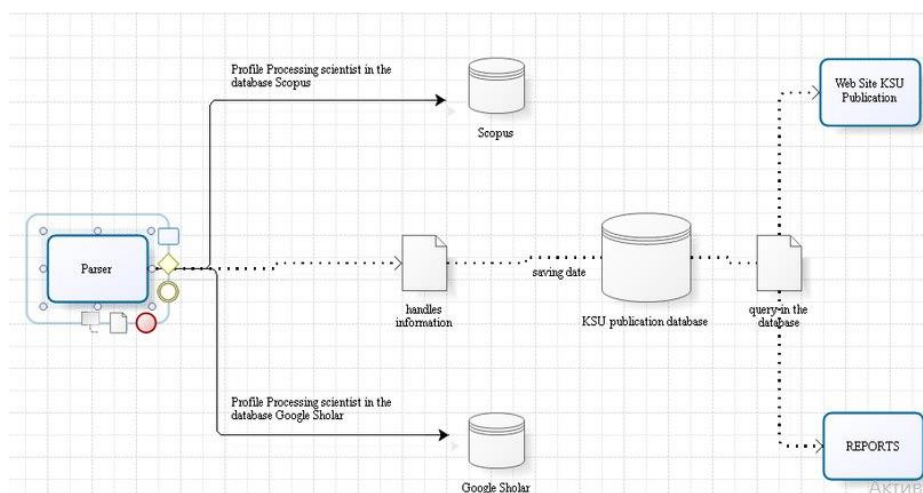


Fig. 1. The interaction of key system components

Thus, the user can view general information of scientific activity of scientist, scientific group, certain university or scientific organization, as well as the consolidated rating. Scientist, registered in the system is able to receive notifications about changes in their scientometric indicators. The system administrator can generate a general statistical report of their organization.

Assumptions and Constraints

In the current version of the system it is implemented the ability to handle scientist indicators on Scopus data and Google Scholar. The algorithm of automatically search for links to profiles of Ukrainian scientists is developed, the algorithm of automatic distribution profiles of scientists on the name of the organization in which they work is implemented, the automatic generating of department ratings, faculties and research teams is implemented, the ability to send messages to e-mail scientists about changes of academic indexes.

Scientometric indices on which ratings are based in the system are:

1. h-index (Scopus&Goggle Scholar). The h-index is based on the highest number of papers included that have had at least the same number of citations;
2. citations (Scopus& Goggle Scholar). Numbers of total citations of documents that are indexed by the system;
3. i10-index (Goggle Scholar). Numbers of total citations by documents that have ten or more citations;

At present, about 3,000 profiles of scientists in Scopus has been processed by the system, of which 680 have been identified as the profiles of Ukrainian scientists. Automatic processing of the found profiles allowed constructing the rating of Ukrainian scientists on their indices in Scopus. By sorting the results of belonging the scientists to the university (e.g. KSU), it was implemented the ability automatically generate ratings of chairs, faculties and scientific researches of the university groups (Fig. 2).

Name	Belonging	h-index	citation	docum
Shapoval, Galina S.	Institute of Bioorganic Chemistry and Petrochemistry ...	9	citations by 520 documents	83
Shapoval, Anatoly N.	Institute of Plasma Physics, Kharkov, Kharkov, Ukrai...	9	citations by 162 documents	58
Shy	NameDepartment	h-index	citation	document
Use	Кафедра ботаніки	5	citations by 8 documents	15
Sha	Кафедра інформатики, програмної інженерії та еко...	2	citations by 9 documents	18
Sha	Кафедра фізики та методики її навчання	2	citations by 16 documents	9
Poli	Кафедра екології та географії	2	citations by 13 documents	6
	Кафедра алгебри, геометрії та математичного ана...		citations by 0 document	5
	Кафедра органічної та біологічної хімії		citations by 0 document	2

Fig. 2. Example of system work

The highest number of publications (on 10.02.16) has such scholars as - Oleg Shishkin (581), Leonid Levchuk (463) and Vladimir Gun'ko (322).

The analysis of the scientific activity of KSU scientists' shows the greatest number of publications has the teachers of Chair of Informatics, Software Engineering and Economic Cybernetics (98). And the most h-index has the teachers of the Chair of Botany (5).

The construction of similar ratings according Goggle Scholar, it is currently possible only in the presence of links on it's, as distinct from Scopus, the author himself should register in the system. There is more complicated the ability to search scientists. Thus, we have been processed the records, links have been provided by the University scientists. Now for viewing and analyzing there is available indicators of scientists of Faculties of Physics, Mathematics and Informatics of KSU, Faculty of Pre-School and Elementary Education of KSU, general Chair of Philosophy and Social and Humanities Sciences.

The next stage of development and improvement of the system will:

- automatic integration and analysis of information on scientometric indicators scientist in the case of duplication of its profile in these scientometric database;
- improving the algorithm of processing information on scientometric indicators of organizations, scientific collectives in case of misspelling or change their names;
- improving the algorithm of finding links on the profiles of scientists according their belonging to the country;
- the ability to automatically compare the indexes of scientific activities of scientists, research groups, organizations and the structural divisions.

Analysis of the use

There are two user groups allocated in the system: the administrator of the system on the part of the establishment; user.

The category of "user" is the staff of institutions, scientists, as well as the rest of Internet users, who can view the information provided on the Web-site of the system.

As example, Consider the algorithm of the system work with Scopus in details:

The parser takes a reference to the scientific profile from the database system and loads the appropriate page of Scopus. After that, two parallel streams are run – processing of scientometric indicators of scientist and processing of information about

his articles. Once when processing of the whole page is over, there is an inquiry about the presence records under consideration "name" in the database system. If the name is, it updates the information about scientometric indicators and publications of the scientist. Otherwise - in the database record is created about the author by assigning a unique identifier to him, and information about his articles and scientometric indexes is entered into the appropriate tables. After the upgrading all the database system the administrator and scientists registered in the database get e-mail with information about changes of their indexes.

4 Tools and Technologies

Developing of solutions requires the use of certain products and technologies:

- JSON. It is used in the system for the exchange of data for third-party systems. Thus, our system can be a source of data for other resources. It implements the data exchange via json requests.
- asp.net and framework Entity. It is used to implement Web- Site of System.
- Library of html align pack. This library is used for processing of Scopus pages. It uses PATCH requests and then adds the results to the database. In the previous version of the system the regular expressions were used. The use of html align pack is greatly affected on her productivity.

One of the most important algorithms used in the system is Levenstein algorithm [9].

This algorithm is used for solving the problem of determining belonging the scientist to a particular organization, which arises at changing of the organization's name, its spelling errors in the article, the change of scientists their place of work, etc.

Let's consider the algorithm in detail:

Algorithms of fuzzy search are also known as similarity search or fuzzy string search are the basis of the spelling checker systems and full of search engines like Google or Yandex, This algorithm is an extremely useful feature of any search engine. However, its effective implementation is much more complex than implementing of a simple search by exact match. The most commonly used metric is the Levenshtein distance or edit distance, its algorithms calculation can be found at every turn.

Thus, we compare the author's field of membership of the organization specified in Scopus with many possible names of organizations in the system database. This takes into account the possibility of errors.

Conclusions

The work is developed by processing system of scientometric indicators of scientist on the basis of its profile in Scopus and Google Scholar systems. The main difference between the systems developed by us from others is the ability to automatically build research teams rankings, organizations and entities to which the scientist applies. In

addition, the algorithm of automatically search and group profiles of scientists for their attitude to this or that state, organization, is already have developed.

The personal profile of each scientist collected information about his scientometric and bibliometric indicators is a list of his publications, displays statistics of scientific work - change the number of publications, citations, h-index, etc. Graphical display of the dynamics of scientific work was implemented for the research teams, organizations and their subdivisions.

Today the system is used to calculate indicators of scientific activity of Kherson State University and its structural units - departments, faculties, Specialized Academic Council, etc.

The next stage in the development of the system, we see in the realization of its interaction with other scientometrics systems and databases. Also, one of the most important and necessary features, we consider the need for implementation of the comparison options of several organizations, research groups and scientists.

The implementation of the algorithm of automatic search of references to the scientific profile of the membership of a particular country and the organization, and improve the efficiency of the algorithm allows us to speak about the possibility of sampling and processing of large amounts of information. Thus, in the next version of the system it is supposed to build a data warehouse on the principles of Big Data and Map Reduce. That, in turn, will generate ratings of the scientific activities of scientists, scientific groups and organizations with minimal resources and time-consuming.

References

1. Bibliometrics of Ukrainian Science, [Http://nbuviap.gov.ua/bpnu/index.php?page_sites=pro_proect](http://nbuviap.gov.ua/bpnu/index.php?page_sites=pro_proect)
2. Abstracts database Scopus, http://health.elsevier.ru/electronic/product_scopus/
3. Scientometric database, <http://www.nbu.gov.ua/node/1367>
4. Science Citation Index for scientists, <http://index.petrsu.ru>
5. Spivakovsky, A., Vinnik, M., Tarasich, Y.: Web Indicators of ICT Use in the Work of Ukrainian Dissertation Committees and Graduate Schools as Element of Open Science. In: Yakovyna, V., Mayr, H.C., Nikitchenko, M., Zholtkevych, G., Spivakovsky, A., Batsakis, S. (Eds.) ICTERI 2015. CCIS, vol. 594, pp. 3–19. Springer, Heidelberg (2016)
6. Spivakovsky, A., Klymenko, N., Litvinenko, A.: The problem of architecture design in a context of partially known requirements of complex web based application “KSU Feedback”. *Inf. Technol. Educ.* **15**, 83–95 (2013)
7. Spivakovsky, A., Vinnik, M., Tarasich, Y.: To the Problem of ICT Management in Higher Educational Institutions. *Inf. Technol. Learn. Tools* **39**, 99–116 (2014). (In Ukrainian)
8. Spivakovska, E., Osipova, N., Vinnik, M., Tarasich, Y.: Information competence of university students in Ukraine: development status and prospects. In: Ermolayev, V., Mayr, H.C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) ICTERI 2014. CCIS, vol. 469, pp. 194–216. Springer, Heidelberg (2014)
9. Lowenstein, V. : Binary codes with correction for deletions, insertions and substitutions of character. Reports, USSR Academy of Sciences 163.4 (1965).
10. International Projects, <http://www.kspu.edu/About/DepartmentAndServices/DSAICI/internationalprojects>

Adaptation of Legacy Fortran Applications to Cloud Computing

Eugene Tulika¹, Anatoliy Doroshenko¹, and Kostiantyn Zhereb¹

¹ Institute of Software Systems of National Academy of Sciences of Ukraine,
Glushkov prosp. 40, 03187 Kyiv, Ukraine
eugene.tulika@gmail.com, doroshenkoanatoliy2@gmail.com,
zhereb@gmail.com

Abstract. We propose an approach to the semi-automatic transformation of legacy Fortran applications for execution on cloud computing platforms. An architecture is proposed based on web-services choreography, which allows unlimited scalability of the system and reduces overhead on message passing. The approach is tested on an example program from the quantum chemistry field.

Keywords. virtualization, cloud computing, scalable parallelism, web-services choreography.

Key Terms. ConcurrentComputation, DataGrid, HighPerformanceComputing, FormalMethod, SoftwareSystem

1 Introduction

Fortran language exists since the 1950s and during this time positioned itself as one of the best tools for scientific research. Language has huge support from the software industry, new libraries and compilers are released that allow using modern technologies and standards. Examples are: compiler from Portland Group [1] for Fortran supporting GPGPU and CUDA; High Performance Fortran Forum [2] creates compilers and standards for high performance Fortran; OpenMP and MPI for parallel and distributed computation allow to use Fortran for modern cluster computing; Intel popularizes Fortran [3] in order to support proprietary Intel compilers and programs; Coarray Fortran language fork became a standard in Fortran 2003 and allows to support computation on distributed arrays [4]; there are many Fortran libraries [5] for numerical analysis. Besides, conservative policy for backward compatibility makes code written on old standards compatible with the new compilers. All this makes Fortran an attractive platform for scientific research. However, Fortran suffers from outdated standards that make it hard to write an efficient code for computations with distributed memory. Another problem is a great amount of legacy code which was written without taking into account distributed architectures.

Cloud computing became popular because of the need to reduce the cost of computations. The difference between cloud computing and other parallel approaches is usage of scalability (ability to support more resources) vs. performance (running fast

on given resources). Scalable systems can contain components which individually perform poorly. Cloud computing is focused on reducing the cost and time/performance balance using cheap components, and therefore allows using more resources.

This paper describes work in progress aimed at parallelizing Fortran scientific applications for the cloud platform. We propose a methodology for semi-automatic parallelization of legacy applications. The paper also describes an architecture for running distributed applications on the cloud.

Related parallelization approaches for porting legacy applications to cloud platform have been described, such as Pydron [6] and Bio-Cirrus [7]. The contribution of this paper is 1) using rewriting rules technique to automate transformation steps and 2) using service choreography instead of orchestration to reduce overhead.

2 Description of the Studied Application and Problem Statement

In this paper, we discuss our approach of transforming legacy scientific applications based on an example Fortran application from the quantum chemistry field. Some properties we discuss are specific to this application, such as time distribution of subroutines (Section 2) or the structure of control flow graph (Section 3). However, our approach can be applied in the more general case.

The example application calculates the geometry of electron orbitals [8]. The application was initially optimized for single-core performance, without any parallelism. The sequential processing time has quadratic increase depending on the input size.

The first step of our approach requires identifying the most promising subroutines for parallelization. The example program consists of data input, computation subroutines, logging intermediate results into the file and data output. Profiling the application before optimization shows the time distribution for the main computational steps:

- Data input from files and initialization – approximately 1% of time;
- Subroutine `hcore` calculates integrals for every atom – 60% of time;
- Subroutine `iterc` – optimization of the geometry of molecule, 30% of the time.

Therefore, the main focus of parallelization would be spent in subroutines `hcore` and `iterc`. Time performance of algorithms in both subroutines has a linear dependency on the input size. Each of two subroutines is applied independently for every atom in the input, which allows parallelization. However, `iterc` depends on the results of `hcore`, because the calculation of the Fock operator requires integrals computed in `hcore`.

The generic question of Fortran parallelization for distributed architecture is widely studied [16], [17]. As a part of this paper, Asynchronous Network technique with the usage of web-service choreography is applied to the parallel algorithms. This provides a robust framework for the development of applications which can be distributed across cloud infrastructure without additional efforts, while usage of MPI for the highly scalable applications requires manual control over data placement and inter-process communication.

3 Equivalent Code Transformation for Scalable Parallelization

On the next step, we try to find and parallelize independent loops that allow unlimited scalability. Structurally, the application can be modeled with the following graph:

$$I \rightarrow A([a_{1..n}]) \rightarrow B([b_{1..n}]) \rightarrow O \quad (1)$$

Graph vertices I , A , B , O stand for the sequential steps of computation: I – data input and allocation of the memory, A – calculation of the integrals for every atom (`hcore`), B – calculation of the Fock operator for every atom (`iterc`), O – final steps and data output. Graph edges correspond to the control flow between parts of the program, $a_{1..n}$ and $b_{1..n}$ stand for input data of subroutines. The goal of this step is to transform program to the parallel form:

$$I \rightarrow A_{1..n}([a_{1..n}]) \rightarrow B_{1..n}([b_{1..n}]) \rightarrow O \quad (2)$$

We transform sequential loop A , into n parallel processes $A_{1..n}([a_{1..n}])$. Each process A_i will perform calculations on a single segment of data a_i . The same transformation will be applied to B as well. After manual analysis of the data dependencies, it was identified that there are no dependencies between iterations of the loop in code fragment A , and processes $A_{1..n}$ can be invoked in parallel, same applicable to the processes $B_{1..n}$. However, B had a dependency on the results of A and in order to address this dependency barrier synchronization will be used between parallel processes A_i and B_i .

After selection of the model for parallel computation, we should ensure that code fragments invoked in parallel do not have any side effects. This is done by replacing subroutines with pure functions (introduced in Fortran 95 standard). The following changes in code are needed:

1. Create FUNCTION instead of SUBROUTINE
2. Remove IMPLICIT statements – all local variables of the function should be explicitly declared.
3. Remove COMMON BLOCKs (global variables) – all corresponding global variables should be passed as an inputs/outputs of the function, and reads/writes to global variables performed by the calling code.
4. Remove read and write operations – all such operations should be invoked before and after the call to the pure function.

4 Automatic Code Transformation Using TermWare

TermWare [9] is a tool for automatic code transformation which can be applied to the task of transformation to pure functions. TermWare allows describing transformations in a declarative way which simplifies its development and makes them reusable. In the earlier work [10] TermWare was used to build high-level algebraic models of Fortran code and perform transformations on them. This paper uses a similar approach but focuses on the transition of the subroutines to the pure functions.

As an example of transformation we use a set of rules to transform IMPLICIT statements to explicit declarations of variables:

1. `_MarkPure(Subroutine($name,$params,$return,$body))`
`-> Function($name,$params,$return,_MkImp($body))`
2. `_MkImp([$x:$y]) -> [_MkImp($x):_MkImp($y)]`
3. `_MkImp(NIL) -> NIL`
4. `_MkImp(Declare($var,$type,$val) ->`
`Declare($var,$type,$val) [check($var, $type)]`
5. `[_MkImp(Assign($var,$expr)): $y] [isUnchecked`
`($var)] -> [Declare_MARK($var, $type) : [Assign`
`($var,$expr) : $y]] [inferType ($var,$type)]`
6. `[$x:[Declare_MARK($var,$type) : $y]] ->`
`[Declare_MARK($var,$type):[$x:$y]]`
7. `Function ($name,$params,$return,`
`[Declare_MARK($var,$type):$y]) -> Function`
`($name,$params,$return,[Declare($var,$type):$y])`

Rule 1 triggers transformation, marking the body of the function with the marker term `_MkImp`. Rules 2 and 3 walk through the body of the function and expand the `_MkImp` marker to all operations. Rule 4 memorizes the variables which have explicit declaration using the method `check($var, $type)` from the facts DB. Rule 5 finds variables without explicit declaration using method `isUnchecked($var)`. For these variables it determines the type with the method `inferType($var,$type)` and adds declaration marked with `Declare_MARK($var,$type)`. To determine the variable type, the method checks the variable name against IMPLICIT statements in Fortran code, as well as default convention that declares variables starting with 'i'-'n' as INTEGER and all others as REAL. Rule 6 moves this declaration to the beginning of the function, and rule 7 removes the mark. As a result, term `Declare($var,$type)` is generated, which later is transformed to the declaration of the variable in the code.

As an example of rules application, consider a simple procedure of square matrix multiplication (Table 1).

Table 1. Initial and transformed source code for removing IMPLICIT declarations

Initial code	Transformed code
<pre> SUBROUTINE MATR_MULT(N,A,B,C) INTEGER,INTENT(IN) :: N REAL*8,INTENT(IN) :: A(N,N),B(N,N) REAL*8,INTENT(OUT) :: C(N,N) DO I=1,N DO J=1,N S = 0.0D+00 DO K=1,N S=S+A(I,K)*B(K,J) END DO C(I,J)=S END DO </pre>	<pre> FUNCTION MATR_MULT(N,A,B) INTEGER,INTENT(IN) :: N REAL*8,INTENT(IN) :: A(N,N),B(N,N) REAL*8 :: MATR_MULT(N,N) INTEGER :: I, J, K REAL*8 :: S DO I=1,N DO J=1,N S = 0.0D+00 DO K=1,N S=S+A(I,K)*B(K,J) END DO MATR_MULT(I,J)=S END DO </pre>

END DO

END DO

In the initial code, some variables are used with no declaration. After transformation, all the variables are declared. Also, note that the syntax for SUBROUTINE is different from the FUNCTION. But such changes should not be described as additional rules. A simple substitution of the term “Subroutine” with the term “Function” is sufficient. During the code generation phase, all necessary changes are added automatically, which is one of the advantages of the TermWare and high-level algebraic models [10].

5 Transition to Distributed Application Executed on Cloud

Previously discussed transformation steps are not specific to any given parallel platform. Starting from this section, we discuss additional steps needed for the cloud platform. In order to transform application to distributed architecture, its source code has to be transformed to support network calls. Functions A_i and B_i are converted into web-services – separate programs with HTTP interface which can be invoked remotely. The body of the program is transformed into transaction script which invokes remote web-services and aggregates results. In order to use HTTP calls from the Fortran, *libcurl* library is used with the C interface [11]. Transformations of the functions to the separate web-services is done with the Java wrapper. Data for invocation of the remote services is composed in the transaction script. The script collects all input data of the program and sends messages to the remote services.

Cloud platform operation system, such as CloudStack, provides APIs to do scaling – provisioning of the nodes with the predefined configuration on demand. This capability is used by transaction script: after reading the input data and extracting the number of atoms, it makes a call to API of the cloud operating system and requires a startup of the necessary amount of nodes of needed type. In the simplest case, processing of N atoms will require $2N$ nodes, one for each A_i and B_i . However, the number of nodes, time they are actively working and the size of the used memory affects the cost of computation. For optimization of the cost the optimal parameters of configuration should be chosen, so that cost is minimized. In [12] a method of performance optimization of the service-oriented program was proposed based on load estimation. A similar approach can be used for minimization of the cost.

Approach when transaction script calls remote web-services in a service-oriented architecture is called Orchestration. Usage of Orchestration has disadvantages. Paper [13] shows that usage of separate transaction script increases the amount of calls between processes in most patterns of message passing in distributed systems, which increases overhead on data processing.

6 Transition to Choreography

Our goal is to reduce message passing overhead and eliminate a single point of failure represented by transaction script, by using choreography. From the perspective of distributed systems modeling, Choreography could be represented as an asynchronous

network [18]. The asynchronous network consists of the set of processes that communicate with each other. Communication can include: direct message exchange between nodes; broadcast when a node sends a message to each node including itself; multicast when message is sent to the subset of nodes. During the transition to choreography, transaction script is eliminated and its responsibilities are distributed between services. Unlike Orchestration or MPI where transaction script is waiting for the results of web-services execution, services in Choreography don't know about each other and send results of execution to the communication channel. During the execution, the service takes into account the state of the process and type of inbound message in order to determine its position relatively to other services. Having understood its position and taking into account communication protocol, service decides on what type of message should be sent back to the channel when the result is ready. The protocol has the following format: "If current process role is A_i and message M_1 is received as input, procedure F should be invoked and message M_2 should be passed to the processes $A_{2..m}$ ". This format is identical to the description of finite state machine.

Communication Protocol can replace the part of transaction script responsible for service invocation and results processing. Part of responsibilities related to data input is transferred to web-service itself. Code fragment I performing initial data processing is implemented by the new service I_1 which plays the role of starting point in the machine description. Invocation of tail fragment O which outputs the result should be implemented by a separate service O_1 , which will be executed after synchronizing all the processes $B_{1..n}$ and will do the post-processing of the data and data output.

In order to follow the protocol, every web-service should have a controlling mechanism which executes state machine by the description of the protocol. Controlling mechanism is written in Java. Choreography execution is initiated by an initial event, e.g. the file with input data has been uploaded to the cloud data volume. Process I_1 in the waiting state receives this event from the file system and invokes Fortran code responsible for initial data processing. After the event is processed, process I_1 sends broadcast message with the results of invocation to all the processes of the asynchronous network. In the simplest case, this message is received by all the processes $A_{1..n}$, $B_{1..n}$, I_1 , and O_1 . The processes $B_{1..n}$, O_1 , and I_1 ignore this message because B_i and O_1 processes do not have enough information yet to start, and process I_1 already finished its part of the work. It means that only processes $A_{1..n}$ start execution.

An important aspect of choreography is the implementation of barrier synchronization. In [15] global and local synchronizers are described that can be used for synchronization of the processes of the asynchronous network. If global synchronizer is used, then the separate process should be responsible for controlling the conditions of the barrier. The local synchronizer is controlling just its neighbors. In this case message "process reached barrier" is gradually distributed through the network which allows reducing the amount of interactions between services.

In order to achieve better resource utilization using Choreography, following considerations should be taken into account. If certain web-service, such as I_1 , is expected to consume fewer resources than others, it should run on the smaller node (with less RAM and CPUs). If some web-services are never expected to run simultaneously, such as A_i and B_i , they should re-use the same node. Ideally, it should be the same service with a same controlling mechanism which can play different roles in the protocol. This will guarantee that minimal amount of nodes are started at any point of

process execution. Choreography allows reducing the number of messages passed between transaction script and web-services. It also allows better resource utilization having non-blocking requests between web-services. Besides, it provides integration framework with the generic rule sets defined which reduce the amount of code needed to be written to convert the application into distributed system.

7 Testing of the Approach

In order to verify the proposed approach, we used a simplified task with the same model of the data dependency – Gaussian elimination. Testing was performed on the Amazon cloud platform and compares two different configurations of the system. Both configurations have the same amount of processors – 8, and the same amount of RAM – 32Gb. The first configuration consists from one server AWS m4.2xlarge (26 ECUs, 8 vCPUs, 2.4 GHz, Intel Xeon E5-2676v3, 32 GiB memory, EBS only) and is used to run the sequential program. Second configuration consists of four servers AWS m4.large (6.5 ECUs, 2 vCPUs, 2.4 GHz, Intel Xeon E5-2676v3, 8 GiB memory, EBS only) and it is used to run the service-oriented application (with choreography). We have measured the time spent on the processing of the square matrices of different sizes. Comparison of execution time is in Fig. 1. For smaller matrix size, the sequential program runs faster because of overhead in a parallel program. However, for larger matrices the execution time of sequential program grows faster, and it becomes less efficient.

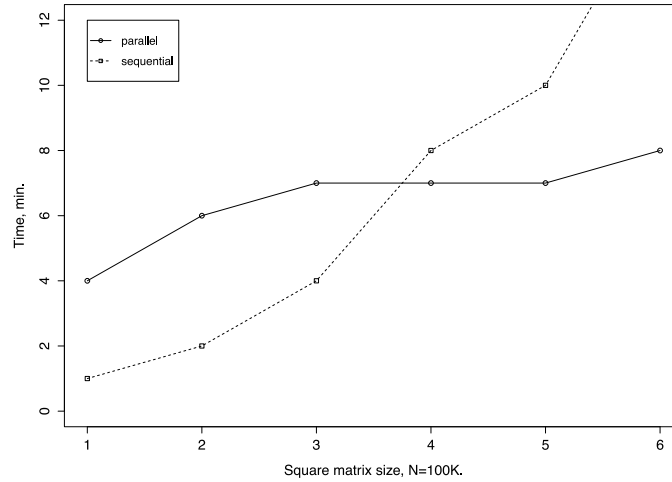


Fig. 1. Comparison of the execution time of sequential and service-oriented versions

8 Conclusion

The paper describes the work in progress of scaling legacy Fortran code using cloud platforms. Proposed architecture uses choreography of web-services which allows

unlimited scalability and reduces overhead on message passing. Scaling exercise is performed on application from quantum chemistry field for calculation of atoms orbitals. One of the main results of the paper is a methodology for adjustment of the legacy source code to the cloud infrastructure, including transition steps to distributed scalable architecture. Our future research directions include automating additional transformation steps using TermWare framework, applying our approach to different applications, as well as testing different cloud configurations to find the most efficient ways of parallelizing legacy applications.

References

1. PGI Compilers & Tools, <http://www.pgroup.com/products/pvf.htm>
2. High Performance, <http://hpff.rice.edu>
3. Fortran is more popular than ever; Intel makes it fast, <https://software.intel.com/en-us/blogs/2011/09/24/fortran-is-more-popular-than-ever-intel-makes-it-fast>
4. Coarrays in the next Fortran Standard, <ftp://ftp.nag.co.uk/sc22wg5/N1751-N1800/N1787.pdf>
5. Netlib Repository, <http://netlib.org>
6. Müller, S. C., Alonso, G., Amara, A., and Csillaghy, A.: Pydrion: Semi-automatic parallelization for multi-core and the cloud. In: 11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14), pp. 645-659.(2014)
7. Karlsson, T. J. M., et al.: Bio-Cirrus: a framework for running legacy bioinformatics applications with cloud computing resources. In Advances in Computational Intelligence, pp. 200-207. Springer Berlin Heidelberg. (2013)
8. Doroshenko, A., Khavryuchenko, V., Iegorov, V., Suslova, L.: Modeling for quantum chemistry computations (in Russian). Upravliayushchie sistemy i mashiny N 5, pp. 83-87. (2012)
9. Doroshenko A., Shevchenko R.: A Rewriting Framework for Rule-Based Programming Dynamic Applications. Fundamenta Informaticae 72 (1-3), pp. 95-108. (2006)
10. Tulika, E., Zhreb, K., Doroshenko, A.: Fortran Programs Parallelization Using Rewriting Rules Technique (in Ukrainian). Problems in Programming N 2-3, pp. 388-397. Kyiv (2012)
11. Libcurl – the multiprotocol file transfer library, <http://curl.haxx.se/libcurl>
12. Tulika, E.: Performance Optimization in SOA Using Load Estimation and Load Balancing (in Ukrainian). Problems in Programming, N 2-3, pp. 193-201. Kyiv (2010)
13. Barker, A., Weissman, J.B. and Van Hemert, J.I.: Reducing data transfer in service-oriented architectures: The circulate approach. Services Computing, IEEE Transactions on Services Computing, N 5(3), pp. 437-449. (2012)
14. Peltz, C.: Web services orchestration and choreography. Computer, (10), pp. 46-52. (2003)
15. Lynch, N.A.: Distributed algorithms. Morgan Kaufmann, San Francisco (1996)
16. Golub G., Ortega J.: Scientific Computing: An Introduction with Parallel Computing, 215-236. Academic Press (1989)
17. Schelter S., Boden C., Schenck M., Alexandrov A., Markl V.: Distributed matrix factorization with MapReduce using a series of broadcast-joins. In Proceedings of the 7th ACM conference on Recommender systems (RecSys '13). ACM, New York, NY, USA, pp. 281-284. (2013)
18. Barker A., Walton C., Robertson D.: Choreographing Web Services. IEEE Transactions on Services Computing, N 2(2), pp. 152-166. IEEE Computer Society. (2009).

Application of the Method for Concurrent Programs Properties Proof to Real-World Industrial Software Systems

Taras Panchenko¹

Taras Shevchenko National University of Kyiv,
64/13, Volodymyrska st., Kyiv, 01601, Ukraine,
tp@infosoft.ua,
WWW home page: <https://www.facebook.com/tpanchenko>

Abstract. Software correctness is an actual topic in many industries nowadays. Significant system properties should be checked in order to use these software safely. The properties proof task is applicable to previously developed systems also, if the property is new one or when the system is transferring to another runtime environment – for example, to multi-core processor. The method for software properties proof in interleaving concurrent environment with communication via shared memory was developed to solve the problem of simultaneous check the required property over the family of programs being run arbitrary times in parallel – instead of doing separate proofs for every number of program instances being run concurrently. Application of this method to real-world industrial tasks is demonstrated in this work.

Keywords: software correctness, shared memory concurrency, safety property, interleaving concurrency, formal methods

Key Terms: Characteristic, Environment, Process, Research, Development

1 Introduction

Software correctness problem (or programs' properties check & proof) is important and challenging in many industries and for different reasons [1]. The problem of software correctness has transformed over the time but still remains the actual one. The huge variety of program system's properties needs to be proven (or checked and warranted somehow) in order to use this software safely. First of all, it concerns safety-critical systems, but also these questions are important for many other software systems:

- do software program behaves as expected? (software validation)
- do software give us the results we need when it completes the work? (safety properties verification)
- is software system processing the request in a timely manner? (program liveness verification)

There are three main approaches to solve these problems [2]:

- testing,
- model checking and
- theorem proving.

While testing techniques itself can't cover all the use cases, even behavioral, they can't warrant the total safety of the system. Moreover, Dijkstra in [3] insists on the main testing limitation – it can show an error, but it can't demonstrate the absence of errors. The two other approaches mentioned above tries to do that. But, the model checking has well-known state explosion problem [4] while pure theorem proving is rather hard technique, even for professionals. One of the advanced approaches address the problem is symbolic model checking which incorporates the pros of both latter. Detailed comparative analysis can be found in many sources, for example in [2].

Software correctness problem is actual in many cases and for different industries:

- when we deal with safety-critical (or life-critical) systems whose failure or malfunction may result in death or serious injury to the people OR loss or damage to the equipment or property OR environmental harm (aircraft, space, nuclear industries etc.),
- concerning any system especially important or mission-critical ones (financial, banking, real-time systems etc.) – to prove its correct behavior (or its correct results of work) and other formal properties,
- to check the properties of existing (already developed) systems in new circumstances – to be sure if previously known properties keep *true* for the system being migrated to another environment with changed characteristics (for example, transferring the system from single-CPU environment to multi-processor or multi-threading architecture).

So, methods for program properties check and proof are required in wide range of situations.

2 The Task and the Method

2.1 Problem Overview and Complexity

Classical questions of the program safety (partial correctness) and liveness (total correctness) are complicated by many contemporary programming techniques like object-oriented programming (OOP), concurrency, Web Services, closures and so on. For example, OOP actualizes type checking problem due to dynamical but strong typing, signature problems due to polymorphism and inheritance (plus virtual methods and dynamic linking) and visibility in the scope problems (or accessibility in the scope, i.e. kind of security). Many of them can be addressed to [1], where OOP is deeply investigated from the start and closely integrated into the method. Web Services usage, for example, arouses questions

on accessibility, reliability and discoverability (including semantic specifications like RDF) etc. But we will concentrate on concurrency and its' challenges in this work.

Main troubles of concurrency are non-determinism and inter-process communications complexity. The definitive difference between parallel and sequential worlds is that the state of program (shared memory data or messages in communication channels) can be changed between two consequent statements of one of concurrently running processes by another one, being executed in interleaving manner. In other words, the (global) state can be influenced by side effect of another concurrent process, which is impossible behavior for pure sequential programs.

Concurrency as the ability to execute more than one operation somehow at the same time can be formalized via the inter-process communication (IPC) mechanism. There exist two main approaches for IPC: message passing and shared memory. While the first approach is investigated in details, for example, in [5], the latter is rather less researched [2]. Moreover, the classical Hoare logic for reasoning on programs [6] can't be directly applied to interleaving shared memory concurrency because of possible side effect of parallel programs to each other [7].

2.2 Interleaving Concurrency with Shared Memory

Thus here we will concentrate on interleaving concurrency with shared memory environment, which is present de facto in many cases (runtime environments) in different industries and in various systems (hardware as well as software) all over us:

- hardware multi-processor and multi-core architectures (of course with shared memory), primarily SMP-architectures (Symmetric Multi-Processing),
- supercomputing, namely with UMA and NUMA (Uniform and Non-Uniform Memory Access) architectures,
- multi-tasking Operating Systems,
- software services and applications running in parallel on hardware with shared memory (like UMA, NUMA),
- servers (hardware as well as software ones, like a reactive system executing request – response cycles):
 - Web Server
 - Application Server
 - SQL Server (DataBase Management Systems as well as Data Warehouse Systems)

All samples mentioned above have common characteristics of environment:

- interleaving concurrency – means that control can be passed to any other concurrent pretender to CPU time at any time moment (also known as 'process context switch' system operation),

- shared memory – is the mean for IPC, where each process can write the values to and read the values from the so-called shared variables (at any time moment) in order to communicate with each other,
- it represents the most wide MIMD model (Multiple Instruction Multiple Data) in Flynn's taxonomy.

2.3 The Model of Environment

The formal model for interleaving concurrency with shared memory proposed in [8, 11] is the special subclass of IPCL (Interleaving Parallel Compositional Languages) [8]. Such programs will have the following form: $P_1^{k_1} || P_2^{k_2} || \dots || P_n^{k_n}$ where P_j – are sequential programs running in parallel (in interleaving concurrent mode) k_j times, and $||$ – is the interleaving concurrency operation sign. Proposed construction is an adequate model for program systems being executed in interleaving concurrent runtime environment with shared memory [8, 2].

We should notice also that proposed model follows the recommendations ("best practices") and considers warnings (avoidances) by C.A.R. Hoare [5] about parallel execution operation (like fork), that, in general, multithreading is an incredibly complex and error-prone technique, not to be recommended in any but the smallest programs and, thus, concurrency can be afforded only at the outermost (most global) level of a job, and its use on a small scale is discouraged.

2.4 Program Properties

The problem formulation and the method for program properties proof can be found in [8, 2]. Two types of properties defined there are:

1. Hoare-style: $\{PreCondition(S)\} Program \{PostCondition(S)\}$
2. Invariant-style: $Inv(S)$, which keeps *true* throughout all execution time

where $PreCondition(S)$, $PostCondition(S)$ and $Inv(S)$ are predicates over the current state (S).

The state S includes current "instruction pointers" (labels or marks) of each sequential sub-program running in parallel as well as the data – shared and local for every such sub-program. These two property types are the kinds of safety properties, while the liveness properties (first of all – termination problem) are out of our scope here because of its algorithmically undecidability in common case.

Here are some examples of such properties:

- race condition analysis,
- critical section condition,
- "transactionality" (integrity, consistency) of the system,

3 Known Alternative Approaches Overview

We should note that the right way which addresses software correctness problem is formal specification combined with stepwise refinement from an abstract model (specification) to the specific one (final code) with verification of these steps. In this case we obtain correct programs by construction. Testing techniques can be complementary only to other, more precise or formal methods, because of impossibility to test things totally. Formal methods can also be useful for proving properties of already existing systems – for example, while being transferred to another run-time environment with different characteristics or when we need to be sure in some (probably new) properties of the system and we can't (or don't want) replace current system (for some new system with these properties being proven).

Starting our consideration from famous Petri Nets regarding verification of parallel programs, we should note that this formalism was developed for concurrent processes behavior modelling and not for verification purposes. So, it can not cope with state-based properties check and proof, and fits for reasoning about flow-like properties only (data-flow or concurrent processes behavior).

Although there is a broad range of approaches to handle this issue – most of them have remarkable disadvantages in terms of using them on practice as they are too complicated or too theorized. Moreover, some of them are not applicable to cope with real tasks in general. For instance, without specifying the details, original Owicki-Gries method [9] requires the quadratic number of verifications relating to the program operators count. While the extended version of the rely-guarantee Owicki-Gries method [7] needs the implementation of additional variables as well as non-evident formulating of rely- and guarantee- conditions in order to tackle this task. In TLA [10] (Temporal Logic of Actions) Lamport offers to construct a model which is not much easier than the two previous ones. Moreover, TLA is characterized with a big difference between the program and its proving formula. In such a way IPCL [8,2,11] might be one of the most efficient solutions. While IPCL is up to solve the verification problem of the parallel software, in fact, it describes the so-called serializability mechanism. Though ultimately, we will work with sequential processes which steps will be interrupted by parallel running programs in an unpredictable way.

Many approaches have been adopted to deal with shared memory concurrency, but majority of them have different disadvantages [2]. One of the lacks is that known methods can be applied to a fixed program whilst we often have deal with the family of the same fixed structure program running arbitrary instances of it in parallel. So do Web and SQL servers, which run any arbitrary number of the same scripts (requests, queries) concurrently. So does operating system, executing various count of the same services and applications at any given time slice.

Consequently, we rather need the method for reasoning over the family of programs than the method which operates with a fixed program when the run-time environment is shared memory interleaving concurrency. In the latter case

we need to reason over many programs (each number of program copies running concurrently) instead of parameterized form of one of them.

4 The Method and Its Applications

The method for program properties proof in compositional nominative languages IPCL with interleaving concurrency and shared memory communication mechanism is described in [8, 2, 11] in details.

It operates over IPCL, which includes common compositions: ";", "if-then-else", "while-do", vector assignment operator $((a, b) := (c, d))$ and the interleaving concurrency "—", which has obvious structural operational (not pure compositional) semantics [8, 2, 11]. Because of it, IPCL is similar to any common universal (structural, imperative) programming language. Thus, every program (from C, PHP, C#, Java, JavaScript, even SQL etc.) could be rewritten in IPCL without substantial problems – to apply the method.

Main stages of the method are the following [8]:

- to label (to mark-up) the program in IPCL according to its syntax tree – each particular operator should have a label (a mark),
- to specify the notion of a state – the vector, which components include:
 - labels (marks) of every particular sequential sub-program of the whole IPCL-program,
 - the single shared data of the whole program,
 - and local data for every particular sequential sub-program (if present),
- to construct transition system (i.e. the model of the program execution, or operational semantics), which fix all possible transitions between states (really, macro-transitions or transitions schema because of its countable quantity) – this could be done algorithmically by the program syntax tree using labels and states fixed above [8],
- to fix start and final states of the transition system (of the program),
- to formulate the required precondition and postcondition over start and final states respectively,
- the main step is to formulate invariant of the system – it should incorporate the program behavior as a whole, in a single predicate, and it could be completed by human only for now,
- to prove that
 - each of macrotransitions keeps invariant true,
 - precondition on starting states implies invariant,
 - invariant on final states implies postcondition.

We will not go into details of the method, but concentrate on examples of the method applications.

Here we note that by the years the method mentioned above was successfully applied to prove some properties of the industrial software systems, which are the banking systems mainly. Let us discuss those examples.

Here we consider some of the method applications to the classical algorithms and tasks (for demonstration purposes) as well as to the commercial (industrial) software systems.

4.1 Methodological Samples

First, we would like to mention some "classical" examples:

- parallel addition to shared variable consistency [12],
- Peterson's Algorithm for mutual exclusion correctness proof [13].

In first article it is proven that $i := i + 1 || i := i + 1$ increases the value of i by 2. The main result is that the proof is two times shorter than in any other known formal techniques [12].

Correctness of the well-known Peterson's Algorithm (for mutual exclusion) is the subject of the latter paper. The fact that two concurrent processes can not appear in their critical section at the same time is proved in [13]. The proof again is a half of length of else known classical one (for example, in TLA).

More of that, the more generous task was discussed there – because the proof was made simultaneously over the family of each number of program run concurrently in first case, and it could be done in the second case.

4.2 Presentation System

Infosoft e-Detailing 1.0 is commercial presentation support system designed to hold the almost synchronous online+offline presentations with one lecturer (manager) and one-or-more listeners (clients). The usage of this system basically lies in switching slides on a manager's device which is almost immediately followed by an automatic switching to the same slide on each of the clients' devices connected to the current presentation session.

The most important from the correct functioning point of view is to make sure that every client will see the same slide that manager has switched to. Work of the system consists of cycles, namely switching a slide on manager's device and then switching a slide on all of clients' devices. The amount of such cycles is unlimited, the only stopping criteria is that everyone has left their presentation session. Typical cycle would look like this: manager sends to the server, and clients are reading from it, the index of a current slide (currently using HTTP + AJAX + Long Poll technologies) – in such a way the asynchronous slide replication is achieved on all the devices.

In [14] the safety property proof of Infosoft e-Detailing 1.0 software system using correctness proof methodology [8] in IPCL is presented. Operational semantics of the system with interleaving concurrency was described by means of transition system, program Invariant as well as Pre- and Post- conditions were formulated in accordance with the methodology. Application of the method to the real-world system and its usage simplicity were demonstrated in [14].

The detailed proof was given in [15] also. Thus partial correctness of an Infosoft e-Detailing 1.0 software system was proven.

4.3 Electronic Exchange

In the case with Currency Stock Electronic Exchange for Joint-Stock Company "State Oschadny Bank of Ukraine" compositional-nominative languages and the

method were used to formulate properties of the system (that each bid/ask offer is presented to stock exchange participants and that buy/sell deal is closed only once for each offer, with only one dealer, in despite of their concurrent work) and then to prove it [16]. We will not go into details here because of more interesting next example.

4.4 Remittances Payment System

Now let us consider the system for international remittances from Vigo Remittance Corp. paying out developed for Joint-Stock Company "State Oschadny Bank of Ukraine".

In this case the fact that no money remittance can be paid out twice needs to be proven [17]. The task was stated, transitional system according to the method was built for the model with simplified state (i.e. without local data), and the program invariant was formulated and proved to keep *true* over the software system at any given time. Conclusions about the convenience and adequacy of method application to prove the correctness of parallel systems were made as well as correctness property of the banking system for remittances payments was proven, namely that each money transfer can be authorized to pay out only once (due to authorization procedure) [17].

The simplified state [18] modification of the method is applicable when there is no local data within processes. In this case the state itself as well as the reasoning over the system can be significantly simplified [18].

5 Conclusions

Application of the method for concurrent programs, communicating via shared memory, properties proof to real-world industrial software was demonstrated in this work.

There are many questions left out of scope:

- software tools for automatic support for reasoning over such families of programs,
- more deep research of the proposed methodology through more industry samples (i.e. real-life software systems),
- the absence of native support of OOP in IPCL and compositional languages for now,
- non-compositional nature of interleaving concurrency etc.

But now we can affirm that the method developed is applicable and usable to proof the properties of real industry software systems.

Partial correctness of the software systems, namely Infosoft e-Detailing 1.0, Currency Stock Electronic Exchange for Joint-Stock Company "State Oschadny Bank of Ukraine" and the system for international money transfers from Vigo Remittance Corp. paying out according to an initial problem statement has been

proven using the correctness proof methodology [8, 11] in an IPCL language. Considering the difficulties in the process of such proofs in parallel environments, one can state:

- correctness proof method in IPCL is well suited for the verification of parallel programs or the software correctness proof in terms of safety properties;
- the method allows shortening the proof at the expense of choosing an adequate abstraction level [19] due to universality of a compositional nominative approach [19, 20] and by fixing the appropriate basic function set of semantic algebra.

Taking into account the flexibility of the methodology, existence of 'simplified state' model for reasoning in some cases [2, 12], and universal nature of the approach [21], it can be recommended for program properties proof (particularly safety property or partial correctness) for wide range of industrial software which is executed in interleaving concurrency environment with shared memory interaction, primarily for server-side software of client-server complexes, because of the method's usability and advances. The same conclusion was obtained in [14, 15, 17, 22] also.

6 Acknowledgments

Firstly, I would like to acknowledge my scientific mentors, Mykola S. Nikitchenko and Volodymyr N. Red'ko for their time of long discussions and attention to all my questions.

I would also like to thank the anonymous reviewers, who provided important comments and suggestions that improved this paper.

But especially I would mention and thank my dear co-authors of referenced works, namely Nataliia Polishchuk, Mykyta Kartavov, Yuliia Ostapovska and Andrii Zhygallo. These results were obtained due to their huge work as well.

References

1. The RAISE Method Group. The RAISE Development Method. BCS Practitioner Series. Prentice Hall. 493 p. (1995)
2. Panchenko, T. Compositional Methods for Software Systems Specification and Verification (PhD Thesis) [in Ukrainian]. Kyiv. 177 p. (2006)
3. Dijkstra, E. Why Correctness must be a Mathematical Concern. The Correctness Problem in Computer Science. By Boyer, R.S., Moore, J.S., (eds.) London: Academic Press. 1–8 (1981)
4. Clarke, E., Wing, J. et al. Formal methods: state of the art and future directions. ACM Computing Surveys. Vol. 28, no. 4. 626–643 (1996)
5. Hoare, C.A.R. Communicating Sequential Processes. Prentice Hall International. 238 p. (1985)
6. Hoare, C.A.R. An Axiomatic Basis for Computer Programming. Communications of the ACM. Vol. 12, no. 10. 576–583 (1969)

7. Xu, Q., de Roever, W.-P. and He, J. The Rely-Guarantee Method for Verifying Shared Variable Concurrent Programs. *Formal Aspects of Computing*. Vol. 9, no. 2. 149–174 (1997)
8. Panchenko, T. The Method for Program Properties Proof in Compositional Nominative Languages IPCL [in Ukrainian]. *Problems of Programming*. No. 1. 3–16 (2008)
9. Owicki, S. and Gries, D. An Axiomatic Proof Technique for Parallel Programs. *Acta Informatica*. Vol. 6, no. 4. 319–340 (1976)
10. Lamport, L. Verification and Specification of Concurrent Programs. A Decade of Concurrency. deBakker, J., deRoever, W., Rozenberg, G. (eds.). Berlin: Springer-Verlag. Vol. 803. 347–374 (1993)
11. Panchenko, T. The Methodology for Program Properties Proof in Compositional Languages IPCL [in Ukrainian]. *Proceedings of the International Conference "Theoretical and Applied Aspects of Program Systems Development" (TAAPSD'2004)*. Kyiv. 62–67 (2004)
12. Panchenko, T. Parallel Addition to Shared Variable Correctness Proof in IPCL [in Ukrainian]. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physical and Mathematical Sciences*. No. 4. 187–190 (2007)
13. Zhygallo, A., Ostapovska, Yu. and Panchenko, T. Peterson's Algorithm for Mutual Exclusion Correctness Proof in IPCL. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physical and Mathematical Sciences*. No. 4. 119–124 (2015)
14. Polishchuk, N., Kartavov, M. and Panchenko, T. Safety Property Proof using Correctness Proof Methodology in IPCL. *Proceedings of the 5th International Scientific Conference "Theoretical and Applied Aspects of Cybernetics"*. Kyiv: Bukrek. 37–44 (2015)
15. Kartavov, M., Panchenko, T. and Polishchuk, N. InfoSoft e-Detailing System Total Correctness Proof in IPCL [in Ukrainian]. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physical and Mathematical Sciences*. No. 3. 80–83 (2015)
16. Panchenko, T. Using the Formal Specifications for Development of Electronic Stock Exchange of Oschadny Bank [in Ukrainian]. *Problems of Programming*. No. 1-2. 161–167 (2002)
17. Ostapovska, Yu., Panchenko, T., Polishchuk, N. and Kartavov, M. Correctness Property Proof for the Banking System for Money Transfer Payments [in Ukrainian]. *Problems of Programming*. No. 2-3. 119–132 (2016)
18. Panchenko, T. Simplified State Model for Properties Proof Method in IPCL Languages and its Usage with Advances [in Ukrainian]. *Proceedings of the International Scientific Conference "Theoretical and Applied Aspects of Program Systems Development" (TAAPSD'2007)*. 319–322 (2007)
19. Nikitchenko, N. A Composition Nominative Approach to Program Semantics. Technical Report IT-TR: 1998-020. Technical University of Denmark. 103 p. (1998)
20. Redko, V. Compositions of Programs and Composition Programming [in Russian]. *Programming*. No. 5. 3–24 (1978)
21. Panchenko, T. Formalization of Parallelism Forms in IPCL [in Ukrainian]. *Bulletin of Taras Shevchenko National University of Kyiv. Series: Physical and Mathematical Sciences*. No. 3. 152–157 (2008)
22. Kartavov, M., Panchenko, T. and Polishchuk, N. Properties Proof Method in IPCL Application To Real-World System Correctness Proof. *International Journal "Information Models and Analyses"*. Sofia, Bulgaria, ITHEA. Vol. 4, No. 2. 142–155 (2015)

Safety-critical Certification of FPGA-based Platform against Requirements of U.S. Nuclear Regulatory Commission (NRC): Industrial Case Study

Vladimir Sklyar

National Aerospace University “KhAI”, Kharkiv, Ukraine
vvslyar@ukr.net

Abstract. Instrumentation and Control systems play important role in operation and maintenance of Nuclear Power Plants. A challenge in such systems implementation is certification and licensing against national safety regulatory requirements. A considered case describes certification of Instrumentation and Control platform of Ukrainian company Radiy against the United States nuclear safety requirements. General framework is described. Research activities of the project are presented.

Keywords: certification, licensing, FPGA, NPP I&C system

Key terms. MathematicalModeling, MathematicalModel, SoftwareSystems

1 Introduction

Nuclear energy is still one from the essential sources in energy agenda of many countries. In Ukraine, for example, up to 50% of energy is generated at Nuclear Power Plants (NPP). Instrumentation and Control (I&C) systems play important role in NPP safety and security assurance as well as in effective control of energy production. Safe and cost-effective operations of NPPs require the modernization of I&C systems to cope with obsolescence and age-related degradation. A computation core of the most parts of I&C systems are generic programmable platform also named as Programmable Logic Controllers (PLCs) [1].

Research and Production Corporation (RPC) Radiy (Kirovograd, Ukraine) has a long history of working with operating NPPs and installing new I&C systems during turn-key projects. RPC Radiy provides a wide variety of I&C solutions ranging from full-scope turn-key modernization projects to reverse engineering and printed circuit board-level. Also like-for-like replacements and equipment to solve ageing and obsolescence problems are implemented for both safety and non-safety applications. RPC Radiy uses Field Programmable Gate Array (FPGA) technology in its digital platform to implement customized solutions for NPPs I&C systems. RPC Radiy’s proven technological expertise has been demonstrated in over 100 systems installed to date.

RPC Radiy's I&C systems have been installed in safety related systems of all operating NPP sites in Ukraine and Bulgaria [2].

FPGA-based safety I&C platform RadICS is the main product of RPC Radiy. The RadICS Platform is designed to be functionally and physically similar to currently installed I&C systems. Such platform capabilities include input processing, customizable control logic, and output processing. The RadICS Platform continuously monitors system status through signals that are received from field sensors. It performs logic computations to create control commands. It also converts control commands to output signals that are applied to field actuators. The RadICS Platform has a modular and scalable design that can be configured to meet the needs of safety I&C applications in NPPs [3].

A big part of efforts to provide I&C systems for utility is licensing efforts. I&C systems have to comply with state-of-the-art standards. A licensing permission process lies in consideration of appropriate documentation related with I&C system design. A challenge is each country has its own set of regulation requirements. So I&C vendors face to make a new licensing case with penetration to any new market. For I&C vendors there are the most challenging licensing barriers at the United States (U.S.) nuclear market. From the one hand the U.S. operate more than one hundred reactors what is the biggest nuclear fleet in the world. From the other hand the U.S. Nuclear Regulatory Commission (NRC) implements one from the strongest regulatory framework in the world, and it is a reason why the biggest part of the nuclear community respects the U.S. NRC licensing permission.

In 2015 RPC Radiy started a project to certify RadICS platform against the U.S. NRC requirements. This paper contains description of the project framework and states some obtained results.

An essential part of the mentioned certification process is a deep University-Industry Cooperation (UIC) conducted between RPC Radiy and National Aerospace University (Kharkiv, Ukraine). Academicians are involved in all parts of RadICS platform certification what is one from the main factor of successful project running. Conclusion of this paper contains a list of researches directed to support of industrial certification activities.

Available references in the investigated area are mainly technical reports available from the U.S. NRC documentation system ADAMS (www.nrc.gov/reading-rm/adams.html).

2 FPGA-based I&C Safety Platform RadICS

The RadICS Platform is a state-of-the-art digital product specifically designed for safety-related control and protection systems of NPPs. A modular and distributed FPGA-based architecture is one from the RadICS Platform features. There is a set of general purpose building blocks (modules) that can be configured and used to implement application specific functions of I&C systems (see Fig. 1).

The basic architecture of the RadICS Platform consists of instrument chassis containing a logic module, as well as up to 14 other Input/Output (I/O) and fiber optic

communication modules. Logic module gathers input data from input modules, execute application specific logic, and update the value driving the output modules. Logic module is also responsible for gathering diagnostic and general health information from all I/O modules. The I/O modules provide interfaces with field devices (e.g., sensors, transmitters, and actuators). The functionality of each module is defined by the logic implemented in the FPGA that are part of the above modules. The backplane of the RadICS Platform provides interfaces to power supplies, process I/Os, communication links, and indicators. The internal backplane provides interfaces to the various modules installed within each chassis by means of a dedicated, isolated, point-to-point low voltage differential signaling (LVDS) interface.

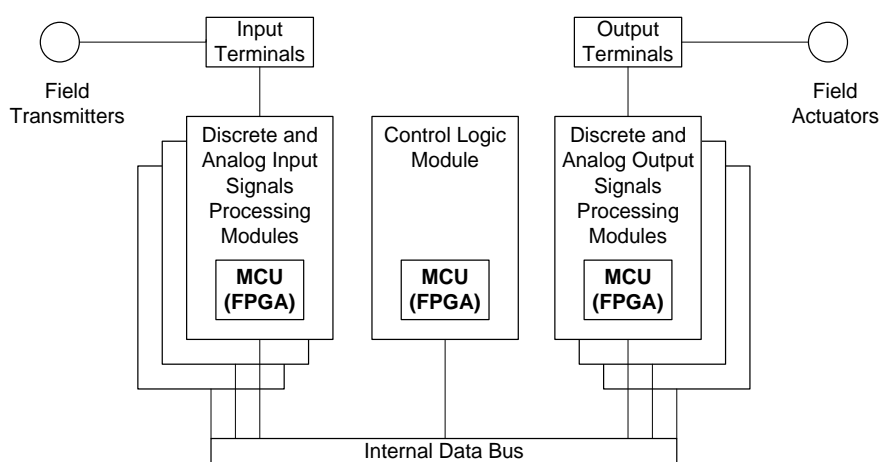


Fig. 1. Typical structure of Instrumentation and Control system based on Microcontroller Unit (MCU) or Field Programmable Gates Array (FPGA)

For application development, RPC Radiy provides a tool called Radiy Product Configuration Toolset (RPCT). This tool can be used to configure logic for various applications using the Application Functional Block Library (AFBL).

In addition, the RadICS Platform includes extensive on-line self-surveillance and diagnostics at various levels, including control of FPGA power, watchdog timer, cyclical redundancy check (CRC) calculations, and monitoring of the performance of FPGA support circuits, I/O modules, communications units, and power supplies.

Safety Life Cycle (LC) of the RadICS Platform is presented on Fig. 2. The RadICS Platform LC implements specific stages of FPGA design development and verification. Specific technique of fault insertion testing has been performed for both hardware and software parts. This LC complies with requirements traceability concept which requested the following:

- Every requirement has a child (either a lower level requirement or a solution);
- Every lower level requirement or solution has a higher level requirement (and opposite, an orphan represents unjustified functionality);
- Every requirement has been tested.

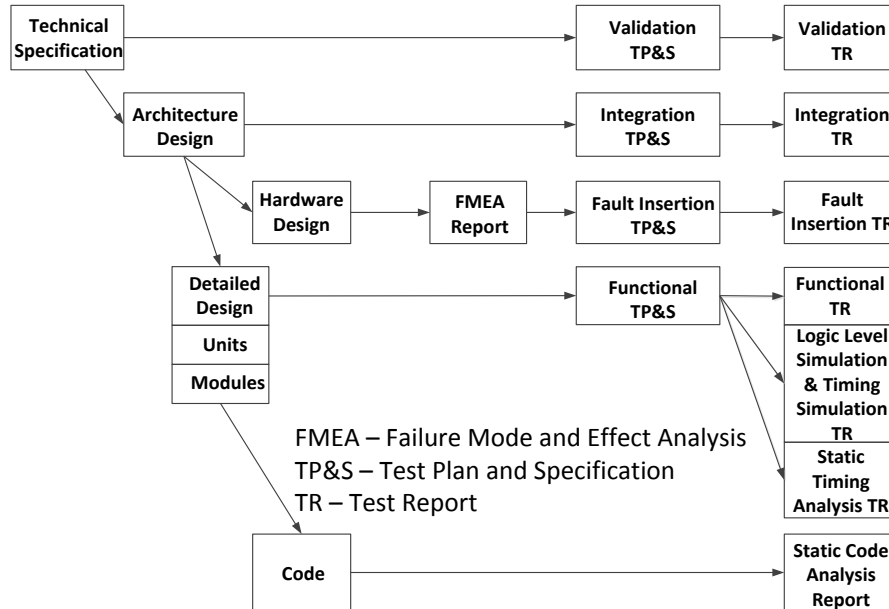


Fig. 2. Safety Life Cycle of FPGA-based RadICS Platform

3 General Approach to Certify I&C Safety Platform in the U.S.

Firstly RPC Radiy established RadICS in 2012, as a wholly owned Limited Liability Company (LLC). The Company RadICS business focus is the design and delivery of I&C systems for NPPs using the RadICS Platform equipment. Company RadICS, based together with RPC Radiy in Kirovograd, Ukraine, is responsible for all RadICS-based application project activities except the manufacturing of the RadICS Platform equipment. Such approach allows to focus certification efforts only on target processes for safety assurance.

After that the basic U.S. licensing strategy is to demonstrate that the generic RadICS Platform and the associated quality and software life cycle processes comply with U.S. nuclear safety requirements. As it is mentioned above, the RadICS Platform has been already licensed in Ukraine and Bulgaria. Difference of the U.S. regulatory requirements was analyzed. Licensing activities workflow was built on the base of results of such analysis, as it is shown in Table 1.

Actions to demonstrate compliance with the U.S. licensing requirements are presented at Fig. 3. An umbrella document for licensing activities is the Topical Report. Some details of the RadICS Topical Report are discussed in the next part.

It should be noticed, the U.S. NRC requires conducting technical meetings for discussion of the provided documents. Since the U.S. NRC has a specific philosophy of I&C systems consideration, additional research have been performed to fulfill this gap (see the Conclusions).

Table 1. Results of Analysis of the U.S. Licensing Specific Part

Difference in requirements	Actions to meet requirements
The U.S. NRC requires to implement a specific Quality Management Program (QMS) in accordance with regulations 10 CFR Part 50 Appendix B	Establish RadICS QMS in compliance with the U.S. NRC requirement. It should be noted U.S. NRC requirements to QMS are different from widely used ISO 9001 QMS, so many efforts are spent to adopt specific requirements in accordance with regulations 10 CFR Part 50 Appendix B “Quality Assurances Requirements for Nuclear Power Plants and Fuel Reprocessing Plants” and ASME NQA-1-2008, “Quality Assurance Program Requirements for Nuclear Facilities”. The RadICS QMS governs the system design, integration, and delivery of I&C systems for NPPs using the RadICS Platform equipment
The U.S. NRC requires to implement a specific process for Commercial Grade Dedication for as named commercial components. RadICS Platform components at the RPC Radiy site are commercial since they are produced under not nuclear ISO 9001 QMS	Dedicate the generic RadICS Platform, which was not originally developed under a 10 CFR Part 50 Appendix B QMS, in accordance with the basic requirements for Commercial Grade Dedication. RadICS is employing the commercial dedication processes described in Electric Power Research Institute (EPRI) Technical Report (TR)-107330 “Generic Requirements Specification for Qualifying a Commercially Available PLC for Safety-Related Applications in Nuclear Power Plants” and EPRI TR-106439 “Guideline on Evaluation and Acceptance of Commercial Grade Digital Equipment for Nuclear Safety Applications”
The U.S. NRC requires to submit the Topical Report which has to represent the main platform features	Develop the RadICS Topical Report. The purpose of the RadICS Topical Report is to demonstrate that the RadICS Platform and the associated quality and programmable logic life cycle process comply with NRC requirements
The U.S. NRC requires to submit a set of documents which support the Topical Report argument	Submit to the U.S. NRC the RadICS Topical Report with sets of relevant documents. These documents have to be submitted in three phases. The main part of such documents have been already developed before
The U.S. NRC requires to provide a representative (QTS)	Design, produce and dedicate from RPC Radiy the Qualification Test Specimen (QTS) with the with the Data Acquisition System (DAS)
The U.S. NRC requires to perform a set of qualification tests for the QTS	Perform QTS Equipment Qualification Testing in one from the U.S. testing laboratory recognized by NRC

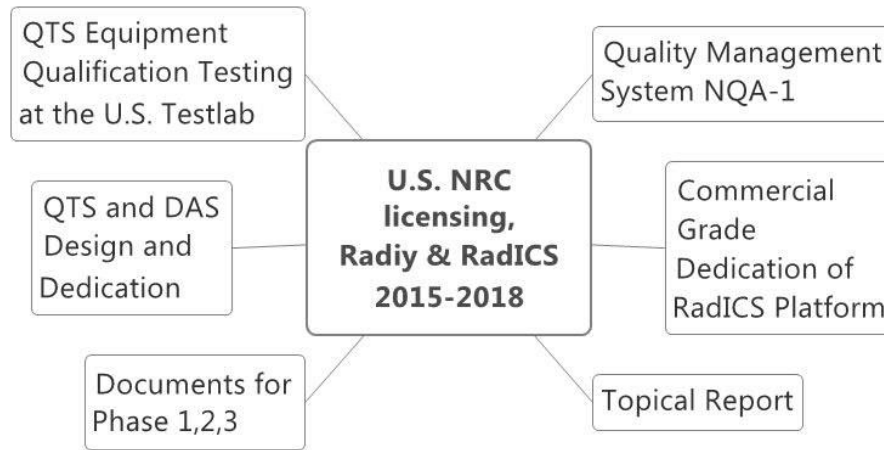


Fig. 3. Parts of Licensing Approach against the U.S. NRC Regulatory Requirements

4 Content of the Topical Report

The RadICS Topical Report is the summary licensing document for the RadICS Platform digital safety I&C platform. It presents design, performance, and qualification information for the RadICS digital safety I&C platform developed by RPC Radiy. The RadICS Platform is a generic digital safety I&C platform dedicated to the implementation of Class 1E (the highest safety class) safety I&C functions in the U.S. NPPs.

The RadICS Topical Report has been divided into 12 chapters and 3 appendices (see Fig. 4). The most important issues of this Topical Report are the following:

- An overview of RadICS development and operational use in international NPPs where it is currently deployed in a variety of digital safety I&C applications. This information is provided to illustrate the safety I&C developments that led to the RadICS Platform;
- An overview of the quality program and the quality process employed to dedicate the generic RadICS Platform hardware and associated programmable logic and develop systems for delivery to U.S. customers;
- An overview of the commercial grade dedication program used to dedicate the generic RadICS Platform hardware and associated platform programmable logic;
- A description of the RadICS Platform operation and how it can be applied in NPP safety-related applications. It also provides descriptions of the hardware and associated generic programmable logic that comprise the RadICS Platform. In addition, details are provided on how digital communications and testability are implemented in the RadICS Platform;
- A description of the hardware development process with associated planning documents and component testing process;

- A description of the RadICS Platform generic programmable logic development life cycle, planning documents, and the verification and validation process. The RadICS programmable logic life cycle processes were examined in more detail as part of the functional safety certification;
- An overview of the generic equipment qualification program for the RadICS Platform. The RadICS qualification “envelope” is designed to meet or exceed the environmental qualification requirements for NPPs in the U.S.;
- An overview of the RadICS approach to platform diversity;
- A summary of a RadICS Platform vulnerability analysis and the secure development and operational environment controls provided by RPC Radiy;
- A conformance summary of the RadICS design and development processes for the key regulatory guidance documents;
- The plant-specific system design guidance for use of the RadICS Platform, including recommended practices and any restrictions;
- A compliance matrix for the U.S. NRC regulatory document DI&C-ISG-04 “Highly Integrated Control Rooms – Digital Communication Systems” with the requirement listed as well as RadICS Platform compliance to each criterion defined;
- A list of the RadICS Platform design documents associated with the Electronic Designs for the RadICS Modules and identifies an initial set of documents planned for submittal to NRC to support the review of the RadICS Topical Report.

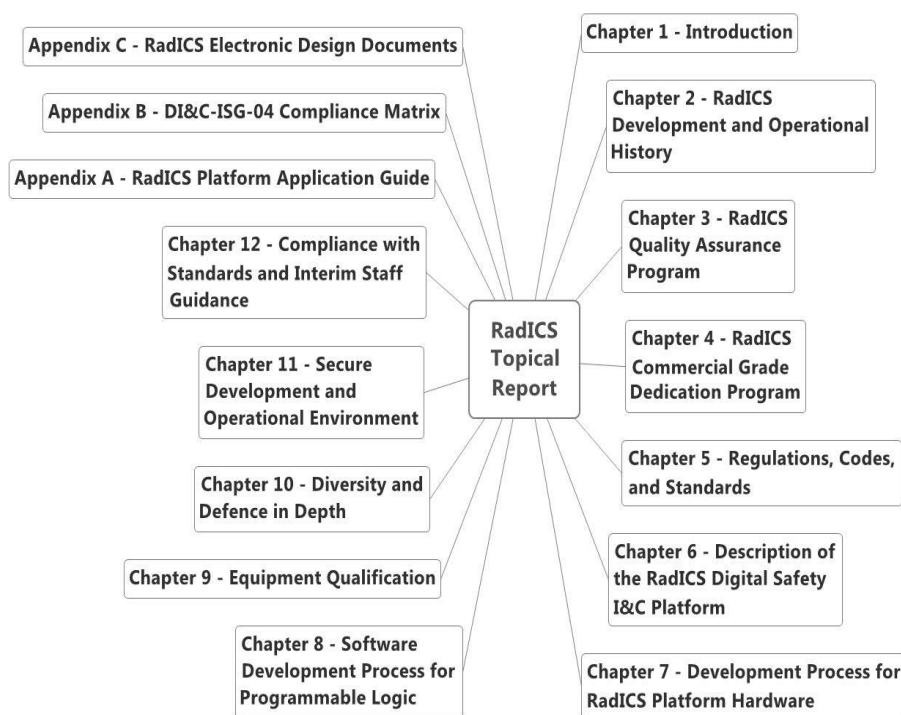


Fig. 4. A Structure of the RadICS Topical Report

5 Conclusions

As well as I&C platform is a high tech product, certification and licensing of such product requires innovative approach followed with research activities. A good basis for such researches provides UIC conducted between RPC Radiy and National Aerospace University. The following research directions have been chosen as priorities to support safety-critical certification:

- Research in FPGA and design tools safety features to support safe use of FPGA chip as a logic control core;
- Research in combination of different testing methodology with different coverage criteria to support effective verification and validation of I&C platform through life cycle;
- Research in efficient power consumption of I&C platform with optimization of used algorithms [4];
- Research in security of I&C platform to protect safety critical application from malware injection;
- Research in Safety and Security Case [5] methodology as a tool to support integral safety evaluation.

References

1. Kharchenko V.S., Sklyar V.V. (Edits). FPGA-based NPP Instrumentation and Control Systems: Development and Safety Assessment / Bakhmach E.S., Herasimenko A.D., Golovyr V.A., Kharchenko V.S., Rozen Yu.V., Siora A.A., Sklyar V.V., Tokarev V.I., Vinogradskaya S.V., Yastrebenetsky M.A. Research and Production Corporation "Radiy", National Aerospace University named after N.E. Zhukovsky "KhAI", State Scientific Technical Centre on Nuclear and Radiation Safety, 2008. 188 p.
2. Nuclear Power Plant Instrumentation and Control Systems for Safety and Security / Yastrebenetsky M., Kharchenko V. (Edits). – IGI Global. – 2014. – 470 p.
3. Multi-Version FPGA-Based Nuclear Power Plant I&C Systems: Evolution of Safety Ensuring by Vyacheslav Kharchenko, Olexandr Siora and Volodymyr Sklyar in the book "Nuclear Power - Control, Reliability and Human Factors" edited by Pavel Tsvetkov, Texas A&M University, USA, 2011.
4. Kharchenko V., Gorbenko A., Sklyar V., Phillips C. Green Computing and Communications in Critical Application Domains: Challenges and Solutions // Proceedings of the 9th Digital Technologies International Conference "DT 2013". – Žilina, Slovakia, May 29–31, 2013. – P. 241-247, on CD-ROM, ISBN 978-80-554-0682-4.
5. T. Kelly, Arguing Safety: A Systematic Approach to Managing Safety Cases, PhD thesis, Univ. of York, 1998.

Simulation model for computerized testing of learning success in quality management systems

Alexander N. Alexeyev ^{1,*}, Nataliya A. Konovalova ¹, Kateryna A. Lozova ¹,
Elena N. Korol ²

¹ Sumy State University, Faculty of Technical Systems and Energy Efficient Technologies,
Sumy, Ukraine

² Sumy State Pedagogical University, the Department of Preschool and Primary Education,
Sumy, Ukraine

alekseev.aleksandr.nik@gmail.com,
konovalova.nataliyall@yandex.ua, katarina_lozovaya@ex.ua,
korol.9@mail.ru

Abstract. A computerized test successfully complements and enhances traditional methods for the assessment of knowledge. This article introduces a simulation model of computerized testing of learning success that is complementary to the existing methods used for knowledge evaluation. The simulation model combines possibilities of computerized testing with mathematical rationale in examiner's decision-making during oral knowledge assessment. Application of the simulation model enables one to make mathematically precise decisions in the majority of standard procedures of test development, during computerized testing, and in the analysis of its results. The new features of computerized testing, introduced here within the framework of the simulation model, help diminish the limitations of computerized testing that arise from the impossibility of utilizing diagnostic potential of a human examiner in traditional testing procedures.

Keywords: simulation model, educational measurement, computerized testing of learning success, key stages of computerized tests, test tasks.

Key Terms: ICTTool, QualityAssuranceProcess, Teaching Methodology, Teaching Process, Technology.

1 Introduction

Fast and accurate evaluation of knowledge formation remains to be a relevant task for long-existing forms of learning. Moreover, it has become increasingly important for the comparatively recently emerged distance learning or blended learning (partial implementation of distance learning technologies into classes that are conducted traditionally). The most important characteristics of different forms of learning remains

* Professor of Department of Manufacturing Engineering, Machines and Tools, Doctor of Pedagogy., Associate Professor

the objective monitoring of students' academic achievements and the construction of effective teaching methods based on that.

The development of theory and practice of computerized tests makes it possible to increase the precision of pedagogical measurements as tools for objective knowledge control. Computerized testing carries out a number of pedagogical functions assigned to tests, hence becoming an effective means for evaluating the results of learning at all stages of education, from an entrance test to a comprehensive final exam.

This article describes a simulation model of computerized tests that on the one hand draws upon modern information and communication technologies and on the other is maximally reliant on the traditions of active participation of an instructor in students' knowledge assessment. Combination of the advantages of computerized testing with mathematical grounding of examiner's decision-making expands the range of effective applications of test-based knowledge assessment. The authors hope that the use of the simulation model developed by them will contribute to the further development of quality management systems at the institutions of higher learning.

2 Antecedents of simulation model for computerized testing of learning success

The classical period of the development of computerized testing theory to a great extent expanded the field of rational usage of computerized tests. In studies by A. Birnbaum, H. Gulliksen, G. F. Kuder, F. M. Lord, M. Novick, G. Rasch and others attempts were made to create an objective tool for observations in the fields of psychology, sociology, pedagogy, and other behavioral sciences. C. Spearman, one of the founders of the classical approach in the testing theory, proposed using methods of physical measurements in psychology. In pedagogy, this approach is called educational measurement. Increase in reliability of results of educational measurement in this period is due to the introduction into the testing theory of certain provisions of mathematical statistics, as well as of the elements of correlation theory aimed to justify the reliability and validity of the tests.

The 1970s witnessed emergence of a new direction in the theory of educational measurement – one connected with the Item Response Theory (IRT). Scholars J. B. Bjorner, B. Gandek, R. K. Hambleton, H. J. Rogers, S. J. Sinclair, M. H. Stone, H. Swaminathan, J. E. Ware, B. D. Wright et al. significantly contributed to the development of this new direction. The mathematical logistic models proposed by G. Rasch and A. Birnbaum were used to construct tests, or educational measurements. The goal of such measurements was an obtainment of numerical equivalents that were identified with the estimates of the measured variable. The measured variable was associated with the level of academic achievement, which was allowed in a certain way to reflect the latent parameter of the test-takers – their level of preparation.

The modern development of theory and technologies of educational measurement happens as a continuation of approaches founded in the previous period. The progress in development of the new testing methods has been driven by the applied and theoretical research of such scholars as F. B. Baker, R. Draney, G. R. Engelhard,

G. G. Kingsbury, D. J. Weiss, and M. Wilson. One of the most dynamically developing directions today is the one related to the design of adaptive tests, where new test questions are chosen based on one's performance on the previous questions. As the information and telecommunication technologies improve, the computerized testing of learning success becomes more and more prevalent in the theory and practice of educational measurement.

At the same time, many researchers (F. M. Bernt, A. C. Bugbee, D. C. Buhr, M. F. Johnson, S. M. Legg, K. C. Moe, R. Sutton and others) note the salient disadvantages of computerized tests that have not been resolved to date. Their findings, the results of our studies, suggest that testing designed based on most of the modern techniques still remains biased. Therefore, if no action is taken, the substitution of oral control with computerized control of learning success would not increase the reliability of educational measurement. Moreover, the exclusion of teachers from the monitoring process does not allow using the invaluable diagnostic capabilities of an instructor.

Nobody but an instructor, through conversation and additional probing questions, can determine whether a student's seemingly expressionless answer means the absence of knowledge on the subject or his or her mere nervousness. The instructor also has more opportunities to formulate questions not only by taking into account the student's responses to previous questions, but also depending on the content of the tested study material. For courses that require unconventional thinking and experiential approach, it is often difficult to create adequate and easily conveyed test questions. Hence such test questions frequently present difficulties for students. On the other hand, the fact that test design is still largely a subjective process also remains to be a problem. At the time of test creation, it is up to each of the test makers to decide upon the requirements for the number and complexity of tests to be included in a given assessment. Obviously, students with the same level of preparation are likely to score differently in such case, with students that had more simple test questions receiving higher grades than those whose test questions were more complex.

The objectivity of the results of computerized tests is also vulnerable to the inconsistency in the definition of evaluation criteria. It is certainly possible to introduce uniform requirements to testing. However, these might still be the same only for a given group of students, whereas in another group of students, or when tested by another instructor, a simple change in the grading criteria might change test results dramatically.

Therefore, with a steady ever-increasing usage of testing in knowledge assessment, there is a pressing need to create a model of computerized control of learning success that would utilize all the advantages of the testing method and would also maximally draw on the experience of active participation of instructor in diagnosing students' learning success, gained in the course of traditional knowledge assessment process.

3 Key stages of control

To solve this problem, the author's team has developed a simulation model of computerized control of learning success, which combines technological capabilities of computer-based testing with mathematical justification used in an instructor's decision-making procedures. In this diagnosis, the identity of the examiner is replaced, as much as it is possible, with his or her mathematical model.

The figure below shows a diagram of a multi-level computerized test, which has advanced measurement capabilities. Similarly to other approaches to the organization of testing procedure, the control is comprised of three phases: test design, test administration, and analysis of test results. The test design and analysis of test results phases rely on well-known theoretical positions, grounded in wide usage of statistical methods to increase of accuracy and objectivity of testing. In the test administration phase, mathematical methods that model diagnostic functions of an instructor are used to increase the reliability of results of educational measurement.

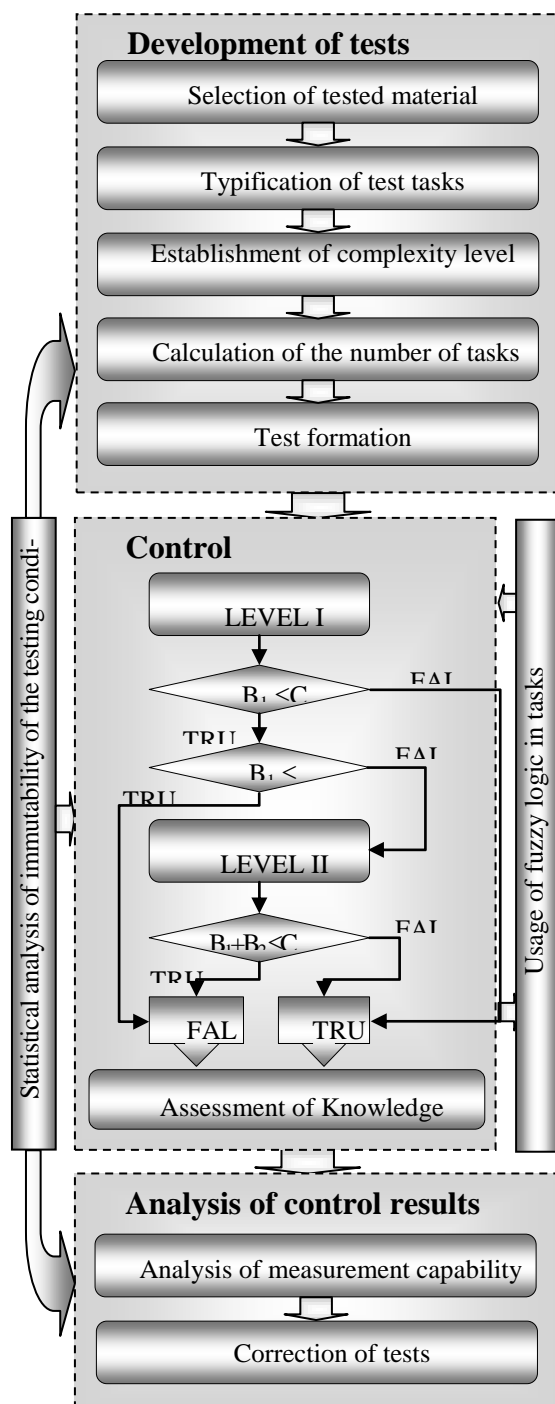


Fig. 1. Simulation model technological scheme

4 Test design

According to the scheme provided here, test design starts with the selection of test material. In this simulation model, it is supposed that this part of test design - similarly to many other testing methods - is done by experts that comprise a group of test makers. At the time of material selection, the experts are first-most guided by the ultimate goals of testing. In consideration of these goals, the experts decide upon the types of knowledge and skills that are most important for the goals set, as well as on the sufficient level of their demonstration by students.

After selecting the test content, test makers proceed to the design of test questions. The tested material is divided into separate parts, on which students can then be tested using sample test tasks. Provisions of the IMS Global Learning Consortium are placed at the basis of classification. These provisions are processed in such a way as to empower an instructor with more possibilities for formulation of test questions that would be maximally close to the content of the tested material. A total of 13 types of standardized test tasks are included into the proposed simulation model.

In addition to the recommendations of IMS, the simulation model contains special types of tasks that enable an instructor to check the extent to which the student's knowledge and skills have been formed. These include tasks on the control and sequence of actions. The design of a test task on control is a set of graphical images that reflect separate states of a certain object, and test takers are evaluated on their ability to manage it. The image shows targets, and the visible or invisible boundaries of these targets correspond to the contours of the object's organs of control. As one uses the pointer of a mouse to click on the required target, a graphical image of the object is substituted with a simulated control action. In a test task on the sequence of actions, object management happens with the mouse-click on control keys. In both types of test tasks, there is an option of setting an allowed interval of time between the mouse clicks.

The adjustment of the complexity of test tasks is possible through the procedures of design and corrective calculations, which are included into the simulation model. Expert assessment, which is accomplished using the method of paired comparisons, lies at the basis of design calculations (which are performed when prototypes of test tasks are created). Execution of such an expertise is most justified when a given test contains many tasks, and hence when it is difficult to preserve a single strategy and to have a comparable level of complexity for each of the tasks in the test. The corrective calculations procedure uses classical approach, which is based on the statistical processing of test results: expert grade estimates are refined taking into account students' performance on the test. It is assumed that the more students have answered a given test question incorrectly, the higher was its level of difficulty.

Once the level of test difficulty is determined, the test maker can move on to the next stage of test design: defining the necessary number of test questions in a given test. In the simulation model, the method of choosing a reasonable number of test questions is grounded in an assumption that it is important to account for both the quantity and the complexity of each task. Here, the total number of test questions is determined in such a way, that the cumulative complexity of one test would be com-

parable to that of another test. (For instance, in order to compare test results in physics and in chemistry, it is critical that the total complexity of tasks for the test in each of these subjects would be comparable.)

As a rule, test tasks have different levels of complexity. This is reflected in the assignment of unequal quantitative characteristics of test complexity measures. Since the tasks selected for a given test are chosen at random, while the method for calculating the number of tasks to be included into a test requires the tasks' cumulative complexity to remain constant, the authors recommend using genetic algorithms to design tests. In accordance with them, the process of test creation is seen as a successive change in the populations of species, whose genomes are random collections of test tasks of varying complexity. To generate different test versions (species of new populations) we apply operators of selection, crossover, mutation and survival. Such cyclical execution of operators is repeated until the total complexity of all tasks in a test does not reach optimal, i.e. as close as possible to the specified one.

5 Conducting iterative control measures

In the simulation model of computerized testing of learning success, the step during which the test is actually carried out is built on the basis of mathematical modeling of diagnostic capabilities of an examiner. Similarly to an oral testing procedure, in which an examiner can deem necessary to continue and ask a student additional questions which would help her determine the student's true level of knowledge, the simulation model provides for both basic and additional examination sessions. The procedure enables such a multilevel control via the employment of an apparatus of statistical analysis that resembles one used in engineering for the development of plans for the selective acceptance control.

Analogous to how the conclusion about the satisfactory quality of products that are manufactured in hundreds of thousands of pieces is made by means of an inspection of just a sample of them, the conclusion about the extent to which students' learning has been successful is evaluated by means of the statistical processing of the results of tests which have a limited number of questions. Comparison of the cumulative number of points received for the test with the values specified for the acceptance and rejection criteria makes it possible to make a final conclusion about the need to have an additional session of control.

If, upon completion of all tasks in the test, a student scores above the acceptance threshold, then his knowledge is evaluated as sufficient for a corresponding grade. Analogously, if a student knows the tested material worse than the rejection threshold, a conclusion is made that the student knows the tested material worse than the level of knowledge required for a given grade. However, when the number of points that a student receives for the test lies within the range of the pre-set acceptance and rejection values, the conclusion is made that it is impossible to determine the student's true level of knowledge and additional sessions of control are then carried out.

To expand the adaptive capabilities of a simulation model, the authors modernized the genetic algorithm for the selection of test tasks for additional test sessions. To

accommodate for such changes, a survival operator is altered and includes a criterion, which takes into account results of the preliminary test sessions. Here, the more poorly the student performs in questions on a certain topic in the previous sessions, the more likely is a question on this topic to show up in the additional testing session.

The testing methodology that is based on a simulation model stands out among most other existing methodologies in that, similarly to an oral exam, it enables a student to express their level of confidence in the correctness of a given answer in case the knowledge they possess does not allow them to give a definitive answer to the test question. Mathematical apparatus of fuzzy logic is used to make this functionality in the simulation model possible. A student that is being tested in the traditional way has to give a definitive answer to the test question by choosing one of several answer choices or by formulating their own answer choice using a limited set of words, letters, numbers, or graphical symbols. When giving an answer, a student has to formulate a response which would contain conclusions about the truthfulness of an expressed judgment using terminology of strict logic and hence has no way to express doubt or specify how far, in their opinion, the answer deviates from truth. Application of the fuzzy logic apparatus, on the other hand, allows a student to operate not only with the classical values of logical variables such as "false" and "truth", but also to use the interim values that smoothly transition from the one extreme value ("false") to the other extreme value ("truth"). This capability hence liberates a student from the necessity to make conjectures about an answer and go beyond their own knowledge on the topic. Such solution thus helps avoid introduction of an additional error into the results of computerized control of learning success.

In the computerized control of learning success nowadays, the prevalent methodology is one in which the resulting grade is assigned through a comparison of the total number of gained points with some linear and, less frequently, nonlinear scale of assessment. Grading scale in such an approach is typically set based on the probability of guessing the right answer or based on the expert assessments. However, both options are not the best ones for the creation of such a grading scale. In the first case, usage of such a scale would be justified if the probability of the randomly picked answer choice being correct does materialize: the student does not know the answer but happens to guess it correctly. Such a grading scale quickly becomes inaccurate if the probability of randomly selected correct answer does not materialize: the student actually knows the answer and hence responds correctly. In the latter case, the student's knowledge of the subject is underestimated in such a grading scale. On the other hand, empirical grading scales are not universal. Here, expertise assessments should be carried out maximally often since the continuously changing conditions, in which the knowledge is being gained, to a large degree predetermine the students' efforts at achieving a given level of knowledge. Therefore, the grading scale used in the simulation model is constructed based on the comparison of test results among students in the class. Similarly to oral testing, when an examiner that has to decide on a grade takes into account not only his assessment of the correctness and fullness of an answer but also other students' answers, the grading scale in the simulation model is based on the distribution of grades in the tested groups of students. To realize such

an approach this study adapted a method for building a five-point criteria scale for grading introduced by T.D. TenBrink.

Considering the fact that the change in the content of material covered in class or organizational and methodological supplements for it have a roughly the same effect on all the students, such method makes additional test sessions unnecessary for the conclusion of the assessment process in new conditions and the assignment of final grades. This is accomplished on the basis of the selective characteristics of the grade distribution parameters.

6 Analysis of test results

The mathematical rationale of the examiner's decision-making process mitigates significantly the disadvantages of computerized testing as of a tool for educational measurement. Additionally, the simulation model includes the stage for the analysis of test results, which rests on the traditional approaches. This stage includes procedures for evaluating measurement capabilities of individual test tasks and of the entire test using adapted for the use in simulation model indicators of distinctive capabilities and reliability. Furthermore, it is suggested to use the probability characteristics of impossibility of the extreme marks, as well as to use the specific for the simulation model criterion of abnormal amount of time spent on test completion. To identify the test items with an unsatisfactory measurement capability in the simulation model, the authors suggest using characteristics of impossible (more than 95%) probability of scoring only at the highest or only unsatisfactorily, and of impossible probability of abnormally spent time on completing the test. Distinctive capability of a test task is measured using the biserial correlation coefficient (discrimination index). The extent to which a test is reliable is characterized by the correlation of marks obtained for different parts of the test. (In the simulation model a change in the approach of dividing the test into parts was made: the selection of tasks is done at random, however in such a way that the total complexity of both parts of the test would be the same).

The level of knowledge and learning effectiveness are integrated indicators of many factors that influence the learning process. Students' performance on tests is dependent on the students themselves, on their instructors, on the methodological and organizational support of the learning process, as well as on other factors. Any changes made to the learning process, including changes to the procedures of knowledge control, can cause distortion to the statistical picture of test results. In the simulation model, most of the decisions rely on the statistical analysis of test results, and hence it is necessary to measure statistical significance of the changes that occurred in the course of the semester with a coefficient of reliability of statistical differences.

7 Conclusion

The simulation model for computerized testing of learning success makes it possible to make mathematically precise design solutions for the majority of standard procedures of test development, implementation and results analysis. The authors do not

deny the fact that any testing procedure, including one on the basis of the simulation model, cannot fully replace an expert examination board, in which subjective evaluation and pedagogical expertise of its individual members make it possible to give overall a fuller and more objective evaluation of each student's knowledge. However, such a method is not always possible in the conditions of today's computer-based learning. Creation of expert committees is further limited by economic considerations and is implemented in the rare cases when different supervisory committees are created to ascertain a student's inability to master a discipline, or in controversial cases, etc. Most universities are forced to find their own ways to make educational process in the environment of market relationships economically feasible and, based on the need to reduce expenses related to the educational process, increasingly switch to various forms of test-based knowledge control. The mathematical justification for the examiner's decision-making procedure, which lies within the framework of the model proposed here, will significantly mitigate the weaknesses of computerized testing as a tool for educational measurement.

8 References

1. Alekseev, A. N. Multi-level computerized control of learning success of quality of knowledge by quantitative parameters. *Computer modeling and new technologies*, 11(3), 43 – 45 (2007)
2. Alekseev, A., Aleksieieva, M., Lozova, K., Nahorna, T. Using Fuzzy Logic in Knowledge Tests ICTERI 2015: 11th International Conference on ICT in Education, Research, and Industrial Applications. – Lviv., 51 – 61 (2015)
3. Baker, F. B. *The Basics of Item Response Theory*, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD (2001)
4. Birnbaum A. Some Latent Trait Models and Their Use in Inferring and Examinee's Ability. In Lord F. M., Novick M. *Statistical Theories of Mental Test Scores*. Addison-Wesley Publ. Co. Reading, Mass, 397 – 479 (1968)
5. Bugbee, A. C., Bernt, F. M. Testing By Computer: Findings in Six Years of Use 1982 – 1988, *Journal of Research on Computing in Education*, Vol. 23, 1, 87 – 100 (1990).
6. Gulliksen H. *Theory of mental tests*. – N-Y. Willey (1950)
7. Hambleton, R. K., Swaminathan, H, Rogers H. J. *Fundamentals of Item Response Theory*. N-Y.: SAGE Publications (1991)
8. IMS Global Learning Consortium. Accessible at <http://www.imsglobal.org/question> Ten-Brink, T. D. *An educator's guide to classroom assessment* - Boston: Houghton Mifflin (2003)
9. Kuder, G. F., Richardson, M. W. The theory of the estimation of test reliability // *Psychometrika*, 1937, v.2, N3, 151 – 160 (1937)
10. Legg, S. M., Buhr, D. C. Computerized Adaptive Testing with Different Groups, *Educational Measurement: Issues and Practice*, 23 – 27 (1992)
11. Lord F. M., Novick M. *Statistical Theories of Mental Test Scores*. Addison-Westley Publ. Co. Reading, Mass (1968)
12. Moe, K. C., Johnson, M. F. Participants Reactions to Computerized Testing, *Journal of Educational Computing Research*, Vol. 4, 1, 79 – 86 (1988)
13. Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests* //With a Foreword and Afterworld by B. D. Wright.- Chicago: The Univ. Press (1968)

14. Spearman C. Correlation calculated from faulty data //British Journal of Psychology, Vol.3, 2, 271 – 295 (1910)
15. Sutton, R. Equity and Computers in the Schools: A Decade of Research, Review of Educational Research. Vol. 61, 4, 475 – 503 (1991)
16. Volkov, N. I., Alexeyev, A. N., Kochevsky, A. N. Composing of test using the software tool SSUQuestionnaire. Education Technologies on Electronic Platforms in Engineering Higher Education: symposium, Bucharest, 297 – 304 (2005)
17. Volkov, N. I., Alexeyev, A. N., Kochevsky, A. N. Software tool for computer-aided testing of students: types of questions Bulletin of the Petroleum and Gas University of Ploiest, 3 (I, LVII), 41 – 51 (2005)
18. Ware, J. E., Gandek, B., Sinclair, S. J., Bjorner, J. B. Item response theory and computerized adaptive testing: Implications for outcomes measurement in rehabilitation. Rehabilitation Psychology, 50, 71-78 (2005)
19. Weiss D. J., Kingsbury G. G. Application of computerized adaptive testing to educational problems // Journal of Educational Measurement, № 21, 361—375 (1984)
20. Wilson, M., Engelhard, G. R., Draney, R, Objective measurement. Theory into practice. England: Ab. Publ. Corporation, Vol. 4, 37 – 50 (1997)
21. Wright, B. D., Stone, M. H. Best Test Design. Chicago: Messa Press (1979)

Method of Evaluation of Electronic Educational Resources Quality for Conformity Assessment (Certification)

Dmytro Fedasyuk¹, Nataliya Hots¹, Mykola Mykyychuk¹, Viktor Kuts¹, Yuliia Dzikovska²

¹ Lviv Polytechnic National University, Lviv, Ukraine
natalia.gots@lp.edu.ua

² Lviv Polytechnic National University, Lviv, Ukraine
pastsvitlo@gmail.com

Abstract. In the article the expediency of quality assessment and certification of electronic educational resources of the higher education institution is justified. The order and scheme of internal certification of electronic educational resources are formed as a result of research. Certification includes the software-technical and complex expertises. Also this article includes a system of criterias for the complex expertise of electronic educational resources (criterias of structural, scientific and methodological assessment). The quality level evaluation method of electronic educational resources is proposed on the basis of the expert assessment of compliance with the requirements. In the article the extension of this algorithm by necessity to take into account the views of consumers and the opportunity to provide a recommendation for getting ISBN are provided.

Keywords: Electronic educational resource, electronic educational edition, certification, Qualimetry, ISBN.

Key Terms. TeachingProcess, KnowledgeManagementProcess, StandardizationProcess, Development.

1 Introduction

The formation of a highly qualified specialist in the modern university is based on student's necessity of processing large amounts of information. Today we can define the following global trends in the informatization of educational process:

- extension of usage of distance education technologies in all fields of training;
- appearance of e-books and e-courses industry;
- steady entry to market new suppliers of technologies for e-learning;
- standardization of electronic content and electronic educational environments;
- development of distant education system of network electronic universities.

The transition to individual learning, the growth of the self-learning, the introduction of distance education, the quality assurance of different forms of educational process need to be supported in the educational process by modern educational literature, which includes advanced scientific developments and innovative technical solutions of world-class.

Therefore, the introduction into the educational process electronic educational resources, that involves the formation of electronic educational and scientific literature fund in the university and electronic educational editions, is important. Electronic educational editions can be viewed as an automated training information system, which includes learning, methodical, research and reference materials for academic subjects. It allows for the use of them in the learning process based on the appropriate software.

It should be noted that the quality of education largely depends on quality of educational manuals and textbooks that are used by students. To control and provide the quality of electronic educational literature of university it is appropriate to certify the electronic educational resources, that means to evaluate their compliance with certain requirements on the content, structure, included information, functionality and training capabilities, advanced scientific statements. Therefore, the development of principles of certification of electronic educational resources with the following definition of their quality level is actual task.

2 Certification of electronic educational resources

Electronic educational resource is a set of teaching, research, information, reference materials and tools developed in electronic form and presented in media of any kind or placed in computer networks, which are reproduced by electronic and technical means and are necessary for effective organization of the educational process, as part that is related to its filling with the qualitative teaching and methodical materials.

2.1 General requirements for electronic educational resources

Information about electronic educational resources is contained in these normative documents [1,2,3].

However, it should be noted that these normative documents contain only general classification information. The widest information regarding the classification of electronic educational editions is presented in the Order of the Ministry of Education and Science of Ukraine [4], which reveals the definition of electronic educational resources, their types, procedure for development and implementation. But unfortunately, these documents do not contain any requirements for the filling and forming of electronic educational resources.

Further international requirements for electronic educational resources are considered.

In the development of international standards for information and communication educational means the cooperation between the International Organization for Stand-

ardization and other international organizations and committees, including the International Electrotechnical Commission, is important. International Standard ISO/IEC 19796: 2005 Information technology – Learning, education and training – Quality management, assurance and metrics [5] is the basis for the description, comparison, analysis, quality management and quality assurance approaches in this area. It serves as a tool for comparison of existing approaches and their coordination based on the total quality model. The key element of the standard is the Reference Framework for the Description of Quality Approaches.

Standard ISO/IEC 19796: 2005 consists of three parts:

- ISO/IEC 19796-1:2005 – Part 1: General approach – the first step towards building a harmonized model for learning quality based on IT;
- ISO/IEC 19796-2 – Part 2: Harmonized quality model – defines the tools and metrics in order to implement a common approach to quality;
- ISO/IEC 19796-3:2009 – Part 3: Reference methods and metrics – expands the boundaries of the Reference Framework for the Description of Quality Approaches, mentioned in the previous part of Standard ISO/IEC 19796-1:2005. It is done by the introduction of harmonized description of the methods and metrics recommended for quality assurance's system implementation, quality management of stakeholders, who are engaged in the design, development, utilization of IT systems for learning, education and training.

But unfortunately, most of the requirements of ISO/IEC 19796: 2005 apply only to the requirements about the quality of design, development, utilization of IT systems for learning, education and training. Therefore we can say that international standards also do not reveal the concept of quality of electronic educational resources, do not include requirements to ensure it.

As noted above, the introduction of electronic educational resources in the educational process is a worldwide trend today. There are various methods in the literature for their quality control [6,7,8,9,10]. In particular, their common disadvantages can be considered:

- the formation of too many requirements which complicates the assessment and determining to what extent evaluation is satisfactory;
- the duplication of requirements content;
- presented methods for monitoring and testing do not systematize the requirements;
- methods of structural expertise allow to evaluate only the structure, not substantive, methodological and scientific filling of electronic educational resources;
- complexity of algorithms for evaluation.

But all these methods are aimed at one goal – to control the compliance of electronic educational resources to specific requirements. This process is called certification in the technical regulation. It is an action of a third party who proves that provided the necessary confidence that the duly identified product, process or service correspond to a specific standard or other normative document [11].

Therefore it is advisable:

1. To ensure and to control the quality it is necessary to form a list of requirements for electronic educational resources and to develop the methodology for internal certification of electronic educational resources in the university based on provisions of existing certification rules.
2. To implement the comparability of the various components of electronic educational resources it is necessary to develop a methodology for evaluation of their quality level on the basis of the theory of Qualimetry as evaluating the degree of conformity of electronic educational resource to established requirements.

Among the existing list of components of electronic educational resources in the normative documents we consider that it is advisable to carry out certification for the following: electronic textbook, electronic learning manual, electronic teaching-methodical complex, electronic display didactic materials (video lessons) and electronic distance learning course.

2.2 Algorithm of certification procedure of electronic educational resources

The purpose of certification of electronic educational resources is to be consistent with current scientific and technical requirements with the subsequent formation of the fund of the educational and scientific literature for the needs of the educational process of the university. At the stage of formation of electronic educational resource there is it's full filling using electronic teaching and methodical resources completely accordance with the criteria of software-technical, educational, scientific and methodical expertises.

Certification of electronic educational resources is offered to carry out in two stages: the first stage includes a preliminary examination, the results of which the expert's opinion is consisted; the second stage involves complex expertise, based on the result of which an expert certification committee decides on the certification of electronic educational resources and prepares recommendations for its entry in the register of electronic educational resources of the university. The procedure for certification of electronic educational resource, that is designed based on [3,4], generally involves the following steps according to Figure 1:

1st stage:

- document preparation and submission of an application for certification of electronic educational resources;
- previous (software-technical) expertise, documentation analysis;
- expert's opinion regarding the performance of electronic educational resources.

2nd stage:

- complex examination of electronic educational resources (structural, scientific and methodological expertises);
- evaluation of the degree of compliance with the requirements;
- analyzing of the obtained results and deciding on the possibility of issuing a certificate of conformity;

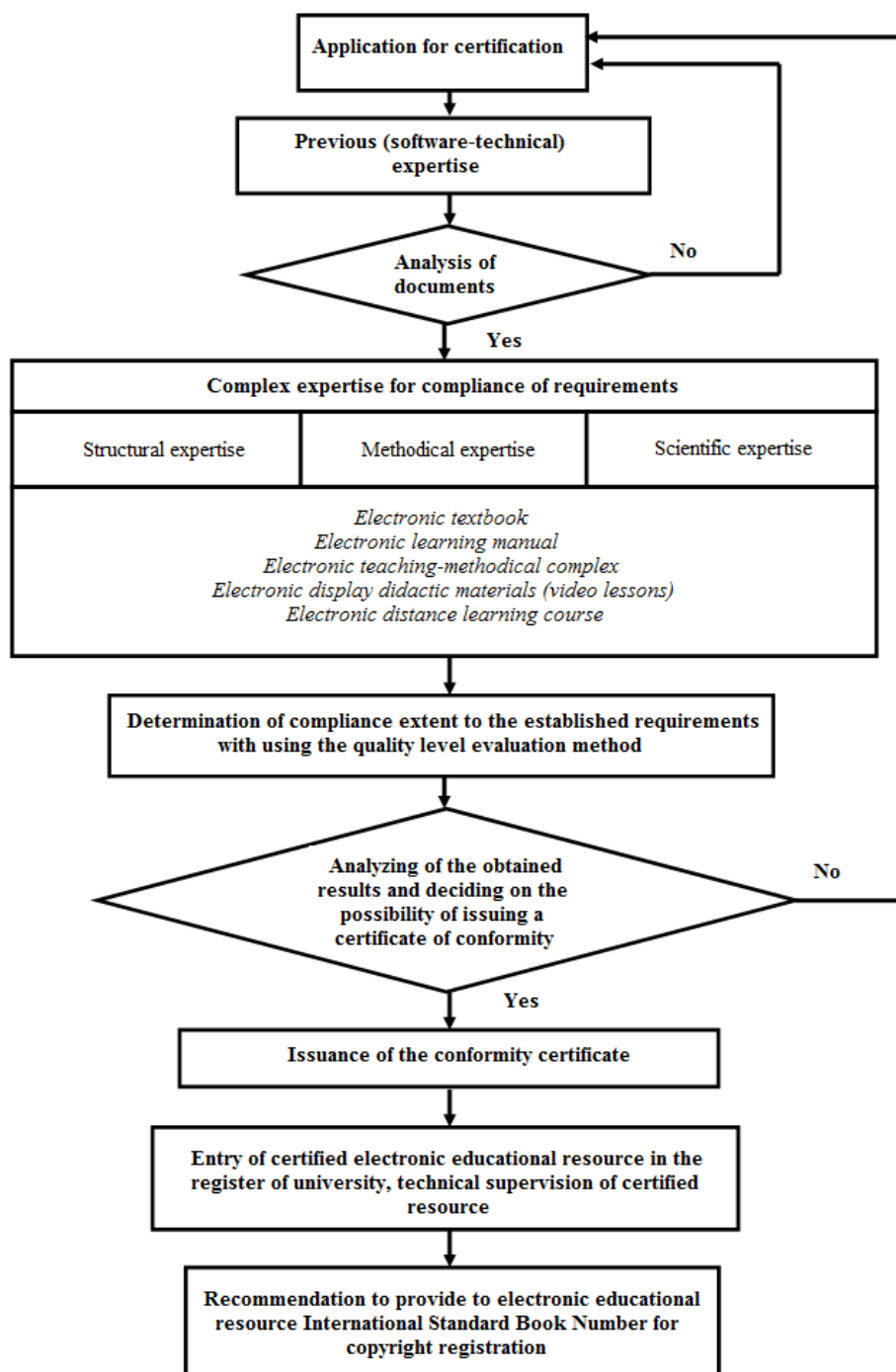


Fig. 1. Procedure of certification of electronic educational resource

- issuing the conformity certificate and entering certified electronic educational resources in the register of teaching electronic educational resources;
- technical supervision of certified electronic educational resource;
- informing about the results of certification.

The content and requirements of each expertise have been developed.

For each of these expertises are developed point scale of assessment of the implementation of the requirements (according to the theory of Qualimetry their volume is 7) and their respective weight coefficients.

To assess the degree of compliance with the requirements it is offered a method that is based on theory of Qualimetry [12], the method of differential quality indicators, taking into account the weight coefficients and using the arithmetic mean weight and weighted ratios parameters for results of certain types of expertise by the model:

$$\left\{ \begin{array}{l} K = \sum_{i=1}^N n_i \cdot k_i, \quad \sum_{i=1}^N n_i = 1, \quad N \leq 7 \\ k_i = \sum_{j=1}^J n_j \cdot k_{ij}, \quad \sum_{j=1}^J n_j = 1, \quad J \leq 7 \\ k_{ij} = \frac{P_{ij}}{P_{ij}^{bas.}} \quad or \quad k_{ij} = \frac{P_{ij}^{bas.}}{P_{ij}} \end{array} \right. \quad (1)$$

where K – generalized indicator of quality; k_{ij} – relative indicators of quality for certain types of expertise; k_i – group indicator of quality; n_i , n_j – weight coefficients; $P_{ij}^{bas.}$, P_{ij} – indicators of quality for basic and the existing levels.

Relative indicators of quality are defined for each parameter in all types of expertise and $\max(k_{ij}) = 1$.

Group indicators of quality are defined for each group of parameters of separate expertise and $\max(k_i) = 1$.

Generalized indicator of quality contains all information about the individual values of relative indicators of all parameters for each group and $\max(K) = 1$. As this is ideal option and in reality some of relative indicators can be less than 1, the decision of positive result of certification of electronic educational resource is offered to accept when the value of generalized indicator of quality is within: $0,8 \leq K \leq 1$ according to Harrington's scale. When $0,63 \leq K < 0,8$, electronic educational resources is sent for revision to correct deficiencies. Electronic educational resource is removed from the list of resources, that are certified, if $K < 0,63$.

The developed method allows the assessment of the degree of compliance of electronic educational resource to established requirements (based on expert evaluation) and identify important indicator – the level of quality.

2.3 Evaluation of the views of consumers about the quality of electronic educational resources

Assessment of quality of electronic educational resources is impossible without taking into account the views of consumers, including people, who are trained. Therefore, we developed example of questionnaire, in which customer requirements are formed. Questionnaires can be made at the end of the study course. The results can be attached to the methodology of evaluation of the conformity degree of electronic educational resources to established requirements.

3 Ownership on the electronic educational resource

As a result of a positive outcome for the certification of electronic educational resources, recommendation for getting the International Standard Book Number (ISBN) can be provided. ISBN is universal identification code that is put on the books and brochures, on different data storages regardless of their mode of production, distribution, circulation and volume. ISBN accompanies a publication from the time of its manufacture, allows the getting into the international database – the Global Register of publishers.

ISBN is also provided to the electronic publications in accordance with SSU/DSTU 7157: 2010 [1], Instruction on providing procedure of the International Standard Book Number in Ukraine [13] and the International Standard ISO 2108: 2005 Information and documentation. International Standard Book Number [14].

Also, it is necessary to add, that today practically the one way to protect the intellectual property in Ukraine is the registration the copyright by State Intellectual Property Service to obtain appropriate certificates for electronic educational resources according to [15,16]. Although registration is not a condition of the legal protection of work, but gives the right to judicial protection, that is the right to file legal action for copyright infringement.

4 Conclusion

Electronic educational resources make possible to present systematically the teaching materials, to make it more accessible for study and open to adjustment and further improvement. Using of electronic educational resources in the educational process will help to provide technologically the process of individualization of learning, the process of implementation of distance education, informatization of educational process in higher educational institutions. The proposed method allows the estimation of the quality level electronic educational resource based on the results of the examination. The result of the certification of electronic educational resources should be formation of electronic fund of modern educational and scientific literature in the university as part of the quality assurance system of educational process for various forms of learning.

In addition, this algorithm is universal and can be applied in any educational institutions, including during the training or retraining of staff, in particular during the certification of training programs for the specialists of nondestructive thermal control using thermal imaging technology.

References

1. SSU/DSTU 7157: 2010 Information and Documentation. Electronic Editions. Main Types and Output Information (Інформація та документація. Видання електронні. Основні види та вихідні відомості). – Effective as of 2010-07-01. – K.: State Committee of Ukraine for Technical Regulation and Consumer Policy (Держспоживстандарт України), 2010. – 13 p.
2. SSU/DSTU 4861:2007 Information and Documentation. Edition. Output Information (Інформація та документація. Видання. Вихідні відомості). – Effective as of 2009-01-01. – K.: State Committee of Ukraine for Technical Regulation and Consumer Policy (Держспоживстандарт України), 2009. – 45 p.
3. SSU/DSTU 3017-95 Edition. The Main Species. Terms and Definitions (Видання. Основні види. Терміни та визначення). – Effective as of 1996-01-01. – K.: State Committee of Ukraine for Technical Regulation and Consumer Policy (Держспоживстандарт України), 1995. – 47 p.
4. Order of Ministry of Education and Science, Youth and Sports of Ukraine from 01.10.2012 # 1060 On Approving Regulation of Electronic Educational Resources (Наказ Міністерства освіти та науки, молоді та спорту України від 01.10.2012 № 1060 Про затвердження положення про електронні освітні ресурси) [Electronic resource]. – Access mode: <http://zakon5.rada.gov.ua/laws/show/z1695-12>. – Title from screen.
5. ISO/IEC 19796-1:2005 Information technology – Learning, education and training – Quality management, assurance and metrics – Part 1: General approach [Electronic resource]. – Access mode: http://www.iso.org/iso/catalogue_detail?csnumber=33934. – Title from screen.
6. Quality Criteria for Digital Learning Resources [Electronic resource]. – Access mode: https://iktsenteret.no/sites/iktsenteret.no/files/attachments/quality_criteria_dlr.pdf. – Title from screen.
7. Regulations on Certification of Educational Electronic Editions for Higher Education Programs (Положение о сертификации учебных электронных изданий по программам высшего образования) [Electronic resource]. – Access mode: <http://istu.ru/component/jdownloads/viewdownload/5/1580?Itemid=0>. – Title from screen.
8. Kravtsov, G.M.: Structure of Quality Management System of Electronic Teaching Resources of University (Кравцов Г.М. структура системы управления качеством электронных ресурсов обучения университета) [Electronic resource]. – Access mode: http://lib.iitta.gov.ua/745/1/KravtsovH_issue10.pdf. – Title from screen.
9. Kravtsov H. Methods and Technologies for the Quality Monitoring of Electronic Educational Resources [Electronic resource]. – Access mode: http://ceur-ws.org/Vol-1356/paper_109.pdf. – Title from screen.
10. Regulation on electronic learning resources of HNTU (Положення про електронні навчальні ресурси ХНТУ) [Electronic resource]. – Access mode: <http://kntu.net.ua/index.php/ukr/Setevye-resursy/Elektronne-navchannya>

HNTU/Normativne-zabezpechennya/Polozhennya-pro-elektronni-navchal-ni-kursi-HNTU.
– Title from screen.

11. DSTU ISO/IEC 17000:2007 Conformity assessment. Glossary of terms i general principles (ISO / IEC 17000: 2004, IDT). – Effective as of 2008-04-01. – K.: State Committee of Ukraine for Technical Regulation and Consumer Policy (Держспоживстандарт України), 2008. – 20 p.
12. Kutz, V.R., Stolyarchuk, P.G., Druziuk, V.M.: Qualimetry: Tutorial (Кваліметрія: навчальний посібник / В.Р. Куць, П.Г. Столярчук, В.М. Друзюк). – Lviv: Publishing House of Lviv Polytechnic National University, 2012. – 256 p.
13. Instructions on the International Standard Book Number (ISBN) in Ukraine (Інструкція про порядок надання Міжнародного стандартного номера книги (ISBN) в Україні) [Electronic resource]. – Access mode: http://www.ukrbook.net/zakony/instr_isbn.pdf. – Назва з екрана.
14. ISO 2108:2005 Information and documentation – International standard book number (ISBN) [Electronic resource]. – Access mode: http://www.iso.org/iso/catalogue_detail?csnumber=36563. – Title from screen.
15. On copyright and related rights. – Law of Ukraine, Resolution of Verkhovna Rada number 3792-12 of 12/23/1993 // Verkhovna Rada of Ukraine (Про авторське право і суміжні права. – Закон України: постанова Верховної Ради № 3792-12 від 23.12.1993 р. // Верховна рада України) [Electronic resource]. – Access mode: <http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=3792-12>. – Title from screen.
16. The Order of State Registration of Copyright and Agreements Concerning Copyright to Work, Cabinet of Ministers of Ukraine of 27.12.2001 # 1756 // Verkhovna Rada of Ukraine (Порядок державної реєстрації авторського права і договорів, які стосуються права автора на твір: постанова Кабінету Міністрів України № 1756 від 27.12.2001 р. // Верховна рада України) [Electronic resource] – Access mode: <http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=1756-2001-%EF>. – Title from screen.

Hybrid Cloud-Oriented Educational Environment for Training Future IT Specialists

Olena G. Glazunova, Tetyana V. Voloshyna

National University of Life and Environmental Sciences of Ukraine, Heroiv Oborony St. 15,
03041, Kyiv, Ukraine
(o-glazunova, t-voloshina)nubip.edu.ua

Abstract. The article considers the issue of creating and utilizing a hybrid cloud-oriented environment for training future IT specialists in higher educational establishments. It solves the problem of designing and creating such an environment, which will provide effective development of IT students' professional competencies as well as Soft Skills. The proposed model of hybrid cloud-oriented environment of a higher educational institution is verified experimentally. Methods of its usage in the educational process was developed.

Keywords. Hybrid cloud-oriented educational environment, Soft Skills, cloud technologies

KeyTerms. ICT Environment, Information Communication Technology, Teaching Methodology, Teaching Process

1 Introduction

The current situation of the development of higher vocational education is associated with the transition to practical implementation of a new educational paradigm that aims to create an integrated system of lifelong learning, to increase student self education in learning process by means of information and communication technologies (ICT), which form self education competence and such skills as self organization and self education.

The professional level of training IT specialists of higher education establishments meets new demands within dynamic development of an information society. A modern IT developer must not only possess professional knowledge and skills, but also gain the so-called "soft skills" (Soft Skills). The study conducted at Harvard and Stanford universities showed that only 15% of career success is provided by the level of professional skills, while the other 85% is Soft Skills [6]. P. Moss and C. Tilly define skills as abilities and qualities of a person, attitude and behavior but not formal

or technical knowledge [16]. Modern employers in IT industry expect that a candidate will have dozens of different skills such as the ability to think creatively and to manage time, communication skills, networking, project management, effective teamwork. It is possible to form professional skills and Soft Skills in future IT specialists in a traditional classroom. A great deal of methods and technologies can solve the problem of effective training of future IT developers, including blended learning, flipped classroom, problem learning, project method and so on. These methods usually use ICT, e-learning environment (ELE), web resources.

The aim of higher education in IT is to provide quality fundamental knowledge that can serve as a solid foundation for rapidly building commercial IT variables and technological superstructures. It is sometimes difficult for higher educational establishments around the world and their faculty to keep up with new trends, to track changes in commercial technologies of companies-vendors. Programming languages and IT technologies (C#, Java, JavaScript, Python, HTML5 and others) that are supplied by commercial companies develop fast and often unpredictably. There are two tasks before university IT education that will help to ensure proper training of students of IT specialties; one of them is continuous professional development of teachers and to involve certified training centres into learning process, the second is to create a learning environment through which students will be able to develop their professional skills and soft skills.

This article discusses the issue of ELE for training future IT specialists, in particular the design of ELE based on cloud technologies as well as the issue of the efficiency of its use.

2 The Presentation of the Main Research and Explanation of Scientific Results

Ukrainian scientists Bykov V. Yu., Bohachkov Y.M., Panchenko L.F. and others studied the issue of information and education environment of an educational establishment. In particular, Bykov V.Yu. defines the concept of e-learning environment as a kind of learning environment that is specifically designed as simulation and formulated, educational and cognitive, organizational, technological and information and communication environment, which provides necessary and relevant conditions for the effective achievement of the objectives of e-pedagogical systems [1, c.169].

Panchenko L.F. [5, c.78] defines information and education environment of a university as an open multidimensional pedagogical reality that includes psychological and pedagogical conditions, modern information and communication technologies and learning techniques and provides interaction, collaboration, personal development of teachers and students while solving educational problems.

E-learning as "the information environment of an educational establishment which is built on the integration of information data on electronic media, information and communication technologies of interaction that include a virtual library of full-text electronic resources, media materials, structured e-learning courses, which are used

on the basis of a new educational system, media for collaboration and learning management system” considers in [7].

G. McCray came to the conclusion that courses which combine online learning with traditional classroom learning enable students to become more active participants of interactions, using different learning styles by providing a variety of content [13].

K. Graham, in his book “The Blended System of Education: Definition, Current Trends and Future Directions”, gives mainly idealized definition of “hybrid” or “blended” learning environment that is blended learning is an approach to combine different methods of training and resources, to utilize them and to accept them in an interactively meaningful learning environment. Students must have an easy access to a variety of learning resources to apply knowledge and skills with the support of a teacher in the classroom or outside it [13].

Synchronic hybrid environments are technologies of full learning environments which allow students to interact online and in the classroom with each other as well as with an instructor [22].

Foreign researchers Sneha D. and Naharaya J. define virtual learning environment as a system to transfer educational materials to students through Web. These systems contain every student’s profile, means of communication, assessment of tasks and cooperation. It may be available inside and outside the campus 24 hours a day, seven days a week [20].

Tools of virtual learning environment (VLE) support e-learning by giving access to training materials, links, online tools (e.g., electronic notice boards and chat rooms), administration and evaluation tools [9].

The Committee of common information systems defines VLE as a set of components, which enable teachers and students to participate in online interactions of various types, including online training. VLE changes the way students study specific subjects [9].

The most popular, convenient and efficient technology of such an environment is the technology of cloud computing. National Institute of Standards and Technology (NIST) defines cloud computing as follows “Cloud computing is a model to provide a convenient “on demand” access to Internet so that information resources e.g., networks, servers, storage, applications and services were easily accessible with minimal effort to manage and to interact with the supplier” [17].

There are two main types of cloud infrastructures. That is internal and external. In an internal cloud, servers and software are used inside the system in order to form a scalable infrastructure that meets the requirements of cloud computing. In an external cloud environments, providers offer services at the request of an educational establishment. IT support, services and experience will be included in the package, which must work only in providing applications and services.

Services for educational cloud computing represent a growing number of relevant services available online, and is the most innovative and fastest growing element of technology and education. It also promises to provide several services, which will be very useful for students, faculty and staff [18].

The role of cloud computing in higher education should not be underestimated, as higher educational establishments can benefit in getting direct access to a wide range of different academic resources, research programs and manuals [8].

Kiran Yadav suggests the following benefits of cloud computing for educational establishments and students [122]:

1. The personal approach to learning. Cloud computing allows a student to have more options in learning. Using an Internet connected device, students can have access to a wide range of resources and software that meet their interests and learning styles.

2. Reduced costs. Cloud services can help educational establishment to reduce costs and accelerate new technologies to meet the changing needs of education. Students can use free office applications, install and maintain these programs upgraded on their computers but at the same time it provides some commercial applications.

3. Availability. Availability of services is the most important and desirable point for the user who uses educational cloud technologies. You can log in and access the necessary information from any place.

4. The absence of additional infrastructure will increase the number of research centers available for students and will create a global learning environment. There is no need in spending time to think about classrooms and laboratories.

Cloud computing environment provides the necessary foundation to integrate platforms and technologies. It integrates teaching and research resources, which have places, using existing conditions as much as possible to meet the demands of teaching and learning [21].

The term “academic cloud” becomes more and more popular which [7] defines as information and communication technology of education which is built in the principles of cloud technologies and aims at providing education services at educational establishments. “Academic cloud” of a university is a cloud-oriented environment of an educational establishment which combines technical, software and technological, information resources and services and which functions on the basis of technologies of cloud computing and provides academic process of a university by means of a local network of an educational establishment and Internet.

Higher educational establishments mostly use hybrid cloud environments to organize learning process by integrating internal and external cloud. Thus, hybrid cloud-oriented educational environment of a higher educational establishment is the system that combines academic cloud of an educational establishment and external academic clouds based on integration of resources into the educational environment of an educational establishment.

At National University of Life and Environmental Sciences of Ukraine a hybrid cloud-oriented environment was designed to train IT major bachelors. This environment combines internal and external platforms (Fig. 1).

EEE of the university provides IT students with:

- Electronic learning course (ELC) for every subject;
- Electronic books;
- Software to do practical and laboratory activities by means of a virtual desktop;
- Environment to improve practical skills in programming (automatic system ejudge).

The main element of this environment is e-learning course (ELC) based on CLMS system platform Moodle, which places different types of learning resources [4, 15]. To teach IT specialists using a virtual learning environment it is necessary to upload

academic videos, video tutorials, video lectures and other video resources (<http://video.nubip.edu.ua>). To provide students with academic and research activities, the university has institutional knowledge repository that contains full-text electronic academic and research resources. It is available at elibrary.nubip.edu.ua and can be used by students for self-study. Students have access to a virtual desktop via appropriate links for laboratory or individual work. With virtual desktop DaaS users are able to access necessary applications. All resources, which support every subject, are integrated into an ELC. Efficiency of such an environment studies in [7]; this research states the efficiency increases by 6%, the consent - by 12%, individual work - by 8%, motivation - by 17%.

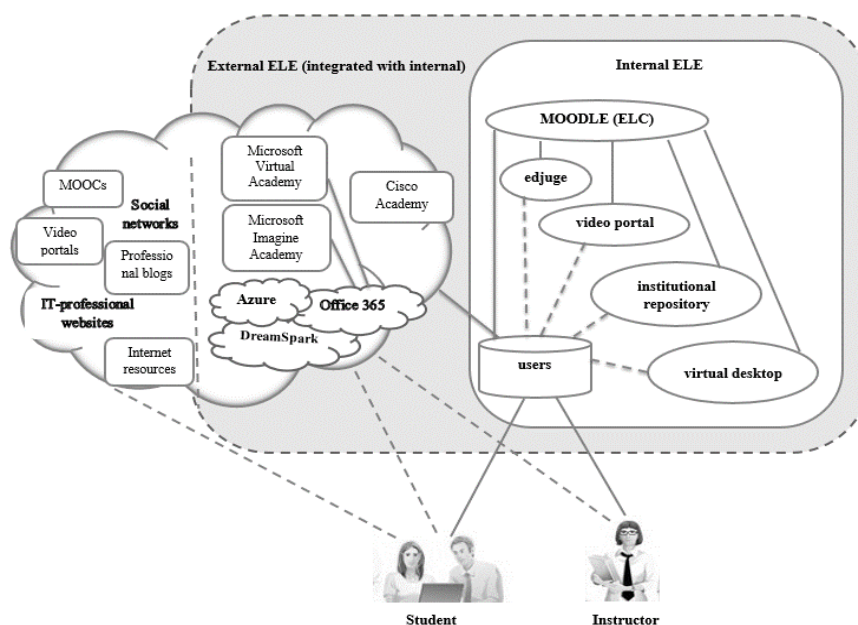


Fig. 1. The Model of hybrid cloud-oriented environment of a higher educational institution

The incentive of constant practice plays a significant role during the process of training future IT specialists in programming languages and standard algorithms. Therefore, automated system ejudge was integrated into ELE of the university which enables students to get a significant amount of programming tasks as individual work and thus provides automated assessment of their progress.

A systematic use of external academic clouds such as Microsoft, Cisco, IBM is significant to form professional skills and Soft Skills in a future IT specialist. The NULES of Ukraine has a license contract with Microsoft Enrollment for Education Solutions. Students and faculty have access to cloud service Microsoft Office 365 which gives access to different software and services on the platform Microsoft Office, business class email, function for communication and management. Besides, students are advised to use a virtual academy Microsoft Virtual Academy (MVA), educational portal, where there is available interactive academic course in

programming (Fig. 4), complements development, Windows Server 2012, Windows 8, visualization and complements developments for HTML5, Windows i Windows Phone, Microsoft Office365, SQL Server, AzureSystem Center and Microsoft Imagine Academy. To provide students with learning software we have access to Microsoft DreamSpark, that enables students to get a free access to tools of projecting and developing software. Platform Microsoft (Windows) Azure enables students and faculty to develop, doing software in storing data which are primarily placed in distributed data servers.

3 Results of Experimental Research

Students of such specialties as “Computer sciences”, “Programming Engineering” participated in this pedagogical experiment. The pedagogical experiment predicted that students of a control group were offered electronic course to study programming, each theme of this electronic course was presented as a resource “Lesson” which is a structured succession of pages. It is possible to place texts, graphics, video, tests etc there. Students were also offered a resource “Video lesson” which as [11] states is the most effective type of resources for students who study IT. This resource was in the form of screen cast of a certain program or practical implementation of software code of scripts with obligatory texting and audio, which is built according to a certain script. The use of this resource enables students to take academic material individually and, if necessary, to revise the performance that is demonstrated in Video Lesson. All this resulted in achieving maximum effect by using all sources of perception and assimilation of information: visual, auditory and kinetic [14].

The experimental group 1 studied using both electronic academic course and ejudge. Students had access to their own tasks and individually solved problems; results then were sent to be assessed. The assessment was automatically done by the system on predefined criteria. It should be noted that students had several attempts to download tasks, but each next attempt reduced a number of points. Also, the tournament had a fixed time to do tasks. The student could view their progress after doing every task, namely whether it was successfully done, the number of solutions, the number of tests. In the case when the task was successfully done a student could see the number of test that they failed. As a result, automated system ejudge evaluated a student’s work according to criteria such as the number of fully done tasks; the number of tasks with one mistake; the number of tasks with two or three mistakes; the number of errors that exceeded the runtime which indicated that students selected the wrong algorithm (Fig. 2).

Experimental group 2 had additional access to Microsoft Office 365, did a distance course in programming at a Microsoft Virtual Academy, Cisco and had a wide range of professional blogs, communities of IT specialists in social networks, open electronic resources to study programming, different Internet resources.

Courses of network academy Cisco gave students an opportunity to learn functioning of hardware and software components, structure of networks, security problems and methods of solving them, obtain skills to collect and set up a computer, to install operating systems, software, and to identify and correct errors connected

with hardware and software (Fig. 3).

Submit a solution										
Настройки Регистрация Инфо Итог Посылки Положение участников										
Отправить вопрос Сообщения										
16:56:53 / RUNNING										
Место	Участник	A	B	C	D	E	F	G	Всего	Штраф
1	KI_P131user	-2		+7			+2	+3	3	469
2	KI_P12user			-5			+	+1	2	184
3	KI_P16user			+1			+2		2	228
4	KI_P14user			+2			+		2	235
5	KI_P16user						+3	+2	2	372
6	KI_P17user	-1					+	-2	1	51
7	KI_P17user	-1	-3	-4		-2	+1		1	67
8	KI_P18user						+1		1	83
9	KI_P10user						+1		1	105
10	KI_P11user			-1			+1		1	112
11	KI_P15user						+1		1	128
12	KI_P13user			+					1	129
13	KI_P12user	-3					+		1	150
14-20	KI_P13user	-1							0	0
14-20	KI_P18user								0	0
14-20	KI_P12user	-3							0	0
14-20	KI_P15user						-1		0	0
14-20	KI_P16user			-2					0	0
14-20	igl								0	0
Total:		11	3	26	0	2	25	11	78	
Success:		0	0	4	0	0	12	3	19	
%:		0%	0%	15%	0%	0%	48%	27%	24%	

Fig. 2. The Page of Tasks results

Cisco Networking Academy

Розділ 3: Монтаж комп'ютера

3.0 Монтаж комп'ютера

3.0.1 Вступ

3.0.1.1 Вступ

3.1 Монтаж комп'ютера

3.1.1 Відкривання корпусу комп'ютера та встановлення блоку живлення

3.1.2 Встановлення системної плати

3.1.2.1 Встановлення ЦП, блоку радіатора і вентиляторів

3.1.2.2 Встановлення ОЗП (RAM)

3.1.2.3 Встановлення системної плати

Монтаж комп'ютера

Вступ

Монтаж комп'ютера є важливою частиною роботи інженера. Вам, як інженеру, слід використовувати логічний і методичний підхід до роботи з компонентами комп'ютера. Іноді вам буде потрібно визначити, чи потребує комп'ютер замовника оновлення або заміни. Важливо, щоб ви розвивали професійні навички з встановлення компонентів, діагностики несправностей і їх усунення. У цій главі йдеться про важливість сумісності обладнання з програмним забезпеченням. Крім того, в ній розглядаються відповідні системні ресурси для роботи з обладнанням та програмним забезпеченням замовника.

Fig. 3. Course «IT Essentials: PC Hardware and Software»



Fig. 4. Course «Introduction to Programming with Python»

Using social networks, IT specialists are able to obtain new knowledge individually because they have open access to professionally-oriented information that is covered in magazines, newspapers, books, videos, blogs, etc., to fast share information with peers who are users of social networks and have common professional interests; to discuss issues on information technology. In addition to social networking sites, there are special professionally orientated in IT sites which contain a large number of manuals, code samples, links to download software, discussion forums, blogs, etc.

The outcomes of progress in “Algorithmic and Programming” of control and experimental groups were measured by means of tests; individual work and motivation was measured by means of observation and surveys. The results of the experiment are presented in Table 1. According to the results of the experiment, individual work increases significantly when students solve problems, fulfil other tasks. Students of experimental groups are more motivated and ready to solve non-standard tasks.

Table 1. The results of the experiment using the system of training of future IT specialists

Indicator	Control group	Experimental group 1	Experimental Group 2
Academic progress (average), maximum – 100	64,8	77,5	79,6
Individual work (high, intermediate, low), %	17; 35; 48	29; 47; 24	48; 35; 17
Motivation (high, intermediate, low), %	15; 55; 30	34; 58; 8	44; 48; 8

So, this hybrid cloud oriented environment for students of IT specialties, which combined possibilities of electronic learning environment of the university (internal) and external services of Microsoft and Cisco, where the university gained its part of “academic” cloud (externalities) made it possible to develop Soft Skills together with

developing professional skills, namely personal effectiveness (group 1) and communication skills (group 2) according to the classification by Dluhonovych N.A. [2]. Managerial and strategic skills were also developed in groups 1 and 2. If we add to this classification critical thinking skills and information management skills (Group 5), which Indian researcher V. Saravanan [19] highlights, we will receive Soft Skills which are presented in table 2.

Table 2. Classification of Soft Skills

Personal progress (group 1)	Communicative skills (group 2)	Managerial skills (group 3)	Strategic skills (group 4)	Skills to manage the information (group 5)
Ability to set and to achieve the set goals	Ability to team work	Ability to organize a group	Strategic planning	Critical thinking
Time management	Ability to communicate effectively	Ability to form a team	To make strategic solutions	Ability to analyze the information
Resistance to stress	Inter social communication	Ability to form a system of communication in a team	Ability to take risks	Ability to synthesize data
Responsibility	Ability to solve conflict problems	Ability to motivate team members	Ability to delegate responsibility	Ability to evaluate information
Creativity	Ability to held talks	Development of leadership skills		Ability to make decisions
Analytical thinking	Ability to persuade	Formal and non formal leadership		Lifelong learning skills
Ability to present	Ability to make group decisions			

The task to identify a number of indicators of students' personal progress, communicative and managerial skills was set as well as the task to identify the ability of managing the information according to the classification of Soft Skills in Table 2. In particular, in order to determine whether the student is able to manage their time, the groups were given tasks to accurate state the type, beginning time and finish time of the work. In order to identify formal and informal leaders in the group, their abilities to form a group a sociometric technique developed by J. Moreno was used [3]. Students were offered to answer some questions of a socio metric card; the number of options was limited. According to these results the index of grouping was identified which revealed internal emotional atmosphere of a group. Students of

Experimental groups 1 and 2 demonstrated qualities to turn the idea into the ability faster than Control Group. It shows more developed features of personal progress, communication, ability to influence the surrounding people, the ability to foresee the outcome, to manage the process.

Conclusions

The conducted research resulted in designing and utilizing a hybrid cloud oriented environment that integrates the components of university academic cloud such as e-learning courses, electronic tools and electronic manuals, video resources, virtual desktop and environment for automated assessment of tasks in programming; with academic components of Microsoft and Cisco clouds and external cloud services. The efficiency of such hybrid clouds while teaching IT students programming was tested by means of the pedagogical experiment which showed as effective progress (in average by 14%), so the development of Soft Skills necessary for career success of future IT specialists.

References

1. Bykov V. Modeli orhanizatsiinykh system vidkrytoi osvity: monohrafiia / V. Bykov. – K.: Atika, – 648 s. (2009)
2. Dluhunovych N. A. Soft skills yak neobkhidna skladova pidhotovky IT-fakhivtsiv / Visnyk Khmelnytskoho natsionalnoho universytetu, №6 (219) s. 239-242 (2014)
3. Moreno Dzh. Sotsyometryia. Эспериментальный метод в наука об obshchestve. M: "Ynostrannaia lyteratura" (1958)
4. Morze N. V. Hlazunova E. H. Sposoby yntehratsyy tsyfrovoho kontenta v elektronnom uchebnom kurse [Elektronnyi resurs] / N. V. Morze, E. H. Hlazunova // Obrazovatelnye tekhnolohyy y obshchestvo. – Tom 17. – № 1. – S. 537–545. – Rezhym dostupu: http://ifets.ieee.org/russian/depositary/v17_i1/pdf/14.pdf (2014)
5. Panchenko, L.F. Informatsiino-osvitnie seredovyshche suchasnoho universytetu : monohrafiia / L. F. Panchenko. – M-vo osvity i nauky Ukrainy, derzh. zakl. "Luhans. nats. un-t im. Tarasa Shevchenka". – Luhansk : LNU imeni Tarasa Shevchenka. – 279 s. (2010)
6. Sotsialni kompetentsii molodykh fakhivtsiv: rezultaty realizatsii mizhnarodnoho proektu TEMPUS "IMPRESS" [Elektronnyi resurs]. – Rezhym dostupu : <http://www.edu-trends.info/tempus-impress/>
7. Teoretyko-metodychni zasady proektuvannia ta zastosuvannia systemy elektronnoho navchannia maibutnikh fakhivtsiv z informatsiinykh tekhnolohii v universytetakh ahrarnoho profilu [Tekst] : dys. d-ra ped. nauk: 13.00.10 / Hlazunova Olena Hryhorivna ; Natsionalna akademiia pedahohichnykh nauk Ukrainy instytut informatsiinykh tekhnolohii i zasobiv navchannia.– 450 ark. – ark. 164-166 – Kyiv (2015)
8. Bell J., Sawaya S., Cain W. Synchronodal classes: Designing for shared learning experiences between face-to-face and online students // In International Journal of Designs for Learning. – 5 (2014)
9. Currier Sarah. Libraries and e-learning: be inspired by INSPIRAL // Library and Information Research News. – 26(82), pp. 4-15. Spring (2002)

10. Garrison R. D, Kanuka H. Blended learning: Uncovering its transformative potential in higher education // *The Internet and Higher Education*. – Volume 7, Issue 2, 2ndQuarter, S. 95-105. (2004)
11. Glazunova O.G. Types of academic internet-resources for it students' individual work management / O. G. Glazunova, T. V. Voloshy`na // *Informacijni tehnologiyi v osviti*. – Vy`p. 21. – S. 78-86. (2014)
12. Kiran Yadav. Role of Cloud Computing in Education // *International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization)* Vol. 2, Issue 2, February. –s. 3108-3112 (2014)
13. McCray G. E. The hybrid course: Merging on-line instruction and the traditional classroom // *Information Technology and Management*. – 1. – S. 307-327 (2000)
14. Morze N. Glazunova O. Designing of Electronic Learning Courses For IT Students Considering the Dominant Learning Style / N. Morze, O. Glazunova // *Information and Communication Technologies in Education, Research, and Industrial Applications / Communications in Computer and Information Science / – Volume 469*. – Springer. – P. 261–273. Rezhym dostupu: <http://link.springer.com/chapter/10.1007/978-3-319-13206-8> (2014)
15. Morze N., Glazunova O. What Should be E-Learning Course for Smart Education [Електронний ресурс] / Natalia V. Morze, Olena G. Glazunova // *ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer / CEUR Workshop Proceedings*. – Vol-1000. – P. 411–423. – Rezhym dostupu : <http://ceur-ws.org/Vol-1000/ICTERI-2013-MRDL.pdf> (2013)
16. Moss P., Tilly C. Soft skills and race: An investigation of black men's employmentproblems. *Work and Occupations*. 23(3).– S. 252 (1996)
17. National Institute of Standards and Technology [Elektronnyi resurs]. – Rezhym dostupu : <http://www.nist.gov/>
18. Richard M. Felder, Rebecca Brent. The abc's of engineering education: abet, bloom's taxonomy, cooperative learning, and so on // *American Society for Engineering Education*. – Session 1375. (2004)
19. Saravanan V. Sustainable Employability Skills for Engineering Professionals / *The Indian Review of World Literaturein English*, Vol. 5 No.II – July. (2009)
20. Sneha J.M, Nagaraja G.S.Virtual Learning Environments –A Survey // *International Journal of Computer Trends and Technology (IJCTT)*. –volume 4, Issue 6. – s. 1705-1709 (2013)
21. Srinivasa Rao, Ratna Kumari Challa. Adoption of Cloud Computing In Education and Learning // *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 2, Issue 10, s. 4160-4163 (2013)
22. Webb H., Gill G., Poe G. Teaching with the case method online: Pure versus hybrid approaches // *Decision Sciences Journal of Innovative Education*. – 3 (2). – S. 223-250 (2005)

Information, Communication, and Modeling Technologies in Prosthetic Leg and Robotics Research at Cleveland State University

Yuriy Kondratenko^{1,2}, Gholamreza Khademi¹, Vahid Azimi¹, Donald Ebeigbe¹,
Mohamed Abdelhady¹, Seyed Abolfazl Fakoorian¹, Taylor Barto¹,
Arash Roshanineshat¹, Igor Atamanyuk², and Dan Simon¹

¹ Department of Electrical Engineering and Computer Science
Cleveland State University, Cleveland, Ohio, USA
d.j.simon@csuohio.edu, y.kondratenko@csuohio.edu

² Department of Intelligent Information Systems
Petro Mohyla Black Sea State University, 68-th Desantnykiv str. 10,
54003 Mykolaiv, Ukraine
y_kondrat2002@yahoo.com

Abstract. This paper analyzes the role of information and communication technology (ICT) and computer modelling in the education of engineering students. Special attention is paid to research-based education and the implementation of new modelling methods and advanced software in student research, including course work, diploma projects, and theses for all student categories, including Doctoral, Master's, and Bachelor's. The paper concentrates on the correlation between student research and government priorities and research funding. Successful cases of such correlations with specific description of computer modeling methods for the implementation of prosthesis and robotics research projects are presented based on experiences in the Embedded Control Systems Research Laboratory in the Electrical Engineering and Computer Science Department in the Washkewicz College of Engineering, Cleveland State University, USA.

Keywords: ICT, computer modeling, research-based education, student project, prosthesis research, governmental priority

Key Terms: Academia, Research, MathematicalModeling, ComputerSimulation, Experience

1 Introduction

Information and communication technologies (ICT), mathematical modeling, and computer simulation play a significant role in higher education. Most advanced educational systems in the world are oriented toward the implementation of educational processes of modern ICT and software for modelling and simulation in various fields of human activity, including science, engineering, and technology. This approach is required for the efficient training of students at various levels: undergraduates, gradu-

ates, and doctoral students. Many international conferences on ICT and its applications for education are devoted to the use of computer modeling, open-source software, pedagogical e-learning, web-based e-learning, course-centered knowledge management and application in online learning based on web ontology, on-online learning in enterprise education, simulation languages, modeling and simulation for education and training, improving education through data mining, 3D software systems, 3D visualization, wireless communication, experimental teaching of program design, different approaches in teaching programming, web-based computer-assisted language learning, and so on.

It is important that university and IT-industry participants of conferences try to find efficient solutions for the abovementioned computer-modeling-based educational problems. For example, participants from 178 different academic institutions, including many from the top 50 world-ranked institutions, and from many leading IT corporations, including Microsoft, Google, Oracle, Amazon, Yahoo, Samsung, IBM, Apple, and others, attended the 12th International Conference on Modeling, Simulation and Visualization Methods, MSV-2015, in Las Vegas, Nevada, USA.

If IT industry today supports higher education, then tomorrow's IT-based companies, government research agencies, and national laboratories will obtain the high-quality graduates that they need. New achievements in ICT require continuous tracking by educators, and implementation in education.

Successful introduction of ICT to higher education based on research-oriented education and training is considered and analyzed in this paper. The focus is on the role of computer modeling and simulation in prosthesis and robotics research for increasing student quality, including grading their practical skills, and including efficient professor-student interactions.

2 Literature Analysis and Problem Statement

Many publications are devoted to teaching methods and approaches based on ICT and computer modelling, for increasing the efficiency of their interrelation: qualitative modeling in education [3], computer simulation technologies and their effect on learning [21], opportunities and challenges for computer modeling and simulation in science education [31], web-based curricula [4] and remote access laboratories, computer-based programming environments as modelling tools in education and the peculiarities of textual and graphical programming languages [15], interrelations between computer modeling tools, expert models, and modeling processes [41], efficient science education based on models and modelling [9], educational software for collective thinking and testing hypotheses in computer science [23], and others.

Many publications are devoted to improving teaching efficiency for specific courses by introducing modern ICT and computer modelling technologies. In particular, modelling supported course programs, computer-based modelling (AutoCAD, Excel, VBA, etc.) and computer system support for higher education in engineering [8]; software to enhance power engineering education [32]; computer modeling for enhancing instruction in electric machinery [20]; computer modelling in mathematics

education [37]; GUI-based computer modelling and design platforms to promote interactive learning in fiber optic communications [42]; RP-aided computer modelling for architectural education [33]; teaching environmental modelling; computer modelling and simulation in power electronics education [24]; and a virtual laboratory for a communication and computer networking course [19].

Special attention in the literature [5] is paid to the role of ICT and modeling technology in education and training in the framework of research-based curricula. This educational approach deals first with educational directions such as robotics, mechatronics, and biomechanics (RMBM) [12, 30, 38]. The correlation of RMBM with ICT and modeling are underlined by results such as: a multidisciplinary model for robotics in engineering education; integration of mechatronics design into the teaching of modeling; modelling of physical systems for the design and control of mechatronic systems [38]; biomechanical applications of computers in engineering education [30]; computerized bio-skills system for surgical skills training in knee replacement [6]; computer modelling and simulation of human movement [22]; computer modelling of the human hand [17]; and design and control of a prosthesis test robot [26, 27].

The main aims of this paper are given as follows.

(a) Description and analysis of research-based education based on the experience in the Embedded Control Systems Research Laboratory at the Electrical Engineering and Computer Science Department at the Washkewicz College of Engineering at Cleveland State University (CSU), USA, with a focus on undergraduate, graduate, and doctoral student participation in prosthesis and robotics research, which is funded by the US National Science Foundation (NSF);

(b) Analysis of applied ICT and modeling technologies and advanced software, as well as their implementation in student research, including course work, diploma projects, and Doctoral, Master's, and Bachelor's theses;

(c) Focus on the correlation between student research and government science priorities based on successful cases of ICT and advanced modelling implementation in US government-funded prosthesis research, with particular focus on undergraduate, graduate, and doctoral student participation in prosthesis and robotics research.

The rest of this paper is organized as follows. Section 3 presents a general description of the prosthesis research project granted by the US NSF. In Section 4 the authors consider the implementation of ICT in prosthesis and robotics research at CSU. The paper ends with a conclusion in Section 5.

3 NSF Project “Optimal Prosthesis Design with Energy Regeneration” for Research-Based Education

CSU's research project “Optimal prosthesis design with energy regeneration” (OPDER) is funded by the US NSF (1.5M USD). Professors and students from the Department of Electrical Engineering and Computer Science, and the Department of Mechanical Engineering, are involved in research according to the project goals, which deal with the development of: (a) new approaches for the simulation of human limb control; (b) new approaches for optimizing prosthetic limb control, capturing

energy during walking, and storing that energy to lengthen useful prosthesis life; (c) prosthesis prototype development.

The human leg transfers energy between the knee, which absorbs energy, and the ankle, which produces energy. The prosthesis that results from this research will mimic the energy transfer of the human leg. Current prostheses do not restore normal gait, and this contributes to degenerative joint disease in amputees. This research will develop new design approaches that will allow prostheses to perform more robustly, closer to natural human gait, and last longer between battery charges.

This project forms a framework for research-based education. Doctoral, graduate, and undergraduate students are involved in research such as: the study of able-bodied gait and amputee gait; the development of models for human motion control to provide a foundation for artificial limb control; the development of electronic prosthesis controls; the development of new approaches for optimizing prosthesis design parameters based on computer intelligence; the fabrication of a prosthesis prototype and its test in a robotic system; the conduct of human trials of the prosthesis prototype.

The role of student participation in all aspects of the research is significant for increasing their qualifications for their careers, for presentations at conferences, for publishing in journals, and for research with professors who can help them be more successful in building their future careers in industry or academia. In the next section we describe the student contribution to prosthesis and robotics research at CSU.

4 Student contributions to prosthesis and robotics research

Seven cases of student research in the framework of the OPDER project are described in this section.

Evolutionary Optimization of User Intent Recognition for Transfemoral Amputees.

Powered prostheses can help amputees handle multiple activities: standing, level walking, stepping up and down, walking up and down a ramp, etc. For each walking mode, a different control strategy or control gains are used to control the prosthesis. It is important to infer the user's intent automatically while transitioning from one walking mode to another one, and to subsequently activate the suitable controller or control gains. Pattern recognition techniques are used to address such problems.

In this research, mechanical sensor data are experimentally collected from an able-bodied subject. Collected signals are processed and filtered to eliminate noise and to handle missing data points. Signals reflecting the state of the prosthesis, user-prosthesis interactions, and prosthesis-environment interactions are used for user intent recognition. Principal component analysis is used to convert data to a lower dimension by eliminating the least relevant features. We propose the use of correlation analysis to remove highly correlated observations from the training set.

We use K-nearest neighbor (K-NN) as a classification method. K-NN is modified and optimized with an evolutionary algorithm for enhanced performance. In the modified K-NN, the contribution of each neighbor is weighted on the basis of its distance to the test point, and the history of previously classified test points is considered for classification of the current test point. This modification leads to better performance

than standard K-NN. Optimization techniques can be used to tune the parameters and obtain a classification system with the highest possible accuracy. We choose biogeography-based optimization (BBO) as the evolutionary optimization algorithm for this purpose. The optimization problem is to minimize the classification error.

We use MATLAB to implement user intent recognition. BBO is a stochastic algorithm, so it requires several runs to optimize the parameters. The optimization process may take multiple days, so we use parallel computing to reduce the optimization time from 7.77 days to about 20 hours [11]. To test the proposed method, multiple sets of experimental data were collected for various gait modes: standing (ST), slow walking (SW), normal walking (NW), and fast walking (FW). Fig. 1 illustrates the experimental setup for able-bodied subjects. Hip and ankle angles, ground reaction force (GRF) along three axes, and hip moment, comprise the six input signals which were used for user intent recognition. Fig. 2 shows an example of test data for a walking trial lasting approximately 18 seconds, which included different walking modes.



Fig. 1. Experimental setup: data collection for able-bodied subjects

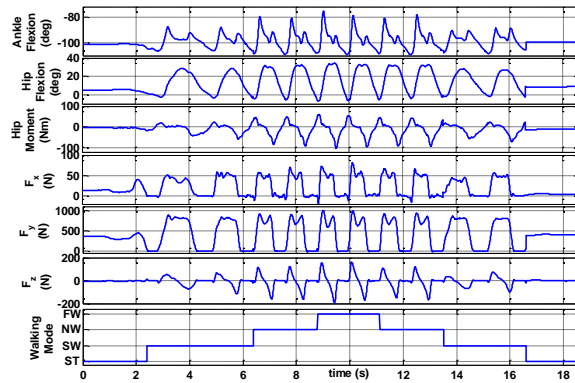


Fig. 2. Sample test data showing four different gait modes and transitions: ST (standing), SW (slow walk), NW (normal walk), and FW (fast walk)

Fig. 3 shows the performance of the classifier using both simple K-NN and optimized K-NN. Classification error for optimized K-NN is 3.6% which is improved from 12.9% with standard K-NN.

In conclusion, K-NN was modified to enhance the performance of a user intent recognition system. An evolutionary algorithm was applied to optimize the classifier parameters. Experimental data was used for training and testing the system. It was shown that the optimized system can classify four different walking modes with an accuracy of 96%. The code used to generate these results is available at <http://embeddedlab.csuohio.edu/prosthetics/research/user-intent-recognition.html>. Further details about this research can be found in [11].

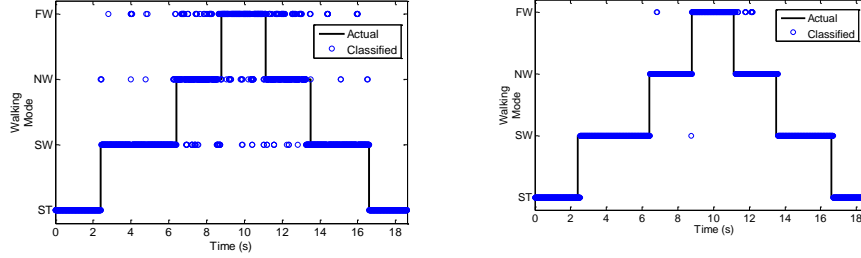


Fig. 3. Classifier results for optimized K-NN 3.6% error (right), improved from 12.9% with standard K-NN (left)

Stable Robust Adaptive Impedance Control of a Prosthetic Leg. We propose a nonlinear robust model reference adaptive impedance controller for a prosthetic leg. We use an adaptive control term to compensate for the uncertain parameters of the system, and a robust control term so the system trajectories exhibit robustness to variations of ground reaction force (GRF). The algorithm not only compromises between control chattering and tracking performance, but also bounds parameter adaptation to prevent unfavorable drift. The acceleration-free regressor form of the system removes the need to measure joint accelerations, which would otherwise introduce noise in the system. We use particle swarm optimization (PSO) to optimize the design parameters of the controller and the adaptation law. The PSO cost function is comprised of control signal magnitudes and tracking errors.

The prosthetic component is modeled as an active transfemoral (above-knee) prosthesis. This model has a prismatic-revolute-revolute (PRR) joint structure. Human hip and thigh motion are emulated by a prosthesis test robot. The vertical degree of freedom represents human vertical hip motion, the first rotational axis represents angular thigh motion, and the second rotational axis represents prosthetic angular knee motion [26, 27]. The three degree-of-freedom model can be written as follows [36]:

$$M\ddot{q} + C\dot{q} + g + R = u - T_e \quad (1)$$

where $q^T = [q_1 \ q_2 \ q_3]$ is the vector of generalized joint displacements (q_1 is the vertical displacement, q_2 is the thigh angle, and q_3 is the knee angle); u is the control signal that comprises the active control force at the hip and the active control torques at the thigh and knee; and T_e is the effect of the GRF on the three joints.

The contribution of this research is a nonlinear robust adaptive impedance controller using a boundary layer and a sliding surface to track reference inputs, in the presence of parameter uncertainties. We desire the closed-loop system to provide near-normal gait for amputees. Therefore, we define a target impedance model with characteristics that are similar to those of able-bodied walking:

$$M_r(\ddot{q}_r - \ddot{q}_d) + B_r(\dot{q}_r - \dot{q}_d) + K_r(q_r - q_d) = -T_e \quad (2)$$

where q_r and q_d are the state of the reference model and the desired trajectory respectively. Since the parameters of the system are unknown, we use a control law [36]

$$u = \hat{M}\dot{v} + \hat{C}v + \hat{g} + \hat{R} + \hat{T}_e - K_d \text{sat}(s/\text{diag}(\varphi)) \quad (3)$$

where the diagonal elements of φ are the widths of the saturation function; s and v are error and signal vectors respectively; \hat{M} , \hat{C} , \hat{g} , \hat{R} , and \hat{T}_e are estimates of

M, C, g, R , and T_e respectively. The control law of Eq. (3) comprises two different parts. The first part, $\hat{M}\dot{v} + \hat{C}v + \hat{g} + \hat{R}$, is an adaptive term that handles the uncertain parameters. The second part, $\hat{T}_e - K_d \text{sat}(s/\text{diag}(\varphi))$, satisfies the reaching condition and the variations of the external inputs T_e .

We use PSO to tune the controller and estimator parameters. PSO decreases the cost function (a blend of tracking and control costs) by 8%. We suppose the system parameters can vary $\pm 30\%$ from their nominal values. Fig. 4 compares the states of the closed-loop system with the desired trajectories when the system parameters vary. The MATLAB code used to generate these results is available at <http://embeddedlab.csuohio.edu/prosthetics/research/robust-adaptive.html> [2].

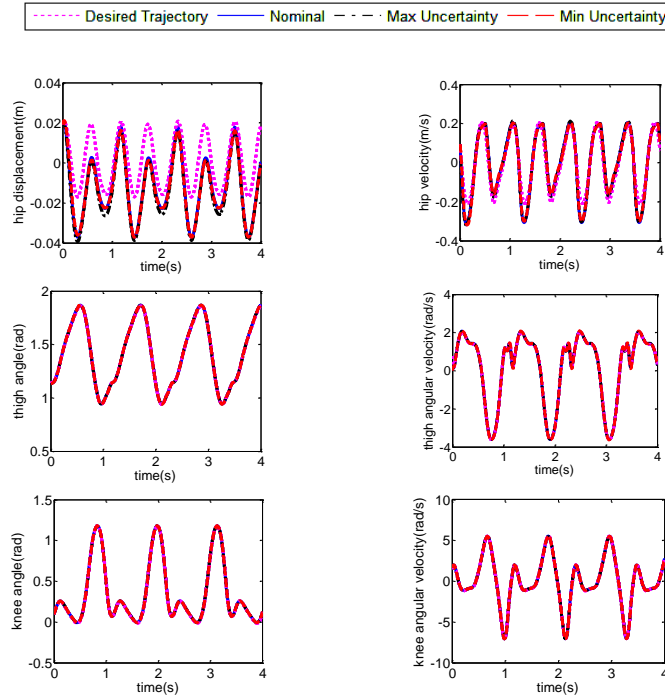


Fig. 4. Tracking performance for the joint displacements and velocities

Hybrid Function Approximation Based Control for Prosthetic Legs. The combination of a prosthesis test robot and a prosthesis and how their respective controllers could be combined to yield a coupled stable controller is addressed in this research. The prosthesis test robot was assumed to be controlled by a regressor-based controller while the prosthesis was assumed to be controlled by a regressor-free controller. We address this problem by first defining a framework on which two controllers could be combined where the controllers are indirectly dependent on each other. We propose a

theorem that yields a stable robotic system by the combination of the prosthesis test robot and the prosthesis leg.

The mathematical proof depends on using the open loop dynamics of the system to develop the closed loop system dynamics using the control law developed in the theorem. We then employ a Lyapunov function to verify the stability of the robotic system with the proposed controller. We also evaluate the transient response of the system by evaluating the upper bounds for both the Lyapunov function and the error vector.

We use MATLAB/Simulink to model the robotic system and then simulate the system's behavior when the proposed controller is applied; see Fig. 5 and Fig. 6.

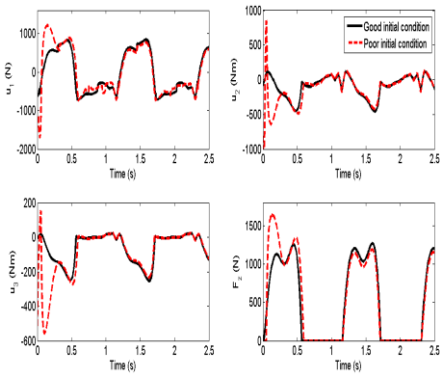


Fig. 5. Plot of joint angle trajectories

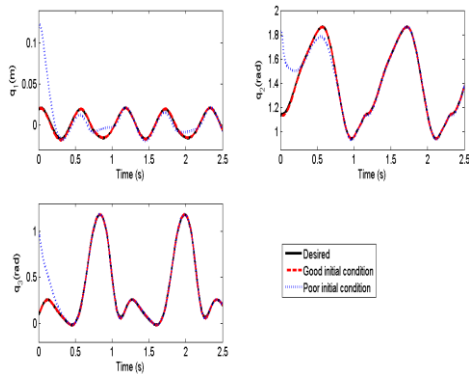


Fig. 6. Plot of control signals and vertical ground reaction force

Results show that the controller is able to drive the system to a desired state. Fig. 5 shows good tracking of the reference trajectories which is desired. However, Fig. 6 shows that the control signals u_2 and u_3 are too large to be implemented on the robotic system in real-time as it will lead to damage of equipment; additional research is needed to reduce the control signal magnitudes.

In conclusion, the simulation results show that the combination of two different robotic systems with different control schemes is possible, which is further verification of the stability proof. The simulation results help us investigate implementation of an environmental interaction controller to trade off tracking accuracy and reaction force magnitudes, hence reducing the control signal magnitudes.

The MATLAB code that was used to generate these results can be downloaded from <http://embeddedlab.csuohio.edu/prosthetics/research/hybrid-fat.html> [7].

System Identification and Control Optimization of a Prosthetic Knee. A Mauch SNS knee has been attached to an EMG-30 geared DC motor as our active leg prosthesis. The Mauch SNS knee is a widely-used passive prosthesis; we have modified it by removing the damper connection and driving it with our DC motor. Our work provides a conceptual approach for the system identification, control optimization, and implementation of an active prosthetic knee during swing phase.

To apply velocity control to the system, Proportional-Integral-Derivative control (PID) is used due its effectiveness in a wide range of operating conditions, its func-

tional simplicity, and its ease of use with embedded systems technology. The goal is to investigate the behavior of PID parameters with respect to shank length. To achieve this goal we have to find a model for the prosthetic leg. We use heuristic algorithms and gradient algorithms to identify model parameters and tune the PID controller. Particle Swarm Optimization (PSO), BBO, and Sequential Quadratic Optimization (SQP) [16, 18, 29, 34] are used for identification and tuning.

Hardware setup includes a PC connected to a Quanser© DAQ card. MATLAB with Quanser Quarc software for real-time connectivity, and DAQ hardware; see Fig. 7. The DAQ system delivers an analog control signal to a servo amplifier to drive the EMG30 DC motor. The encoder sends signals through two digital channels. We use a quadrature encoder which has the ability to sense rotational direction.

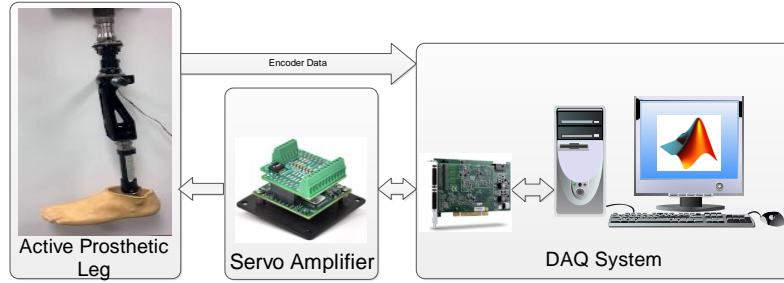


Fig. 7. Hardware Setup

Numerical differentiation is usually used to obtain angular velocity by differentiating the encoder signal [39]. This technique leads to a distorted signal due to encoder resolution. So a Kalman filter is instead designed to estimate the angular velocity.

The DC geared motor and the Mauch SNS joint are described mathematically [10]. Simulink is used to implement the models. In order to find model parameters, each optimization algorithm executes 20 times. The DC motor model and Mauch knee joint model are combined to form the active prosthetic leg model. We also conducted a sensitivity analysis test for PSO and BBO.

The active prosthetic knee model and PID are used to build a closed-loop feedback system. To investigate PID controller parameter behavior with respect to shank length, we use optimization algorithms to tune controller parameters (K_p , K_i and K_d).

Results show that for model parameter identification, PSO gives the best optimization results, and BBO gives better average overall performance than SQP. For PID tuning, BBO achieves the best average overall performance, but PSO shows the fastest average convergence. Finally, we see that increasing shank length results in an increase in the optimal proportional gain, and a decrease in the optimal differential and integral gains as shown in Fig. 8.

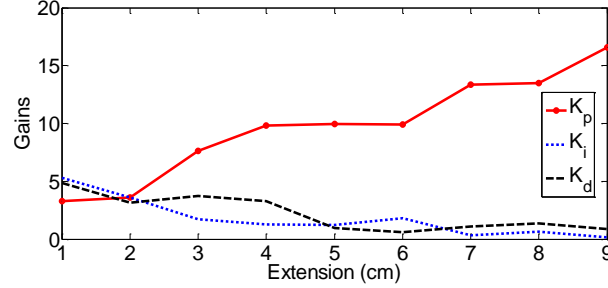


Fig. 8. PID parameter changes with respect to shank length

Ground Reaction Force Estimation with an Extended Kalman Filter. A method to estimate GRF in a robot/prosthesis system is presented. The system includes a robot that emulates human hip and thigh motion, and a powered prosthesis for transfemoral amputees, and includes four degrees of freedom: vertical hip displacement, thigh angle, knee angle, and ankle angle. A continuous-time extended Kalman filter (EKF) [35] estimates the states of the system and the GRFs that act on the prosthetic foot.

The system includes eight states: q_1 is vertical hip displacement, q_2 is thigh angle, q_3 is knee angle, q_4 is ankle angle, and their derivatives. Horizontal and vertical GRF is applied to the toe and heel of a triangular foot. The ground stiffness is modeled to calculate GRF. The initial state $x(0)$ is obtained from reference data, and we randomly initialize the estimated state $\hat{x}(0)$ to include estimation error. The diagonal covariance matrices of the continuous-time process noise and measurement noise are tuned to obtain good performance.

Results are shown in Fig. 9. Although significant initial estimation errors are present for displacements and velocities, the EKF converges to the true states quickly.

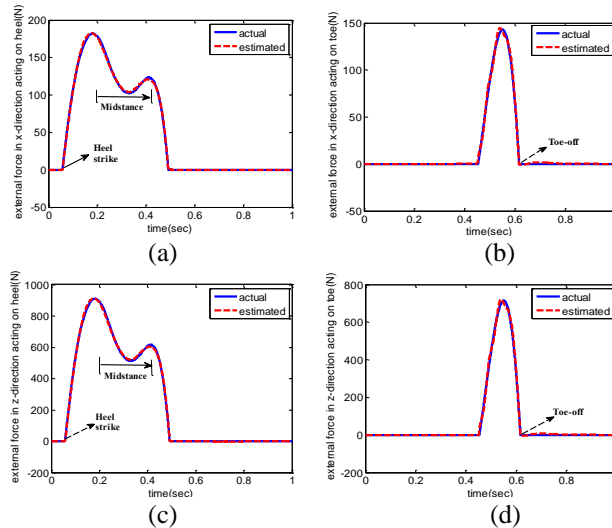


Fig. 9. Horizontal and vertical ground force (GRF) estimation

Electronic Energy Converter Design for a Regenerative Prosthetics. Prosthetic models use ideal electromechanical actuators for knee joints, which do not include energy regeneration. In order to focus on energy regeneration, a voltage source converter is designed to interface an electric motor to a supercapacitor.

A converter was designed to resemble a typical H-bridge motor driver. The voltage converter control system allows power to flow from the motor to the capacitor (motor mode) and from the capacitor to the motor (generator mode). During motor mode, the voltage converter's control system modulates the voltage applied to the motor using two circuits; one with the capacitor connected (powering the motor from the capacitor) and one with the capacitor disconnected (shorting the motor connection through the H-bridge). During generator mode, the voltage converter control system changes the impedance connected to the motor using two circuits; one with the capacitor connected (charging the capacitor) and one with the capacitor disconnected (allowing the motor to move with less resistance from the electronics). The circuit and motor were modeled with state space equations using MATLAB and Simulink software.

Two controllers were designed for the voltage converter. Both controllers use reference knee torque from control signals in the mechanical model with an ideal actuator at the knee. The first controller, a PD (proportional-derivative) controller, compares reference torque to the torque generated by the motor and voltage converter. The controller uses the comparison between reference and simulation data to determine switching between connecting and disconnecting the capacitor and motor. The switches use measured velocity to determine the direction of motor rotation. The controller uses direction, mode, and torque error to provide correct modulation. The second controller, an artificial neural network, follows the same logic as the PD controller. The controller gains were optimized with BBO. The optimized controller was able to track the reference torque with root mean square (RMS) error of 1.35 Amps as shown in Fig. 10. As can be seen in Fig. 11, the system was able to store 17.6 Joules in the capacitor bank. The results from the motor and voltage converter simulation show that it may be possible to gain energy through a normal stride. The energy gained would allow a prosthesis to operate longer than current powered prostheses.

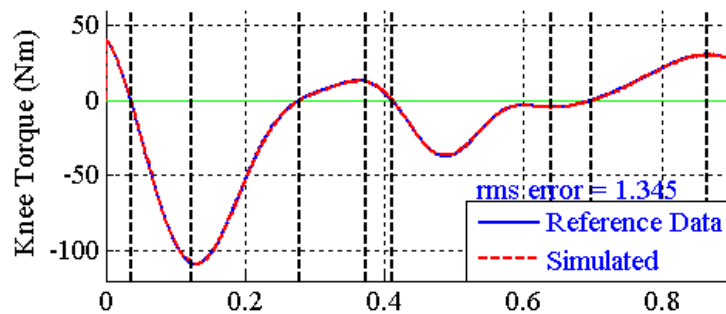


Fig. 10. Tracking a reference current for the knee joint with a motor and voltage converter

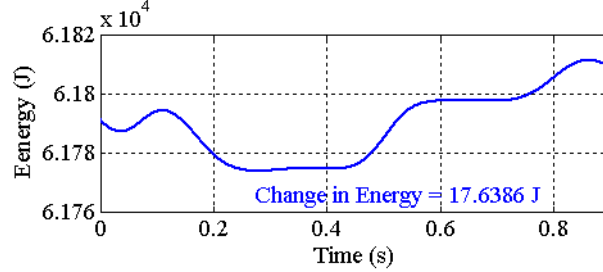


Fig. 11. The energy gained during one stride of gait with a motor and voltage converter

Fuzzy Logic for Robot Path Finding. This research deals with fuzzy logic to find a path for mobile robots that move in environments with obstacles, when the robot does not have prior information about the obstacles.

The radar of the robot returns a fuzzy set based on the distance L_i from obstacle i (see Fig. 12): $\mu_i^\varphi(\varphi_i) = \frac{L_i}{L_{max}}$. The robot finds the angle between its position and the target position, which we call α . If the robot moved in the α direction in an obstacle-free environment it would follow a direct line to the target. However, there are obstacles in the path. To find a safe path around the obstacles, we introduce a Gaussian fuzzy set [13, 14, 25, 40] which has a maximum value at α as follows:

$$\mu_i^\alpha(\varphi_i) = e^{-\frac{(\varphi_i - \alpha)^2}{2\sigma^2}}$$

We combine $\mu_i^\varphi(\varphi_i)$ and $\mu_i^\alpha(\varphi_i)$ to obtain a new fuzzy set, $\mu_i^\psi(\psi_i)$, shown in Fig. 13.

$$\mu_i^\psi(\varphi_i) = \min(\mu_i^\alpha(\varphi_i), \mu_i^\varphi(\varphi_i))$$

The movement direction then is φ , which is the maximum point in $\mu_i^\psi(\psi_i)$, which we call A . If the robot moves in φ_A , it will touch the obstacles. To solve this problem we introduce a new fuzzy set that has the value 1 in a range of 180 degrees around φ_A :

$$\mu_1^\theta(\varphi_i) = \begin{cases} 1 & \varphi_i < \varphi_{A+90} \text{ and } \varphi_i > \varphi_{A-90} \\ 0 & \text{otherwise} \end{cases}$$

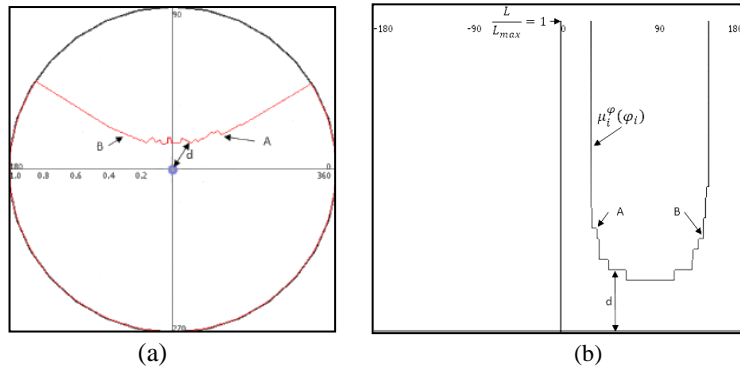


Fig. 12. (a) A polar radar map in the presence of an obstacle, and (b) its transformation to Cartesian coordinates

In the next step we defuzzify $\mu^\psi(\varphi_i) * \mu_1^\theta(\varphi_i)$ using center of mass [28], which is shown in Fig. 13.

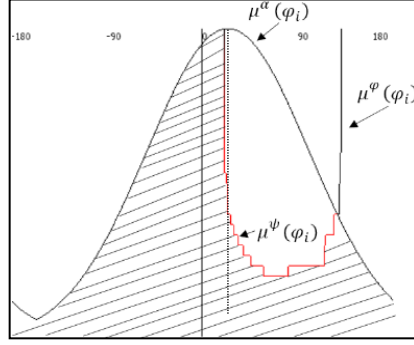


Fig. 13. Highlighted area is $\mu_i^\psi(\psi_i)$

Simulations confirm that the proposed approach provides reliable output. In different layouts and robot positions and target positions, the robot was able to find a path to the target point without touching any obstacles; see Fig. 14.

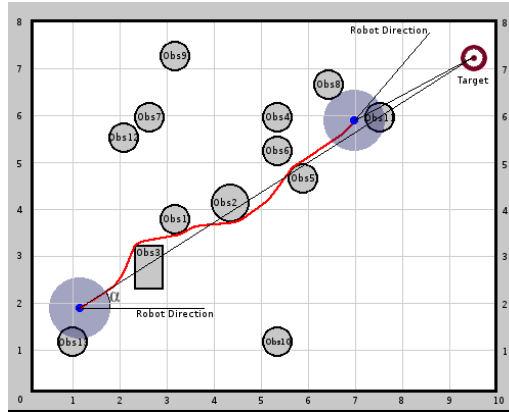


Fig. 14. Fuzzy path planning results: the red line is the robot path from start to target.

5 Conclusions

The authors have described university student training. The description has focused on student participation in the US NSF project “Optimal prosthesis design with energy regeneration” and the application of ICT and modelling technologies.

Several factors play an important role in the results of this paper. Student research requires skill in programming and software, and a broad theoretical knowledge in

computer science, and mechanical, electrical, and control engineering. Students used MATLAB, Simulink, and toolboxes (Optimization, Fuzzy Logic, etc.), and programming in C and C++. The software used for robot trajectory planning research was designed and written by students in C++, and the GUI was designed using Qt and OpenGL. Standard libraries were used to make the software cross-platform.

The most important foundation for student research is theoretical knowledge in fundamental and elective disciplines such as Circuits, Linear Systems, Control Systems, Nonlinear Control, Machine Learning, Artificial Intelligence, Intelligent Controls, Optimal State Estimation, Optimal Control, Embedded Systems, Robot Modeling and Control, Probability and Stochastic Processes, Population-Based Optimization, and Prosthesis Design and Control, which provides a basic understanding of human biomechanics and lower-limb prosthesis design and control. These courses played a vital role in the proper grounding of basic and advanced ICT and control theory for robotic and prosthetic leg research. The facilities at CSU and funding from the NSF significantly helped in furthering student research-based education.

Finally, student participation in government-sponsored research, student exchanges of research experiences with each other, and publication of research results in high-caliber journals and conferences [1, 2, 7, 11, 16, 26], provide students with effective training and self-confidence in their higher education. Research-based education also allows students to obtain practical experience as research assistants, with corresponding responsibilities in the development and implementation of research projects.

Student participation in real-world research significantly influences their engineering and research qualifications by: (a) giving them a strong understanding of ICT and engineering concepts that are covered in corresponding courses; (b) giving them practical experience and the ability to apply theoretical knowledge; (c) giving them the opportunity to learn technical material independently; (d) helping them improve fundamental skills to apply in other research in their future; (e) providing them with a rich interdisciplinary research environment; and (f) providing them with an understanding of concepts both familiar and unfamiliar. Through extensive literature review and actively seeking ways to solve research problems, students are prepared to make meaningful future contributions to the field of ICT and control engineering.

Acknowledgements. The authors thank the Fulbright Program (USA) for supporting Prof. Y. P. Kondratenko with a Fulbright scholarship and for making it possible for this team to conduct research in together in the USA. This research was partially supported by US NSF Grant 1344954.

References

1. Azimi, V., Simon, D., Richter, H., Fakoorian, S.A.: Robust composite adaptive transfemoral prosthesis control with non-scalar boundary layer trajectories. In: Proceedings of the American Control Conference (ACC), Boston, MA, USA (2016)
2. Azimi, V., Simon, D., Richter, H.: Stable Robust Adaptive Impedance Control of a Prosthetic Leg. Dynamic Systems and Control Conf., Columbus, Ohio, USA, October (2015)
3. Bredeweg, B., Forbus, K.D.: Qualitative modeling in education. AI Magazine 24, No. 4, American Association for Artificial Intelligence, 35-46 (2003)

4. Chou, C., Tsai C.-C.: Developing web-based curricula: Issues and challenges. *Journal of Curriculum Studies* 34, No.6, 623-636 (2002)
5. Clements, D.H.: Curriculum research: Toward a framework for "Research-Based Curricula". *Journal for Research in Mathematics Education* 38, No. 1, 35-70 (2007)
6. Conditt, M.A., Noble, P.C., Thompson, M.T., Ismaily, S.K., Moy, G.J., Mathis K.B.: A computerized bioskills system for surgical skills training in total knee replacement. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 221, No. 1, 61-69 (2007)
7. Ebeigbe, D., Simon, D., Richter, H.: Hybrid function approximation based control with application to prosthetic legs. *IEEE Int. Systems Conf.*, Orlando, FL, USA, April (2016)
8. Gáti, J., Kártyás, G.: Computer system support for higher education programs in engineering. In: *Proc. 4th International Conference on Emerging Trends in Engineering and Technology (ICETET)*, pp. 61-65 (2011)
9. Gilbert, J.K.: Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education* 2, No.2, 115-130 (2004)
10. Gonçalves, J., Lima, J., Costa, P., Moreira, A.P.: Modeling and simulation of the EMG30 geared motor with encoder resorting to SimTwo: The official Robot@Factory simulator. In: *Advances in Sustainable and Competitive Manufacturing Systems*. Azevedo, A. (Ed.), Springer International Publishing, 307-314 (2013)
11. Khademi, G., Mohammadi, H., Simon, D., Hardin, E.C.: Evolutionary optimization of user intent recognition for transfemoral amputees. In: *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Atlanta, Georgia, USA, pp. 1-4 (2015)
12. Kondratenko, Y.P., Kondratenko, G.V.: Robotic system with myoelectric adapting for rehabilitation tasks. *ACTA of Bioengineering and Biomechanics* 3, pp.1-6 (2001)
13. Kondratenko, Y., Kondratenko, V.: Soft computing algorithm for arithmetic multiplication of fuzzy sets based on universal analytic models. In: *Information and Communication Technologies in Education, Research, and Industrial Application. Series Communications in Computer and Information Science* 469, Springer, Switzerland, 49-77 (2014)
14. Kondratenko, Y.P., Klymenko, L.P., Al Zu'bi, E.Y.M.: Structural optimization of fuzzy systems' rules base and aggregation models. *Kybernetes* 42, Issue 5, 831-843 (2013)
15. Louca, L.T., Zacharia, Z.C.: The use of computer-based programming environments as computer modelling tools in early science education: The cases of textual and graphical program languages. *International Journal of Science Education* 30, No. 3, 287-323 (2008)
16. Ma, H., Simon, D.: Biogeography-based optimization with blended migration for constrained optimization problems. *Proceedings of the 12th annual conference on Genetic and Evolutionary Computation*, New York, NY, USA, pp. 417-418 (2010)
17. McKee, N.H., Agur, A.M., Tsang, W., Singh, K.S.: Possible uses of computer modeling of the functioning human hand. *Clinics in Plastic Surgery* 32, No. 4, 635-641 (2005)
18. Mo, H., Xu, Z.: Research of biogeography-based multi-objective evolutionary algorithm. In: *Interdisciplinary Advances in Information Technology Research*. Khosrow-Pour, M. (Ed.), 125-135 (2013)
19. Musa, A., Al-Dmour, A., Fraij, F., Gonzalez, V., Al-Hashemi, R.: Developing a virtual laboratory for a communication and computer networking course. *International Journal of Continuing Engineering Education and Life Long Learning* 20, No. 3-5, 390-406 (2010)
20. Nehrir, M. H., Fatehi, F., Gerez, V.: Computer modeling for enhancing instruction of electric machinery. *IEEE Transactions on Education* 38, No. 2, 166-170 (1995)
21. Nemanjic, B., Svetozar, N.: *Computer simulations: Technology, industrial applications and effects on learning*. Nova Science Publishers, Inc., New York (2012)
22. Neptune, R.R.: Computer modeling and simulation of human movement. *Scientific Principles of Sports Rehabilitation* 11, No. 2, 417-434 (2000)

23. Panselinas, G.E., Komis, V.: Using educational software to support collective thinking and test hypotheses in the computer science curriculum. *Education and Information Technologies* 16, No. 2 159-182 (2011)
24. Patil, L. S., Patil, K. D., Thosar, A.G.: The role of computer modeling and simulation in power electronics education. In: *Proc. 2nd International Conference on Emerging Trends in Engineering and Technology (ICETET)*, pp. 416-419 (2009)
25. Piegat, A.: *Fuzzy modeling and control*. Springer, Heidelberg (2001)
26. Richter, H., Simon, D., Smith, W., Samorezov, S.: Dynamic modeling, parameter estimation and control of a leg prosthesis test robot. *Applied Mathematical Modelling* 39, 559-573 (2015)
27. Richter, H., Simon, D.: Robust tracking control of the prosthesis test robot. *Journal of Dynamic Systems, Measurement and Control* 136, No. 3, Paper No. DS-13-1052 (2014)
28. Runkler, T.: Selection of appropriate defuzzification methods using application specific properties. *IEEE Transactions on Fuzzy Systems* 5, No.1, 72-79 (1997)
29. Sayed, M., Saad, M., Emara, H., El-Zahab, E.A.: A novel method for PID tuning using a modified biogeography-based optimization algorithm. *Proc. 24th Chinese Control and Decision Conference (CCDC)*, pp. 1642-1647 (2012)
30. Schonning, A.: Biomechanical applications of computers in engineering education. In: *ASME 2007 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Las Vegas, Nevada, USA, American Society of Mechanical Engineers, pp. 389-393 (2007)
31. Schwarz, C.V., Meyer, J., Sharma, A.: Technology, pedagogy, and epistemology: Opportunities and challenges of using computer modeling and simulation tools in elementary science methods. *Journal of Science Teacher Education* 18, No.2, 243-269 (2007)
32. Shaalan, H.: Using software to enhance power engineering education in a technology program. *IEEE Power Engineering Review* 21, No. 6, 62-63 (2001)
33. Shih, N.-J.: RP-aided computer modeling for architectural education. *Computers & Graphics* 30, No. 1, 137-144 (2006)
34. Simon, D.: *Evolutionary optimization algorithms: biologically inspired and population-based approaches to computer intelligence*, John Wiley & Sons, 2013.
35. Simon, D.: *Optimal state estimation: Kalman, H-infinity, and nonlinear approaches*. John Wiley & Sons (2006)
36. Slotine, J.-J. E., Coetsee, J.A.: Adaptive sliding controller synthesis for non-linear systems. *International Journal of Control* 43, No. 6, 1631-1651 (1986)
37. Teodoro, V.D., Neves, R.G.: Mathematical modelling in science and mathematics education. *Computer Physics Communications* 182, No. 1, 8-10 (2011)
38. Van Amerongen, J., Breedveld, P.: Modelling of physical systems for the design and control of mechatronic systems. *Annual Reviews in Control* 27, No. 1, 87-117 (2003)
39. Yang, W.: *Applied numerical methods using MATLAB*. Hoboken, NJ.: Wiley-Interscience (2005)
40. Zadeh, L.A.: Fuzzy sets. *Information & Control* 8, 338-353 (1965)
41. Zhang, B.H., Liu, X.F., Krajcik, J.S.: Expert models and modeling processes associated with a computer-modeling tool. *Science Education* 90, No. 4, 579-604 (2006)
42. Zhang, Q.: GUI based computer modeling and design platform to promote interactive learning in fiber optic communications. In: *Proc. 37th Annual Frontiers in Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, FIE'07*, pp. S3H-14 – S3H-19 (2007)

University Curricula Modification Based on Advancements in Information and Communication Technologies

Yuriy Kondratenko^{1,2}, Dan Simon¹, Igor Atamanyuk²

¹ Department of Electrical Engineering and Computer Science
Cleveland State University, Cleveland, Ohio, USA
y.kondratenko@csuohio.edu, d.j.simon@csuohio.edu

² Department of Intelligent Information Systems
Petro Mohyla Black Sea State University, 68-th Desantnykiv str. 10,
54003 Mykolaiv, Ukraine
y_kondrat2002@yahoo.com, atamanyuk_igor@mail.ru

Abstract. This paper discusses the main methods for modification of university curricula for graduate students based on advanced research results in information and communication technology (ICT), artificial intelligence, control, and decision making. Special attention is paid to classifications of the approaches and their evaluation. Examples from the Washkewicz College of Engineering at Cleveland State University, and Black Sea State University, show the efficacy of the authors' proposals, approaches, and classification results.

Keywords: computer science, curricula, knowledge transfer

Key Terms: InformationCommunicationTechnology, Educational Process, KnowledgeEngineeringMethodology, Academia

1 Introduction

Many countries are reforming their science and technology systems to implement the recent advanced achievements of distinguished researchers from their own country and abroad in higher education [10, 13, 14]. At the same time, those of us directly involved in the reform process realize that some countries are deficient in devising genuine exchange programs, both within the country and abroad.

During the last few decades, university and industry researchers from different countries have paid much attention to the modification and development of new artificial intelligence techniques and optimization methods, especially for control and decision making systems [17, 21, 26, 27, 33, 38, 43]. Many international conferences, congresses, symposia, and seminars are specifically devoted to the needs of inter-university and university-industry cooperation [22, 40] in the field of soft computing, fuzzy systems, evolutionary optimization, and artificial neural networks, which allow for the solution of many practical tasks that include uncertainty. These research efforts include the development of new theoretical methods, advanced devices and equipment, joint research projects, joint publications, the incorporation of research results in university curricula, and so on, and also include the determination of the

most efficient means by which these goals can be achieved. Many publications in the educational field are devoted to the following topics.

(a) Using modern information and communication technology (ICT) in training processes [7, 8, 23, 24, 39], and developing new teaching methods, infrastructures, and software for ICT education and application;

(b) Creating university curricula [29, 30, 31, 32, 42] with balance between theoretical and practical components, and with possibilities for modification in the near future according to new requirements and developments;

(c) Developing new teaching methods and tools for on-line learning [13, 14];

(d) Using modeling and simulation techniques [34] for the investigation of the different dynamic and uncertain processes in education.

The optimization of fuzzy algorithms and systems opens up opportunities for co-operation and collaboration between scientists from different countries. At the same time, university curricula needs constant modification based on new research results in the fields of ICT and computer science to improve the quality of student training. This modification must take into account the dynamics of society's economic development, regional peculiarities, the increasingly intellectual level of technological and production processes, the complexity of market relations and the labor market, and the globalization and internationalization of societies and educational systems [1, 2, 3, 6, 25]. Such constant modification is possible by introducing new fundamental and elective courses, or by content modification of existing courses, taking into account that the present educational systems in many countries allows for elective courses.

The aim of this paper is to review the analysis and evaluation of modern educational approaches for creating and modifying university curriculum in the sphere of ICT, artificial intelligence, evolutionary optimization, automatic control, decision-making, and intelligent robotics. These educational approaches correspond to the recent results in research and science, and are based on the authors' experience in the Department of Electrical Engineering and Computer Science at Cleveland State University (CSU) in Cleveland, Ohio, USA [44], and the Department of Intelligent Information Systems at Petro Mohyla Black Sea State University (BSSU) in Mykolaiv, Ukraine [45].

The rest of this paper is organized as follows. Section 2 deals with related works and section 3 presents a classification approach to advanced scientific and engineering achievements. Section 4 considers the most efficient methods for curriculum modification. Section 5 is devoted to approaches for knowledge transfer, and Section 6 discusses evaluation. Section 7 provides some concluding remarks.

2 Related Works

Here we consider current research in curriculum development education challenges [35, 41, 42]. Some research is devoted to the modification of well-known educational approaches, but some is devoted to the development of new approaches based on the results in ICT and educational methodology [7, 8, 23, 24]. The overview, analysis, interconnection, and correlation of general education reform and the computer revolution is given in [10, 23, 35, 42]. Previous research includes investigations into the role

of inquiry as an organizing theme for science curricula [1], the anatomy of narrative curricula [4], the advantages of problem-based curricula [6], the correlation of course and curriculum design with learning outcome assessment at the course and curricular levels [5], the suitability, evolution, and impact of online learning, especially in ICT [13], and the issues, challenges, and opportunities for internet-based curriculum and individual courses [2].

Particular attention in the scientific literature has been devoted to the problem of interdisciplinary research and education, the development of integrated engineering curricula [12] with links between distinct disciplines, new approaches for teaching ICT to the next generation of students [25], embedding employability into curricula [16], and efficient approaches to internationalize university curricula [26]. Other research includes investigations into research-based curricula in response to needs from government agencies and members of the research community [3].

Curriculum modification according to current research is common but different (in some aspects) among various countries, so it is important to share best practices at the international level. The education community needs a wide spectrum of approaches and tools to increase the quality of university graduates and to imbue them with modern knowledge at each stage of their training, based on the latest theoretical and experimental research.

3 Research that Significantly Impacts Higher Education

Here we consider several items due to their significant influence on the higher education and training of graduate students. These items are all related to scientific and engineering research, and can be divided in 4 categories.

1. New directions and recent achievements in science and technology:
 - a) New theoretical results, including methods, models, algorithms, principles, approaches, etc. [18, 26, 27, 37, 38, 43] (e.g., biogeography-based optimization, invasive weed optimization, and other evolutionary optimization methods; new reliability assessment methods for critical computer systems; etc.).
 - b) New experimental results that demonstrate theory or scientific phenomena in simple and informative ways (e.g., thermoacoustic engines, ecopyrogenesis, intelligent robots, etc.).
2. Recent products of advanced industrial manufacturing:
 - c) New devices with improved characteristics and properties that allow investigation into scientific phenomena, and that extend the number of experimental modes for experiments related to hydrodynamics, thermodynamics, electrohydraulics, electromagnetics, nanotechnology, and others (e.g., NAO humanoid robot, IMS radiation detection based on gamma ray spectrometry and the Nerva LG robot, etc.).
 - d) New electronic components, sensors, and materials that can be used to design next-generation devices (e.g., new slip displacement sensors, FPGA-based controllers and electronic devices, etc.).
3. New software solutions developed by leading and advanced IT companies:

e) New computer-aided design (CAD) software, including 2D and 3D CAD software, that enables new levels of design processes in computer science, computer engineering, device design, machine design, ship design, and other important manufacturing fields (e.g., Active-HDL, Siemens PLM Software, parcel shipping software, etc.).

f) New information and communication technologies for industrial applications and domestic use, including IoT - the internet of things (e.g., Verizon IoT Solutions, Cisco IoT System, etc.).

4. Recent achievements in education based on the modern information and communication technologies:

g) New software and information technologies for teaching more efficiently, testing student knowledge, modeling object behavior in uncertain environments, control, identification, and decision making in education (e.g., e-learning and e-testing information technologies, etc.).

h) New educational methods using the internet to increase motivation and educational efficiency, to teach students to train themselves, and to apply current international standards to education (e.g., miscellaneous electronics and software - MPLAB from Microchip – for the course “Embedded Systems”, etc.).

Information about the above items can be obtained by students and teachers from various sources, including the following.

a) Publications in the scientific literature, including articles in international and domestic journals, chapters in monographs and edited books, and abstracts in conference proceedings (e.g., the journals Information Sciences, Kybernetes, etc.).

b) Articles in internet journals with both open access and registration access, chapters in e-books, electronic textbooks, and e-monographs (e.g., open access journals in engineering & computer science, Applied Computing and Informatics, etc.), the journal Sensors & Transducers, Elsevier journals, etc.).

c) Patents from various countries, such as the US, Ukraine, and international (multi-country) patents, with detailed information about new methods of signal processing and new technical solutions for devices in various fields of human activity (e.g., U.S. Patent No.8467921, 2013; Ukraine Patent No. 106288, 2016; etc.).

d) Presentations at conferences, congresses, symposia, and seminars, which in many cases include new and first-hand research results and pre-print material (e.g., ACC 2016, WConSC 2016, CDC 2016, etc.).

e) Industrial reports about new results and achievements that are obtained from industry, associations, or industrial consortia (e.g., NAICS Industry Report Collection, Industrial Report on Samsung Electronics' Processor Exynos, etc.).

f) International and domestic exhibitions of new devices and equipment in various scientific arenas, including new ICT technologies (e.g., Smart City Expo, NANOTECH: Advanced Materials & Applications, etc.).

g) International and domestic research projects that are financially supported by agencies such as the US National Science Foundation, NASA, the Fulbright Program (USA), the Tempus and Erasmus Programs (European Union), DAAD, DFG (Germany), Ministry of Education and Science (Ukraine), etc. [44, 45].

h) Newsletters from professional associations with recent information about new achievements, events, and activities, such as IEEE Spectrum newsletters (Institute of Electrical and Electronics Engineers), the IFAC Newsletters (International Federation of Automatic Control), the Sens2B (Sensor to Business) Newsletter, the E-Letter on Systems, Control, and Signal Processing from the IEEE Control Systems Society, the Medallion e-Newsletter from the PBD Honor Society for International Scholars, etc.

i) Web portals of engineering companies, consortia, and professional associations, such as the smart sensors web portal of the IFSA (International Frequency Sensor Association), web portal of Aldec, Inc. (The Design Verification Company), and others.

4 Research-Based Modification of University Curriculum

Usually a university curriculum consists of a list of year-by-year subjects for student learning, which includes vertical and horizontal relationships and correlation [44, 45]. For example, curricula for undergraduate study at the Department of Electrical Engineering and Computer Science (EECS) in the Washkewicz College of Engineering (WCE) at Cleveland State University are shown in Fig.1 (Bachelor of Science in Computer Science - BSCS) and in Fig. 2 (Bachelor in Electrical Engineering - BEE) [44]. CSU's curricula for the BEE, BSCS, and Bachelor in Computer Engineering (BCE) degrees are approved by the Accreditation Board for Engineering and Technology (ABET). Fig. 1 and Fig. 2 include different notations for required courses, core courses, and electives, as well as interrelations between courses, such as Pre-requisite, Co-/Pre-requisite, and Co-requisite.

Analysis of the possibilities for university curriculum modification according to new research results allows us to classify and discuss (using CSU examples) the following educational methods and approaches, which are directed to the improvement of the graduates' qualifications, and which promise to imbue them with modern knowledge in the field of science, engineering, and technology.

New specializations for existing Master of Science programs. New specializations allow CSU to take into account the newest research in science and engineering, and to specify required courses according to new engineering knowledge. For example, specializations in the modern science of nanotechnology has been included in the academic programs of several universities in various countries. The following areas of specialization are offered for graduate study and research in the Master of Science in Electrical Engineering (MSEE) program in the Department of Electrical Engineering and Computer Science (EECS) at CSU.

- a) Communication Systems
- b) Computer Systems
- c) Control Systems
- d) Power Electronics and Power Systems
- e) Nanobiotechnology

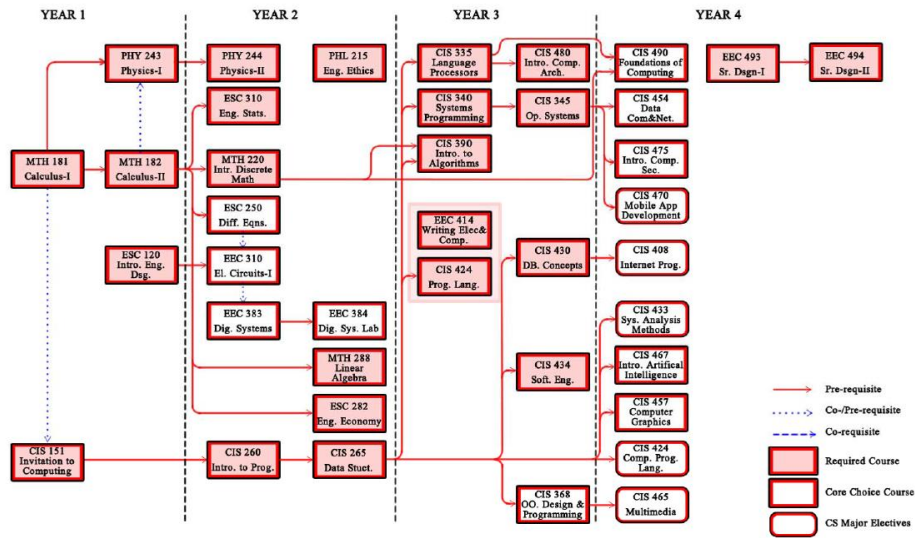


Fig. 1. The BSCS curriculum at CSU

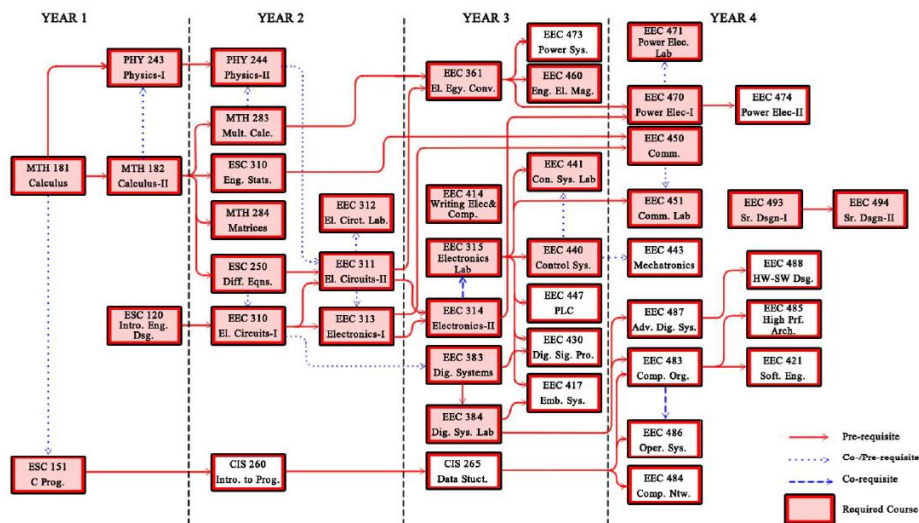


Fig. 2. The BEE curriculum at CSU

Developing a new Master of Science program. The first example here is the Master of Science in Software Engineering (MSSE) program at Cleveland State University, which is the first of its kind in Ohio, USA. The program introduces students to current and best practices and based on the recent achievements in the engineering of

software systems. A distinguishing feature of the program is its emphasis on the architecture, design, quality, management, and economic aspects of software engineering. The program exposes students to new technological developments in an advancing field, and teaches them how to apply their advanced knowledge in the workplace. Graduates meet the modern demands of industry and the needs of information technology professionals, in general, and software engineers, in particular. The second example deals with the computer science track in the Master of Science in Computer Science (MSCS) program at CSU. This track emphasizes the study of computing using the latest technologies, and the graduates of the program are prepared for immediate employment in business, industry, and government, or can pursue higher studies in the discipline. BSSU's MSCS program is preparing the new specialization Methods of Artificial Intelligence.

Doctor of Engineering (DRE) Programs and PhD Programs. PhD Programs in the United States are often theoretical programs which consist mostly of theoretical courses and research. The PhD thesis is a work with a theoretical hypothesis, proposals, and mathematical theorems that are supported, proven, and confirmed by modeling and simulation. The DRE thesis usually includes new models and algorithms for solving specific engineering problems. Compared to PhD programs, DRE programs are more practical, experimental, and industrially oriented.

New elective courses for programs at all educational levels (bachelor, masters, and doctoral). University curricula usually include required and elective courses, with a set of alternatives for electives. In particular, the Master of Science in Electrical Engineering program at CSU includes various sets of elective courses, depending on the specialization. For example, 10 elective courses are available for the Control Systems specialization, including Probability & Stochastic Processes, Embedded Systems, Art and Science of Feedback Control, Advanced Control System Design, System Identification, Nonlinear Systems, Optimal Control Systems, Intelligent Control Systems, Dynamics and Control of MEMS, and Robot Dynamics and Control. Fourteen elective courses are available for the Computer Systems specialization, including Embedded Systems, Software Engineering, Modern Digital Design, Rapid Digital System Prototyping, Formal Methods in Software Engineering, Software Quality Assurance, Software Testing, Software Design & Architecture, High Performance Computer Architecture, Distributed Computing Systems, Computer Networks II, Parallel Processing Systems, Mobile Computing, Secure and Dependable Computing.

Modification of existing courses with new content and teaching methods based on modern software. Here we consider the example of adjusting the content of a Control Systems course by including a new section on Fuzzy Control based on the recent research results, and using the MATLAB Fuzzy Logic Toolbox (Fig. 3) to model control system behavior in different modes and with various disturbances. The teaching approach in the EECS department establishes a spiral framework in which key concepts are revisited at increasing levels of sophistication and interconnection.

New case studies and examples in flexible courses. This approach applies to flexible courses such as Fundamentals of Research Investigations (BSSU, Ukraine), and Writing in Electrical Engineering (CSU, USA). Many engineering examples can be used to teach flexible courses, so it is easy for the instructor to consider new modern

engineering examples while taking into account new achievements in the field of electrical and computer engineering, intelligent information systems, and robotics.

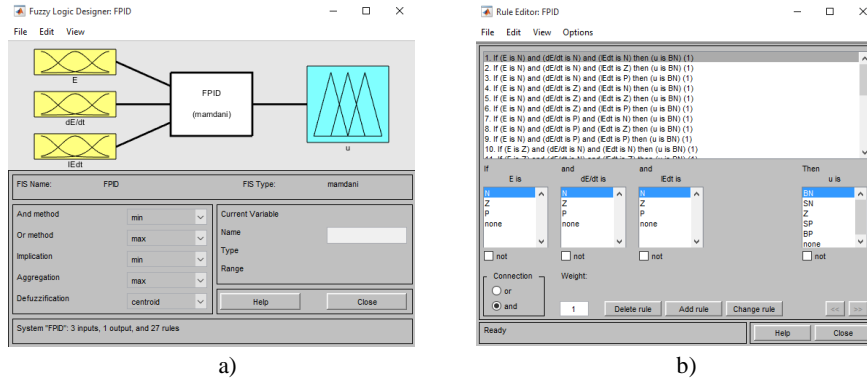


Fig. 3. Design of Mamdani-type fuzzy PID-controller (FPID) using the MATLAB Fuzzy Logic Toolbox: (a) the structure of the FPID controller; (b) Fuzzy rule editor

New research directions for (a) theses, (b) dissertations, and (c) course projects. Topics of student theses, dissertations, and course projects at CSU and BSSU deal with new research in the field of robotics, artificial intelligence, and control systems, as well as with the current research in the corresponding departments. For example, the EECS department (CSU) received research grants from the US National Science Foundation, Cleveland Clinic, Innovative Developments, Ford Motor Company, American Diabetes Association, and Electronics and Telecommunications Research Institute. BSSU received research grants from the European Commission for TEMPUS for the project Model-Oriented Approach and Intelligent Knowledge-Based System for Evolvable Academia-Industry Cooperation in Electronics and Computer Engineering (2013–2016). Thesis and dissertation research topics include Bio-Inspired Optimization of Ultra-wideband Patch Antennas Using Graphics Processing Unit Acceleration, Applications of Sliding Mode Controller and Active Disturbance Rejection Controller to a PMSM Servo System, and Evolutionary Optimization of Atrial Fibrillation Diagnostic Algorithms.

ICT for lectures and demonstrations. Multi-media plots and program code are efficient ways for introducing ICT and software. For example, consider the lectures concerning optimal state estimation methods [37]. MATLAB plots and demonstrations of the pseudo code are presented in Figs. 4 and 5 for the CSU course Optimal State Estimation.

Teaching and learning in academic consortia. Integrated processes between different universities and colleges is a powerful means for education reforms [10, 34]. Academic consortia allow cross registration (multi-vector) continuous education. The terms “cross registration” and “multi-vector education” mean that students are offered the possibility for parallel study at their home University and elective courses according at other universities [19]. The objective of multi-vector education is to create con-

ditions for the study of both foundational courses and elective courses to meet student inclinations, abilities, aspirations, and desires.

Any curricular innovation based on ICT should be supported by software facilities. For example, the relationship between disciplines and software in the EECS department at CSU can be seen in Table 1, Fig. 1, and Fig. 2.

```
function DiscreteKFEx1(N)
% Discrete time Kalman filter for position estimation of a Newtonian system.
% This example illustrates the effectiveness of the Kalman filter for state
% estimation. It also shows how the variance of the estimation error
% propagates between time steps and decreases as each measurement is
% processed.
% INPUT: N = number of time steps.
if ~exist('N', 'var')
    N = 6;
end
T = 5; % time between measurements
sigma = 30; % position measurement standard deviation
R = sigma^2;
P0 = [100 0 0; 0 10 0; 0 0 1]; % initial state estimate uncertainty
% A = [0 1 0; 0 0 1; 0 0 0]; % continuous-time system matrix
H = [1 0 0];
F = [1 T T^2/2; 0 1 T; 0 0 1]; % state transition matrix
x = [1; 1; 1]; % initial state
xhat = x; % initial state estimate
Q = zeros(3,3);
Q(3,3) = 0.01;
posArray = zeros(1, N);
xhatArray = zeros(3, N);
yArray = zeros(1, N);
Pplus = P0;
Varminus = zeros(1, N);
Varplus = P0(1,1);
KArray = zeros(3, N);
for k = 1 : N
    ...
```

Fig. 4: Sample MATLAB code for CSU's Optimal State Estimation course

5 Efficient Knowledge Transfer

In this section we classify the most efficient ways, according to authors' point of view, for knowledge transfer using various combinations: teacher–student, student–student, student–student team, and teacher group–student group. We consider these approaches mostly using examples from CSU.

Invitation of visiting professors. Usually, visiting professors present individual specialties, for example Fuzzy Modeling and Control, Decision Making in Uncertainty, Optimal State Estimation, Intelligent Sensors, Robotics, Biomechanics, Mechatro-

nics, etc. This is an efficient way to give students new knowledge based on research results within the framework of regular classes or special classes.

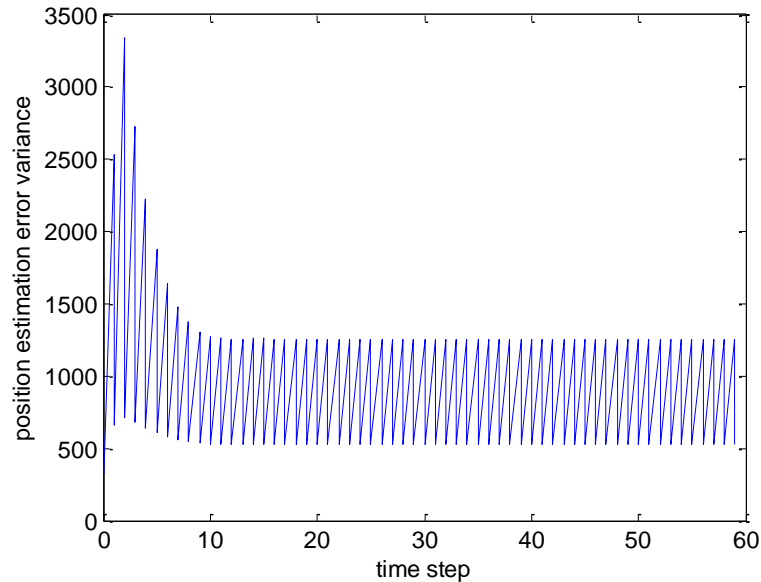


Fig. 5. Sample MATLAB output for the CSU Optimal State Estimation course

Students' participation in research projects and publication with professors.

When students conduct research (in the framework of research grants) with professors they can obtain a lot of new knowledge. Many CSU students are currently involved in research on the US NSF projects “Optimal prosthesis design with energy regeneration” (\$1.5 million), “The game changer: a new model for password security” (\$200,000), Acquisition of a 4G/LTE wireless communications test set” (\$252,000), “A spiral computer engineering lab framework” (\$245,000), and others. Students are heavily involved in the preparation of articles and papers for the publication of research results [17, 27]. This allows faculty to give students knowledge in modern data information processing and skills in formatting and formulating a goal, introduction, main idea, conclusion, references, citations, and so on.

Gathering students into a single research team. This approach broadens the perspective for knowledge transfer when students with different ICT knowledge can gather in one team for executing one or several projects. For example, students who are members of a research team may have various knowledge in using software for evolutionary optimization (genetic algorithms, partial swarm optimization, multi-objective invasive weed optimization, etc.), decision making based on Pareto optimization, sliding mode control, impedance control, fuzzy logic, artificial neural networks for parametric identification, pattern recognition, image processing, and so on. CSU student teams have the possibility to conduct research in laboratories in the EECS Department such as the Digital Systems Lab, Control Systems Lab, Power Systems Lab, Computer Networks & Distributed Systems Lab, Communications &

Table 1. Relationships between disciplines and software in the EECS Department at CSU

Course Code	Course Title	Software
ESC 120	Introduction to Engineering Design	Arduino (an open-source electronics platform based on easy-to-use hardware and software)
EEC 310	Electric Circuits I	PSpice, MATLAB, MultiSim
EEC 311	Electric Circuits II	
EEC 312	Electric Circuits Laboratory	Agilent IntuiLink software (Agilent scopes and signal generators, breadboards, passive components, transformers)
EEC 315	Electronics Laboratory	
EEC 451	Communications Laboratory	
EEC 314	Electronics II	PSpice, MATLAB/Simulink
EEC 384	Digital Systems Laboratory	EDA (electronic design automation) software from Altera and Xilinx (FPGA prototyping boards, logic analyzers)
EEC 487	Advanced Digital Systems	
EEC 488	Hardware-Software Co-design	
EEC 417	Embedded Systems	Miscellaneous electronics and Software (MPLAB from Microchip)
EEC 421	Software Engineering	Eclipse for Java development
EEC 440	Control Systems	PSpice, MATLAB/Simulink
EEC 450	Communications	SystemView by Elanix (design and simulate communication systems)
EEC 483	Computer Organization	(Microsoft Visual C, Quartus II from Altera (software), DE0 from Altera (hardware))
EEC 492	Software Defined Radio	Universal software (USRP) from Ettus (hardware) and GNU Radio (software), and LabView from National Instruments (software)
EEC 492	Computer Security	Quartus II from Altera (software), DE0 from Altera (hardware)
EEC 525	Data Mining	WEKA, RapidMiner, R (o/s software)
EEC 622	Formal Methods in Software Engineering	Model checker SPIN, Visual Studio
EEC 623	Software Quality Assurance	JUnit, GitHub, SPSS (statistical software)
EEC 624	Software Testing	JUnit (open source testing tools)
EEC 626	Software Engineering Project	Java, Mysql, Perl, Python, PHP, Apache, C#, SQL server, ASP.net, development tools: Eclipse, NetBeans, and, Visual Studio
EEC 683	Computer Networks II	Network simulator NS2
EEC 492/592	Kinect Application Development	Website - http://academic.csuohio.edu/zhao_w/teaching.html
EEC 688/788	Secure and Dependable Computing	
EEC 693	iPhone Application Development	
EEC 693	Network Security and Privacy	Various attack and dense tools

Electronics Lab, Software Engineering Lab, and Senior Design Lab. Available software includes: Altera Quartus II 10.1 SP1, Altera Quartus II 13.1, Cisco Packet Tracer, Oracle VM VirtualBox, Microsoft SQL Server 2008, Microsoft Visual Studio 2010, Microsoft Office 2010, Orcad family Release 9.2 lite Edition, ORCAD 16.5 lite, Python 2.7.5, MATLAB R2013b, dSPACE Control Desk 5.1, ModelSim SE 10.0a, Precision Synthesis 2010 a.218, SystemView V6.0, and Agilent Data Capture Application. BSSU students have the availability of Visual Studio, MS SQL Server, MS Windows Server, MS Windows 7, MS Access, MS Visio, MS Project, Free Ware, Moodle, Libre Office, Eclipse, NetBeans IDE, Ubuntu, FreeBSD, Apache, Qt, OmegaT, VirtualBox, Python, Java, JavaFX, C/C++, PHP, JavaScript, HTML5.

Participation of students in conferences, seminars, and research meetings. The goal of any conference is the sharing of knowledge and discussion of recent research results. CSU hosts a weekly Human Motion and Control seminar that includes senior researchers and students giving presentations (approximately half-and-half) on advanced research results with participation from the departments of EECS, Mechanical Engineering, and Engineering Technology. Moreover, every student has the responsibility to report their research achievements on a weekly basis at meetings or seminars of their separate research team, with additional short presentations on new funding, new methods, software, technology, sensors, computer components, and so on. The Annual Research Day of the WCE (CSU) is a scientific event with poster presentations by master's and PhD students, including time for discussion and awards for the best posters.

Cooperation of the university with advanced ICT companies. This approach allows for the possibility of knowledge transfer within the framework of lectures by ICT company representatives at the university, familiarization by the students with new ICT company software, student internships (NASA Glenn Research Center, General Electric, etc.), joint programming projects, certification of students, and creation of student start-up and spin-off companies [11, 15, 22, 40].

IT for papers, articles, and course work preparation. Students have the possibility to learn specific software editing of their manuscripts using different LaTeX or MS Word templates which correspond to specific journal or conference formatting requirements.

Textbooks and manuals for students based on the recent faculty experience. Courses in the EECS department at CSU are based on the faculty's own textbooks, classical / fundamental textbooks, and internet resources. For example, recent research results on optimal state estimation and evolutionary optimization are represented in faculty textbooks [37, 38] with accompanying MATLAB codes which is available on the CSU website [44].

User-friendly pseudo code in published articles and research projects. At the EECS department, every student has free access to pseudo code. Authors and developers transfer their programming achievements by sharing them with students and the world-wide research community. CSU's web site includes pseudo code which is developed by research teams or individual developers.

Internet search systems and databases. To increase their level of knowledge based from recent research results, students can use search systems from databases such as Scopus, Science Direct, Google Scholar, IEEE Xplore, and others;

Research portals like Research Gate, LinkedIn, and others. Students can gain new knowledge by following recent publications from specific authors, by asking and answering questions in dialogues with colleagues, and so on.

Memberships in professional societies. Memberships in professional and scientific societies, like IEEE, IFAC, PBD, and others, gives students a wide spectrum of opportunities for knowledge transfer using the resources of the corresponding society.

6 Evaluation of the Modified University Curricula and Approaches to Knowledge Transfer

An evaluation of the quality of training processes and of the quality of university graduates is a feedback from the implementation of the proposal in Section 4, Research-Based Modification of University Curriculum. Here we propose some indicators for evaluation processes:

a) Awards and participation of students in programming Olympiads, and student research project competitions. For example, a team from CSU's Washkewicz College of Engineering took first place at an international student design competition sponsored by the American Institute of Aeronautics and Astronautics (AIAA). For their winning project they designed, built, and tested an engine air particle separator for an unmanned ariel vehicle using 3D printing technology. BSSU student have been repeated winners in the Aldec, Inc. (USA) Olympiad on C++, VHDL and Verilog [20].

b) The level of published articles by students is indicated by databases such as the Web of Science, Scopus, etc. Other important considerations are the impact factor of corresponding journals, and the rank of conferences with student presentations (international, regional, university, college, department, etc.) [14].

c) Grading of student knowledge and erudition using traditional testing approaches (homework, midterm, term project, final exam) and using advanced software and ICT [9, 36].

d) Tracking graduates and applicants for PhD or DRE programs illustrates research aspirations and the desire for conducting continuing research.

e) Successful employment, for example: (1) CSU graduates work in US companies and industrial corporations such as Rockwell Automation, Phillips, Foundation Software, Winncom Technologies, UTC Aerospace Systems, Swagelok Company, RoviSys, American Railways, United States Postal Service, and others; (2) BSSU graduates work in Canada, France, Germany, Great Britain, Latvia, Netherlands, Norway, Poland, UAE, USA, and Ukraine, including companies such as Camo-IT, Ciklum, eBay, EPAM Systems, GeeksForLess, GlobalLogic, HostingMaks, LinkedIn, Luxof, Microsoft Research, MobiDev, NetCracker, Oracle, TemplateMonster, and others.

7 Conclusions

The authors have described research related to the increasing efficiency of university graduate training by modification of the curricula in electrical engineering, and computer science and engineering, based on the latest achievements and advanced research results in the corresponding fields. Analyzing and classifying the knowledge transfer and knowledge evaluation methods for examination and verification of the proposed curriculum modification approach, the authors have presented many educational examples and successful cases from the Department of Electrical Engineering and Computer Science at CSU (USA) and the Department of Intelligent Information Systems at BSSU (Ukraine). Because of the limited space of this paper, only a few specific use cases have been considered. All discussed approaches can be successfully implemented for graduate student curricula modification in other universities around the world, especially those which do not currently use all of the discussed methods.

Acknowledgements. The authors gratefully thank the Fulbright Program (USA) for providing the possibility to conduct research together in the USA by supporting Prof. Y. P. Kondratenko with a Fulbright scholarship.

References

1. Anderson, R.D.: Inquiry as an organizing theme for science curricula. In: Abell, S., Lederman, N. (Eds.), Handbook of research on science education. Mahwah, NJ: Lawrence Erlbaum Associates, 807–830 (2007)
2. Chou, C., Tsai, C.-C.: Developing web-based curricula: Issues and challenges. Journal of Curriculum Studies 34, No. 6, 623–636 (2002)
3. Clements, D.H.: Curriculum Research: Toward a Framework for "Research-Based Curricula". Journal for Research in Mathematics Education 38, No. 1, 35–70 (2007)
4. Conle, C.: An anatomy of narrative curricula. Educational Researcher 32, No.3, 3–15 (2003)
5. Diamond, R.M.: Designing and assessing courses and curricula: A practical guide. John Wiley & Sons (2011)
6. Dolmans, D.H., Schmidt, H.: The advantages of problem-based curricula. Postgraduate Medical Journal 72(851), 535–538 (1996)
7. Encheva, S., Sharil, T.: Multimedia factors facilitating learning. WSEAS Transactions on Advances in Engineering Education 4, No. 10, 203–209 (2007)
8. Encheva, S., Sharil, T.: Application of association rules in education. In: Intelligent Control and Automation, Lecture Notes in Control and Information Sciences 344, Springer Berlin Heidelberg, 834–838 (2006)
9. Encheva, S., Sharil, T.: Uncertainties in knowledge assessment. WSEAS Transactions on Advances in Engineering Education 5, no. 4, 895–900 (2008)
10. Ferren, A.S.: General education reform and the computer revolution. The Journal of General Education 42, No. 3, 164–177 (1993)
11. Fostering University-Industry Relationships, Entrepreneurial Universities and Collaborative Innovations. Meerman, A., Klieve, T. (Eds). Good Practice Series 2013, University Industry Innovation Network (2013)
12. Froyd, J., Ohland, M.W.: Integrated engineering curricula. Journal of Engineering Education 94, No.1, 147–164 (2005)

13. Hiltz, S.R., Murray, T.: Education goes digital: The evolution of online learning and the revolution in higher education. *Communications of the ACM* 48, No.10, 59-64 (2005)
14. Keane, J.: Teacher vs. computer: Where educators stand in the technology revolution. *The Journal Technological Horizons in Education* 30, No.1, 38-40 (2002)
15. Kharchenko, V.S., Sklyar, V.V.: Conception and models of interaction between university science and IT-industry: S2B-B2S. *Journal KARTBLANSN*, No. 8-9, 40-47 (2012)
16. Knight, P.T., Yorke, M.: Embedding employability into the curriculum. York: The Higher Education Academy. Retrieved (2011)
17. Kondratenko, G.V., Kondratenko, Y.P., Romanov, D.O.: Fuzzy Models for Capacitative Vehicle Routing Problem in Uncertainty. *Proceeding of the 17th Intern. DAAAM Symposium "Intelligent Manufacturing and Automation: Focus on Mechatronics & Robotics"*, Vienna, Austria, pp. 205-206 (2006)
18. Kondratenko, Y. P.: Robotics, Automation and Information Systems: Future Perspectives and Correlation with Culture, Sport and Life Science. In: *Decision Making and Knowledge Decision Support Systems, Lecture Notes in Economics and Mathematical Systems* 675, Gil-Lafuente, A. M., Zopounidis, C. (Eds.), Springer, Switzerland, 43–56 (2015)
19. Kondratenko, Y.P.: The Role of Inter-University Consortia for Improving Higher Education System. *Proceedings of Phi Beta Delta, Smithee, M. (Ed.), Volume 2, No. 1. Phi Beta Delta, Honor Society for International Scholars, USA, May 2011*, pp. 26-27 (2011)
20. Kondratenko, Y., Sydorenko, S.: Cooperation between Ukrainian Universities and Aldec, Inc. (USA) in the field of VHDL and Verilog introduction to design of digital devices. In: *Kondratenko, Y. (Ed.), Proceedings of Intern. Conf. "Higher Education Perspectives: The Role of Inter-University Consortia"*, Mykolaiv, Atol, pp. 150-153 (2004)
21. Kondratenko, Y., Kondratenko, V.: Soft Computing Algorithm for Arithmetic Multiplication of Fuzzy Sets Based on Universal Analytic Models, In: *Ermolayev, V. et al. (Eds.), Information and Communication Technologies in Education, Research, and Industrial Application. Communications in Computer and Information Science* 469, Springer, Switzerland, 49-77 (2014)
22. Kondratenko, Y.P., Kondratenko, G.V., Sidenko, Ie.V., Kharchenko, V.S.: Cooperation Models of Universities and IT-Companies: Decision-Making Systems Based on Fuzzy Logic. *Monograph, Kondratenko, Y.P. (Ed.), Kharkiv: MES of Ukraine, PMBSSU, NASU "KhAI" (2015), (in Ukrainian)*
23. Kondratenko, Y.P.: Revolution in Computer Science and Engineering and Its Impact on Evolution of Higher Education. In: *Revolucion, Evolucion E Involucion En El Futuro De Los Sistemas Sociales. IX Sesion Internacional Celebrada en Barcelona el 11 de Noviembre de 2014. Real Academia de Ciencias Economicas y Financieras, Barcelona, 127-160 (2014)*
24. Kushnir, N., Manzhula, A., Valko, N.: New Approaches of Teaching ICT to Meet Educational Needs of Net Students Generation. In: *ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proceedings of the 9th International Conference on ICTERI-2013, CEUR-WS, Vol. 1000, Kherson, Ukraine, June 19-22, 2013*, pp. 195-208 (2013)
25. Leask, B.: Bridging the gap: Internationalizing university curricula. *Journal of studies in international education* 5, No. 2, 100-115 (2001)
26. Lodwick, W. A., Kacprzyk, J. (Eds): *Fuzzy Optimization: Recent Advances and Applications*, Springer-Verlag Berlin Heidelberg (2010)
27. Lozovyy, P., Thomas, G., Simon, D.: Biogeography-Based Optimization for Robot Controller Tuning. In: *Computational Modeling and Simulation of Intellect, Igel'nik, B. (Ed.), IGI Global, 162-181 (2011)*

28. McCarthy, P.R., McCarthy, H.M.: When case studies are not enough: Integrating experiential learning into business curricula. *Journal of Education for Business* 81, No. 4, 201-204 (2006)
29. Navarro, P.: The MBA core curricula of top-ranked US business schools: a study in failure? *Academy of Management, Learning & Education* 7, No.1, 108-123 (2008)
30. Thornton, R.K., Sokoloff, D.R.: Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula. *American Journal of Physics* 66, No.4, 338-352 (1998)
31. Riordan, J.E., Noyce, P.E.: The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education* 32, No. 4, 368-398 (2001)
32. Rubin, R.S., Dierdorff, E.C.: How relevant is the MBA? Assessing the alignment of required curricula and required managerial competencies. *Academy of Management, Learning & Education* 8, No. 2, 208-224 (2009)
33. Schissler, H., Soysal, E.N. (Eds.): *The nation, Europe, and the world: Textbooks and curricula in transition*. Berghahn Books (2005)
34. Schmidt, M.A.: *Modeling the Adaptive Process of Consortia in Higher Education: A Structural Equation Modeling Approach*. Dissertation for the degree of Doctor of Philosophy, The Florida State University, College of Education, USA (2000)
35. Shackelford, R., et al.: *Computing curricula 2005: The overview report*. *ACM SIGCSE Bulletin* 38, No. 1, 456-457 (2006)
36. Shahbazova, S.N.: *Application of Fuzzy Sets for Control of Student Knowledge*. *Applied and Computational Mathematics* 10, No. 1, 195-208 (2011)
37. Simon, D.: *Optimal State Estimation: Kalman, H-infinity, and Nonlinear Approaches*. John Wiley & Sons (2006)
38. Simon, D.: *Evolutionary optimization algorithms: biologically inspired and population-based approaches to computer intelligence*. John Wiley & Sons (2013)
39. Spivakovsky, A., Vinnik, M., Tarasich, Y.: *Using ICT in Training Scientific Personnel in Ukraine: Status and Perspectives*. In: *ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer*. Proceedings of the 9th International Conference on ICTERI-2015, CEUR-WS, Vol. 1356, Lviv, Ukraine, May 14-16, 2015, pp. 5-20 (2015)
40. *The State of European University-Business Cooperation*. Final Report: Study on the Cooperation between Higher Education Institutions and public and private organizations in Europe. Science-to-Business Marketing Research Centre (2011)
41. Thomas, I.: Sustainability in tertiary curricula: what is stopping it happening? *International Journal of Sustainability in Higher Education* 5, No.1, 33-47 (2004)
42. Tondeur, J., Braak, J.V., Valcke, M.: Curricula and the use of ICT in education: Two worlds apart? *British Journal of Educational Technology* 38, No. 6, 962-976 (2007)
43. Zadeh, L.A., Abbasov, A.M., Yager, R.R., Shahbazova, S.N., Reformat, M.Z. (Eds): *Recent Developments and New Directions in Soft Computing*. *Studies in Fuzziness and Soft Computing* 317, Springer (2014)
44. <https://www.csuohio.edu>; <http://www.csuohio.edu/engineering/eecs/eecs> - Cleveland State University, Electrical Engineering and Computer Science (2016)
45. <http://www.chdu.edu.ua/index.php?page=computerscience&hl=ukr> - Black Sea State University, Computer Science (2016)

The Experience of the Master Classes as a Means of Formation of Readiness of Teachers to Implement Innovation

Nataliya Kushnir¹, Nataliya Osipova¹, Nataliya Valko¹, Olena Litvinenko²

¹Kherson State University

kushnir@ksu.ks.ua natalie@ksu.ks.ua valko@ksu.ks.ua

²Kherson National Technical University

mmkntu@gmail.com

Abstract. The article describes the experience of the master classes for school teachers and university lecturers to use ICT in the educational process. Improving teacher's/lecturer's qualification involves their willingness to follow the advanced technology. However, there are several factors that should be considered in the preparation of training courses for teachers. The authors consider the technology of the master classes as an alternative to long-term courses. This approach has several advantages: short term, productivity at work. During the master-class the main goal is not to teach how to use the technology perfectly but to show its advantages, persuade to use a particular technology, demonstrate the ease of use, and motivate teachers for the further use of the offered technologies in their professional activity. The experience of master classes for teachers of Kherson and Kherson region over the past 5 years was analyzed.

Keywords. Teaching strategies, ICT Tool, Teaching Methodology, Capability.

Key Terms. ICT Tool, Teaching Methodology, Technology

1 Introduction

This article describes the purpose and methodology of the master classes for the successful teachers' training programs on the use of appropriate ICTs and their knowledge of the trends and new approaches in the education system.

The introduction of new educational technologies in the educational process depends primarily on the willingness of a teacher. Without proper motivation of the teacher the best teaching ideas will not be implemented simply by the order. The famous Pareto principle 20/80 can be interpreted as following: only 20 % of teachers supports introducing innovations, 12 % of them are passive ones, and 8 % are active ones who encourage, develop and improve the technology. The majority of teachers

(80 %) oppose any innovation. This position does not depend on the innovation itself. However, trends in society require the restructuring of the education system and active introduction of innovation, which are able to prepare students for a successful life in the information society.

2 Related Work

Programs of involving teachers in learning using ICT are implemented by many researchers. At the same time general researches are conducted, for example, associated with the formation of teachers' ICT competence and aimed at effective implementation in the educational process and the further development of individual approaches (mobile studying, use of social networks in learning process and so on). Thus, in the research [10] a mobile learning project and experience in the use of mobile technologies in the training of teachers is described. The creation of flexible learning solutions that will allow teachers to have access to information using a variety of devices and training support in different situations is shown in the paper.

Another perspective direction of development of education is using social networks in learning [1]. Practical activities in the field of education, i.e. using of social networks in work to support learning, interaction of teachers and students, creating business relationships and knowledge sharing.

Analysis of the studies shows that the use of cloud technology is already normal practice for one third of population. In the future, this number will only increase and leading position will be owned by those companies and organizations that will offer their services using cloud resources, like services that can be customized to individual liking. This also applies to educational technology. Firstly cloud technologies allow universal access to data at any time. Secondly, they contribute to the construction of their own learning trajectory at the right pace.

According to research [4] 50 % of the adult population of Ukraine (over 15) use online resources, and two-thirds of them use social networking services. Modern gadgets (smartphones, tablets, laptops) are owned by 57 % of the population.

Conducted by UNESCO [14] research has shown that without appropriate training educators use technology only to solve their former tasks. In this case there is no transformation and qualitative growth of the level of teaching and learning.

As the recommendations UNESCO offers [14] the following stages of the introduction of ICT in educational practice:

- Demonstrate and model how mobile technologies can improve the quality of teaching, learning and management of the educational process.
- Share research data and analysis of mobile learning programs.
- Encourage dialogue between the main participants, including principals, teachers, students, parents, local authorities and the community on mobile technologies training.
- Provide a consistent concept of how technology, including mobile one, can contribute to the achievement of learning objectives.

However, there are some limitations such as the lack of theoretical and pedagogical foundations, sustainable integration in formal educational contexts and, in particular, the lack of support for teachers and training [2].

3 Problem Setting

Learning each trend is time-consuming process for a teacher. However, it is important to form teachers' sustained interest to the changes in society and the educational system under the influence of digital technologies. The main factors that need to convince teachers are:

- Change the role of ICT in Education (shift from a focus on "ICT for education" to the use of "ICT for education" [15])
- Use of digital devices which are natural part of the life of modern students' generation that meet their educational expectations.
- Emergence of new approaches in education such as SMART, IoT, STEM education (Science Technology Engineering Math), BigData, MOOC, BYOD (Bring Your Own Device), Blended Learning, Flipped Classroom, mobile learning, cloud, gamification of education, etc.
- Change in the spectrum of required specialties on a labor market (now and in the future) and the need for learning throughout life.

Thus, there is a need not only to teach the teachers to use certain software products and services in the work, but to form an understanding of the society development trends and their own willingness to learn new technology. Since 2012 we research the subject of teachers' training in ICT, use new approaches in teaching that meet the educational requirements of the today's youth, including training teachers through master classes [8]. In our view, conducting master classes are the most effective form for solving these tasks.

The purpose of the article is to describe the experience of the master classes as a means of formation of readiness of teachers to implement innovation.

Tasks:

- to analyze the main trends in modern education;
- to reveal teachers' attitude to the use of ICT in educational process;
- to determine the understanding of general trends in the development of education under the influence of ICT;
- to conduct master classes with the use of ICT tools and services for teachers.

In our view, for effective implementation of innovation each teacher should take certain steps:

1. To get knowledge about the nature, advantages and disadvantages of a particular innovation.
2. To accept innovation as the integral part.
3. To develop necessary competences, in particular informational.

4. To start using innovation.
5. To develop and spread innovation

There are both objective and subjective factors that explain the reluctance of most teachers to implement innovation:

- Mastering of innovation requires teacher to spend more time, which is not paid.
- It is necessary to rebuild style of work and usual teaching methods that causes a situation of discomfort.

Analyzing the results of the anonymous surveys (129 respondents), interviews with teachers, we find an explanation which is: "innovation – is just a passing fad", "the Headmaster wants to gain an authority", "I do not have opportunities to master innovation due to lack of free time" and others.

The article [9] considered three levels of information competence (I level – beginner; II level – active user; III level – expert). Based on these levels the authors propose to divide teachers' information competence according to their ability and willingness to use ICT in the learning process. The article also identifies the major reasons of teachers' unpreparedness for using ICT in learning. They include lack of motivation for using ICT; lack of complexity; learning computer skills only without the support of innovative educational technologies; ignoring the characteristics of adult education; neglect of interactive teaching methods; insufficient integration of knowledge and skills of students from different academic disciplines; insufficient formation of Computer Science teachers the skill concept of the 21st century.

However, today in our country there is a situation when teachers are supported by large companies and famous brands for the production of computer hardware and software. Since 2005 in Ukraine education programs were adopted: "Intel® Learning for the Future" and the educational program "Partnership in Learning", "Teachers online" and "Expert Teacher" (Microsoft). Also recently the Ministry of Education of Ukraine announced the cooperation with the company LEGO in the implementation of the robotics into learning process.

Under the impact of these initiatives teachers create their own educational spaces using ICT tools. Many teachers and lecturers are the professional bloggers, websites owners. An example is a community founded by Ivan Ivanov "ICT training of teachers in Ukraine" and its resource "To be able to live. The development of the XXI century skills using digital technology". This project was responded by hundreds of teachers and professors.

Teachers begin to link their professional growth not with deepen and broaden the knowledge of the school subject, but with the improvement of general pedagogical skills and knowledge. A special role in the professional growth of teachers belongs to the pedagogical aspects of ICT use.

Increasing the role of ICT in the educational process contributes to changing pedagogical techniques.

4 History of development of master class system in KSU

Conducting master classes in Kherson State University was founded in 2006 under the Intel program "Teaching for the Future." As a result of questioning teachers of pedagogical universities, who were trained under the program, expressed a desire to use IT in their work and realized that IT will develop an educational approach that focuses on the students' needs.

The cooperation of the Ministry of Education and Science of Ukraine with the Microsoft implementation of the joint program "Partnership in Learning" (Memorandum of Understanding of 28 October 2003. Protocol № 1 of 3 August 2004), to the order of MES # 693 of December 6, 2005. a program to train teachers and students of pedagogical universities 'Basis of Information Technologies' is implemented. Within the cooperation with pedagogical institutions of I-IV levels "Partnership in Learning' program supports 36 hours of seminars "Basics of Information Technology" for teachers of high schools. In 2009, 2 seminars for teachers of KSU were held. 17 university teachers took part in it.

The next step was the introduction of the course "Use of information technologies in educational process" for students of specialties of Physics, Mathematics, and a Primary School Teacher. However, this course is taught only to those teachers of with Computer Science specialization. At the same time, we saw that the students are sometimes hard to see the potential of a particular technology to be used in the educational process. They lacked the practice of working with school students. We developed the idea of organizing joint work of school teachers and students. We have chosen the master class as the best form. It is a modern form of educational training seminar for practical skills on different techniques and technologies to improve the professional level and share best experiences of participants, expanding horizons and introduction to the newest branches of knowledge.

Since 2012, we started cooperating with a number of Kherson schools and conducted master classes for teachers of educational institutions based on them. A feature of these master classes was that the teachers worked with students – future teachers. Students, as representatives of Net generation, easily mastered the technology, and teachers have seen better educational opportunities of new service [13]. Also, during the master classes we use gaming technology, brainstorming, group work, pair work. More detail on this experience in the article in 2014 [7], where the focusing was on the benefits of this technology for students. Each workshop was dedicated to one particular resource, such as didactic possibility of using word clouds; it was lasting for two hours.

Joint work of teachers and students in the format that often hard to implement. This is due to the time issue for every participant: students, teachers, professors. However, today the need for school teacher training is sharply increasing, and that's where the system of master classes can be an effective tool.

5 Setting up the Pedagogical Experiment

According to the importance of the formation of teachers understanding of education development trends under the influence of digital technologies, their willingness to self-development and use of new techniques and services we found the possibility of holding master classes for teachers of Kherson region who are passing courses on the basis of Kherson Academy of Continuous Education. According to Ukrainian law, each teacher must go through training 1 time in 5 years. To do this, they arrive at an appropriate center and have trainings for two weeks.

So, master class "Modern ICT tools and services for teachers of Mathematical and Technology disciplines" was first performed by the teachers of the department of Computer Science, Software Engineering and Economic Cybernetics of Kherson State University as part of the All-Ukrainian scientific-practical web conference with international participation "Innovative Dimension of Development of Mathematics and Technology Education ", held in October 29–30, 2015. This master class was held for 4 groups of Physics, Mathematics and Computer Science teachers in Kherson region. Total in the 2015, 129 teachers participated in the workshops.

Interest in the master classes was shown by the teachers of the Department of Physics and Methods of teaching of the Faculty of Physics, Mathematics and Computer Science in Kherson State University. The effectiveness of the master class significantly was improved through joint participation of school teachers and university professors. This event has become the platform to exchange experience and to find effective ways to use the web services offered in teaching physics in school. In the master class teachers and lecturers of physics took part, they teach future teachers of physics.

It made possible to pay attention to the peculiarities of the implementation of innovative approaches using ICT services in accordance with the methods of teaching school Physics course.

During the master classes our objective was to present teachers common tools that they can use in future professional activity in school regardless of the subject, to develop participants' interest to learning ICT and pedagogical and technological capabilities of the services, the ability and willingness to use them in their professional activities, to create an electronic bank of ideas on the use of ICT services for the solution of pedagogical problems.

Each workshop is designed for 7 hours in the computer lab with a video projector. The master class program includes the following issues:

1. Consideration of conflicts arising due to psychological characteristics of students with using modern and traditional approaches, methods and means of training.
2. Introduction to the concept of teachers 'flipped classroom'.
3. Use of Content management systems (CMS) and Learning Management Systems (LMS) (as examples there are systems used in Kherson State University: KSUOnline [6] and Virtual Kherson University [5]).
4. Highlighting educational opportunities and services (YouTube and International Educational Network of Khan Academy).

5. Create video tutorials for using services EdPuzzle and Blendspace, allowing to search desired videos from popular video sharing like YouTube, Vimeo, Khan Academy, LearnZillion etc. Create exercises, quizzes, additional questions and comments. Services include creating classes and have powerful features statistics.
6. Development of educational games using the service Learningapps.
7. Introduction to Internet services for the research and design activity for the Concept of STEM-education (Interactive works of virtual educational laboratory www.virtulab.net allow to conduct virtual experiments in Physics, Chemistry, Biology, Ecology and other subjects, in two-dimensional and in three-dimensional as well. Globallab – a platform that serves as the international environment of research interaction between students).
8. Introduction to interactive services and chronicles creation Dipity and Tiki-Toki.
9. Using of network services for the organization of modern educational process.
10. At modeling stage teachers were suggested to work in groups and to develop a lesson using Internet services: interactive video tutorial with test tasks; training game with the chosen theme; representing chronicle of historical facts related to the chosen training topic; publishing links to educational materials created in chosen CMS, LMS or blog.
11. Demonstration and discussion of copyright material for the lesson with the master class participants, use of these tools for enhancing learning activities.
12. At the stage of reflection there is a discussion on the results of joint activities; a survey of teachers to identify their satisfaction with the results of master class and intentions on how to use learnt technology in teaching is conducted.

Various online services use in their basis the principles of teamwork, cooperation, openness and accessibility. When choosing a service for the master class, we were guided by the following criteria:

- The feasibility
- Easiness to learn and to use the service
- Interactivity

The successful mastering of the master class theme was based on the productive activities of all participants.

During master classes these materials were used: "Introduction to Information Technology" and "Master Class" published on the website <http://ksuonline.kspu.edu/>.

6 Survey Analysis

For the final phase of the master class a survey was prepared. The survey contained 19 questions of closed type in three blocks:

- general information;
- experience in using ICT in the classroom;
- tendencies of changes in the education system.

The first set of questions concerned the general information about work experience, school location, and classes the teacher works with.

Among the interrogated there were many teachers with great work experience: 88 % have experience of more than 10 years. Majority of teachers work in secondary and high school (1-4 grades – 12 %, 5-11 grades – 88 %).

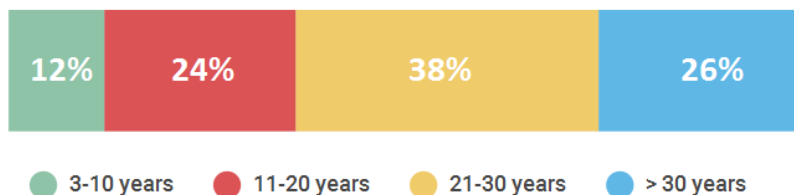


Fig. 1. Representation of teachers regarding work experience

As a master-class was conducted for teachers of training courses, most of the participants came from nearby towns and villages (88%).

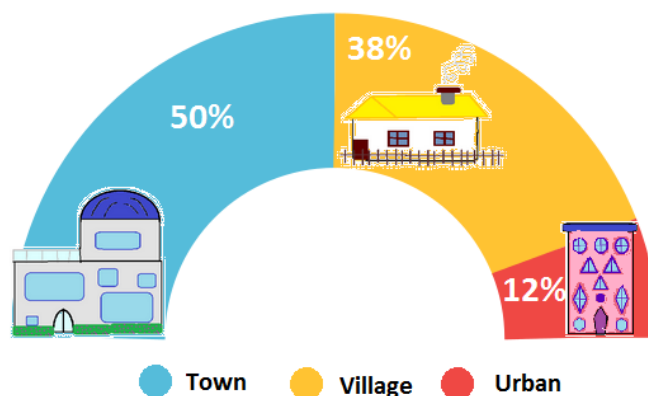


Fig. 2. Representation of teachers regarding school's location

The second set of questions concerned the identification of experiences and purposes of using ICT in professional work. Questions contained answers to the multi-selection (checkbox) and the line to write additional information.

The purpose you use Information technologies on a lesson is: *

- ☐ Never use
- ☐ To represent new topic
- ☐ To increase the motivation
- ☐ To examine students' achievement
- ☐ For assimilation of new material
- ☐ To interact with students
- ☐ To organise or entertain student
- ☐ Other

Fig. 3. Detail of the survey to identify the use of ICT experience

Three-quarters of respondents said that they use different devices in their practice, and a quarter said that they were going to use them, but do not have this opportunity yet. It should be noted that some respondents (3%) noted that their schools require lessons using ICT: teachers have to conduct lessons with an interactive whiteboard in the computer lab or using other ICT.

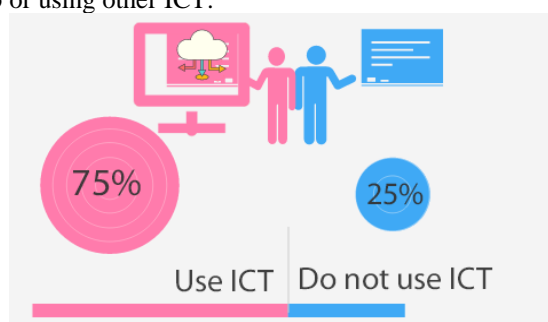


Fig. 4. Use ICT in teachers practice

Analyzing responses of 75% of teachers who use ICT, we found out the purpose of using ICT and distinguished the most popular activities. Traditionally, the most common purpose is to explain new material – 28 % and conducting current and final examination – 25 % (Table 1). 18 % of teachers use ICT in order to increase motivation and interest of students. Another 17 % of respondents use ICT to assimilate the material. None of the teachers gave their own version. The most difficult activity appeared the organization of communication and collaboration among students using digital technologies. From our point of view, this is due to the difficulty for the teacher to assess the contribution of each student in the joint work, the lack of formation of the teachers skills of cooperation organization of students work in groups.

Table 1. Activities with using of ICT

Activity	Answers, %
Explaining new material	28%
Increasing motivation	18%
Conducting current of final examination	25%
Assimilation of new material	17%
Improving communication with students	4%
Organization or entertaining of students	8%
Other	0%

On the question about their experience in using ICT in the classroom the following answers were recieved: laptop / computer (100 %), video (87 %), television (37 %), tablet (25 %), an interactive whiteboard (12 %) (Fig.5). Some respondents chose more than one answer. These data have improved significantly compared with surveys of previous years. They confirm the understanding of the need to use digital technologies in the classroom. However, the technical equipment is not always used effectively. It should also be noted that we do not show the statistics of equipped classrooms in the city and region where the average numbers may be very different. We present results of a survey conducted in November 2015 and January 2016 among teachers who held a master class. The findings of our survey data are correlated with the survey conducted among teachers of Computer Science in February 2015 [11].

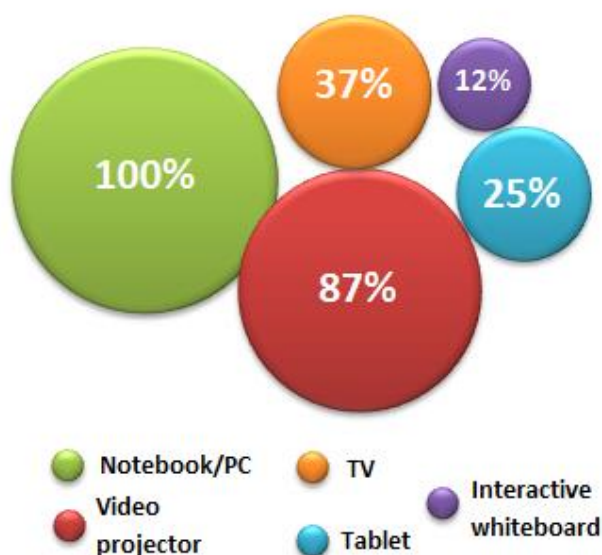


Fig. 5. Teacher's ability to use technical devices in the classroom

However, the teacher said that the possibility of using digital devices by students is less. Thus, almost all students are able to use computers in the classroom, 12 % of teachers give students an opportunity to use interactive whiteboards and mobile phones. Half of teachers (50 %) indicated that students have the opportunity to use tablet on the lesson.

When asked about their own activity in the network the largest number of respondents said they use e-mail (33 %). Among these, only 28 % use social networking, 17 % have blogs, and 11 % – a website. The intersection of these activities as blogging and maintaining the site, was not observed in the responses. Among the responses were others, including the absence of any activity (11 %).

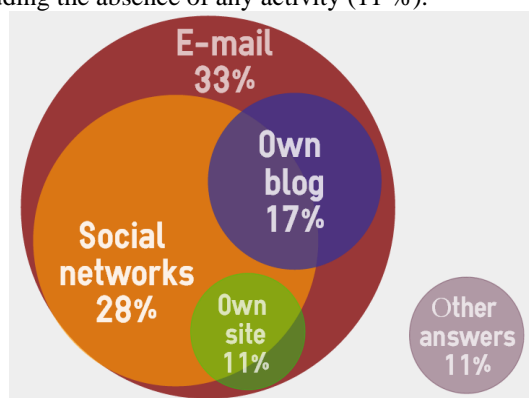


Fig. 6. Services used by teachers

One of the tasks we have set is to show teachers the opportunity to create their own information space. They can use a website or blog. Comparing the results of the answer to this question, we asked teachers in 2012, it should be noted that the percentage of teachers using websites and blogs in their work has increased. So, we can note the increase of use of blogs from 5 % to 17 %, sites – from 2 % to 11 %. E-mail is started using by 7% more. The number of social network users didn't change. After the master class, we monitored the sites and blogs of teachers. The audit of results made it possible to ensure that teachers use the proposed services and implement them in the learning process. The most popular are the video tutorials and training services LearningApps games. For example, the blog of teacher of mathematics Natalia Vas-kovskaya of Nikolskaya school (<http://natali-mir.blogspot.com/p/learningappsorg.html>), teacher Lebedenko Natalia of Gorkivska school (<http://lebedenkonat.blogspot.com/>), teacher Olhovska Natalia of Genichesk school №3 (<http://olkh68.wix.com/portfolio#!blank/gdznn>), teacher of Computer Science of Kherson school №27 Ignatenko Alla (<http://school2777.blogspot.com/>).

The last set of questions was developed with desirable psychometric properties to measure (detect) the level of awareness of current trends of changes in the education system. This block contains 6 questions and it was developed using a four-point Lik-ert-type scale, ranging from "strongly disagree" to "strongly agree". The trends high-

lighted as a key factor in the development of education for the period from 2010 to 2015 were taken as a basis [12].

Teachers tend to connect their professional growth not with the increasing knowledge on a subject but with improvement of general pedagogical skills. Special role in teacher's development belongs to the pedagogical aspects of using ICT *					
	1	2	3	4	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree
Learning Tools are changing. In training student uses tools that are used by a pro-fessional in different fields. So "chalk pedagogy" is replaced by different pedagogical techniques which concern the variety of learning styles of students. *					
	1	2	3	4	
Strongly disagree	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Strongly agree

Fig. 7. A fragment of the third block survey

It was not a set of questions but most likely thesis with which most teachers agree. In fact, all the questions received a response "strongly agree." The unanimous approval was the fourth point. Here is a list of thesis.

1. Teachers tend to connect their professional growth not with the increasing knowledge on a subject but with improvement of general pedagogical skills. Special role in teacher's development belongs to the pedagogical aspects of using ICT the variety of learning styles of students.
2. Learning Tools are changing. In training student uses tools that are used by a professional in different fields. So "chalk pedagogy" is replaced by different pedagogical techniques which concern
3. Teacher's mark is increasingly replaced by students' self-esteem, their mutual and automatic estimation as well. The function of a mark is not to evaluate students perform on the tasks but target students on what is needed to be done to extend academic work and achievement.
4. The teacher does not know and should not know all the answers. Students do not expect that the teacher gives them the answers to the questions, but they hope they get help in finding the answers by themselves. Answers to the questions can be found using many different sources, including the Internet.
5. The educational space expands beyond the classroom. The teacher encourages students to study in various environments, including outside schools both in real and virtual environments. Much of the work takes place in a real world, with the participation of fellow student and other adults other than teachers.

6. Individual work becomes a common activity. Most of the students do academic work together with classmates or within the creative (research) groups, using a computer and the Internet as the main tool of their activities. The teacher is preparing educational materials in digital format, exchanging designs with colleagues inside and outside the school.
7. Expanding the range of knowledge and educational information. The student determines and is an essential source of information and resources. The teacher encourages the materials and tools that can be used by student. More and more teacher start to use examples from real life, and not limited to the scope of textbooks only.
8. Increasing of students' independence and responsibility for the results of educational activities. The teacher teaches only the core disciplines form the students' ability to develop the subject beyond this core independently. Student determines what to do, based on existing knowledge and an idea of what is needed to know. The role of teachers is reduced to a more general management in different types of work that initiate and carry out the students themselves.

The lowest asset was given to the 3d point which is connected with the change of the role of evaluation at school. Teachers' acquaintance with molding evaluation was one of the goals of the "Intel@Teach to the Future" program. Self-esteem – is one of the priorities skills for successful professional development and lifelong learning. Minor teachers' support of this idea is the basis for us to develop this direction in the future (i.e. to add in courses thematic blocks and tasks aimed at development of formative assessment skills, to acquaint teachers with the appropriate on-line services on master classes). Low estimate was given to point 1 and 5 which is also directly related to the work of teachers. However, unanimous approval was given to point 4, which increases student's responsibility for learning achievements.

Important is the desire expressed by teachers in more detailed use of new trends in education in their work.

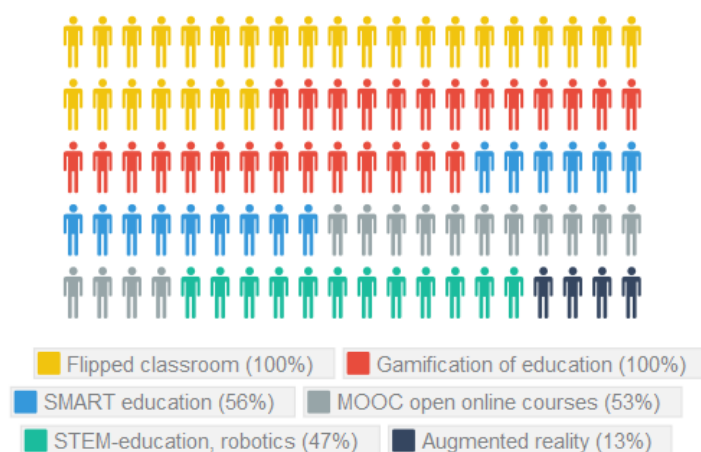


Fig. 8. The level of interest in the development and use of new educational trends

In the last survey, respondents noted what brands and services they would like to explore in more detail and use in their work. All of the teachers indicated that they wanted to use those services in the future which they learnt on a master class. The lowest support was given to STEM-education and Augmented Reality. In our opinion, it is connected with the material investment that needs to be done to implement these trends in education

7 Conclusions and Outlook

Technology of master classes can be a good alternative to long-term courses. This approach is well accepted by teachers and has several advantages like: short term, work productivity, creation of teaching materials using web services. During the master class the main goal becomes not to teach how to own the technology but to show its advantages and persuade to use a particular technology, to demonstrate the ease of its use, motivate teachers to further development of new services and the their use in the professional activity.

An important result is shifting teachers' point of view to changes in the education system under the influence of digital technology, learning and adopting new trends.

References

1. Arhipova, T., Osipova, N., Lvov, M.: Social Networks as a Means of Learning Process. In: Information Technologies in Education. 22, 7-18 (2015) (In Ukrainian)
2. Baran, Evrim.: A Review of Research on Mobile Learning in Teacher Education. Educational Technology & Society. 17.4, 17-32 (2014)
3. Berk, R.A.: Teaching strategies for the net generation. Transformative Dialogues: Teaching & Learning Journal, 3(2), 1–23 (2009)
4. For half of the Ukrainian Internet use; two-thirds of them - in social networks, <http://www.gfk.com/uk-ua/rishennja/press-release/vzhe-polovina-ukrajinciv-koristujutsja-internetom-dvi-tretini-z-nikh-v-socialnikh-merezhakh> (In Ukrainian)
5. Kherson virtual university, <http://dls.ksu.kherson.ua/>
6. KSUOnline, <http://ksuonline.kspu.edu/>
7. Kushnir, N., Manzhula, A., Valko, N.: Future and Experienced Teachers Should Collaborate on ICT Integration. In: Ermolayev, V., Mayr, H.C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) Information and Communication Technologies in Education, Research and Industrial Applications. CCIS, vol. 469, pp. 217-237. Springer, Heidelberg (2014)
8. Kushnir, N., Manzhula, A.: Formation of Digital Competence of Future Teachers of Elementary School. In: Ermolayev, V., Mayr, H.C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) Information and Communication Technologies in Education, Research and Industrial Applications. LNCS. vol. 347, pp. 230-243. Springer, Heidelberg (2013)
9. Morze, N.V.: How to Train Teachers to Computer Technologies Ceased to Be a Miracle in Training? The Computer at School and Family. 6, 10-14 (2010) (In Ukrainian)
10. Seppälä, P., Alamäki, H.: Mobile Learning in Teacher Training. Journal of Computer Assisted Learning. 19.3, 330-335 (2003)

11. Shovkun, V. Training of Future Teacher of Informatics to Work in Modern Information and Educational Environment of School. *Information Technologies in Education*. 23, 136-146 (2015) (In Ukrainian)
12. Soldatova, G., Zotova, E., Lebesheva, M., Shlyapnikov, V. Digital Literacy and Internet Security. *Methodical Manual for Specialists Basic General Education*. Google, Moscow. (2013) (In Russian)
13. Spivakovska, E., Osipova, N., Vinnik, M., Tarasich, Y.: Information Competence of University Students in Ukraine: Development Status and Prospects. In: Ermolayev, V., Mayr, H.C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) *Information and Communication Technologies in Education, Research and Industrial Applications*. CCIS, vol. 469, pp. 194-216. Springer, Heidelberg (2014)
14. UNESCO Policy Guidelines for Mobile Learning. <http://unesdoc.unesco.org/images/0021/002196/219641E.pdf>
15. Unwin, T.: Towards a framework for the use of ICT in teacher training in Africa. *Open Learning: The Journal of Open, Distance and e-Learning*. 20.2, 113-129 (2005)

Systems of Computer Mathematics for Educational Purposes as the Means of Reception of Procedural Knowledge

M. Lvov, L. Shishko, I. Chernenko, E. Kozlovsky

Kherson State University, 27, 40 rokiv Zhovtnya St., 73000 Kherson, Ukraine
{Lvov, Shishko, Chernenko, Evgen}@ksu.ks.ua

Abstract. This article describes systems of computer mathematics for educational purposes with intelligent properties oriented to support practical activities of users - students and teachers.

The article is presented general theoretical and methodological bases, functional requirements to systems of computer mathematics for educational purposes, models of systems of computer mathematics for educational purposes as systems of support of learning processes

Keywords. Systems of computer mathematics for educational purposes, computer software, procedural knowledge

Key Terms. Educational process, computer software.

1 Introduction

One of the key technological problems in the development of informative society in Ukraine is the informatization of social activities, including the computerization of educational and research activities [1, 2].

Over the last years, Ukraine has intensified processes of the informatization of education. In this regard, special relevance acquire general scientific, methodological and technological problems related to the organization of the processes of creation, maintenance and effective use of software for educational purposes (SEP).

Supporting professional mathematical activity is carried out with the help of modern professional systems of computer mathematics (PSCM). Their use provides specialists in theoretical and applied mathematics the ability to solve a wide spectrum mathematics tasks [3-5].

Practice of applications of PSCM in the learning process has shown that they do not solve all the problems related to the efficiency of the learning process.

Activities aimed at the assimilation of the courses in mathematics have a certain specificity. Mathematical practical activities of the student consist in solving mathematical tasks.

Market analysis of SEP in mathematics shows that there are several types of specialized software functional learning tools in mathematics [6-9]:

- educational programs such as e-textbooks, which offer the users of didactic materials and systems testing acquired knowledge;
- software support practical user experience on computing, graphics and geometric constructions;
- step by step solvers of educational tasks.

There are specific tasks that require the development of new specialized SEP in mathematics, based on the technologies of symbolic transformations and algorithms of computer algebra [10-15].

For this reason, it is necessary to create systems of computer mathematics for educational purposes (SCMEP) in which intelligent capabilities PSCM focused on specific tasks support the learning process.

We understand SCMEP as software for educational purposes for exact sciences and other disciplines that use mathematical models and methods of subject domains based on the technologies of character conversion and methods of computer algebra.

General model of SCMEP is the model of intelligent software that corresponds to the forms of the educational process organization. It is focused on all participants of the educational process and on all kinds of learning activities and based on the knowledge of the subject domain.

SCMEP provides the formation of a declarative (factual) and procedural (algorithmic) knowledge.

Thus, the purpose of the given work is the research of general theoretical and methodological bases, the formulation of the functional requirements to SCMEP and the development of a model of SCMEP as systems of support of learning processes based on the analysis of the topical forms and the peculiarities of the processes of learning of the exact disciplines.

2 The outline of the problem

The concept of programmed learning was introduced in the automated training systems and complexes. On the basis of this concept created a lot of computer courses in mathematics, physics and other disciplines.

In most SEP in mathematics [6-9] the lecture part of the course is the maximally advanced and sophisticated one from both methodical and technical points of view. As a rule, the educational material of the lecture is accompanied by a system of control questions and tests. Such computer system provides incomplete, indirect knowledge control. A mathematician should be able to solve the problem; a programmer should be able to write computer programs. These professional skills are formed during active forms of learning: practical exercises and laboratory work. So, this part of the curriculum in many disciplines (mathematics, physics, computer science ...) is central.

The functionality of software for educational purposes in mathematics is limited only by graphical and computing tasks. They do not perform functions which require character conversion and computer algebra methods. From our point of view, these

properties are crucial for creating effective intelligent SEP in mathematics, physics and other exact and natural sciences.

Market introduction of PSCM led to their intensive implementation into the learning process and numerous educational research on the application of PSCM in learning mathematics. Generally, supervisors recommend using of universal computer algebra system - Mathematics, Maple, Mathcad, Derive and etc. to support workshops [3-5, 16-17]. However, from our point of view, the use of PSCM in the learning process is somewhat limited.

First, PSCM are designed for solving mathematical tasks, while SEP in mathematics should support the process of solving mathematical tasks. This specificity is known as the principles of black and white boxes (Fig. 1).

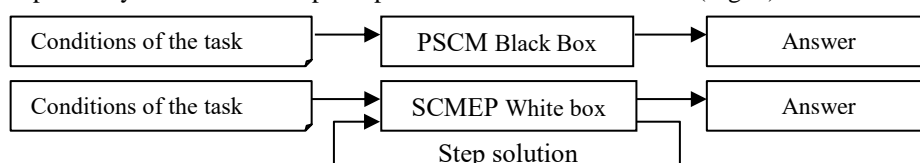


Fig. 1. The principle of black and white boxes.

Second, PSCM does not contain didactic materials. User Manual of PSCM can replace neither a textbook nor a book of mathematical tasks.

Third, the interface of PSCM is not focused on high school students. PSCM usually do not even use the specialized mathematical editor, only a line of editor with limited programming syntax.

Software products for educational purposes are based on the principle of white box do not generate the course of solving a task according to its conditions.

1. Answer Ace, Formulae 1, Universal Math Solver - step by step algebraic tasks' solvers.
2. Bagatrix Geometry Software, Bagatrix Algebra Software, Bagatrix Trigonometry Software, Bagatrix Graphing Software - step by step solvers of mathematical problems with capabilities of graphics support.
3. KwikTrig – step by step trigonometric problems' solver.
4. WebAlmir - an integrated software environment for learning the basics of linear algebra [18-20].

In the summary overview of educational software for mathematical disciplines, we are highlighting the following.

There are several functional types of specialized software for learning mathematics on the global market of software. First of all, these are teaching programs such as electronic books, which offer users didactic materials and systems of testing of acquired knowledge. Second, there are software tools of support a user's practical work with computing, graphics and geometric constructs. Third, there are step by step solvers of educational tasks. The most specialized software tools can combine only few of these functional types. Yet, each of these software tools is focused on different stages and on different categories of participants of the educational process.

The universal computer algebra systems, which are positioned as being suitable for use in the educational process, are operating on the principle of black box. They contain no electronic teaching materials (textbook, taskbook, etc.). Besides, they focus more on the application during the studying of applied sciences.

The practice of using PSCM in the learning process has shown that they do not solve all the problems related to the efficiency of the learning process. There are such specific tasks that require the development of new algorithms and specialized software systems for educational purposes, based on the technologies of character conversion and algorithms of computer algebra [10-15].

3 The model of learning process

The rapid growth of the professional requirements for professional widespread complex professionally-oriented information systems penetration into all areas of the global information and communication networks and other factors determine the following fundamental changes in educational technology [21]:

- a wide use of effective information technologies focused not only on classroom work - the work of the teacher and the students, but also on the teacher's individual work with a student, the student's individual work, both in the classroom and beyond;
- a wide use of distance learning not only in higher education but also in secondary and vocational schools;
- an individualization, increased internationalization of the learning process;
- widely used forms of remote access to knowledge, creating entirely new conditions for the organization of independent work.

The concept of information systems for educational purposes (ISEP) is in a comprehensive automation of support of the learning process, the main participants of which are the teacher as a professional, a carrier of knowledge in the field of didactics and teaching methods, in the subject area, and the student as an object of study.

The main goal of ISEP is a comprehensive providing of learning process.

The main tasks of ISEP are:

MT1. Ensure the relevance and accessibility of learning tools.

MT2. Support the process of transferring of new knowledge.

MT3. Support the knowledge control (feedback).

In accordance with the general schemes of the application of ICT in the learning process, we offer conceptual methodological requirements for ISEP:

MR1. ISEP must match the form of the learning process.

MR2. ISEP should focus on all participants of the learning process.

MR3. ISEP should focus on all kinds of training activities.

MR4. ISEP must be based on knowledge of the subject area.

Thus, the problem of the study can be defined as the study of general theoretical and methodological bases, the functional requirements of mathematical models and methods, technologies and tools of the SCMEP creation, that satisfy conceptual and methodological requirements 1-4 and solving the tasks MT1-MT3.

4 The information models for providing a process of learning a discipline

The class-form of organization of learning process for a certain discipline is regulated by curriculum course. This document is the basis for determining the content of documents and didactic materials such as:

- Thematic lesson's plan of discipline,
- Textbooks and educational manuals,
- Taskbooks,
- Libraries of supporting compendiums (visual aids),
- Summaries of lessons,
- Exercisebooks,
- Examination tasks (tickets),
- Methodical manuals,
- Glossaries and reference books.

An important role in informational support of a learning process, regardless of the form of organization, plays technical means of training. There are a blackboard, a video projector, a ruler, a calculator, laboratory equipment, computer, etc.

Models of didactic content ISEP. Information basis for modeling software ISEP is structural and logical schemes (SLS) of subject domain - speciality, discipline, a training module.

SLS of subject domain is a directed graph without cycles, which is the basic unit of domain knowledge and logical connections between them. The last one determine, in particular, the sequence of the study of knowledge units. In our opinion, the SLS should represent a three-level hierarchy:

"speciality" - "discipline" - "training module"

Models of a representation of the level of a learning module are described in knowledge engineering. They are semantic networks, frame models, object-oriented models. A well-known model of mathematical knowledge is semantic (algebraic calculating) Tiugu network [22]. Next, we review the mathematical models of representation of the procedural knowledge in SCMEP.

Model of information support of mathematical discipline is shown on Fig. 2.

The basic structural unit of informational support of a learning process is a training module. SLS of module in mathematical discipline defines:

- The contents of didactic materials;
- Signature of a learning module;
- A list of mathematical models of a learning module;
- A list of elementary transformations of models of a learning module;
- A list of types of training tasks of a learning module.

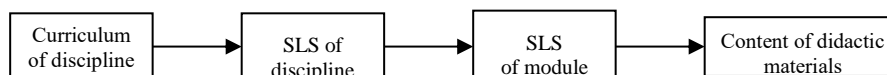


Fig 2. The structure of a knowledge representation of an academic discipline

The signature of learning module. Mathematical theories, which are taught in the module, use mathematical symbols (operations, predicates, functions). For example, Trigonometry module introduces the characters of trigonometric functions and inverse trigonometric functions, a symbol of the constant π . The list of these symbols consists of the signature of a learning module. The complete signature of this module consists of own signature of the module and signatures of modules which this module depends on in SLS of discipline.

Mathematical models of learning module. The subject of the study is the formal definitions of mathematical objects. In the training module Trigonometry it is, for example, the formal definition of trigonometric expressions, trigonometric equations, etc. Graphical tools play an important role in the presentation of mathematical models. Generally mathematical models of applied learning modules can and should be presented in the form of graphs, figures, charts, etc. Therefore, the formal definition of mathematical objects includes the information about the parameters of their graphical representations.

Educational tasks of learning module. The main subject of the study in the learning module of mathematics and exact disciplines is educational tasks; the list of their types is defined in the curriculum of a discipline. These are standard tasks. To solve them is the requirement of the state standard (the educational professional program. The formal definition of an educational task P includes the model of task $M(x_1, \dots, x_n)$, the condition of task $\varphi(x_1, \dots, x_n)$ and a question $Q(x_{j_1}, \dots, x_{j_m})$ to it ($P = \langle M, \varphi, Q \rangle$). The definition of educational task can be interpreted as:

Given $M(x_1, \dots, x_n)$ and $\varphi(x_1, \dots, x_n)$. Find $Q(x_{j_1}, \dots, x_{j_m})$.

For example, the task: Build a tangent to the graph F of the function $y = \frac{x+1}{x}$ at point A with the abscissa $x_1 = 1$ is submitted by the model:

$$M = F(y = f(x)) \ \& \ A(x_0, y_0) \ \& \ L(y - y_0 = f'(x_0)(x - x_0)) \ \& \ (y_0 = f(x_0)),$$

$$\text{condition } \varphi = (f(x) = \frac{x+1}{x} \ \& \ (x_0 = 1) \text{ and question } Q = ?L.$$

Let's note, that the model and questions of the task contain the model of the function $y = f(x)$, the points $A(x_0, y_0)$ and the straight (tangent) $y - y_0 = f'(x_0)(x - x_0)$. Letter designations F, A, L mean that these models have graphic images designated by the appropriate letters on the chart.

Elementary transformations of models. The process of solving of an educational task is defined as a sequence of steps, above each of which one of the elementary transformations $M \ \& \ \varphi \ \& \ Q$ is carried out. For each learning module specific transformations are defined. A complete list of elementary transformations of this learning module consists of specific transformations of this module and transformations of the modules which this module depends on in SLS of discipline.

The model of didactic content is shown on Fig. 3.

Thus, the formal definition of the learning module (Subject Domain) is defined by the quartet $SD = \langle \Sigma, MM, ET, Task \rangle$.

If the learning module in the SLS discipline depends on the modules SD_1, \dots, SD_k , then $\bar{\Sigma}_{SD} = \bigcup_{j=1}^k \bar{\Sigma}_j$. Similarly $\bar{ET}_{SD} = \bigcup_{j=1}^k \bar{ET}_j$. $\bar{\Sigma}_{SD}$, $\bar{\Sigma}_j$ are the complete signatures of modules (set of all symbols of functions and operations defined in this module). These parities are a formal definition of the modular structure of the learning discipline and speciality.

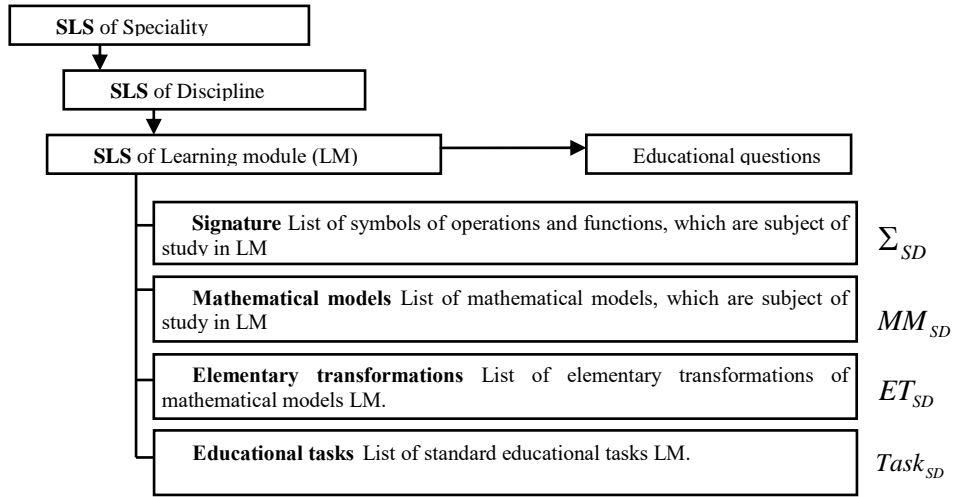


Fig. 3. Model of didactic content of ISEP for speciality

5 Models of the management of a learning process

The analysis of the learning process allows building basic models of learning management: lessons (Fig. 4), lectures, workshops, control works, editing didactic material and consultation.

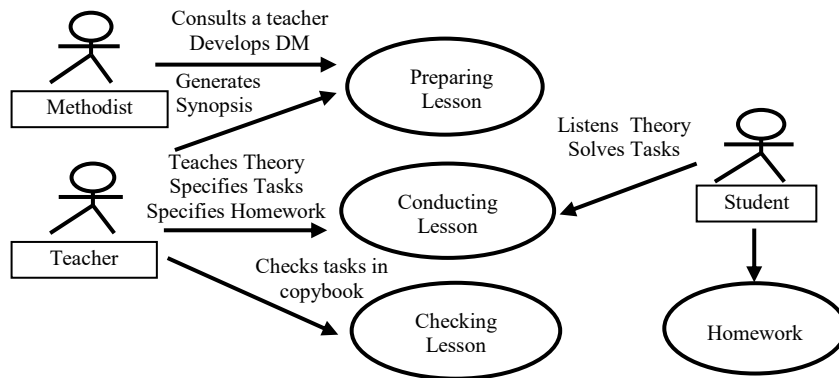


Fig. 4. Model business precedents "Lesson"

First of all, these models take into account all stages of learning - from the preparation to the supply of a relevant precedent's results, secondly, they show participants' learning activities.

The teacher's responsibility is the organizing of a lesson as a business-precedent. It means preparation and conducting of the lesson, reviewing the results of it. A student is responsible for homework, participation and completion of the lesson.

6 Activity environments as a means of support of practical mathematical activity

The content of an educational discipline for a class-lesson form of organization of a learning process has the following structure:

Academic discipline (*name*)

Subject (*number, name*)

Theoretical questions (*number, name*):

Content // Definition, Properties, Theorems, Proof, Examples ...

Practical learning objectives:

Content // Types of educational tasks. Methods for solving of educational tasks.

Thematic tests (*Number, Name*)

Content // Options for thematic control work

It is possible to extract the group of materials that contain educational information on the corresponding area of knowledge from all the (electronic) didactic materials in the learning discipline. These are a Textbook, a Taskbook, a Library of Supporting Compendiums, a Solving Environment, an Educational Manual, a Guide, Summaries of Lessons, an Exercisebook, a Library of Practical Lessons, a Testing Environment, etc. All of these didactic materials is intended to provide a certain form of a lesson, a type of educational activity and oriented on a specific category of users as well. The functionality of these modules is described in [23-25].

Practical mathematical student's activity is the main form of educational activity in the study of subjects, based on mathematical models and methods. Its educational goal is to build the course of solving educational tasks without getting answers. That is why object-oriented ISEP in mathematics, such as SCMEP, must support the process of mathematical problem solving. From the technological point of view, the information support of the solving of educational tasks is possible, only if its solution is in a specialized software module - activity environment (AE). AE is the main software modules of SCMEP. There are some works devoted to the mathematical methods of its construction [23, 26, 27-29, 30, 31, 32, 33].

The main function of the Solving Environment (SE) is to support the solving process of mathematical task in different modes.

Here are some basic functional requirements to the SE: the ability to export of a learning task from Taskbook or Exercisebook; the ability to save solving process of the task (fully or partially) in a Exercisebook; the ability of keyboard input of condition of an educational task of one of the standard types; the ability to select one

of the modes of solving the educational task; the ability to review a solving process of the educational task.

One of the most important aspects of practical mathematical support of the student is the validation of the performance of his actions at various stages of solving task - from the stage of the construction of mathematical models to the finishing stage of checking of the process solutions or answers. The second important aspect of the support is the automation of routine activities associated with the calculations. The third aspect is providing a convenient system of tips to a student in various stages of solving the problem like a generation of a mathematical model of a problem, process or step of its solution, answers.

Practical mathematical activity of the teacher should also be supported. The first aspect of this support is checking of the process of the problem solving. The system has to check the correctness of the process of the problem solving, done earlier by the student (mode of control work).

The second aspect of teacher's support is the automation of testing of students' knowledge. The specific activity environment is to verify the knowledge of basic mathematical rules and formulas (special test that uses mathematical tests).

The system should also ensure an effective management of the learning process in whole, supporting the interaction of a teacher and a student.

SCMEP must also provide the user with the appropriate mathematical tools (calculator, graphing tools, etc.).

Testing Environment. Testing is the main most common technology of knowledge control in systems for educational purposes. However, in practice the system of testing verification are oriented on declarative knowledge. The problem of checking the procedural knowledge needs to be resolved.

A characteristic feature of mathematical test is that the response must be in the form of mathematical (logical) expression, and the checking of answers is the checking of the semantic correctness of the expression.

Graphs. An important methodological role in the study of mathematics has graphic construction. The topic "Charts of a function" is a major cross-cutting theme in the course of school algebra. That is why graphical constructions' environments should be included as well in the SCMEP.

Calculator. The main function of the Calculator is to support the solving of learning mathematical task. It is the generation of the response or process of the solution and the answer. This environment implements the classic function Solve(), Simplify() professional systems of computer mathematics. The environment should also support algebraic and arithmetic calculations, including approximate calculations for processing of the results of laboratory experiments.

7 Implementation of the concept of systems of computer mathematics for educational purposes

The scientists of the department of computer science, software engineering and economic cybernetics of Kherson State University are engaged in the implementation of the concept SCMEP.

The researches were carried out during the process of the scientific and technical work on government contracts and state program "Information and communication technologies in education and science". They were based on the practical experience received during the development of ISEP and performed by the order of the Ministry of Education and Sciences (copyright certificates [34-38]):

- Program-methodical complex «TerM VII» of the support of a practical mathematical learning activity.
- Software tool "Library of Electronic Visual Aids Algebra 7-9 grade for secondary schools in Ukraine."
- Pedagogical software tool "Algebra, Grade 7".
- Software tool for educational purposes "Algebra, Grade 8".

The system approach was offered for the construction of models and program realizations of SCMEP as for the development of SCMEP that are listed above. The approach can be used at the construction of a wide class of program systems for educational and scientific purposes.

Currently several software products are developed using the same technology. «TerM VII» is one that has English interface.



Fig. 5. General view of the window «TerM VII».

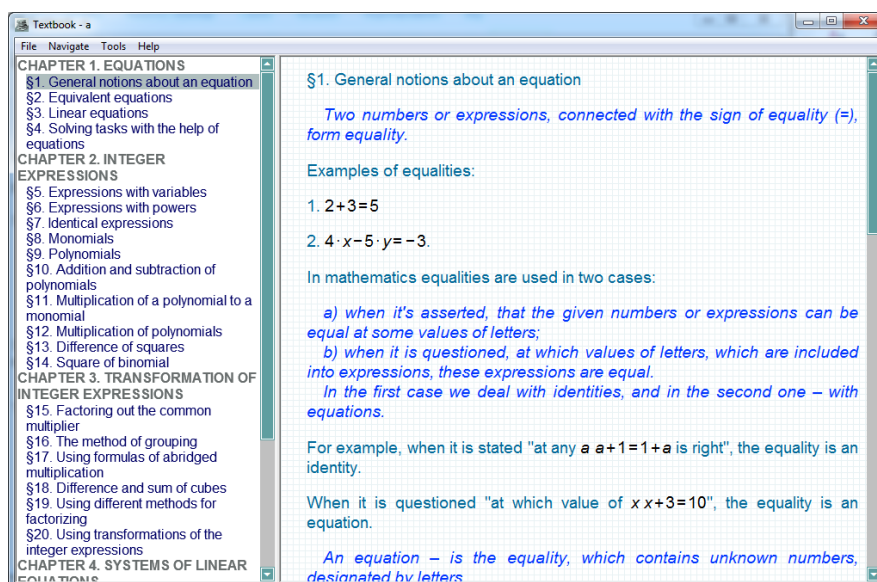


Fig. 6. General view of the page of Textbook

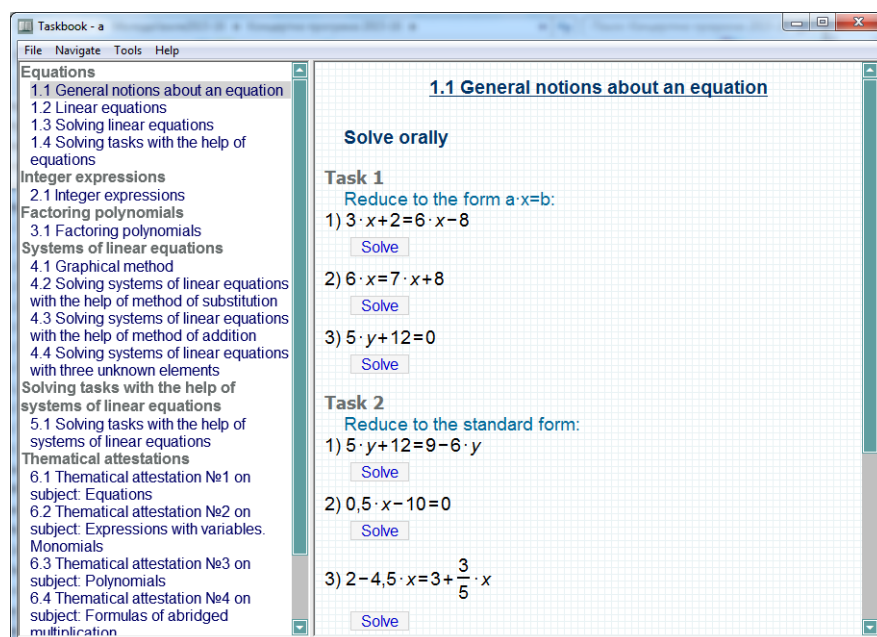


Fig. 7. General view of the window of TaskBook

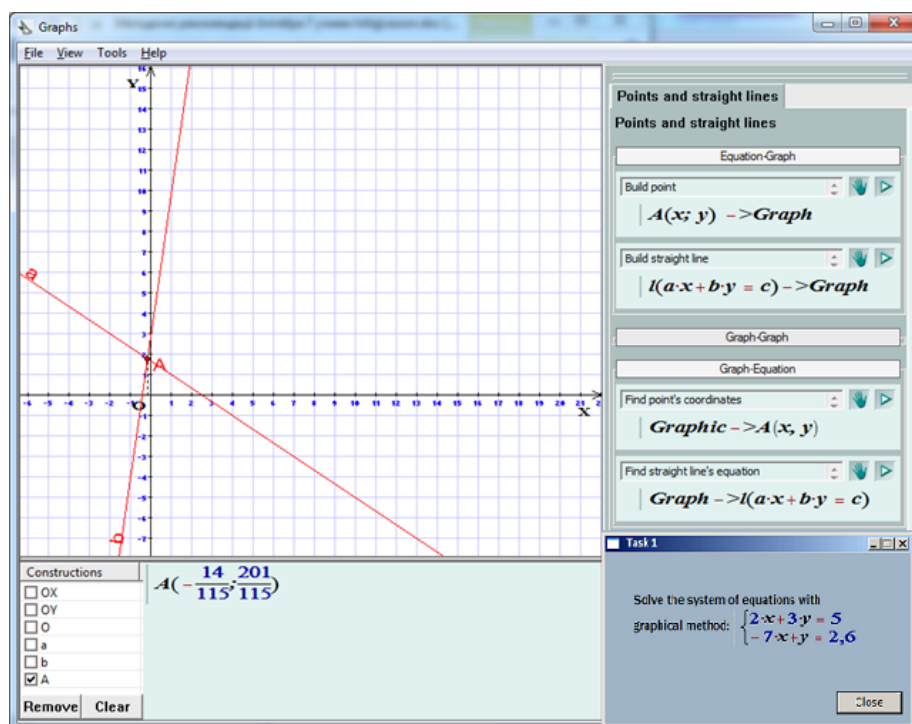


Fig. 8. Solving task in the window of AE "Graphs"

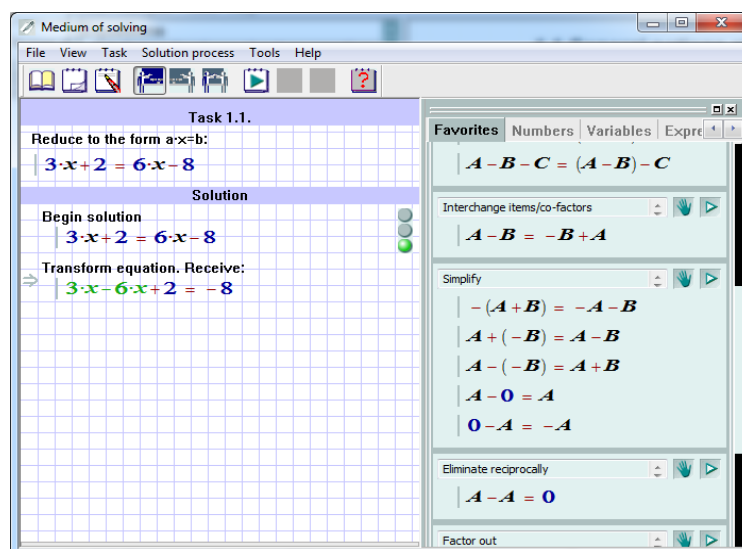


Fig. 9. View of the AE "Solving Environment" with a loaded task

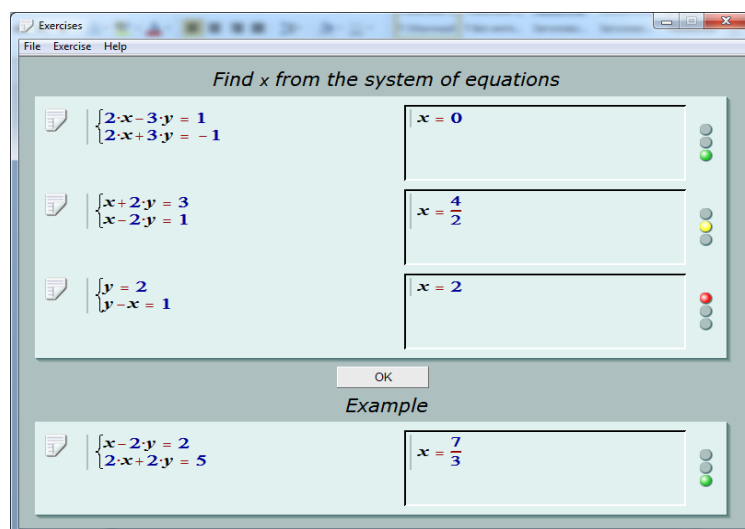


Fig. 10. View of the AE “Testing Environment” with a loaded test.

All SCMEP developed in the framework of this study are used in the educational process in Ukrainian schools. They successfully passed the procedure of the certification, the relevant Commission of Ministry of Education and Science of Ukraine, the established procedures of approbation in secondary schools of Ukraine and recommended for the use in the educational process.

8 Conclusion

ISEP is a system of the learning process support. Its main objectives are to ensure the relevance and accessibility of learning tools, support the processes of new knowledge transferring and the process of knowledge control (feedback).

ISEP should correspond to the contemporary forms of learning process in schools of various levels, be focused on all participants in the learning process, all forms of workshops and activities of participants of the learning process. ISEP must be subject-oriented, based on the knowledge of a corresponding subject domain.

The system of training SCMEP contains appropriate didactic materials and the activity environment for disciplines based on mathematical models and methods.

Practical mathematical activity of a student is the main form of learning activities in the study of subjects, based on mathematical models and methods. Its aim is solving of mathematical learning problems.

Activity environments are computer learning tools designed to support practical mathematical activities of participants of the learning process.

ISEP is the system distributed to the working areas of the participants of the learning process. Each of the area in its composition has appropriate tools of control of the training sessions and types of activities, personalized system of learning tools and tools of the interaction of participants of the learning process support.

9 References

- 1 Natsionalna doktryna rozvytku osvity. Zatverdzheno Ukazom Prezydenta Ukrainy vid 17 kvitnia 2002 roku № 347/2002 // Osvita Ukrainy. – 2002. – № 33. –18 s.
- 2 Dodonov O.Gh. Derzhavna informacijna polityka i stanovlennja informacijnogho suspiljstva v Ukraini / O.Gh. Dodonov, O.S. Ghorbachyk, M.Gh. Kuznecova // Strategichna panorama. – 2002. – № 1. – S. 166–170.
- 3 Karpuhina S. V. Personalizirovannoe obuchenie algebre i nachalam matematicheskogo analiza s ispol'zovaniem komp'juternoj sistemy "Mathematica" : dis. ... kand. ped. nauk: 13.00.02 / Karpuhina Svetlana Viktorovna. – Rjazan', 2009. – 243 s.
- 4 Poznjak Ju.V., Zemskov S.V, Kuleshov A.A. Ispol'zovanie sistemy Mathematica dlja sozdanija jelektronnyh uchebnikov po estestvennym i tehničeskim naukam na primere bazovogo kursa vysshej matematiki.-Jelektronnyj resurs.
- 5 Alekseev E. R. Reshenie zadach vychislitel'noj matematiki v paketah Mathcad 12, MATLAB 7, Maple 9 / E.R. Alekseev, O.V. Chesnokova. – M.: NT Press, 2006. – 496 s.
- 6 Zhaldak M.I. Matematika s komp'juterom: posobie dlja uchitelej / M.I. Zhaldak, Ju.V. Goroshko, E.F. Vinnichenko. - K.: RUNC „DINIT”, 2004. – 251 s.
- 7 Zhaldak M.I. Komp'juter na urokakh gheometriji: posibnyk dlja vchyteliv / M.I. Zhaldak, O.V. Vitjuk. – K.: RUNC „DINIT”, 2003. – 168 s.
- 8 Zhaldak M.I. Pedagogichnyj potencial komp'juterno-orijentovanykh system navchannja matematyky / M.I. Zhaldak // Komp'juterno-orijentovani systemy navchannja: zb. nauk. Pracj.- K.: NPU im. M.P.Draghomanova. – Vypusk 7. – 2003. – S. 3–16.
- 9 Rakov S.A. Matematychna osvita: kompetentnisnyi pidkhid z vykorystanniam IKT: Monohrafiia / S.A.Rakov. – Kh.: Fakt, 2005. – 360 s.
- 10 Kalinina N.A., Khagai J.E. Educational information environment on the base of Computer Algebra Systems // Cbornik trudov konferencii "IT+SE'99"; trudy Mezhdunarodnoj nauchnoj konferencii "Informacionnye tehnologii v nauke i obrazovanii". - Gurzuf, 1999. - S. 159-160.
- 11 Levinskaja M.A. Intellektual'naja sistema s vizual'nym preobrazovaniem, sopostavleniem i vychisleniem formul. / M.A.Levinskaja, V.E.Zajcev // Tezisy dokladov IX Mezhdunarodnoj shkoly-seminara "Novye informacionnye tehnologii" –M.: MGI-JeM, 2001. – S. 388–390.
- 12 Levinskaja M.A. Refal na derev'jah kak mashina logicheskogo vyvoda v intellektual'noj obuchajushhej sisteme. / M.A.Levinskaja // Materialy mezhdunarodnoj nauchno-prakticheskoy Internet-konferencii "Informacionnye tehnologii v nauke i obrazovanii". – Shahty, 2001. – S. 80 – 82.
- 13 Podkolzin A.S. Komp'juternoe modelirovanie logicheskikh processov. Tom 1. Arhitektura i jazyki reshatelja zadach. /A.Podkolzin. –M: Fizmatlit. –2008, 1024 s.
- 14 Sergushicheva A.P. Modelirovanie sinteza intellektual'nyh testov sredstvami formal'noj produkcionnoj sistemy./ A.P.Sergushicheva, A.N.Shvecov.// Tezisy dokladov X Mezhdunarodnoj konferencii "Matematika. Komp'juter. Obrazovanie". – Pushhino, 2003. – S. 66.

- 15 Aksenov M.V. Tehnologija razrabotki jekspertno-obuchajushhih sistem, orientirovannyh na obuchenie tochnym disciplinam: dis. ... kand. tehn. nauk: 05.13.01 / Aksenov Maksim Vladimirovich. – Sankt-Peterburg, 2004. – 110 c.
- 16 Diakonov V. P. Maple 9 v matematyke, fizyke y obrazovanyy / Diakonov V. – M.: SOLON-Press, 2004. – 688 s.
- 17 Alekseev E. R. Reshenye zadach vichyslytelnoi matematyky v paketakh Mathcad 12, MATLAB 7, Maple 9 / E.R. Alekseev, O.V. Chesnokova. – M.: NT Press, 2006. – 496 s.
- 18 Spivakovsky O.V. Teoriia y praktyka vykorystannia informatsiinykh tekhnolohii u protsesi pidhotovky studentiv matematychnykh spetsialnostei: Monohrafiia / O.V.Spivakovsky. – Kherson: Ailant, 2003. – 229 s.
- 19 Spivakovsky O.V. Prohramno pedahohichniy zasib «Svit liniinoi alhebry» / O.V. Spivakovsky // Visnyk Khersonskoho Derzhavnoho Tekhnichnoho Universytetu. – Kherson: KhDTU, 2003. – Vyp. 3 (19). – S. 402–405.
- 20 Spivakovsky O.V. Vykhidni polozhennia pobudovy metodychnoi systemy navchan-nia liniinoi alhebry na osnovi komponentno-orientovanoho pidkходу / O.V. Spivakovsky // Dydaktyka matematyky: problemy i doslidzhennia: Mizhnarodnyi zbirnyk naukovykh robit. – Vyp. 25 – Donetsk: Firma TEAN, 2006.- 260 s. (Mizhnarodna prohrama «Evrystyka ta dydaktyka tochnykh nauk»)
- 21 Lvov M.S. Pedahohichni tekhnolohii ta pedahohichno-orientovani prohramni systemy: predmetno-orientovanyi pidkhid. / O.V.Spivakovsky, M.S.Lvov, H.M.Kravtsov ta in. // Kompiuter u shkoli ta simi:– №2 (20), 2002 – S. 17–21; №3 (21), 2002 – S. 23–26; №4 (22), 2002 – S. 24–28.K
- 22 Tyugu Je.H. Konceptual'noe programmirovanie / Tyugu Je.H. – M.: Nauka, 1984.– 255 s.: il.
- 23 Lvov M.S. Konceptcija prohramnoji systemy pidtrymky matematychnoji dijajlnosti./ M.S.Lvov // Komp'juterno-orijentovani systemy navchannja: Zb. nauk. praci/ K.:NPU im. M.P.Draghomanova.– Vyp. 7.–2003.– S.36–48.
- 24 Lvov M.S. Term VII – shkilna systema kompiuternoi alhebry. / M.S.Lvov // Kompiuter u shkoli ta simi. – 2004. – №7.– S. 27–30
- 25 Lvov M.S. Shkilna systema kompiuternoi alhebry TerM 7-9. Pryntsypy pobudovy ta osoblyvosti vykorystannia. / M.S.Lvov // Naukovyi chasopys NPU im.Drahomanova, seriia №2. Kompiuterno-orijentovani systemy navchannia: zb.nauk. prats/ redkol. –K.:NPU im.Drahomanova.–#3(10)–2005. s. 160–168.
- 26 Lvov M.S. Proektirovanie logicheskogo vyvoda kak poshagovogo reshenija zadach v matematicheskikh sistemah uchebnogo naznachenija / M. S. Lvov // Upravljajushhie systemy i mashiny. – 2008. – №1. – S.25–32.
- 27 Lvov M.S. Konceptcija, arhitektura i funkcional'nost' gibkoj raspredelennoj programmnoj sredy uchebnogo naznachenija dlja srednej shkoly. Rabochee mesto metodista /M. Lvov //Upravljajushhie systemy i mashiny. –2009. –№ 6. –S. 71-78.
- 28 Lvov M.S. Raspredelennye programmnye sredy uchebnogo naznachenija. Podsystema upravljenija uchebnym processom / M.S. Lvov // Upravljajushhie systemy i mashiny. – 2010. –№1. – S. 66 – 71.
- 29 Lvov M.S. Realizacija obchyslenj v alghebrakh chyslovykh mnozhyn v matematychnykh systemakh navchaljnogho pryznachennja / M.S.Lvov // Upravljajushhie systemy i mashiny.– 2010. – №2. – S. 39 – 46.

- 30 Lvov M.S. Matematicheskie testy v sistemah komp'yuternoj matematiki uchebnogo naznachenija. / M.S.Lvov // Upravljajushhie sistemy i mashiny. – 2011. – №6. – S.60–67
- 31 Lvov M.S. Matematychni modeli ta metody pidtrymky khodu rozv'jazannja navchalnykh zadach z analitychnoji gheometriji/ M.S.Lvov // Iskustvennyj intelekt. – №1. – 2010. –C. 86–92.
- 32 Lvov M.S. Integhrovane proghramne seredovyshhe vyvchennja kursu analitychnoji gheometriji dlja VNZ. koncepcija, arkhitektura, funkcionalnistj / M.S.Lvov // Naukovi praci nacionaljnogho universytetu kharchovykh tekhnologhij.– №30.– Kyjiv: NUKhT, 2010. – S. 106 – 109.
- 33 Shishko L.S., Chernenko I.E. Pro orhanizatsiiu kontroliu znan yak zvorotnoho zviazku v systemakh kompiuternoi matematyky navchalnogo pryznachennia / L.S. Shishko, I.E. Chernenko // Informatsiini tekhnolohii v osviti: Zb. nauk. pr. Vyp 16. - Kherson: Vyd. KhDU, 2013. – S. 29-42.
- 34 A.s. 12190 MONU. Derzhavnyi departament intelektualnoi vlasnosti. Zbirka kompiuternykh proghram «Proghramno-metodychnyi kompleks «TerM VII» pidtrymky praktychnoi navchalnoi matematychnoi diialnosti» osvitnoi haluzi «Matematyka» dlja zahalnoosvitnykh navchalnykh zakladiv / Lvov M.S., Spivakovsky O.V., Kravtsov H.M., Kot S.M.,Herasymenko K.S., Peschanenko V.S., Hrabovsky A.Y., Khoruzhenko A.O., Tatochenko V.I., Nihirniak D.I., Bieiko Y.V., Letichevsky O.A., Chuhaenko O.V., Lvova N.M. zaiavl. 08.02.2005. Biul. № 7. – 2005.
- 35 A.s. 12440 MONU. Derzhavnyi departament intelektualnoi vlasnosti. Zbirka kompternykh proghram Proghramno-metodychnyi kompleks «TerM VII» pidtrymky praktychnoi navchalnoi matematychnoi diialnosti.Versiia 2.0, reliz 03.» osvitnoi haluzi «Matematyka» dlja zahalnoosvitnykh navchalnykh zakladiv / Lvov M.S., Spivakovsky O.V., Kravtsov H.M., Kot S.M.,Kreknin V.A., Peschanenko V.S., Hrabovsky A.Y., Khoruzhenko A.O., Tatochenko V.I., Nihirniak D.I., Bieiko Y.V., Letichevsky O.A., Chuhaenko O.V., Lvova N.M., Lvova M.M.; zaiavl. 10.03.2005. Biul. №7. – 2005.
- 36 A.s. 19840 MONU. Derzhavnyi departament intelektualnoi vlasnosti. Proghramnyi produkt «Biblioteka elektronnykh naochnosti Alhebra 7–9 klas dlja zahalnoosvitnykh navchalnykh zakladiv Ukrainy» osvitnoi haluzi «Matematyka» dlja zahalnoosvitnykh navchalnykh zakladiv/ Lvov M.S., Spivakovsky O.V., Kreknin V.A., Peschanenko V.S., Hrabovsky A.Y., Malyshev K.I., Dubina H.H., Trokhymchuk O.I.; zaiavl. 12.03.2007. Biul. №12. –2007.
- 37 A.s. 32722 MONU. Derzhavnyi departament intelektualnoi vlasnosti. Kompiuterna proghrama «Pedahohichnyi proghramnyi zasib «Alhebra, 7 klas» osvitnoi haluzi «Matematyka» dlja zahalnoosvitnykh navchalnykh zakladiv/ Peschanenko V.S., Lvov M.S., Spivakovsky O.V., Kreknin V.A., Shishko L.S., Chernenko I.E., Vinnyk M.O.; zaiavl. 06.04.10. Biul. №21. – 2010.
- 38 A.s. 32723 MONU. Derzhavnyi departament intelektualnoi vlasnosti. Kompiuterna proghrama «Pedahohichnyi zasib navchalnogo pryznachennia «Alhebra, 8 klas» osvitnoi haluzi matematyka dlja zahalnoosvitnykh navchalnykh zakladiv» / Peschanenko V.S., Lvov M.S., Spivakovsky O.V., Kreknin V.A., Shishko L.S., Chernenko I.E., Hrabovsky A.Y., Kozlovsky Ye.O., Vinnyk M.O.; zaiavl.06.04.10. Biul. №21. – 2010.

Educator's e-Portfolio in the Modern University

Nataliia Morze¹, Liliia Varchenko-Trotsenko¹

¹Borys Grinchenko Kyiv University, 18/2 Bulvarno-Kudriavska Str, Kyiv, Ukraine,

n.morze@kubg.edu.ua, l.varchenko@kubg.edu.ua

Abstract. The most effective higher education quality indicators are openness and transparency of all the University activities results, including the results of educational and scientific work of every teacher and student.

That is why modern electronic university learning environment should provide open decentralizing component that would contain a webpage with every educational process participants' qualitative and quantitative indicators of educational activity – an e-portfolio. The motivation to filling a portfolio by every teacher can be university ranking which takes into account in certain proportion all teachers' activities and has moral and material consequences. Indicators of a portfolio should include those of a priority for university development at a certain time and be taken into account on different ratings, including international and Ukrainian. There are different ways of creating and filling an e-portfolio. One of them is the Wiki-technology use that ensures openness and transparency of the performance of the teacher, each structural unit and the university as a whole, used at Borys Grinchenko Kyiv University. E-portfolio on the university Wiki-portal should be considered a generalization of the electronic educational environment various components of the university, in particular, LMS, Moodle, creation and usage statistics of faculty e-learning courses, system of student survey about the quality of every teacher and discipline, filling institutional repository with research and methodological publications, relevant catalogues on teachers' training competition winners, participation in international scientific research projects and more.

Keywords: e-portfolio, institutional repository, Google Scholar, Wiki-portal, quality of education, e-learning environment of the university.

Key Terms: Information Communication Technology, Teaching Process, Teaching Methodology.

1 Introduction

European Association for Quality Assurance in Higher Education (ENQA) has proposed the standards for determining the quality of higher education: internal and external [8], newly-created and based on the experience of Evaluation of the educational quality in the countries of the Western Europe. Standards for internal quality assurance in higher education relate to such areas of activity:

- Policy of the educational institution and the procedures for quality assurance;
- Approval, monitoring and periodic review of programs and diplomas;
- Assessment of students;
- Quality assurance of teaching staff;
- Educational resources and support for students;
- Informational systems;
- Public access to the information.

Indicated standards and recommendations set the common European format to quality assurance and for creation of a single European educational environment. Ukraine is a participant of such an environment that is why higher education institutions have to implement these standards, as stated in the Law of Ukraine on higher education [4].

The portfolio has started to be in use recently in Ukraine. But the single interpretation of the portfolio functions has not been formulated yet, each one uses portfolio in a different way, there is no approach needed to formulate the structure of the final portfolio of scientific-pedagogical staff members, what reflects all aspects of his work – his reflection, reflection of qualitative and quantitative indicators of professional development, results' presentation to the colleagues and students, analysis of administration work and so on. Modern society focuses on quality in everything, including in education and teacher plays a crucial role in shaping the quality of universities. Therefore, measuring the results of scientific and pedagogical employee is very important in our time.

The aim of the article is a description of e-portfolio structure of scientific-pedagogical staff member, as a possible tool for measuring the results of his performance and its further usage in the activities of the higher education institution. Defining the model for formation of e-portfolio of scientific-pedagogical staff members and its usage in the activities of the university.

2 USAGE OF E-PORTFOLIO OF SCIENTIFIC-PEDAGOGICAL STAFF MEMBER IN ACTIVITY OF A MODERN UNIVERSITY

Let's have a detailed look at the standard of "Quality assurance of teaching staff" and define the instruments for measuring the results of activity of the scientific-pedagogical staff member of a modern university. Recommendations of this standard are: these universities should have certain procedures and criteria that would certify

that professors who work with students are qualified and have highly professional level to carry out their duties. These procedures and criteria on which the check is based should be in open environment. The recommendations stated that professor “is the most important learning resource available for the majority of the students”. Therefore, it is important that professors had some level, know their subjects, applied ICT, professionally developed and engaged publication activity. Also “they have access to information about how their work is evaluated”. Universities should use such “procedures for the selection and appointment that allow verifying that the new professor has at least a basic level of competence as a minimum”. The university has to provide all conditions for effective activity. Let us define the parameters of this standard:

- Professors’ surveys;
- Students’ surveys;
- Availability of ICT competency standards for professors;
- Public reporting of faculty members, departments, institutes;
- System for in-service training for professors;
- System of rating indicators of professors’ performance.

E-portfolio can be an instrument for measuring the quality of the teaching staff by assessing the quantity and quality of professors’ performance. Quantitative indicators of activity quality can be: print and electronic publications, participation in projects, conferences, grants, etc. Qualitative indicators are professional internship, training with the purpose of qualification improvement, scientific school, peers reviews, certification and so on.

To display teacher’s activity pedagogical stuff forms e-portfolio space, administration can inspect the data in the portfolio, make appropriate conclusions regarding the activities of a particular employee and the university as a whole, as a result - to improve quality, enhancing the activity of professor. E-portfolio is formed using components of the electronic environment of the University, which systematically reflects the activities of the university as a whole and is the support for all university activities. The structure of the electronic environment is determined by the needs of higher education. The main component provides training through the use of e-content. For information and represent the university created the site. Publication activities for teachers and students are electronic publications, electronic resources of the library is a repository. Platform designed to implement educational technologies aimed at active students and teachers, all members of the educational process is a wiki portal, which also allows you to place the open e-portfolio. Important indicators of scientific and pedagogical employee are also profiles of open and scientometric databases - Google Scholar, Russian Science Citation Index etc. There are many interpretations of the term “portfolio” and different variations of its structures. Portfolio is a “calling card” of scientific and pedagogical employee containing information on various aspects of his activities, information on scientific work, professional development, training activities and personal information. E-portfolio - a portfolio of scientific and pedagogical employee based on electronic resources. The goal of the portfolio depends on its structure, if the portfolio is a tool for measuring the performance of scientific and

pedagogical employee, it should reflect all the aspects of the scientific and pedagogical employee affecting the overall presentation of the university.

When determining the portfolio structure it was taken into account the European standards for quality assurance, methods of forming the ranking of higher education institution in Ukraine “TOP-200 Ukraine”[6] (the quality of scientific-pedagogical potential, the quality of education, international recognition), a set of measures to stimulate the publishing activity of the employee of Borys Grinchenko Kyiv University [7] and the indicators for rating of professors of Borys Grinchenko Kyiv University. The main components of e-portfolio include:

- Personal data;
- Research activity;
- Professional Development;
- Teaching.

Let us have a closer look at these components on the schemes below (Figure 1):

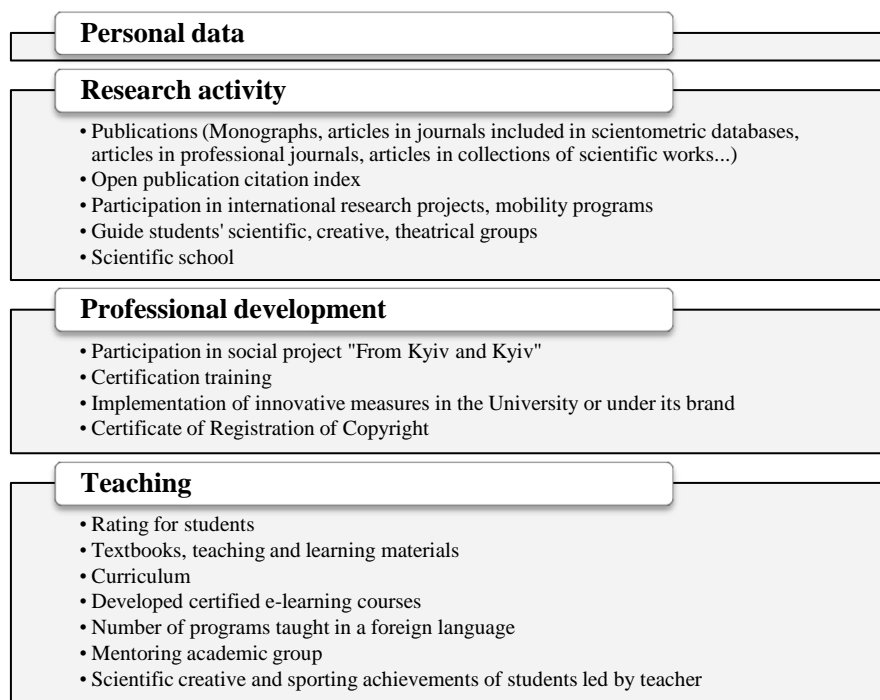


Fig. 1. Structure of e-portfolio of scientific-pedagogical staff member

All these components form a complete representation of scientific-pedagogical activity of a staff member. E-portfolio has the following advantages:

- Update information about yourself
- The check of whether the content is correct and right

- Analysis to create a rating
- Ability of scientific-educational staff member to present himself “well” in the Internet
- Provide an actual resume.
- Conducting a reflection of the own activities.

Let us build a model of e-portfolio of scientific educational staff member and its usage in activity of the higher education institution (Figure 2):

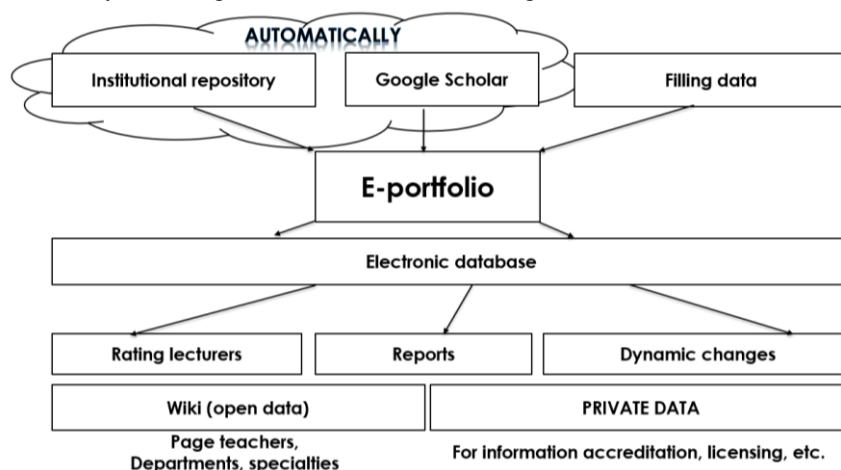


Fig. 2. Model of e-portfolio development and use

The data represented in the e-portfolio is transferred automatically from the institutional repository of University and Google Scholar profile.

Institutional repository (<http://elibrary.kubg.edu.ua>) – is a separate resource, created to review the scientific publications of professors in a free access. The main function of Institutional repository is accumulation, systematization and storage of electronic products of intelligent scientific audience, providing open access to the means of Internet technologies, dissemination of research materials in the world scientific and educational space.

After placing personal materials in the institutional repository, metadata is transferred into the appropriate fields of e-portfolio of scientific-pedagogical staff member. Institutional repository is registered in Google Scholar that makes it possible to have free access to all publications in repository and to see citation index of authors and related articles.

Google Scholar allows easy to perform advanced search of academic literature. The aim of Google Scholar (<http://scholar.google.com.ua>) is to streamline the articles with assessing the full article, author and issue, in which the article was published and citation index of articles in other academic literature.

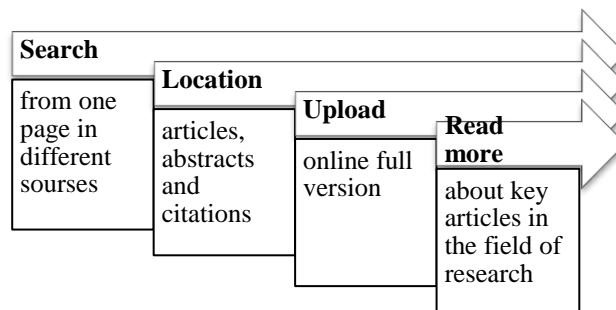


Fig. 3. Using of Google Scholar

The advantages of using Google Scholar for universities are:

- Creation of a profile;
- Monitoring the citation of publications;
- Names of professors appear in search results;
- Citation index;
- Statistics;
- Library.

Citation index of Google Scholar will automatically be transferred to the appropriate fields in e-portfolio of scientific and pedagogical employee.

Professor can fill own data on site <http://e-portfolio.kubg.edu.ua> (Figure 4).

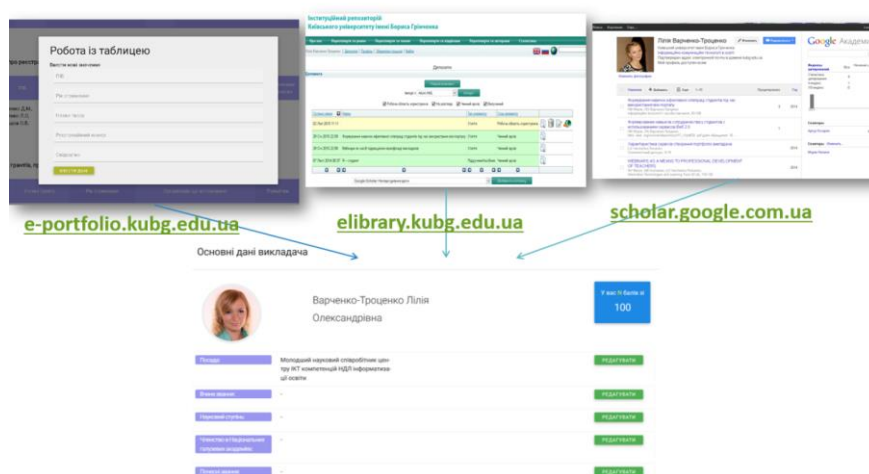


Fig. 4. Algorithm of filling e-portfolio

This forms a complete and relevant portfolio that performs the function of data integration of scientific and pedagogical staff member. Administration can analyze the

data in the right format and draw conclusions. Based on e-portfolios, professors can form the general framework of the university (Figure 5):

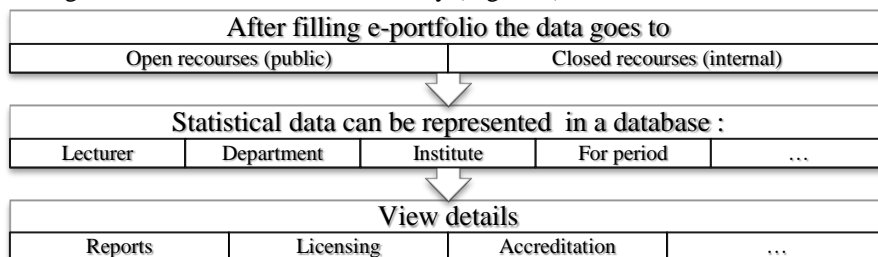


Fig. 5. Use of e-portfolio

E-portfolio can be transformed into the full report of a researcher activity (Figure 6):

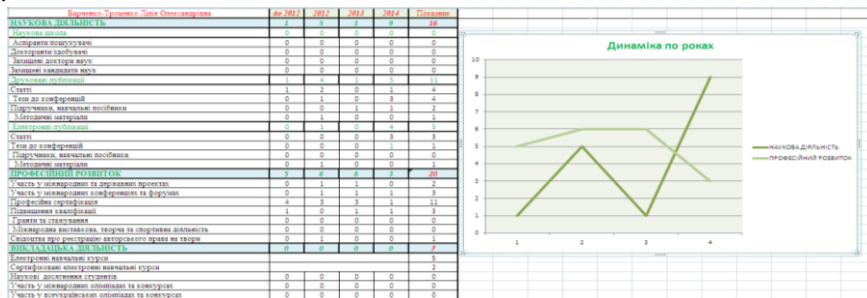


Fig. 6. Example of professor's e-portfolio

In a convenient way the reports of departments and institutes also can be formed. The part of e-portfolio data becomes open and is transferred to the personal pages on the Wiki-portal of Borys Grinchenko Kyiv University, and some part remains closed – to collect materials for accreditation, licensing, etc. (Figure 7).

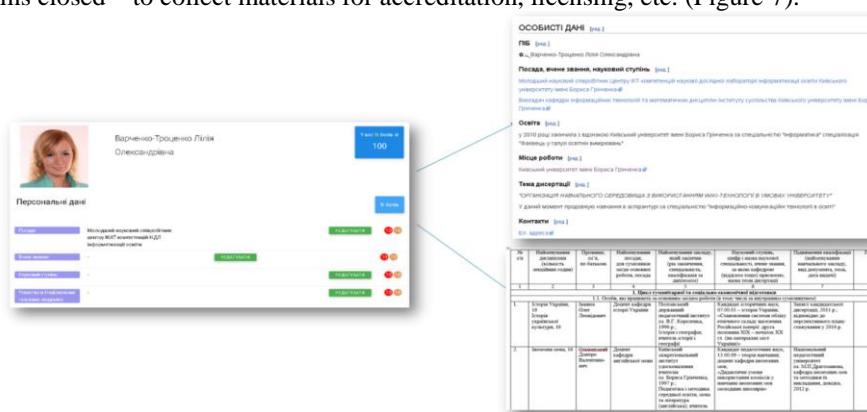


Fig. 7. Example of forming open and closed data

Webometrics is also very important for universities. It provides representation of academic and research institutions in the Internet and stimulation of open access to scientific information is displayed in the global Webometrics ranking of universities based on an analysis of the official website of the university.

Indicators of this rating include:

- Presence – number of web pages
- Impact – links to Universities web pages
- Openness – institutional Wiki repository and portal
- Excellence – scientific articles.

Analyzing key Webometrics rating indicators we can make conclusions that displaying of lecturers data and their work in an open space will have a positive impact on performance of a university in Webometrics.

For e-portfolio to become an instrument for assessment, professors have to fill their e-portfolio conscientiously. Survey was conducted for professors of Borys Grinchenko Kyiv University, the question was: “Do you have your own e-portfolio”, 21% replied that they have – those researchers should only bring its portfolio to the corresponding structure; 35% are ready to build, but they only need to be taught, 45% should be motivated. Most professors see the benefits of using e-portfolio to display their professional activity. In addition, professors determined the advantages of e-portfolio usage (figure 8):

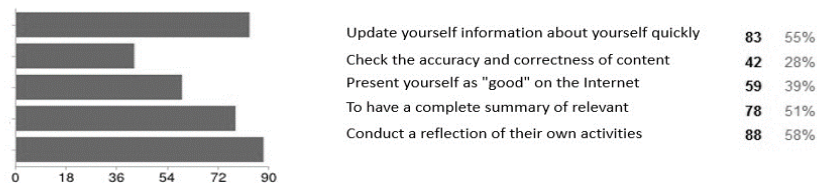


Fig. 8. Results of professors' answer the question about advantages of e-portfolio usage

This means that professors are positive about forming their personal e-portfolios. E-portfolio in the University is shown in Figure 9.

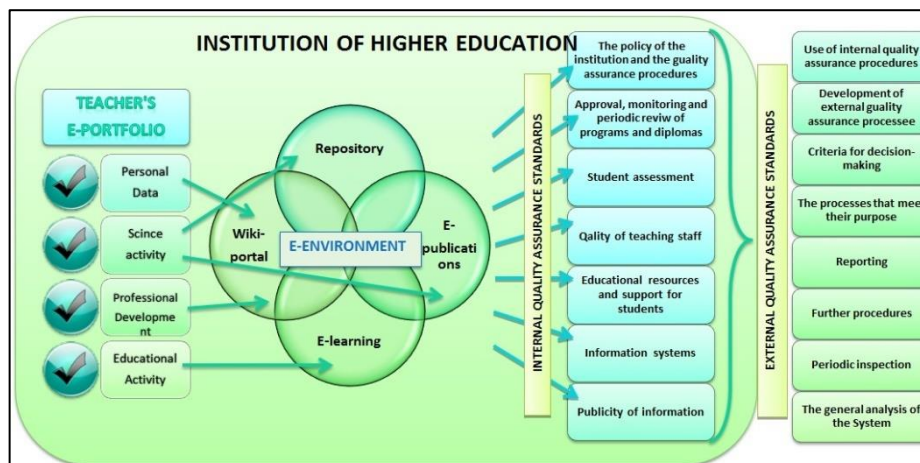


Fig. 9. E-portfolio in the University

Therefore the teacher's e-portfolio is one of the indicators of education quality in universities. The motivation to filling a portfolio by every teacher can be university ranking which takes into account in certain proportion all teachers' activities and has monetary and reputational consequences. Indicators of a portfolio should include those of a priority for university development at a certain time and be taken into account on different ratings, including international and Ukrainian.

3 Conclusions

The use of Wiki-technology provides openness and transparency of the performance of the employee, each structural unit and the university.

E-portfolio on Wiki-portal of the university should be considered as the use of the university various components of the electronic educational environment, in particular LMS Moodle and creation of statistics and use of e-learning courses by faculty, system of students survey about the quality of training by each teacher every discipline, filling institutional repository on research and methodical publications, relevant registers on training teachers of the winners of competitions, participation in international scientific and research projects, etc.

Activity of every university scientific-pedagogical staff member affects the formation and functioning of the electronic environment of the educational institution that is created to provide quality training and development of scientific and educational potential. As it presents different aspects of teaching activity, e-portfolio can be used with the purpose of stimulation and measurement the quality of scientific-pedagogical staff of the institution.

References

1. Kalinicheva, G.: Implementation of the European system of quality assurance in Ukraine: Experience and Prospects: Scientific Bulletin of Chernivtsi University. Collection of scientific works, vol. 2, pp. 66-70 (2009)
2. Kuzmenchuk, I.: Portfolio as individual trajectory of professor's in-service training // Teaching almanac, vol. 6 (2010)
3. Kyrychenko, V.: Professor's portfolio as a result of generalization and systematization of its educational achievements: Electronic collection of scientific works of Zaporizhzhia regional institute of postgraduate education, vol. 3 (2013)
4. Law of Ukraine "On higher education", available at: <http://zakon4.rada.gov.ua/laws/show/1556-18> (2014)
5. Morze, N., Varchenko, L., Smyrnova-Trybulska E.: Ways of formation of effective students' collaboration skill based upon the usage of WBT: International Journal of Web Based Communities, vol.11(1), pp. 25-41 (2015)
6. Ranking of higher education institution in Ukraine "TOP-200 Ukraine" (2015)
7. The set of measures to stimulate the publication activity of Borys Grinchenko Kyiv University stuff, available at: <http://kubg.edu.ua/informatsiya/naukovtsyam/dokumenti/kompleks-zakhodiv-dlia-stymuliuvannia-publikatsiinoi-aktyvnosti-spivrobitnykiv-kyivskoho-universytetu-imeni-borysa-hrinchenka.html> (2014)
8. The Standards and Guidelines for Quality Assurance in the European Higher Education Area – ESG, available at: <http://www.enqa.eu/index.php/home/esg/>
9. Viktorov, V.: Basic criteria and indicators of education quality: Higher education of Ukraine, vol. 1, pp. 54–58 (2006)
10. Zagorodnii, A.: European system of quality assurance: High School, vol.4, pp. 15–22 (2006)

The Hybrid Cloud-based Service Model of Learning Resources Access and its Evaluation

Mariya Shyshkina

Institute of Information Technologies and Learning Tools of the National Academy of
Pedagogical Sciences of Ukraine, M.Berlinskoho Str., 9, Kyiv, Ukraine
marple@ukr.net

Abstract. Nowadays, innovative technological solutions for learning environment organization using cloud computing and ICT outsourcing have shown promise and usefulness. Problems of providing access to electronic educational resources and their configuration within a cloud-based university learning environment have given rise to diverse research trends. This article outlines the conceptual framework of the study by reviewing existing approaches for electronic educational resources access organization within cloud-based settings. The hybrid service model of learning resources access is described and proved. The problems of quality evaluation of learning resources in the cloud-based learning environment are outlined. Indicators of cloud-based learning resources quality evaluation are proposed. An empirical estimation of the proposed approach and current developments of its implementation are provided.

Keywords: learning environment, cloud computing, electronic resources, quality, hybrid model, university.

Key Terms: ICTInfrastructure, Model, TeachingProcess.

1 Introduction

The development of a modern university learning environment is supported by emerging ICT on the basis of advanced network infrastructures, especially within cloud-based settings. Cloud computing (CC) technology is used to enhance multiple access and joint use of educational resources at different levels and domains, combining the corporate resources of the university and other learning resources within a united framework. Progress in the area has provided new insights into the problems of educational electronic resources access and configuration within the learning environment, bringing new models and approaches. The promising trend of research is concerned to the hybrid service models [4, 19, 24]. It has given rise to research of better ways of introducing innovative technology.

A set of different service models may be elaborated and combined to provide access for the cloud-based learning components within the hybrid environment's architecture. Quality evaluation of cloud-based components is a promising way to choose and approve the most appropriate delivery settings.

The purpose of the article is analyse the ways of educational electronic resources delivery within the hybrid cloud-based settings, and to substantiate and validate the quality evaluation indicators and approach to the cloud-based learning components design.

The *research method* involved analysing the current research (including the domestic and foreign experience of the application of cloud-based learning services to reveal the concept of the investigation and research indicators), examining existing models and approaches, estimating the current state of quality research development, considering existing technological solutions and psychological and pedagogical assumptions about better ways of introducing innovative technology, and conducting pedagogical experiments, surveys and expert evaluations.

2 Problem Statement

The challenges of making the ICT infrastructure of the university environment fit the needs of its users, taking maximum advantage of modern network technologies, and ensuring the best pedagogical outcomes, have led to the search for the most reasonable ways of e-resources access within the environment framework. The cloud-based learning resources have many progressive features including better adaptability and mobility, as well as full-scale interactivity, free network access, a unified structure among others [4, 21, 23]. So, the modelling and analysis of their design and deployment in view of the current tendencies of modern ICT advance have come to the fore.

Among the priority issues there are those concerning existing approaches and models for electronic educational resources delivery within the hybrid cloud-based setting; the cloud-based learning components quality assessment techniques; quality research indicators substantiation and validation; evaluation of current experience of cloud-based models and components use.

3 State of the Art

According to the recent research [4, 9, 15, 20, 21], the problems of implementing cloud technologies in educational institutions so as to provide software access, support collaborative learning, implement scientific and educational activities, support research and project development, exchange experience are especially challenging. The formation of the cloud-based learning environment is recognized as a priority by the international educational community [18], and is now being intensively developed in different areas of education, including mathematics and engineering [2, 8, 28, 30].

The transformation of the modern educational environment of the university by the use of the cloud-based services and cloud computing delivery platforms is an important trend in research. The topics of software virtualization and the forming of a unified ICT infrastructure on the basis of CC have become increasingly popular lines of investigation [8, 20]. The problems with the use of private and public cloud

services, their advantages and disadvantages, perspectives on their application, and targets and implementation strategies are within the spectrum of this research [7, 8, 28].

There is a gradual shift towards the outsourcing of ICT services that is likely to provide more flexible, powerful and high-quality educational services and resources [4]. There is a tendency towards the increasing use of the software-as-a-service (SaaS) tool. Along with SaaS the network design and operation, security operations, desktop computing support, datacentre provision and other services are increasingly being outsourced as well. Indeed, the use of the outsourcing mechanism for a non-core activity of any organization, as the recent surveys have observed happening in business, is now being extended into the education sector [9]. So, the study of the best practices in the use of cloud services in an educational environment, the analysis and evaluation of possible ways of development, and service quality estimation in this context have to be considered.

The valuable experience of the Massachusetts institute of technology (MIT) should be noted in concern to the cloud based learning environment formation in particular as for access to mathematical software. The Math software is available in the corporate cloud of the University for the most popular packages such as *Mathematica*, *Mathlab*, *Maple*, *R*, *Maxima* [30]. This software is delivered in the distributed mode on-line through the corporate access point. This is to save on license pay and also on computing facilities. The mathematics applications require powerful processing so it is advisable to use it in the cloud. On the other case the market need in such tools inspires its supply by the SaaS model. This is evidenced by the emergence of the cloud versions for such products as Sage MathCloud, Maple Net, MATLAB web-server, WebMathematica, Calculation Laboratory and others [2, 8]. Really there is a shift toward the cloud-based models as from the side of educational and scientific community, and also from the side of product suppliers. The learning software actually becomes a service in any case, let it be a public or a corporate cloud.

There are many disciplines where it is necessary to outsource the processing capacity: for example, the computer design for handling vast amounts of data for graphics or video applications. This is also a useful tool used to support the collaborative work of developers, as the modern graphical applications appear to be super-powerful and require joint efforts [7]. There is a research trend connected to the virtual computing laboratories (VCL) [16, 29] delivered in the cloud-based paradigm. This trend is inherent in the field of informatics, and learning resources for processing and sharing are needed.

Nowadays there are various universal cloud consumer applications, in particular MicrosoftOffice 365, Google Apps and others which gain an appropriate use in educational process [9, 26]. There is also a wide range of cloud services such as online photo and video editors, web pages processors, services for translation, check spelling, anti plagiarism and many others which are now available [26].

There is a principal transformation of approaches in concern to services supply within the cloud based infrastructure. It is considered to be a new stage of the service oriented models development [10, 27]. There is a branch of research devoted to the service oriented infrastructure in this actual perspective. The issues of service oriented

architecture development are described in [10]. The problem of turning software into a service is also posed [27]. For example, more powerful approaches for services integration appear while services compositions are used as building blocks in a process of elaboration of programming code [16]. The CC development brought the term the *service orchestration* into scientific discussion while number of web services can be combined to perform the higher level business process to manage and coordinate execution of the component processes [11]. In this regard the notion of the global software development (GSD) is considered as novel trends overcome geographical limits [11]. There is a significant revise of approaches to ICT services elaboration and this is concerned to its integration and composition.

Another set of problems is concerned with the hybrid service models and infrastructure solutions combining different public and corporate services on the united platform. Due to this approach, access to educational software set on a cloud server or in a public cloud is organized. This trend is now especially promising for the sphere of education [8, 19]. The challenge regarding novel technological solutions and their impact guide the search for the most reasonable method of implementation.

An essential feature of the cloud computing conception is dynamical supply of computing resources, software and hardware its flexible configuration according to user needs. So comparison of different approaches and cloud models of software access is the current subject matter of educational research [7, 8, 26, 28]. Despite of the fact that the sphere of CC is rather emerging there is a need of some comparison of the achieved experience to consider future prospects [28]. Also the problems of software choice in the learning complexes to be implemented in a cloud arise. This leads to the problems of cloud-based learning resources quality evaluation techniques and research indicators substantiation.

In the cloud-based learning environment, new ways of EER quality control arise. There are specific forms of the organization of learning activity related to quality estimation. For example there are e-learning systems based on the modelling and tracking of individual trajectories of each student's progress, knowledge level and further development [31]. This presupposes the adjustment, coordination of training, consideration of pace of training, diagnosis of achieved level of mastery of the material, consideration of a broad range of various facilities for learning to ensure suitability for a larger contingent of users. The vast data collections about the students' rates of learning are situated and processed in the cloud [31]. There are also collaborative forms of learning where the students and teachers take part in the process of resource elaboration and assessment; this is possible in particular by means of the SageMathCloud platform [2].

Thus, in view of the current tendencies, the research questions are: how can we take maximum advantage of modern network technologies and compose the tools and services of the learning environment to achieve better results? What are the best ways to access electronic resources if the environment is designed mainly and essentially on the basis of CC? What are the most reasonable approaches to validate quality evaluation criteria? This brings the problem of the cloud-based learning components modelling, evaluation and design to the forefront.

4 Pedagogical Aspects of Electronic Resources Delivery and Quality Research Indicators

Cloud computing technology is now one of the leading trends in the formation of the information society. It constitutes an innovative learning concept and its implementation significantly affects the content and form of different types of activities in the sphere of education [4, 12, 20].

Along with the emergence of cloud computing, the number of objects, developments and domain applications are continually growing, which indicates the rapid spread of the innovation [23]. The concept of *the cloud-based learning environment* is now in line with the wider trend; that is to say, the ICT environment of the university, where some didactic functions as well as some fundamentally important functions of scientific research are supported by the appropriately coordinated and integrated use of cloud services [23]. The *aim* of the cloud-based learning environment formation is to meet the users' educational needs. To do this, the introduction of cloud technology in the learning process should to be holistic and carried out according to the principles of *open education*, including meeting the following needs: the mobility of students and teachers, equal access to educational systems, providing qualitative education, and forming and structuring of educational services [3, 23].

The main elements of the cloud computing conception, including the types, application service models, essential features, ICT architecture and others, are reflected in the structure of the modern educational organizational systems [5]. Therefore, a number of concepts and principles that characterize the development and application of CC-based services are significant in the consideration of the educational environment design.

The concept of *electronic educational (learning) resources* (EER) appears to be the centre of attention. In particular, at the Institute of Information Technologies and Learning Tools of the National Academy of Pedagogical Sciences of Ukraine the conception that provided the definition of electronic educational resources (EER) its classification, and the ways it can be applied has been developed and proposed [5].

According to the definition given in [5, p.3], "The electronic educational resources are a kind of educational tool (for training, etc.) that are electronically placed and served in educational system data storage devices which are a set of electronic information objects (documents, documented information and instructions, information materials, procedural models, etc.)".

The elaboration of the electronic learning resources should be considered as a specific activity, which is linked to the mandatory need to take into consideration the psychological and pedagogical aspects of building an educational system methodology, the design of an open computer-based learning environment, and the involvement of the scientific and pedagogical staff, including the best teachers and educators [4].

Cloud Service – is a service that provide “network access to a scalable and elastic pool of shareable physical or virtual resources with self-service provisioning and administration on-demand” [12, p.6]. These services are used to supply the electronic

educational resources that make up the substance of a cloud-based environment, and to provide the processes of elaboration and use of the educational services.

Electronic resources appear to be both the objects and the tools of activity for a learner; therefore, these resources are used to maintain certain functions that are realized in the learning process. By the *educational service* we mean a service provided at the request (in response to an inquiry etc.) of a user that meets some service function carried out by the organization or institution (service provider, outsourcer) [4].

There are also four *service deployment models* for cloud computing application that reflect the mode of the cloud infrastructure set up in a particular organization: *the corporate cloud* is owned or leased by the organization; *the cloud community* is a shared infrastructure used by a community; *the public cloud* is a mega-scale infrastructure that may be used by any person under some payment terms; *the hybrid cloud* is a composition of one or more models [4, 18].

The *hybrid service model* is to combine various approaches for learning resources access within the cloud-based settings.

The EER of the public cloud can occupy the role of software for general purposes such as office applications, systems support processes for communication and data exchange and others, and also the special software designed for educational use [15, 26]. The number of EERs is increasing and this trend is likely to intensify. By means of CC-based tools, a significant lifting of restrictions on the implementation of access to qualitative leaning resources may be achieved. Now, these questions are not a matter of future perspective, they need practical implementation. For this purpose, the problem of the design and delivery of electronic educational resources in the cloud-based environment in particular within the hybrid service modes is a complex one and not only should technological needs be considered, but also the pedagogical aspects.

Due to the significant educational potential and novel approaches to environmental design, its formation and development, these questions remain the matter of theoretical and experimental studies, the refinement of approaches, and the search for models, methods and techniques, as well as possible ways of implementation [4].

To carry out research and experimental activities and the implementation and dissemination of the results, the Joint research laboratory of the Institute of Information Technologies and Learning Tools of the NAPS of Ukraine and the Kherson State University was created in 2011 with the focus on issues of educational quality management using ICT [32].

As part of the programme of joint research work, the Kherson State University was approved as an experimental base for research on the definition and experimental verification of the didactic requirements and methods of evaluating the quality of electronic learning resources in the educational processes of the pilot schools [32]. The purpose of the experiment carried out was to identify and experimentally verify the requirements and methods of evaluating the quality of the electronic learning resources used in the educational process in secondary schools [32].

The quality evaluation of EER in the cloud-based learning environment is a separate line of work in the Laboratory's research. In this case, there are different approaches and indicators. The access organization has been changed so the models

of learning activity have been changed also. There are the following questions: What features and properties have to be checked so as to measure the pedagogical effect of the cloud-based approach? With regard to the pedagogical innovation, what are the factors influencing pedagogical systems, their structure and organization? Is the improvement in learning results achieved due to the cloud-based models? In this context, the quality of EER is a criterion for estimating the level of organization and functioning of the cloud-based learning environment.

With regard to this, the following *hypothesis* is to be posed: the design of the learning environment on the basis of cloud models of access to learning resources contributes to the improvement of the quality of these resources and the improvement of the processes in this environment and their organization and functioning, resulting in an improvement in learning results.

The method of electronic resources quality estimation was developed and used in the Joint laboratory of EER quality control [14, 32]. In this case, the different quality parameters have been detailed and selected [14]. It is important that the psychological and pedagogical parameters are estimated in the experimental learning process, while the other types of parameter such as technological or ergonomic may be estimated out of this process.

The prospective way of the estimation of the quality of learning resources is by means of the cloud-based environment. As the resources are collectively accessed, there is a way to allow experts into the learning process so they may observe and research their functioning. This is a way to make the process of quality estimation easier, more flexible and quicker. The process of estimation becomes anticipatory and timely. The estimation may be obtained just once along with the process of EER elaboration, and it is very important to facilitate the process [25].

There are several groups of quality criteria to be taken into consideration and checked in the process of complex quality assessment of ICT-based learning tools. Generally, there are main groups such as: psychological and pedagogical indicators; and ergonomic and technological indicators [14, 32]. There is no single set of criteria clearly acknowledged to be unambiguous. There is a problem of criteria substantiation as the didactic and methodical, and psychological aspects of educational use of ICT, are hardly regulated and standardized [22]. Still, there are research works devoted to the problems of quality evaluation in this field, where the system of quality criteria has been substantiated and proved experimentally [14].

Another kind of problem is connected to the cloud-based learning tools' quality estimation. Specific kinds of criteria are valuable in this case with regard to innovative features of advanced learning settings. So, for the purpose of this study, the quality criteria were scrutinized to reveal the most significant. There were also two groups of indicators selected: the pedagogical and psychological; and technological. It is not feasible to take into account all possible criteria inherent to cloud-based tools' application. There are a lot of technical and technological aspects to be considered, such as portability, sustainability, security, and others. Not to underestimate the importance of all relevant features, the study is concerned with those quality aspects that are valuable only in the case of educational use of ICT-based tools.

Therefore, among the variety of technological parameters, this study focuses on those that are important for pedagogical study in relation to the introduction of emerging ICT. Technological innovations cause shifts in pedagogical approaches and transformations of target, content, and methodological aspects of educational systems [4]. So, the quality of emerging ICT tools must reflect the prospected shift in learning technologies and improvement of pedagogical outcome.

Among the technological indicators of cloud-based educational resource quality evaluation, are those concerned with *ease of access*, showing if electronic resource access organization is easy and convenient; *the intuitive clarity of the interface*, reflecting if the user interface is clear and easy to learn; *responsiveness*, meaning performance in real time work; *sustainability*, concerning capability of functioning while working with the resource from any computer via a browser; *support of collaborative work*, encompassing facilities to support collaboration in the learning process; *ease of integration*, meaning suitability to be incorporated into a single environment along with other resources; and *usefulness*, covering overall utility (feasibility) of resource use.

The pedagogical and psychological criteria of the cloud-based learning resource quality evaluation should be the same as any other educational electronic resource in many respects. However, this is an important step in the quality evaluation process that cannot be neglected while investigating any tool aimed at learning. The set of psychological and pedagogical indicators for the research are as follows: the *scientific clarity* of the content; *accessibility* of the content, delivered by the resource; *fostering the intellectual development* of a learner while working with the resource; *problem orientation* of content and functioning; *personalization* in the learning process; *adaptability* as suitability for most of the possible user contingent; *methodical usefulness*, meaning the most appropriate support for learning methods; *professional orientation* as providing learners' professional development; and *feedback* connection.

As the criteria set is identified and formed, there is a need to provide validity of every indicator that may be achieved by this research.

5 The Hybrid Service Model of Learning Software Access

To research the hybrid service model of learning software access, a joint investigation was undertaken in 2013–2014 at the Institute of Information Technologies and Learning Tools of the NAPS of Ukraine and Drohobych State Pedagogical University named after I. Franko. At the pedagogical university, the experimental base was established where the cloud version of the Maxima system (which is mathematical software), installed on a virtual server running Ubuntu 10.04 (Lucid Lynx), was implemented. In the repository of this operational system is a version of Maxima based on the editor Emacs, which was installed on a student's virtual desktop [24]. In this case, the implementation of software access due to the hybrid cloud deployment in Scenario 3 was organised.

In Fig.3, the configuration of the virtual hybrid cloud used in the pedagogical experiment is shown. The model contains a virtual corporate (private) subnet and a public subnet. The public subnet can be accessed by a user through the remote desktop protocol (RDP). In this case, a user (student) refers to certain electronic resources and a computing capacity set on a virtual machine of the cloud server from any device, anywhere and at any time, using the Internet connection.

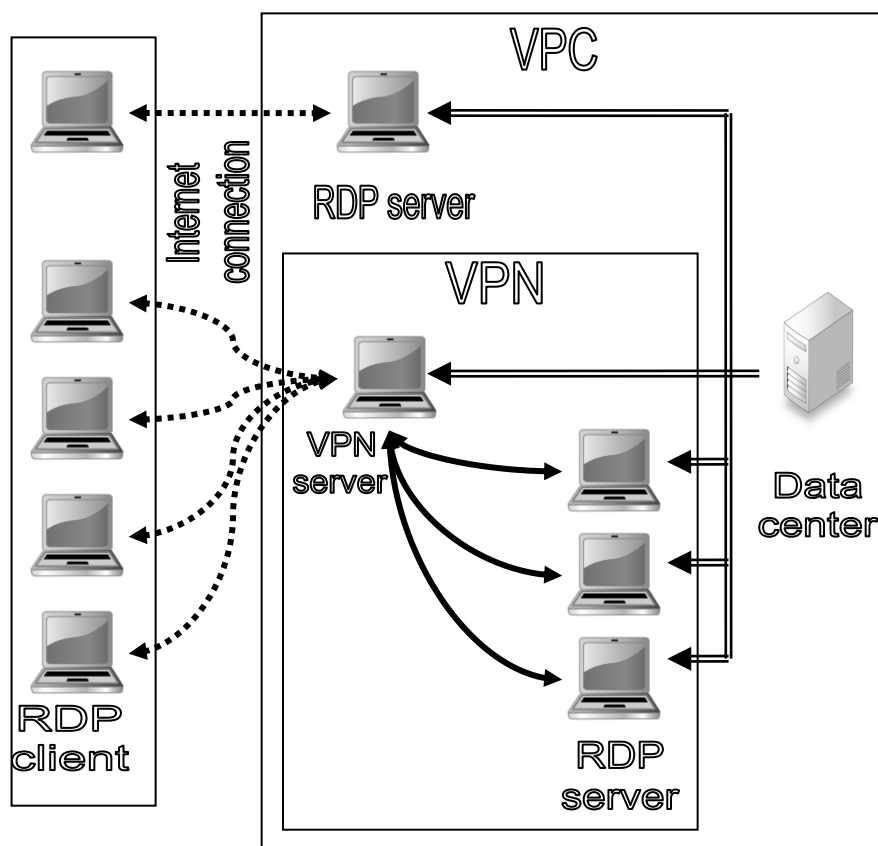


Fig. 1. The hybrid service model of the learning resources access.

In this case, a user's computer is the RDP-client, while the virtual machine in the cloud is the RDP-server. In the case of a corporate (private) subnet, a user cannot apply to the RDP-server via desktop because it is not connected to the Internet directly. Computers in the corporate subnet have Internet access via the VPN-connection, i.e. the gateway. Thus, these computers cannot be accessed from any device, but only from the specially configured one (for example, a computer in the educational institution or any other device where the VPN-connection is set up) (Fig.1).

The advantage of the proposed model is that, in a learning process, it is necessary to use both corporate and public learning resources for special purposes. In particular, the corporate cloud contains limited access software; this may be due to the copyright being owned by an author, or the use of licensed software products, personal data and other information of corporate use. In addition, there is a considerable saving of computational resources, as the software used in the distributed mode does not require direct Internet access for each student. At the same time, there is a possibility of placing some public resources on a virtual server so the learner can access them via the Internet and use the server with the powerful processing capabilities in any place and at any time. These resources are in the public cloud and can be supplied as needed.

6 Implementation and Evaluation

In the joint research experiment held at Drohobych State Pedagogical University named after I.Franko, 240 students participated. The aim was to test the specially designed learning environment for training the operations research skills on the basis of the Maxima system. During the study, the formation of students' professional competence by means of a special training method was examined. The experiment confirmed the rise of the student competence, which was shown using the χ^2 – Pearson criterion [24]. This result was achieved through a deepening of the research component of training. The experiment was designed using a local version of the Maxima system installed on a student's desktop.

The special aspect of the study was the expansion of these results using the cloud version of the Maxima system that was posted on a virtual desktop. In the first case study (with the local version), this tool was applied only in special training situations. In the second case study (the cloud version), the students' research activity with the system extended beyond the classroom time. This, in turn, was used to improve the learning outcomes.

After that another research was held in Ternopil National Pedagogical University named after V.Gnatyuk (2014-2015). The aim of this experiment was to test the use of the cloud-based component in the learning process. 48 students participated in this experiment. There was the experimental group of 24 students who used the cloud-based component with Maxima system. It showed increase of the students' percentage with the high level of ICT competence from 16% to 75%. It was significantly different from the level of ICT competence of those students who did not used this component (from 14% to 20%), which was justified by the Fisher criterion.

The cloud-based learning component used in the experiment has undergone a quality estimation. The method of learning resources quality estimation developed in the joint laboratory of educational quality management with the use of ICT [14] was used and adapted for this study. The 20 experts were specially selected as having experience in teaching professional disciplines focused on the use of ICT and being involved in the evaluation process. The experts evaluated the electronic resource by two groups of parameters. The first group has contained 7 technological parameters:

ease of access; the clarity of the interface; sustainability; support of collaborative work, ease of integration; mobility; and usefulness. The second group has contained 9 psychological and pedagogical parameters: the scientific clarity; accessibility; fostering the intellectual development; problem orientation; personalization; adaptability; methodical usefulness; professional orientation; and feedback connection.

“Expert evaluation of the EER quality can be considered sufficiently reliable only when a good consistency of expert answers. Therefore, the statistical processing of the results of experts evaluations should include an analysis of consensus of experts. Concordance method is used to assess the degree of consensus of experts on the factors: weights of EER types, parameterization of EER quality indicators, and average factor of EER quality” [14, p.322].

Experts were asked to complete the table 1 for peer review of EER quality parameters validity. The values of the weighting factors were selected from 10 point scale. The results of the survey of experts are presented in Table 1.

Table 1. Expert data on validity of EER quality parameters.

Ex pert	ERR Quality Parameters															
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16
1	9	10	10	9	8	6	8	7	9	10	10	9	8	6	8	7
2	10	10	10	10	10	10	10	10	10	10	10	10	9	9	10	10
3	7	6	6	8	8	6	8	9	7	6	6	8	8	6	8	9
4	9	10	10	9	8	6	8	7	9	10	10	9	8	6	8	7
5	8	9	10	10	10	10	10	10	8	9	10	10	10	10	10	10
6	10	9	10	10	9	8	8	8	10	9	10	10	9	8	8	8
7	10	9	8	8	8	8	8	10	10	9	8	8	8	8	8	10
8	9	9	8	9	8	9	9	9	9	9	8	9	8	9	9	9
9	10	9	7	8	8	6	8	9	10	9	7	8	8	6	8	9
10	7	6	6	8	8	6	8	9	7	6	6	8	8	6	8	9
11	9	10	10	9	8	6	8	7	9	10	10	9	8	6	8	7
12	10	10	9	9	10	10	10	10	10	10	9	9	10	10	10	10
13	9	10	10	9	8	6	8	7	9	10	10	9	8	6	8	7
14	9	10	10	9	8	6	8	7	9	10	10	9	8	6	8	7
15	8	9	10	10	10	10	10	10	8	9	10	10	10	10	10	10
16	10	9	10	10	9	8	8	8	10	9	10	10	9	8	8	8
17	10	9	8	8	8	8	8	10	10	9	8	8	8	8	8	10
18	9	9	8	9	8	9	9	9	9	9	8	9	8	9	9	9
19	10	9	7	8	8	6	8	9	10	9	7	8	8	6	8	9
20	7	6	6	8	8	6	8	9	7	6	6	8	8	6	8	9

Table 2. Expert data on EER quality parameters (ranked)

Expert	ERR Quality Parameters (ranked)															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	9	13	14	10	5	1	6	3	11	15	16	12	7	2	8	4
3	7	1	2	9	10	3	11	15	8	4	5	12	13	6	14	16
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
5	1	3	5	6	7	8	9	10	2	4	11	12	13	14	16	16
6	11	7	12	13	8	1	2	3	14	9	15	16	10	4	5	6
7	13	11	1	2	3	4	5	14	15	12	6	7	8	9	10	16
8	5	6	1	7	2	8	9	10	11	12	3	13	4	14	15	16
9	15	11	4	6	5	2	10	14	16	12	3	7	9	1	8	13
10	11	10	1	4	12	9	7	6	5	8	3	2	16	14	13	15
11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
12	12	14	13	10	6	2	7	4	11	16	15	9	5	1	8	3
13	8	1	3	9	10	2	12	15	7	5	4	11	13	6	14	16
14	8	1	7	2	11	9	3	10	12	4	15	13	14	5	16	6
15	1	4	11	8	7	10	6	13	2	3	5	9	12	14	15	16
16	14	8	12	11	7	2	3	1	13	10	15	16	9	4	5	6
17	14	11	2	1	4	5	3	13	15	12	6	7	9	8	10	16
18	13	11	3	10	1	12	9	7	6	8	2	5	4	15	14	16
19	15	11	3	7	5	1	10	12	16	14	4	6	9	2	8	13
δ_i	0,39	0,43	0,33	0,41	0,38	0,33	0,44	0,60	0,62	0,54	0,52	0,63	0,67	0,51	0,73	0,83

Concordance coefficient W is calculated according to the formula proposed by Kendall [13]

$$W = \frac{12S}{m^2(n^3 - n)} \quad (1)$$

$$\text{Here } S = \sum_{i=1}^n \Delta_i^2 = \sum_{i=1}^n \left\{ \sum_{j=1}^m x_{ij} \right\}^2, \quad (2)$$

m – number of experts, n – number of objects of examination (the quality parameters, x_{ij} – assessment of the i -object by j -expert. Coefficient of concordance may vary between 0 and 1. If $W = 1$, all experts gave the same assessment for all parameters, if $W = 0$, the evaluations of experts are not consistent.

Using the formula (1) we calculated that coefficient $W = 0.189$ and it is significantly different from zero, so we can assume that among experts there is objective concordance. Given that the value of $m(n-1)W$ is distributed according to χ^2 with $(n-1)$ freedom degree, then

$$\chi_W^2 = \frac{12S}{m \cdot n \cdot (n+1)} = 52.8. \quad (3)$$

Comparing this value with the tabulated value χ_T^2 for $n-1 = 15$ degree of freedom and significance level $\alpha = 0,01$, we find $\chi_W^2 = 52,8 > \chi_T^2 = 30,5$. Therefore, the hypothesis of concordance of expert evaluations is confirmed according to Pearson.

Thus, the results of pedagogical experiment confirmed the assumption that the method of expert evaluations can be the basis for the cloud-based EER quality evaluating.

The problem was: is it reasonable and feasible to arrange the environment in a proposed way? For this purpose there were two questionnaires proposed to expert concerning two groups of parameters. The 20 experts estimated 16 parameters (there were 7 technological and 9 psychological and pedagogical among them). A four-point scale (0 (no), 1 (low), 2 (good), 3 (excellent)) was used for the questions.

The results of the evaluation of the technological parameters are shown in Fig.2.

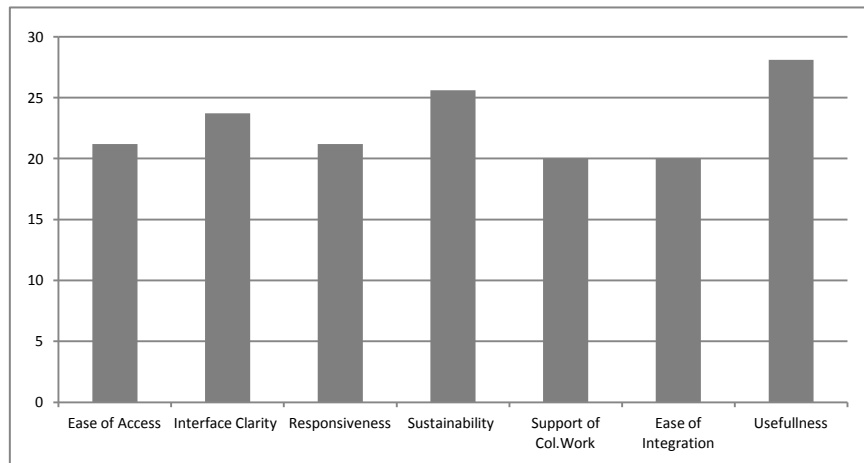


Fig. 2. The results of the cloud-based learning resource technological quality parameters evaluation.

The resulting average value was calculated for every parameter: “Ease of access” = 2.1, “Interface clarity” = 2.4, “Responsiveness” = 2.1, “Sustainability” = 2.56, “Support of Collaborative work” = 2.0, “Ease of Integration” = 2.0, “Usefulness” = 2.8, the total value was 2.3.

The results of psychological and pedagogical parameters evaluation are shown at Fig.3.

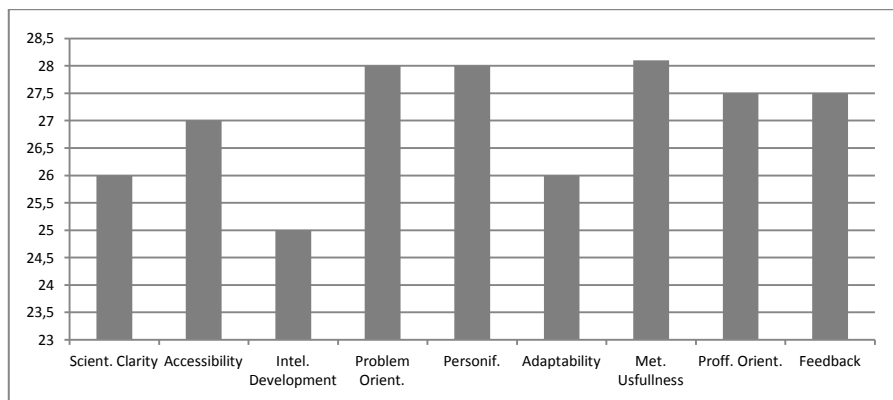


Fig. 3. The results of the psychological and pedagogical quality parameters evaluation

The resulting average values for every parameter are: “Scientific clarity” = 2.6, “Accessibility” = 2.7, “Fostering the intellectual development” = 2.5, “Problem orientation” = 2.8, “Personalization” = 2.8, “Adaptability” = 2.6, “Methodical usefulness” = 2.81, “Professional orientation” = 2.75, “Feedback connection” = 2.75. The total value was 2.71.

The value of the weight factor of *i*-type resource indicator may be calculated using the Table 2 by the formula (4) [14].

$$\delta_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{k_{ij}}{k_{iM}} \quad (4)$$

n_i – quantity of experts, k_{ij} – ranked *j*-parameter of quality, k_{iM} – the maximum value of *j*-parameter of quality.

Then the average criterion of EER quality may be calculated as follows:

$$K = \frac{k}{m} \sum_{i=1}^m \delta_i \bar{x}_i \quad (5)$$

m – quantity of quality indicators, \bar{x}_i – average value of *i*-indicator of quality, k – weight coefficient.

The resulted average criterion of EER quality $K=2.59$. This characterises the resource quality as sufficient for further implementation and use.

The advantage of the approach is the possibility to compare the different ways to implement resources with regard to the learning infrastructure. Future research in this area should consider different types of resources and environments.

7 Conclusion

The introduction of innovative technological solutions into the university learning environment contributes to unified learning infrastructure formation and the growth of access to the best examples of electronic resources and services. ICT use is promising regarding learning settings that can advance and develop the tendencies of CC progress. For example, there are tendencies of using the cloud-based models of environment design, applications virtualisation, unifying infrastructure, integrating

services, increasing the use of electronic resources, expanding collaborative forms of work, widening the use of the hybrid models of ICT delivery and increasing the quality of electronic resources. The hybrid service model proved to be a reasonable framework to deliver and research the cloud-based learning resources and components of the university educational environment.

References

1. Amazon Virtual Private Cloud. User Guide, API Version 2013-07-15, (2013)
2. Bard, G.V.: Sage for Undergraduates. AMS, (2015)
3. Bykov, V.: Models of Organizational Systems of Open Education. Atika, Kyiv (2009) (in Ukrainian)
4. Bykov, V.: Cloud Computing Technologies, ICT Outsourcing, and New Functions of ICT Departments of Educational and Research Institutions. Information Technologies in Education, 10, 8–23, (2011) (in Ukrainian)
5. Bykov, V., Lapinskii V.: The Methodological basis for creating and implementation of the electronic learning tools. Computer in school and in family, 2(98), 3–6, 2012.
6. Buyyaa, R., Chee Shin Yea, Venugopala, S., Broberg, J., Brandicc, I.: Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. Future Generation Computer Systems, 25(6), 599–616, (2009)
7. Cusumano M.: Cloud computing and SaaS as new computing platforms. Communications of the ACM, 53(4), 27–29 (2010)
8. Doelitzscher, F., Sulistio, A., Reich, Ch., Kuijs, H., Wolf, D.: Private cloud for collaboration and e-Learning services: from IaaS to SaaS. Computing, 91, 23–42, (2011)
9. The Future of Cloud Computing: 4th Annual Survey 2014. The North Bridge Future Of Cloud Computing Survey In Partnership With Gigaom Research, <http://bit.ly/2014FutureCloud> (2014)
10. Gold, N., Mohan, A., Knight, C., Munro, M.: Understanding service-oriented software, Software, IEEE, 21(2), 71 – 77, (2004)
11. Hashmi S.I., Clerc V., Razavian M., and others: Using the Cloud to Facilitate Global Software Development Challenges. 2011 Sixth IEEE International Conference on Global Software Engineering Workshops, (2011)
12. ISO/IEC 17788:2014(E) Information technology – Cloud computing – Overview and vocabulary. First edition 2014–10–15, (2014)
13. Kendall M. Rank Correlation Methods, Charles Griffen & Company, London, (1948)
14. Kravtsov H.M. Methods and Technologies for the Quality Monitoring of Electronic Educational Resources / In: Batsakis, S. et al. (eds.) Proc. 11-th Int. Conf. ICTERI 2015, Lviv, Ukraine, May 14-16, 2015, 311– 325, CEUR-WS.org/ Vol-1356, ISSN 1613-0073, P.311–325, online CEUR-WS.org/Vol-1356/paper_109.pdf, (2015)
15. Lakshminarayanan, R., Kumar, B., Raju, M.: Cloud Computing Benefits for Educational Institutions. In Second International Conference of the Omani Society for Educational Technology, Muscat, Oman: Cornell University Library, <http://arxiv.org/ftp/arxiv/papers/1305/1305.2616.pdf> (2013)
16. Maamar, Z., et al.: An approach to engineer communities of web services: Concepts, architecture, operation, and deployment. International Journal of E-Business Research (IJEER), 5(4), 1–21, (2009)

17. Matheson C., Matheson D.: Access and Accessibility in E-Learning. Applied E-Learning and E-Teaching in Higher Education. Ed. by Donnelly R., McSweeney F., Hershey New York, 130–151, (2009)
18. Mell, P., Grance T.: Effectively and Securely Using the Cloud Computing Paradigm. NIST, Information Technology Laboratory, 10–7–2009, (2009)
19. Qing Li, Zeyuan W., Weihua Li, Jun Li, Cheng Wang, Ruiyang Du.: Applications integration in a hybrid cloud computing environment: modelling and platform. Enterprise Information Systems, 7(3), 237–271, (2013)
20. Smith, A., Bhogal J., Mak Sharma: Cloud computing: adoption considerations for business and education. 2014 International Conference on Future Internet of Things and Cloud (FiCloud), (2014)
21. Shyshkina, M.: Innovative Technologies for Development of Learning Research Space of Educational Institution. Information Technologies and Society, 1, http://ifets.ieee.org/russian/depositary/v16_i1/pdf/15.pdf (2013) (In Russian)
22. Shyshkina, M.: Advanced Technologies of E-learning in Engineering Education / In: M.Auer, M.Huba (eds.) Proc. of the 14th Int. Conf. ICL2011,), Piestany, Slovakia, September 21–23 2011, 565–568, (2011)
23. Shyshkina, M.: Emerging Technologies for Training of ICT-Skilled Educational Personnel. Communications in Computer and Information Science, Berlin-Heidelberg, Springer-Verlag, 412, 274–284, (2013)
24. Shyshkina M. U. P. Kohut, I. A. Bezverbnyy. Formation of professional competence of computer science bachelors in the cloud based environment of the pedagogical university. Problems of modern teacher preparation, Uman, FOT Zhovtyy O.O., 9, part 2, 136–146 (2014) (in Ukrainian)
25. Shyshkina, M.P., Spirin, O.M., Zaporozhchenko, Yu.G.: Problems of Informatization of Education in Ukraine in the Context of Development of Research of ICT-Based Tools Quality Estimation. In: J. Information Technologies and Learning Tools, vol. 27, № 1, <http://journal.iitta.gov.ua/index.php/itlt/article/view/632/483> (2012) (in Ukrainian)
26. Tuncay, E.: Effective use of cloud computing in educational institutions. Procedia – Social and Behavioral Sciences, 2(2), 938–942, (2010)
27. Turner, M., Budgen, D., Brereton, P.: Turning software into a service. Computer, 36 (10), 38–44, (2003)
28. Vaquero L. M.: EduCloud: PaaS versus IaaS cloud usage for a n advanced computer science course, IEEE Transactions on Education, 54(4), 590–598, (2011)
29. Vouk, M.A., Rindos, A., Averitt, S.F., Bass, J. and others: Using VCL technology to implement distributed reconfigurable data centers and computational services for educational institutions. VCL/Reconfigurable. Data Centers & Clouds/NCSU/V19-Draft Feb–2009 1–27, (2009)
30. Wick D.: Free and open-source software applications for mathematics and education. Proceedings of the twenty-first annual international conference on technology in collegiate mathematics, 300–304, (2009)
31. Zhang, J., and others: A Framework of User-Driven Data Analytics in the Cloud for Course Management. Proceedings of the 18th International Conference on Computers in Education, S. L. Wong et al., Eds., Putrajaya, Malaysia, Asia-Pacific Society for Computers in Education, 698–702, (2010)
32. Zaporozhchenko, Yu., Shyshkina M., Kravtsov G.: Prospects of the development of the modern educational institutions' learning and research environment: to the 15th anniversary of the Institute of information technologies and learning tools of NAPS of Ukraine. Informational Technologies in Education, 19, 62–70, (2014).

Monitoring of Efficiency of Feedback Systems Use on the Base of Kherson State University

Spivakovska Evgeniya, Vinnyk Maksym, Tarasich Yulia, Kuchma Olesya

Kherson State University, 27, 40 rokiv Zhovtnya St., 73000 Kherson, Ukraine

(Spivakovska, Vinnik, YuTarasich, OKuchma) @ksu.ks.ua

Abstract. The article deals with solution of some problems connected to development of feedback services while surveying students on educational environment at a higher education institution. Our research was carried out by the Department of Informatics, Software Engineering and Economic Cybernetics of Kherson State University. During 6 years (2009-2015) in the mentioned above Department, students' survey regarding their satisfaction with an educational process and lecturer' assessment by students' had been carried out. In the process of research, students from the 1st up to the 4th years of study of the Department of Informatics, Software Engineering and Economic Cybernetics were surveyed. All the respondents were divided into two groups: interested and disinterested ones during the survey execution. Introduction of the service "KSU Feedback" at Kherson State University on the base of the Department of Informatics, Software Engineering and Economic Cybernetics had a positive impact on creation of an educational environment where higher education institution is a corporation for serving the students.

Keywords: feedback, training, services, KSU Feedback, social polls, survey, quality of education.

Key Terms: InformationCommunicationTechnology, TeachingProcess, ICTInfrastructure

1 Introduction

Today, higher education institutions all over the world compete for their influence on educational markets not only in their own countries, but also in the countries located on other continents. The main strategic resource in this competition is a quality of educational services and use of IT for organization of educational and managerial processes. An education institution, that will be able to create the best conditions and resources for training with IT means, comes out to a qualitatively new level of modern world university.

The main kind of activity of education institutions is creating and providing educational services. Educational services, as we see, is a purposeful systematic process of transfer and receipt of the system of knowledge, information, skills and

abilities and result of intellectual, cultural, spiritual and socio-economic development of society and state. [1, 8, 15].

In order to improve increasingly a quality of educational services, it's preferred (and often necessary) to be able to react to changes of a real situation, that is managing object should receive information from the controlled object and, depending on its condition, one way or another, change managing influence. The Feedback is used to transfer information on the condition of the controlled object [2]. Thus, feedback is a mechanism for further operation of the training system, which in connections performs a function of the correction of information perception.

One of examples of the Feedback effectiveness, may be used a polling system in education institutions of Europe which is there as a separate piece of culture used to learn in schools by this time. Standards of feedback providing in Ukraine are rather undeveloped in comparison with countries of Europe. The main precondition, in this case, is a capability to accept criticism and objectivity of assessment as an ability to find and analyze information from various sources or different people.

If user's anonymity and service's simplicity in use is mainly realized by technical means, consequently, satisfaction and user's wish to take the survey aren't connected directly with system's technical characteristics, in our opinion, in a certain extent, is a key factor in this service's use. So, customer's satisfaction using the Feedback service is proportional to qualitative and quantitative indicators of its use.

The Feedback offers a complete picture of needs of each individual, which makes the most effective solution to a particularly set problem, improvement of directions of interaction between lecturers and students, when they are the most accurately defined.

At present, there is a great variety of social services that can be used, both for social polls, and for training as well [4,12,14]:

1. .social networks;
2. blogs;
3. postal services, and also free services for blogging conducting.

Students' surveys concerning their satisfaction level of organization of educational process and teaching job assessment by students is carried out on the base of the Department of Informatics, Software Engineering and Economic Cybernetics since 2009 using KSU feedback system - "Feedback" (hereinafter - feedback) developed by students of the Department - Berezovskiy D. and Tetenok S. under the guidance of professor Spivakovskiy O. V.

But, as many other services, this system requires not only technical support and improvement but constant evaluation of efficiency, prospect and consequences of its use, which is the main task of our research.

The purpose of this article is to analyze quantitative and qualitative indicators of the feedback service KSU Feedback use at Kherson State University.

2 Analysis of recent researches and publications

Feedback theory in educational process with its roots gets to E. L. Torndayka works (1911). It touches an issue of feedback model construction in an educational process, as well as efficient means and methods of its realization. The Feedback is used in the

educational process mainly as a key factor in order to improve the knowledge and to obtain new skills (Bangert-Drowns, Kulik). In addition to its impact onto achievements, the Feedback may also be considered as an important factor for motivation training (Lepper & Chabay, 1985; Narciss & Huth, 2004). Nevertheless, for training, history of the Feedback is not very optimistic and simple. According to Cohen (1985), the Feedback "... is one of the most educationally powerful and the least understandable functions in the pedagogical design". The main goal of the Feedback, which is carried out by a lecturer or a computer, in the class or elsewhere, that is to increase quality of education and productivity too, which causes formation of exact target concepts and skills (Albertson, 1986; Azevedo & Bernard, 1995; Narciss & Huth, 2004; VanLehn 1982) [4].

As a condition of activization of an educational activity is an availability of the Feedback between students and a lecturer, that corresponds to general theory of management systems of education. In the papers of many authors (R.F. Abdeyev, V.P. Bespal'ko, A.A. Bratko, D.I. Dubrovskiy, E.I. Mashbits, Askew Susan, P. Garber, B. Cox, M.S. Lvov and others), who researched informative aspects of training process, high didactic importance of the Feedback between students and a lecturer is noted. It is based on the information, provided by the channel of the Feedback. The lecturer can manage the process of training materials receiving and learning.

In the management theory, for general case, there are defined requirements to information coming through channels the Feedback: fullness, authenticity, efficiency. Applied modern means of ICT allows providing the fulfillment of all the conditions. As it was described in papers of B.E. Starichenko, N.Davidovich, R.Yavicha, P.Partington, J.Brennan, J.Valerie proper organization of informational and educational resources and usage of modern means of communication (first of all, networking) not only improves informational support and educational process management, in the framework of traditions for higher education institution forms of training organization, but also creates its particularly new forms: distant lectures, seminars and forums, distant consultations, forums of disciplines, means of distant control and self-control and wiki-resources.

Scientific heritage of essence and role of information technologies in higher education insitutions' management is connected with the names of local scientist and educationalists: O.V.Spivakovskiy, V.Yu.Bykov, G.M.Kravtsov.

The main point of the majority of researches, conducted in this field, is the fact that qualitative Feedback can significantly improve processes and training results.

Current research is a continuation of scientific-trial work conducted by Kherson State University in 2003-2012 according to an agreement in the framework of the State Programme "Computerization of Ukraine" - Designing and development of Internet technologies and software of remote system testing, Development of methods and technologies of designing flexible and distributed pedagogical software environments, Creation of e-documentation bank on distance learning for higher education, Creation of Internet portal of distance learning ECDL for higher educational establishments (ECDL), as well as conducted by us researches concerning technical component realization of the Feedback services and their use, for example at KSU [5, 6, 7], readiness of students to use IT in the educational process and beyond, and to construct ICT infrastructure of higher educational establishments [12, 13, 14].

3 Analysis of the existing systems feedback

Nowadays, we have analyzed the existing systems. The most common systems for surveying and analysis of the Feedback was selected through a search engine “Google”, “Bing”, “Yandex” and “Yahoo” using the keywords “Survey”, “Survey system”, “Feedback”, “Feedback system”, «Organization of the Feedback», «Poll system», «The Feedback», “Personal feedback”, etc. The following services were considered “ObjectPlanet”, “Murvey”, “QuestionPro”, “CollegeSurveyServices”, “Survey”, “SurveyMonkey”, “PollDaddy”, “Wufoo”, “Surveygizmo” by the shown above criteria:

1. Simplicity of registration, creation and modification of the questionnaires;
2. Type of the questionnaires spreading;
3. Support of differentiation of the target audiences;
4. Convenience of the results' storing;
5. Availability of the means to generate reports;
6. Cost of using the service and so on.

The simplicity is meant the minimum possible number of the steps to complete, correct realization of the particular option. For example, for the analysis of registration were chosen 2 criteria - the number of fields and the approximate spending time.

Type of the spreading questionnaires - the way in which the questionnaire gets to the respondent. This is usually a direct link to the survey, but also for this we use the social networks and other web-resources.

The support for target audiences is important when analyzing the results. The survey is conducted anonymously, but differing, for example, the respondents by the country location, the method of the authentication, etc.

The services are usually supported by the hierarchical structure of saving the results. The essence of the survey, questions, questionnaires are shared. The support for the filters, the availability of the search on the questions and questionnaires, the possibility of the multiple interface languages is the additional options, which are realized, for example, “SurveyMonkey”.

There is the minimum means for analysis of the results in each reporting service, but the multifunctional, interactive module to generate the reports with the supports of the export in the formats, the filters, etc. is in the complete, usually paid, versions. As for prices, some services are free (“Murvey”, “Survey”), some have the basic free functionality, but they are limited to the number of surveys, the completeness of the capacity for analysis, etc. (“Easypolls”, “QuestionPro”), or paid with the temporary trial-version (“Opinio”, “SurveyMonkey”).

All systems, which are considered, have some means for presenting the results of the questionnaire, because it is the main purpose of the reviewed services – the analysis and the processing of the collected information. But in each product of the module for the analysis and the reporting of the information has the characteristic properties, so consider them more.

“Easypolls” is the product “ObjectPlanet Inc.”, which focused at the generation of the surveys for the sites. The main feature of the service is simplicity to use. Therefore,

the reports and the analysis are also minimal. "Opinio" is a large-scale project "ObjectPlanetInc.", which focused at the enterprise and thousands survey.

"Murvey" – a web product "ObjectPlanet Inc." for the conduct and the management of the surveys. There is the possibility to set the period of the survey, to spread the questionnaires through the link in the menu «Report». You can see the results in the form of a histogram or a pie chart. \

"QuestionPro" is a professional tool, which contains a powerful tool for the reports: it is possible to apply the filters, to view the location of the respondents and the type of the device, which was used in charge, you can view the pie charts and the histogram of the responses, the time, which was spent on response and so on. Each of the blocks is optional and customizable in menu.

"College Survey Services" is the product CollegeSurveyServicesInc, which focused to assess the training courses and the generation of the reports.

A characteristic difference "Survey" is the evaluating the particular service, address to which you need indicate when registering. In other word, there is a standard questionnaire about the quality of the product, which cannot be edited. The answers to the questions are optional. Each question in the report is presented in the form of a line chart. You can also review the answers of the particular respondent, there are the identification by the time of the response and the mac-address device.

"SurveyMonkey" offers a very compact and a informative reports. You can select the type of presentation of the results, apply the filters on time, the number of the respondents, completion and more.

There is an export in the format .PDF, .XLS, .CSV, or SPSS, the possibility of tracing the history detailed of the answers in the full version is.

The main difference «PollDaddy» is that the questionnaires can be easily embedded in external websites. The detailed reports, the filters, the export are only available in the full version.

Another type of the survey was separated by the ratings.

"Wufoo" is a project, which owns by "SurveyMonkey" and focused to build the variety of the online forms: the forms for collection of the data, the registration, the contact forms and the surveys. You can use the templates, upload your own files and so on.

To create a report you need:

1. to fill in the basic configuration - the name and description;
2. to choose the data - all forms (as the case, the surveys) or some particular form;
3. to add widget is a graph, chart, number, text or table and choose the layout of location of the components (layout);
4. to configure the widget shows the possible properties when choose a particular item.

This form of presentation of the report is a very convenient, because you can independently generate a report of any complexity.

The feature "SurveyGizmo" is distribution on the data analysis (Data Explorer) and the report (Report). The data analysis contains the filters for date, the location of the respondents, the number of full, partial results and overall views.

World universities such as Princeton University [16], Newcastle University [17], University of Jyväskylä [18], National University of Singapore [19], University of

Melbourne EyeCare [20], University of Mysore [21], University of Sunderland [22], University of Kuopio [23], Loughborough University [24], University of York [25] and other higher educational establishments are the examples of realization and usage of the feedback systems reviewed in the process of the research.

The majority of the reviewed systems are “closed” for study purposes and analysis, as the Web-resource contains only brief annotation about its functional possibilities, or simply has a form of user’s authentication. The detailed analysis of the functional and specificity of feedback system usage mentioned above has been given by the following higher educational establishments, such as Princeton University, University of York, University of Kuopio, National University of Singapore and University of Jyväskylä.

It is important to remark, that the main peculiarity of the feedback system usage of the majority of the reviewed universities is its implementation into all the fields of activity of the higher educational establishments, starting with students’ survey and ending with survey for the visitors of the institutions’ web-resources, attendees of libraries, employees, etc.

Transparency of surveys’ results, their constant update and organization of additional feedback by using Web-resources, electronic mail and constant update and improvement of services shows the relevance of their usage by the mentioned higher educational establishments and positive attitude to survey passing by all the participants of training and administrative processes.

Consequently, most of the systems of the construction of feedback are coping very well with its responsibilities within its class.

Besides the above services considered, some educational institutions and the organizations were developed and a number of its own services of feedback with enough large difference in characteristics (anonymous, open access/registration, orientation to the user, the availability of free fares, etc.).

4 Brief description of “KSU Feedback” system

“KSU Feedback” system is a tool for management of organizations wishing to introduce Feedback into a cycle of decision-making process. This service allows in a user-friendly form to store, to aggregate and to analyze information on Feedback. The essence of this service is in conducting of anonymous or ordinary poll following clear criteria among strictly defined set of respondents.

Objectivity of evaluation is achieved using potential of an anonymous poll. Remote voting is also possible in any suitable place, which reduces an impact of interested ones onto a respondent’s answer. Due to a system of disposable and unique keys, organizers of a poll may determine a group of people who can participate in an evaluation process.

System of key generation is a special service, an environment used to regulate sets of keys, a tool for fast printing. Every key opens an access to vote in certain polls. The key turns to be invalid after its first use, and also may expire depending on the expiration date, set by the organizer.

It is worth noting that all collected data are automatically accumulated and can be presented in the form of various graphs and diagrams.

By reason that all the calculations are done by computer, organizers of a poll cannot influence on counting results. The service also provides a wide set of tools for:

1. Organization of data storage
2. Analysis of results;
3. Distribution of access levels by the poll organizers;
4. Effective teamwork;
5. "KSU feedback" system is used for the following purposes:
6. Lecturers of the department - to receive feedback from the students and to actualize programs and methods for holding training courses;
7. Students' autonomous bodies - to analyze opinions of students about the activities carried out;
8. University senates - to evaluate quality, efficiency and topicality of reports;
9. Department of statistics - to obtain consolidated reports (showing its dynamics) about an adaptation level of junior level university students and vocational guidance of upperclassmen;
10. Scientific-practical conferences - to gather public opinion on the conference, including organizational questions and issues connected to the meaning content, etc.

Holding seminars and conferences using "KSU Feedback" made these activities more open and interactive. At the moment, there is a permanent version of the system (<http://feedback.ksu.ks.ua>), and developing process of other versions for extending the area of application is also carried out.

The advantages of Feedback over traditional system may be the following:

- Fast creation of any number of surveys;
- Provision of the maximum objectivity;
- Simple organization of polls, due to remote voting possibility;
- Instant data processing and results obtaining;
- Low probability of "human factor" influence in obtaining results;
- Access control to the results.

The disadvantages are the following:

- Additional resources such as computers and the Internet are required to hold monitoring;
- Presence of "authorized representative", who distributes the keys among the respondents;
- Difficulties in organizations;
- Close limits of voting time and place;
- Provision of the poll objectivity, as the respondent cannot be convinced in an anonymity of the answer.

5 Research methods

Conducting research on the subject makes use both theoretical and empirical research methods. Thus, the research of the ratio of the students to using the services of feedback in the learning process is impossible without the analysis, comparison and synthesis, abstract approach to determining the basic regularities of the using of such services, logical approach to the description of possible implementations and their main characteristics and requirements. The main means of getting results is conducting the survey and the analysis of the indicators of readiness, interest and satisfaction of university students to use the service of feedback “KSU Feedback” in the learning process.

Considering the possibility of direct interaction with the participants of the educational process and the regularity of the survey, as a group of the respondents was chosen the students 1-4 years of study of the department of Computer Science, Software Engineering and Economic Cybernetics of Kherson State University.

The survey of students was conducted in writing traditional method. It is worth noting that the stages and the algorithm of traditional method fully consistent the algorithms of the service of feedback “KSU Feedback”, and therefore it ensured the comfort and the clarity (understandability, simplicity) embedded the survey as usual for students.

The first research was conducted in 2013. The main purpose of the survey was to determine the relationship of the students to the services and the conduct surveys. The questionnaire with 4 questions was proposed for our research. The total number of the respondents is 79, representing 53% of the students.

In 2015 it was re-conducted the research, which focused to confirmation and clarification of the results, which were received in 2013.

The proposed questionnaire includes 17 questions, which were directed to the research of the criteria such as:

Understanding the student of the concept “Feedback”, the main functions of the system, proposed the criteria of the evaluation;

Detection the readiness and the desire to passing the relevant surveys;

Search strong (positive) and weaknesses (negative) side of the system and ways to improve and the analysis of prospects for future using.

The total number of respondents is 209, which is 63% of the students of the department. All respondents taking part in the second research were divided into two groups - interested and uninterested in passing survey.

The analysis of the results of research included a comparison of survey results two above mentioned group.

6 Statistical analysis of the obtained results

First step of the research was to compare the results obtained in 2013 and in 2015, in order to determine changes in students’ attitude to KSU Feedback poll system. The results of this comparison are shown in the table 1.

Table 1. Comparison of the obtained results in 2013 and in 2015.

Content of questions	Year of study	Result	
		Yes	No
Have you ever ignored feedback?			
	2013	72%	17%
	2015	46%	49%
If you do not take part in surveys, then for which reasons (select one or more correct answers)	Common Criteria		
		Lack of interest	Uncertainty in effectiveness
	2013	11%	26%
	2015	32%	4%
	Different criteria:		
	2013	Take it partially	36%
		Tired of questions	7%
		Do not see the final rating	7%
		It's impossible to appraise the lecturer	5%
		Your answer	5%
		Do not trust in anonymity of the service	3%
	2015	Lack of self-discipline	22%
		Lack of time	21%
		Always taking it	18%
Would you like to continue using this service?		Yes	No
	2013	52%	32%
	2015	67%	14%
You always objectively assess lecturer's work, isn't it?		Yes	No
	2013	65%	35%
	2015	91%	9%

According to the survey's results in 2013, it turned out that 72% of those polled at least once ignored feedback. In 2015, this number has decreased by 26%. (Fig. 1.)

Among the main reasons of ignoring in the first survey a majority of the students had chosen the uncertainty in the effectiveness and wish to take it only on certain disciplines. In the second survey, students who were not interested in taking it, picked - the lack of interest - 20,83%. Students who were interested, chose lack of self-discipline (13,89%), as well as lack of time and lack of interest (11,11%). A great

percentage (22,22%) suggested their own answer, namely - took the Feedback - 18%; and uncertainty in the effectiveness - 4%.

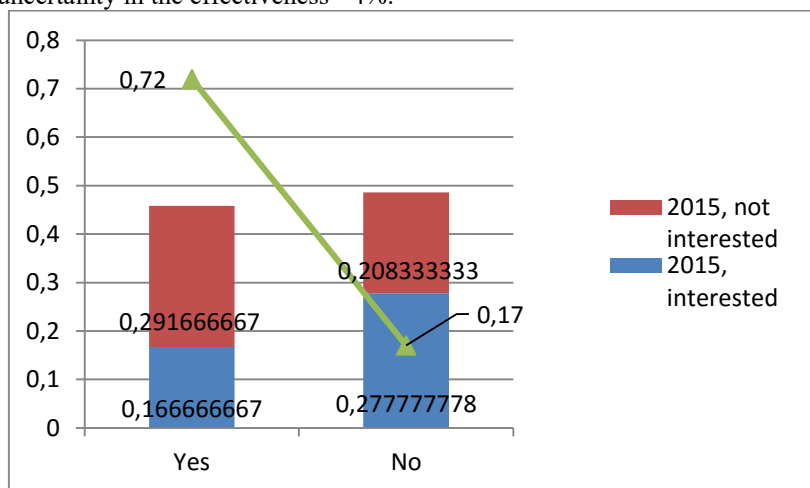


Fig. 1. The results of the survey . The answer to the question "Have you ever ignored the Feedback?"

The number of respondents who always objectively assess the lecturer in 2013 is equal to 65%. In 2015 this number increased by 26%. Also, in comparison with 2013 the number of respondents who want to take the Feedback increased by 15%. At the same time, all the students, who were interested, consider it necessary. (Fig.2)

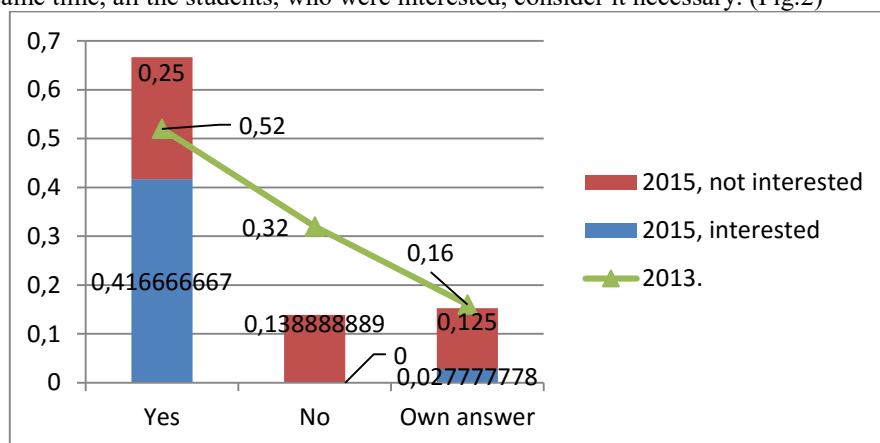


Fig. 2. The results of students' survey. The answer to the question "Is it necessary to carry out the Feedback?"

To assess student's attitude to the Feedback and level of their readiness to the Feedback in the second research we added greater number of questions, directed on the study of these factors.

Table 2. Specification of the results obtained in 2015.

Content of an issue		Choices of answers				
Are you afraid that an objective lecturers' assessment may bring itself to negative consequences?		Yes		No		
	Interes ted	11%		35%		
	Not interes ted	13%		39%		
In your opinion, when is an optimal time to conduct the Feedback?		Before the examina tions	During the examinati ons	Right after the examinati ons	Awhile after the examinati ons	In mid-terms
	Interes ted	14%	0%	17%	6%	17%
	Not interes ted	8%	0%	15%	4%	17%
In a school, lyceum, gymnasium, etc. Have you seen similar technologies to the Feedback?		No	Rarely		Regularly	Often
	Interes ted	23%	15%		10%	3%
	Not interes ted	22%	13%		5%	3%
What is the best way to organize the Feedback conducting?		Add the reminder			Make keys distribution	Your answer
	Interes ted	12%			35%	12%
	Not interes ted	10%			26%	3%
Is it necessary to be taking the Feedback?		Surely, as an examinations			Upon a request	
	Interes ted	14%			32%	
	Not interes ted	13%			33%	
		Yes			No	

Content of an issue		Choices of answers	
Do personal motives influence on your assessment?	Interested	7%	38%
	Not interested	6%	44%

Another important part of our research was to identify which positive and negative sides of the feedback can be seen exactly by the students. The result is shown in the table 3.

The largest of the positive sides, according to the respondents, is a possibility to express his/her opinion about the lecturer and to assess its work. The greatest negative factor in the use of the service-"biased assessment." As the least positive effect, determined by the students, is an opportunity to improve educational process. It should also be noted, that the least negative factors are the lack of opportunities to revise the results of the students, negative consequences and unattractive interface.

Table 3. Advantages and disadvantages of KSU Feedback in respondents' opinion

Positive features in the use of the service	
	Quantity of the respondents (%)
An opportunity to express their points of view about the lecturer, to assess his work	23
Anonymity of the survey	11
A lecturer through a student's eyes	4
An opportunity to improve the educational process	2
Negative features in the use of the service	
Biased evaluation	14
Many of the issues and criteria, resulting loss of time	10
Absence of lecturers' reaction concerning the remarks and students' preferences	5
Closure of the Feedback at only one department	5
Unavailability of revising student's results	4
Negative consequences	3
Unattractive interface	3

Also it was offered to the students to describe methods of improvement and deterioration of the Feedback. To improve, in students' opinion, it is necessary to:

- Update criteria, add encouragement, make the Feedback taking upon request, simplify questions, change an interface, make the process of taking it more organized, increase an interest, add lecturer's to the list from other departments, make service from outside of the university, to make our key distributing, add a possibility of revising the results by the students, expand to other faculties and universities, lecturer's real punishment, who were assessed by the students badly.

Can lead to worsening, according to students:

- To deanonymize, obsessive reminder, obligatory to take the Feedback before the examinations.

In the column "your variant" students were answering the following:

- To simplify criteria and questions, make to the site more understandable, change the design, motivate the students to take the Feedback, reduce the scale of assessment.

One of the options of the Feedback upgrade is to update criteria for evaluation. According to the results of the survey, it was revealed that students want to assess the lecturers according to the following criteria:

- Sincerity, kindness, ability to make the material clear for students, sense of humor, usefulness of the material in everyday life and an ability to see many solutions to problems, lecturer's attitude to corruption.

7 Interpretation of the research results

Conducting a traditional survey allowed us to conduct a survey more organized and get the answers more than 50% of students, who study at the department of computer science, software engineering and economic cybernetics. The traditional type of survey was also chosen as one corresponding algorithm of the described and researched our system familiar to students (recall that provide the feedback means "feedback" is carried out at the department for the last 6 years). In addition, we considered that feedback was the object of our research the use of such services during the survey would not able to obtain the necessary results, or would affect their accuracy despite certain psychological, social and organizational factors. But we spent a lot more time and organizational resources.

The failure to pass a representative number of the students to leave the feedback is the biggest problem of a incorrect display of the real picture of the quality of the educational process. We consider a sufficient number of students to passing the survey at least 60% of the total number of students of the academic group.

It is an important the passage the feedback as many as possible students to receive the objective results. Thus, the result of the first research is only 17% of students passed the feedback every time. In 2015 this number increased to 32%. And most of them were interested in passing our survey. One of the main reasons for neglect in 2013 was the uncertainty in the effectiveness (26%). In the second research agreed with that only 4%.

The most important reason was the lack of interest (32%), and most of the respondents of this group were not interested in passing survey.

There was found 52% of students wishing to pass feedback in the first research. For two years this percentage rose to 67%. Among the interested students no one gave a negative answer to the question about the need for feedback. About 15% had a thought: doing only when necessary for the teacher; doing, but not often; annually, etc.

The results of two researches have shown that more than half of the students always express an objective opinion on the teachers. You can verify the results of answers to the question: Do you always objectively evaluate teachers? As a result of re-research such students became 26% more.

Several questions were added to display a more complete picture in the second research.

We considered that the results and the quality of the feedback effect passing "Culture survey." As we can see in Table 2, 45% of respondents first encounter with technology surveys at the University. The total 20.84% regularly or often used earlier the services, which measured their level of satisfaction. This confirms the above opinion about the low level of culture surveys in Ukraine.

One of the main positive qualities of the system is the anonymity. So 73% of students have no fear of what the evaluation of teachers can have negative consequences. The personal reasons do not affect the evaluation of teachers in 81.94% of students. Thus, the response when the passage feedback is objective.

The students note the lack of organization and interest is the main reason of the failure to pass feedback. These problems can be solved by conducting the survey of students necessarily for example exam, but it is supported by only 26.39%.

Most students (65.27%) expressed the view that it is necessary to conduct feedback either immediately after the session or in the middle of the semester. Conducting the survey immediately after the session, as it is conducted today, supported by 31% of respondents. The largest number (33%) considers that the feedback is best done in the middle of the semester, did not support the idea to conduct survey during the session (0%). All interested students think the need to conduct feedback. And 45.83% of the students consider it necessary to ignore the passage of feedback.

The main aspects on which we should pay more attention to, and which are based on the survey results include:

1. high percentage is not interest in the passing in this survey;
2. high level of the students who ignored the feedback;
3. about 23% have fear about the negative consequences after the passage of feedback;
4. the results of survey influenced by personal reasons - about 12 percent of the students;
5. the desire of the students to change the organization of the passage of feedback;
6. the desire of the students is simplify the question and make their own questions;
7. another.

Conclusions

During the research, low “level of culture” of the polls was revealed among a group of students. It points out lack of experience of the Feedback system use at schools of Kherson and Kherson region, as the majority of students of the Faculty of Physics, Mathematics and Informatics of KSU are graduates of those higher educational establishments in particular. The only possible ways to solve this problem is:

- to devote more time to first year students’ learning more about KSU Feedback, to conduct surveys among first-year students to learn more about this system’s assignment and develop readiness to taking it;
- adaptation and implementation of KSU Feedback system into training process at schools and lyceums of Kherson and Kherson region.

To our mind, knowledge obtained throughout practical use of KSU Feedback system at the Faculty of Physics, Mathematics and Informatics of Kherson State University is positive. It gives an opportunity to lecturers to find out an objective opinion about him/herself and make changes in the training process taking into consideration students’ wishes.

In order to make KSU Feedback more efficient is necessary to attract to taking the survey as many students as possible. This can be done by improving the system. Consequently, in the end of the survey, we have found out that the main directions to change KSU Feedback system should be questions’ updates, additions to students’ proposed criteria and also an opportunity to review the results. Besides, in our opinion, it’s necessary to revise the process of survey conducting for improving its organization and to choose the best time, in students’ mind, to carry on the Feedback.

The results of our survey showed sufficiently high percentage of the students, satisfied with KSU Feedback system. It’s important to underline, that for the past 2 years this percentage had increased. It shows a necessity to continue the research of the present service in order to make it more efficient and expand ranges of its use.

Perspectives for further research. In the future, it’s planned to improve present system, which is based on the results we’ve obtained, and its further monitoring as well. Besides, it’s necessary to reveal causes of negative tendencies in the surveyed group and to conduct an individual research about Department’s lecturers attitude to this service.

References

1. Vaytsel’, S.R.: Efficient feedback. How to generate and transfer your ideas, M (2008)
2. Chesnok, I.I.: Problem of self-actualization in psychology, M.(1977)
3. Partington P. Student feedback – context, issues and practice accounts. Sheffield: CVCP - 1993.
4. Valerie J. Shute. Focus on Formative Feedback, <http://projects.ict.usc.edu/dlxxi/materials/Sept2009/Research%20Readings/Shute%202008%20Focus%20on%20formative%20feedback.pdf>
5. Spivakovsky, O.V., Alfiorova, L.M., Alfiorov, E.A.: Functions and university structure as a difficult mechanism, that serves educational interests. J. Informational Technologies in Education. 12, 21-25 (2012)

6. Spivakovsky, A.V., Berezovsky, D.A., Tityenok, S.A.: Architecture and functionality of software complex "KSU FEEDBACK". J. Informational Technologies in Education. 5, 40-53 (2010)
7. Spivakovsky, A.V., Berezovsky, D.A., Tityenok, S.A.: Functionality of the KSU FEEDBACK 3. J. Informational Technologies in Education. 11, 09-18 (2012)
8. Bagdikian, S.V.: Educational Services: research of conception, classification, basic characteristics. J. Economics. Management. Entrepreneurship. 22(1), 199 – 205 (2010)
9. Kreknin M.E. Mackintosh and Education: Informatics and IT: Manual: Information and management. Feedback. (1999, 2002)
10. CSESSUMS.COM. An unlearning journal, <http://www.csessums.com/2010/04/50-tactics-and-resources-to-support-integrating-technology-into-your-curriculum/>.
11. Kushnir, Nataliya, Anna, Manzhula, Nataliya, Valko. "Future and Experienced Teachers Should Collaborate on ICT Integration." Information and Communication Technologies in Education, Research, and Industrial Applications. Springer International Publishing. pp. 217-237 (2014)
12. Spivakovsky, A.V.; Vinnik M.O.; Tarasich, Yu. H.: Construction of ICT infrastructure of higher educational establishments: problems and ways of solution. J. Informational technologies and training tools.1, 99-116 (2014)
13. Spivakovska, Eugenia, Osipova, Nataliya, Vinnik, Maksim, Tarasich, Yulia. Information Competence of University Students in Ukraine: Development Status and Prospects. Information and Communication Technologies in Education, research, and Industrial Applications. Springer International Publishing, pp. 194-216 (2014)
14. Denysenko, V, Vinnyk, M., Tarasich, Yu.: An Analysis of the readiness of IT Specialties Students to Use Information Technology in the Educational Proces. Society, Integration, Education. Rezekne Higher Education Institution. pp.75-83 (2014) 1
15. Spivakovsky, Alexander, Lyudmila, Alferova, Eugene, Alferov: "University as a Corporation Which Serves Educational Interests." ICT in Education, Research, and Industrial Applications. Springer Berlin Heidelberg. pp. 60-71 (2012)
16. Feedback Systems: An Introduction for Scientists and Engineers, <http://press.princeton.edu/titles/8701.html>
17. Newcastle University feedback system, <https://apps.ncl.ac.uk/sdu>
18. Survey Description for the Course Feedback System of the University of Jyväskylä, <https://www.jyu.fi/yliopistopalvelut/surveys/surveys-and-feedback-staff/coursefeedback/surveydescription>
19. Student Feedback Exercise, <http://www.nus.edu.sg/registrar/event/feedbk.html>
20. Clinical Feedback System, <https://www.university-eyecare.org.au/resources/ocas.php>
21. Department of Higher Education online Student Feedback System, <http://www.uni-mysore.ac.in/departement-higher-education-online-student-feedback-system>
22. Student feedback (University of Sunderland), <http://www.sunderland.ac.uk/studentlife/student-feedback/>
23. The student feedback system in the University of Kuopio, https://www2.uef.fi/documents/13384/1401124/feedbacksystem_english.pdf/f5a6fadc-6e47-43e4-b40d-0c1ab71ed494
24. Student Feedback (Loughborough University), http://www.lboro.ac.uk/services/registry/pqtp/aqphandbook/5_student-feedback/
25. Online Feedback System(University of York), <https://www.york.ac.uk/it-services/services/online-feedback/#tab-1>

Tips and Pitfalls for Blended Learning: Redesigning a CS Curriculum Using IT

Leo van Moergestel, Ander de Keijzer, and Esther van der Stappen

HU University of Applied Sciences Utrecht, Utrecht, the Netherlands,
leo.vanmoergestel@hu.nl,
<http://www.hu.nl>

Abstract. This paper describes the process of introducing blended learning in a CS educational program. The methodology that has been used as well as the motivation for the choices made are given. The first results compared with results from previous courses that used a more classical teaching approach are given. These results show that the new methodology proves to be promising and successful. The successes of the new program as well as the problems encountered are discussed with their possible solution.

Keywords. blended learning, course development, didactic models in practice

Key Terms. TeachingMethodology, ICTEnvironment, TeachingProcess

1 Introduction

The rapid change in insights in how students learn and the technical possibilities has lead to innovative educational technologies. E-learning and MOOCs are examples of recent developments in Technology-Enhanced Learning (TEL). At the Utrecht University of Applied Sciences a project was started to design a new curriculum for bachelor students. There were several reasons for this project. The first reason was the fact that due to a reorganisation of several educational CS programs, there was a need for a new curriculum. The other reasons were new insights and the development of new concepts in blended learning as well as the strategic view of our university on future education. The combination of these reasons turned out to be a good starting point for the project.

One of the goals of the project was to construct this new curriculum using new concepts in higher education. Some of these concepts came into existence because of advances in information and computing technology. Nowadays all of our students have high speed access to the internet which opens the opportunity for video, and other multimedia applications that can be used by students at home in a so-called distant learning situation. By combining several kind of possibilities to learn for students a so-called blended learning environment comes into existence. This paper will discuss the approach that has been followed to use this blended learning approach in our course re-design in combination with the

considerations and motivations for the choices made. A comparison with results from earlier courses is also given.

This paper is organised as follows: In the next section the concepts that are used will be explained. This is combined with a discussion of related work. Next the approach of curriculum re-design is presented in the section "Curriculum re-design". The section "Implementation" is dedicated to the implementation of the first part of the new bachelor program and the choices that have been made about teaching concepts. The first results are presented in the section "Results" followed by a discussion where the pitfalls and problems encountered are treated. The tips to avoid the pitfalls or solve the problems are also given. Finally a conclusion and a bibliography will end the paper.

2 Concepts and related work

The project is based on several innovative concepts in relation to education. In this section the concepts are introduced. In a later section an explanation will be given why these concepts have been used in our innovation project.

2.1 Blended learning

Blended learning [4] was introduced as a mix between face-to-face and online learning, giving rise to the challenge of virtually limitless design possibilities.

The idea of blended learning is that the learning content and subjects should be offered to the student in several ways. The student can choose the type of explanation that fits herself the best. Also a combination of learning possibilities might help a student to understand the subjects. Whitelock & Jelfs [9] opened a journal special issue on this topic with three definitions:

- the integrated combination of traditional learning with web-based online approaches;
- the combination of media and tools employed in an e-learning environment; and
- the combination of a number of pedagogic approaches, irrespective of learning technology use.

Of these, the first is perhaps the most common interpretation. For our approach the third definition fits the best, but we also heavily rely on web-based approaches.

2.2 Flipping the classroom

Flipping the classroom is an educational teaching method where lectures are replaced by self-study. This self-study is supported by moments of interaction between teacher and student where students can ask questions and put the theory into practice. In Figure 1 this situation is depicted. What is not shown in

the figure is the fact that in the flipped class students should do their homework before the moment of interaction with the teacher and other students. The reason to start using this method is mostly based on the observation that the concentration of the student is limited to a small amount of time. Working on a problem related to theory that has been studied before gets the student more involved. A definition of flipping the classroom is given by Lage e.a. [5]: *Invert-*

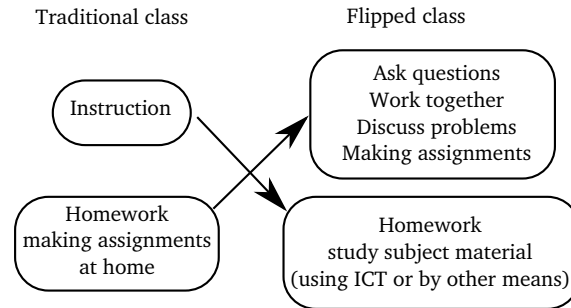


Fig. 1. Flipping the classroom

ing the classroom means that events that have traditionally taken place inside the classroom now take place outside the classroom and vice versa. An overview of flipping the classroom is given by [1].

2.3 4C/ID-model

The 4C/ID-model is a didactic model that is specially suited for training students that are determined to have a professional career in the industry [6]. 4C/ID is abbreviation for 4 components instructional design. The main idea is that students will always work on real-life products that are closely related to their future professional work. In the starting phase of their training students are confronted with situations where most problems are already scoped and analysed by the teachers. Students have to fill in the gaps. Later on, students are given situations where they have to do more work themselves, while in the final part of their training they should be able to solve problems in complex situations by themselves. In all cases they are working on a real-life product fitting their educational program. The 4 components of the model are:

1. Learning task (backbone of the education program).
2. Supportive information.
3. Procedural information.
4. Part-task practice.

In Figure 2 a schematic overview of the concept is shown. Students are working on learning tasks (1) represented by big circles in Figure 2. At the beginning of

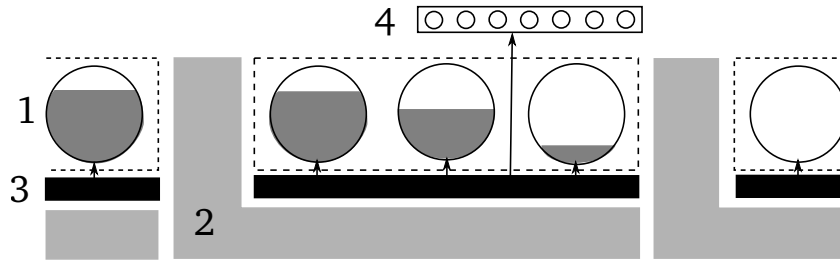


Fig. 2. The 4C/ID model [7]

their study, students start with simple tasks and students will work on complex tasks at the end. The simplicity and complexity is controlled by giving much support in the beginning and reducing this support at the end. In Figure 2 this is shown by the level the circles are "filled". The learning tasks are part of a project (the dashed boxes, containing learning tasks) that will result in a product. The type of product the students will work on is related to their future professional career. Supportive information (2) is given at the beginning of every project. The supportive information gives the students a solid background to handle the problems they will encounter. This mostly theoretical information is also given during the time students are working on their tasks. The procedural information (3) is presented in a "just in time" manner to the students at the moment they really need it to complete the learning task. Practising with new concepts is also included in the model. In Figure 2 this is represented by small circles (4). A thorough treatment of the concept can be found in [7].

2.4 Technology-enhanced learning

Two of the concepts mentioned in this section are closely related to IT. Flipping the classroom and blended learning heavily rely on IT and internet technology, while the 4C/ID model is a concept that can be implemented without any IT support.

3 Curriculum re-design

Our main goal is to build a curriculum that uses new but proven concepts and fits in the educational concept that was chosen in the preliminary phase of the project. The concepts should be proven, because a situation where students are treated as guinea pigs in an experiment should in all cases be avoided. The quality and high standards used in the past should be guaranteed. The research presented in this paper has its main focus on the first part of this new curriculum. This first part consists of three courses of 5 European credits (EC's) [8] each. The courses are attended by around 500 first-year IT students.

3.1 Initial situation

One year ago there were four different IT-based professional bachelor programs at the Utrecht University of Applied Sciences. These four programs were:

- Software engineering (SE)
- Computer science and embedded systems (CSES)
- Business IT and management (BIM)
- System and network engineering (SNE)

These trajectories are now combined in one bachelor study with four different profiles. Combining bachelor studies was stimulated by the government for all existing bachelor studies. This will make it easier for students to select a certain study and specialize for a certain profile later. Combining these four given trajectories may seem like a simple reorganisation, but the students are offered more possibilities to select a set of courses leading to the final diploma. Cooperation between the different types of IT engineers is also more stimulated. This type of cooperation is often needed in the IT industry, so training students to cooperate during their study in multidisciplinary teams will support their professional career.

3.2 Design principles

As a design principle we used the term MOS for the three properties every course should adhere to. MOS stands for:

- Motivation - use of new technology and didactical approaches should align more with students preferences. Working on real assignments (as in the 4C/ID model) and deeper discussions should motivate students more.
- Orientation - give students a good overview over the broad IT-landscape and professional roles therein. Guide them towards choosing one of the four specializations within the Bachelor program.
- Selection - have the right level of difficulty in content and assignments so that passing 10 of the 12 courses more or less predicts finishing the study within the regular duration of four years.

Both orientation and selection are prescribed by regulations of the Dutch government [3].

3.3 Involvement of the teaching staff

At the beginning of the project, ideas and advice from the teaching team of around 40 persons were collected. Presentations and discussion sessions were organised. During two weeks discussion posters were available where all people involved could post their comments and remarks. The whole teaching and management staff was able to contribute ideas and proposals. This resulted in the following set of requirements:

- The same program for all students during the first half year.
- No instruction to big groups, but classes with a maximum of 32 students
- Showcases of real-life products as the main drive for the curriculum
- Special attention to the development of professional skills by the students during their training. Professional skills are skills that one might expect from a professional IT engineer apart from the specific domain related knowledge and skills, such as professional written and oral communication, leadership, (team)planning and ethical skills.

3.4 Involvement of stakeholders

From the beginning the IT industry as well as the students were involved. The plans for the curriculum were presented to representatives of the IT industry and students from all years of the four-year curriculum. By using the feedback of these meetings, the new curriculum adheres to the requirements and expectations of the software industry as well as the students. Every course under development was also checked this way.

The developers of different courses had several meetings to guarantee the coherence between the different courses that were given in parallel. All courses started with an explanation of this coherence and why this specific course had its place in the bachelor program.

4 Implementation

For the time being, only the first year will be considered. The first year curriculum consists of four periods of ten weeks. At the end of the year every student should have a personal assessment to check her knowledge and practical capabilities learned so far.

4.1 Learning model

For instruction a selection of one the following possibilities has to be made:

1. Blended learning: use all kind of teaching techniques to train the student. Let the student decide which one fits her the best.
2. Problem driven education: students are confronted with a problem en should discover by themselves what knowledge and information they need to solve the problem. When they have questions, the teacher will get involved.
3. Project based education: a project is the central part of the training and should control all other educational methods involved.
4. Classical approach, giving theoretical training supported by practising the learned material.

Considering the new possibilities in IT, the blended learning option has been chosen. By definition, all other methods could be included, but the problem driven possibility had been used is the past and turned out to be not so successful.

A pure classical approach is not apt for our education, because the type of student in our institute is more interested in practical problem solutions, than pure academic knowledge. Pure project based education that has been used in the past, seems to be missing some essential aspects. It turned out to be difficult to find a set of projects that covers all end-goals of a professional IT engineer. A choice has been made to combine blended learning with flipping the classroom. This approach did not need the instruction sessions for a huge number of students, that turned out to be not so successful in the past.

Next a choice has to be made between 4C/ID and the classical approach. In our case the 4C/ID model has been chosen because our institute focusses on training bachelor students to act as a professional worker in the IT industry. The teaching team had the experience that these students are more motivated when they are working on real products [6]. This in contrast with the pure academic approach where the student is trained to be a scientist. In the academic approach the coherence with the work in the IT industry is less evident. The 4C/ID also fits well in the concept of blended learning and flipping the classroom. Students are working on projects and will also be instructed to guarantee their theoretical knowledge.

The choices made also fit well with the design principles regarding motivation as explained in Section 3.2.

4.2 Selection of courses

To give the students the possibility to orient on the possibilities in a later phase of the bachelor program, all three courses should more or less be tied to the final four tracks the student has to select during the first half year. Table 1 shows the relationship between the courses and the four tracks. The orientation on the

Table 1. Relationship between track and course

Education	Computers	Programming	Organisation
BIM			✓✓
SE		✓	✓
SNE	✓		✓
CSES	✓	✓	

tracks is now guaranteed. A mark means that the course has a significant relation with the track mentioned. A missing mark does not mean that the course is not significant for the track, but that it is not a main focus. Another reason why these three courses are combined in the first block of the study is that it is a nice way to demonstrate the layered structure that plays an important role in many IT concepts. At the top is IT in organisations, The middle layer is presented by programming and finally the bottom layer (hardware, operating systems and

networking) is covered by computer systems and networks. Having a model and three courses to start with, we will present the results in the next section.

5 Results

The new program started in September 2016 and the first period consisted of the three aforementioned courses. These courses had the same learning goals as courses we had in our old model so a comparison with courses in the previous year is legitimate. Some minor differences exists that will be discussed later. As mentioned and motivated in the previous section the courses are:

- introduction in programming
- computer systems and networks
- IT in organisations

In Table 2 the results for the three courses are given for both the blended and classical learning (previous year). Nb is the number of students doing the final exam for the blended course, Nc is the number of students doing exam for the previous classical course, pass is the number of student that passed the exam (including a second chance exam) and avg. is the average score for the exam on a scale from 1 to 10. The average score has not significantly changed, while the percentage of students passing the exam has increased.

Table 2. Results

Course	Nb.	pass	avg.	Nc.	pass	avg.
IT in organisations	474	372	5,7	413	247	5,7
Programming	486	399	6,1	425	330	6,2
Computer systems & networks	490	351	5,4	454	293	5,3

Figure 3 shows the results of the percentage of students that successfully completed the course in the regular time after a second chance had been offered. In our institute, every course will have a second exam every year. In the same figure, the results for the previous year have been added. From the figure it becomes clear that the yield has significantly increased for all courses. The most successful course so far was programming, though this course was also the most successful one in the previous year. There is also a slight difference in the content of the program, mainly due to the fact that Python has been used instead of PHP or Java may have influenced the outcome. The learning goals did not change, so it is fair to state that the new blended course was more successful than the year before. All other courses show a better outcome than the year before. The other two courses are quite similar in content and learning goals, so the results can be compared. One should keep in mind that the new curriculum was not

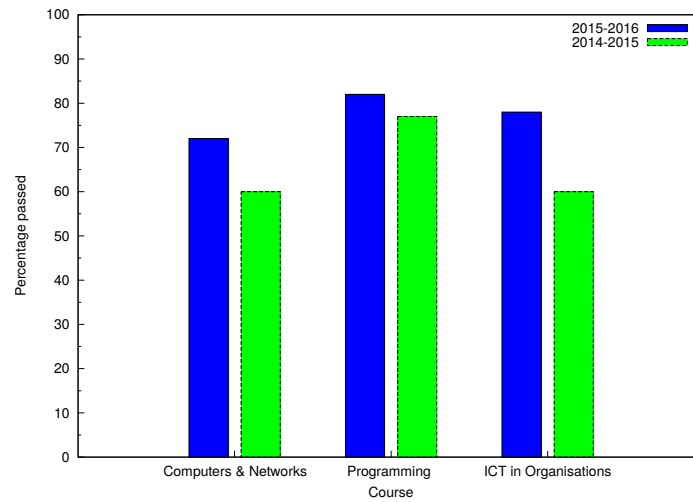


Fig. 3. Success rate per course

only based on blended learning and the 4C/ID model, but also based on the fact that we did not separate the different specialisations from the start. Also the fact that there were no classes with more than 32 students should be mentioned here. In the previous year there were instruction classes for over 100 students. So the positive effect should be considered as the consequence of all changes in the curriculum.

6 Discussion

In this section, some of the problems encountered during development and introduction of the new curriculum are discussed as well as the successes achieved.

6.1 Pitfalls

Though the concepts presented were supported by the majority of teachers in our institute, some remarks should be made here:

1. The amount of work for the developers of the courses was much more than expected in the beginning. Selecting good online material took much more time, because the developers should check the quality and adequateness carefully [2]. Also the match of the online material with the learning goals turned out to be cumbersome. The alternative would have been developing material at our institute, but that would have cost even more time.
2. The flipping the classroom concept only works well if there is a possibility to monitor the self-study phase of the students. Students should study in advance and do some self tests. The results of these tests will give a clue

to the teacher which parts of the teaching material should be treated at the beginning of the classroom meeting before students will work on instructions and tasks related to the material. In our situation monitoring was not yet possible, so we had to ask the students what problems they encountered. This might lead to the situation that a group of students decides to pay less attention to the self-study, knowing that the teacher cannot check and almost certainly will explain the difficult parts at the beginning of the meeting.

3. There was quite a big difference in applying the concepts as proposed by different persons teaching the same course. Many more meetings would have helped to solve this problem, but this would also put a burden on the people involved.
4. Students were not used to the new educational concepts. More effort should be taken to introduce students to the new environment as well as explaining the system.

The problems mentioned here are solvable. In the situation described under number 1 and 3 it means making more time available for the teaching staff to prepare the courses and having the possibility to discuss the method and help each other to use good practices. Currently the solution for the problem described under number 2 is under development. Monitoring activities per individual student is necessary and as an Institute of ICT we are strongly involved with the software used to create the blended learning environment. The problem described under number 4 should be solved by explaining the students the concept more carefully. The fact that students know that their home-activities for the course are monitored may also help to activate the students.

6.2 Successes

There is also a number of successes to report on the new program.

1. The new learning paradigm turned out to help a bigger number of students to complete the courses within due time. Given the fact that some important improvements still can be made, we expect an even better result for the years to come.
2. Most students in IT expect an IT training to be IT-based. Older ways of instruction are sometimes considered by students to be outdated or belonging to the previous century. This fact also motivates students.
3. Another positive point is the fact that from the beginning both students as well as representatives from the software industry were involved in the construction of the new curriculum. This resulted in a curriculum that was a wide support both by industry as well by the student community.
4. A feature of the new course is multidisciplinary cooperation between students in project. This help students in getting a broader view on the IT domain and getting a better understanding of what other disciplines in that field mean and accomplish.

5. In the new situation students will select a specific specialisation after half a year of their study. This makes it much easier for students to find out what specialisation fits the best with their interest and capabilities. In the older setup there was also a half year with almost similar programs, but the switch to another discipline was more complicated. Actually the student had to switch while in the current course she only decides what specialisation she will choose after half a year.

Given the fact that the pitfalls we encountered are solvable, the switchover to the new educational approach can be considered as a success so far.

7 Conclusion and discussion

In this paper we described a project to renew a CS curriculum. In this curriculum concepts like 4C/ID, blended learning and flipping the classroom play an important role. The concepts that have been used are introduced and explained as well as the implementation and first results. The first results are promising.

IT-based solutions offer great opportunities for educational renewal, but care has to be taken about the way they are used. Also the time and work involved was in our case underestimated leading to a result that is not yet at the level that we had in mind at the start. However, as mentioned before the results so far are promising and give support to the idea that this is a good approach to keep students involved and motivated with their study at the educational institute.

References

1. Bishop, J., Verleger, M.: The flipped classroom: A survey of the research. Proceedings of the American Society for Engineering Education (ASEE) Annual Conference (2013)
2. Chauhan, J., Goel, A.: An analysis of video lecture in mooc. Proceedings of the 11th International Conference on ICT in Education, Research and Industrial Applications (ICTERI), Lviv pp. 35–50 (2015)
3. Dutch Government: Wet kwaliteit in verscheidenheid hoger onderwijs (Dutch). Staatsblad van het Koninkrijk der Nederlanden (Dutch) (298), 1–23 (2013)
4. Garrison, D.R., Kanuka, H.: Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education* 7(2), 95–105 (2004)
5. Lage, L., Platt, G.: The internet and the inverted classroom. *The Journal of Economic Education* 31(11) (2000)
6. Merriënboer, J., Clark, R., de, C.M.: Blueprints for complex learning: The 4c/id-model. *ETR&D* 50(2), 39–64 (2002)
7. Merriënboer, J., Kirschner, P.: Ten Steps to Complex Learning. Lawrence Erlbaum Associates Inc (2007)
8. Wagenaar, R.: An introduction to the european credit transfer and accumulation system (ects). EUA, Bologna Handbook (2006)
9. Whitelock, D., A., J.: Special issue on blended learning. *Journal of Educational Media* 28(2-3), 99–100 (2003)

Application of 3D Printing and 3D Scanning Technologies in Educational Activities

Kseniia Yashyna¹, Oleksandr Krupnyk¹, and Oksana Karpenko²

¹ Dniprodzerzhynsk state technical university, Dniprobudivska str., 2,
Dniprodzerzhinsk, Ukraine,

YashinaKsenia85@gmail.com, krupnik_a@ukr.net

² Kyiv State Maritime Academy named after hetman Petro
Konashevych-Sahaydachniy, Frunze str., 9, Kyiv, Ukraine
kapro_2004@ukr.net

Abstract. The article analyzes the spheres of 3D printing and 3D scanning modern technologies application, defines the perspectives of 3D technologies development, and proves the expediency of these technologies application in educational activities of a modern institution of higher education. The authors developed the training course curriculum that allows the students not only to acquire the knowledge and practical skills in the field of 3D printing (scanning), but also to get acquainted with the basics of entrepreneurship and principles of 3D technologies transfer. The article gives the matrix that describes in detail the competences, acquired by the students after getting acquainted with the course materials, developed by the authors.

Keywords: Higher Education, Knowledge Transfer, 3D Printing, 3D Scanning, Matrix of Competence, Training Course.

Key Terms: Academia, Knowledge Technology, Knowledge Transfer, Management.

1 Introduction

Nowadays 3D printing and 3D scanning technologies, originated around three decades ago, gain more and more popularity. 3D technologies are actively used not only by scientific specialists, but also by many companies, producing and designing goods of a wide range. A sharply increasing interest in 3D technologies is explained by the possibility to essentially reduce, with their help, the expenses of complicated technical goods production due to reducing the costs of making sketches, models, prototypes [1–5]. 3D printing is actively used in architecture, building, small-scale production, functional testing, medicine, and manufacture of wearing apparel, footwear, jewelry, toys and souvenirs [6]. The experts call 3D printing "a salvation of economy" and "a breakthrough in medicine"; emphasize ecological compatibility and easiness of use of modern 3D devices [7].

According to some forecasts, by 2050 every third family in the developed world will have a 3D printer at home; and by the end of the 21st century the copying machines, designed for analyzing and reproducing any structure with the precision up to an atom, will be created [8].

3D printing and 3D scanning technologies are actively used in the educational process for producing visual aids. Besides, the penetration of 3D printing in more and more spheres entails the growth of demand for qualified specialists in this field. Therefore, 3D printing technology and equipment application in a modern institution of higher education (IHE) will allow:

- to increase scientific potential;
- to significantly improve the IHE competitiveness at the global level;
- to train highly-qualified personnel.

A tremendous advantage of 3D technologies application is a significant increase of students' interest in the educational process. Modern postgraduate students and young scientists need skills of working with 3D equipment for their successful career. The preparation and organization of qualification courses, teaching the principles of work with modern 3D equipment, by the IHE is very promising.

3D technologies and equipment are actively applications by Norwich University (the USA), Exeter University (the United Kingdom), Staten Island Technical High School (Island), University of Sunderland (England), Stellenbosch University (South Africa), Massachusetts Institute of Technology (the USA), Budapest University of Technology and Economics (Hungary). At the same time, in Ukraine nowadays there are no institutions of higher education, using these technologies in the educational process and research work.

Thus, the development of an innovative course, dedicated to 3D technologies and permitting to acquire skills of working with 3D equipment, and its introduction into the educational process of a modern Ukrainian institution of higher education is the task of a current importance.

2 The Training Course "3D Printing Technology Transfer"

2.1 General Description and Structure of the Course

At the Dniprodzerzhynsk state technical university (DSTU) and the Kyiv State Maritime Academy (KSMA) the 3D equipment was purchased within the framework of realization of the international project "Knowledge Transfer Unit From Applied Research and Technology-Entrepreneurial Know-How Exchange to Development of Interdisciplinary Curricula Modules" (Reference Number: 544031-TEMPUS-1-2013-1-AT-TEMPUS-JPHES, Project Duration: 36 month, 12/2013 11/2016) [9–11].

Moreover, the KTU staff of the Dniprodzerzhynsk state technical university and the Kyiv State Maritime Academy jointly with the representatives of UdG

(Spain) developed a curriculum of the multidisciplinary course "3D PRINTING TECHNOLOGY TRANSFER". As a result of getting acquainted with this course materials the students (post-graduate students) will not only study the modern technologies of 3D printing and scanning, their history, used materials, a sphere of application, and get practical skills of making prototypes, but also master the principles of 3D technologies transfer, the basics of entrepreneurship; get acquainted with the instruments, allowing to carry out the effective 3D technologies transfer in Ukraine and abroad.

The relevance of the training module.

The module is oriented at a wide range of students, who want to have a confident understanding of the current concepts of knowledge economy, knowledge transfer, modern principles of 3D printing technology, and approaches to the implementation of 3D printing technology transfer. Nowadays 3D printing technology is effectively used in various fields of engineering. The transfer should ensure handing of 3D production from developers to customers.

The global scientific and technological transformations are the main world trends of modern society development. They determine a transition from the primary industrial economy to the postindustrial knowledge-based economy. The knowledge transfer is a multidimensional and multi-branch process of transmission of different forms and types of knowledge. The application of modern innovative technologies transfer, such as 3D printing technology, is the most efficient.

The application of modern knowledge transfer principles in 3D printing technology area helps developers to attract potential customers (investors) and to achieve maximum economic benefit.

The aim of the module is a creation of students knowledge system about knowledge transfer and 3D printing technology transfer.

Target group: Phd students, graduates who have a bachelor degree or a master degree.

Didactical approach

There are lectures, practical tasks, teachers consultations (both individual and collective) and students' self-guided work with an overall guidance of teachers with the aim to achieve the objectives. Students receive methodological support in the form of a program (curriculum) of the discipline, and a manual. It contains lectures and recommendations for self-guided work, a list of recommended literature and other useful recourses.

Time requirements / duration: 3 months

2.2 Curriculum of the Course

The developed course consists of 5 training modules:

Part 1: KNOWLEDGE TRANSFER

Module 1: *Basics of Knowledge & Innovation Transfer*

Session 1: *Types of knowledge transfer*

The knowledge transfer concept. The history of the European knowledge transfer system. A university as a basis of the knowledge-based society. The paradigm

of the knowledge triangle: education science innovation. The current state of knowledge transfer in Ukraine. Systems of internal and external knowledge transfer.

Session 2: Technology transfer

The definitions of embodied and disembodied technologies. The components of knowledge transfer: transfer of legally described technologies; transfer of know-how; transfer of physical products and economic benefits; transfer of skilled labor. Technology transfer as the main component of knowledge transfer.

Session 3: Innovation transfer

The definition of innovation. The definition and nature of innovation as an object of transfer. The mechanisms of innovation transfer. The methods that help to determine the costs of the object of transfer.

Module 2: Management of Knowledge Transfer

Session 1: Knowledge transfer infrastructure

The concept of a knowledge transfer unit. Target groups, mission, organizational structure and staffing of a knowledge transfer unit. The list of services of a knowledge transfer unit. Capacity building of a knowledge transfer unit.

Session 2: Knowledge transfer management

General description of project management. The main forms of organizational structure of a project. Common approaches to project planning. Practical principles of project management. The structure of a project. Resource planning, costs and project budget.

Session 3: International knowledge transfer

The definition of international knowledge transfer. Forms of explicit and implicit knowledge. Phases of internal knowledge transfer.

Module 3: Advanced topics of Knowledge Transfer

Session 1: Acknowledgment with Erasmus+ and Horizon 2020 EU Programs

The definition of international knowledge market. Single European space of Higher Education. Single European space of Research. EU programs Erasmus + and Horizon 2020.

Session 2: Acknowledgment with National Transfer Technology Network (NTTN)

European and Ukrainian Technology Transfer Networks. The principles of work in NTTN. The ways of preparation of the description of innovative solutions; technologies, ready for introduction; promising scientific projects and ideas. The methods of search for scientific innovative projects and offers.

Part 2: 3D PRINTING TECHNOLOGY

Module 1: Basics of 3D printing and 3D scanning

Session 1: The history of 3D printing technology

Prototyping. Stereo lithography Apparatus as the first device for 3D printing. Laser 3D printer. Inkjet 3D printer. Solidscape company. PolyJet technology.

Session 2: 3D printing technologies

StereoLithography Apparatus, SLA. Selective Laser Sintering, SLS. Selective Laser Melting, SLM. Electron Beam Melting, EBM. Multi Jet Modeling,

MJM. Laminated Object Manufacturing, LOM. 3DP. Fusing Deposition Modeling, FDM.

Session 3: *Materials for 3D printing*

ABC-plastic. Acrylic. Beton. Hydrogel. Paper. Gypsum. Wood fiber. Ice. Metal powder. Nylon. Polycaprolactone (PCL). Polycarbonate (PC). Polylactide (PLA). Polypropylene (PP). Polyphenylsulfone (PPSU). High Density Polyethylene (HDPE). Chocolate.

Session 4: *3D scanning technologies*

Methods of scanning: contact and noncontact. Active scanners. Passive scanners.

Session 5: *Critical issues of 3D printing*

Module 2: *Application of 3D printing and scanning*

Session 1: *The fields of 3D printing application*

Architecture. Building. Small-scale production. Functional testing. Medicine. Education. Manufacture of wearing apparel. Manufacture of footwear. Jewelry. Packaging design. Printing toys and souvenirs. Geographic Information Systems.

Session 2: *3D printing software*

Provision of information for 3D printing. File formats. Software for 3D printing: paid and free. Basic principles of 3D models. Preparation of models for 3D printing.

Session 3: *The fields of 3D scanning application*

Engineering analysis. Digital analysis. Digital archiving. Architecture. Medicine. Examples of using of 3D scanning technology.

Session 4: *Scanning and printing of a 3D model*

Scanning with a digital camera and scanner. Using special software. Optimizing the Printing Process. Scanning and Printing with a Higher Resolution. Modeling and Printing with Precision. Manipulating Meshes and Bridges.

Session 5: *Creating a project and printing a 3D model*

Different ways of 3D prototyping and mesh grid creation. Slicing. Shells and surface layers. Reviewing the print results.

2.3 Competences

Acquaintance with the course materials will give students the opportunity to possess the following competencies:

General competences

1. Use of English language
 - (a) use specific English terminology in the area of Management of Knowledge Transfer in oral speech;
 - (b) use specific English terminology in the area of Management of Knowledge Transfer in writing;
 - (c) make presentations in English with the help of specific terminology in the area of Management of Knowledge Transfer;
 - (d) Read English literature in the area of Management of Knowledge Transfer.

2. Gather and select information efficiently
 - (a) independently gather and select information from specific Web-resources;
 - (b) independently gather and select information from specific literature;
 - (c) independently gather and select information from news, specific conferences and seminars.
3. Use information and communication technologies.
 - (a) use information and communication technologies for different tasks;
 - (b) select and use the most appropriate information and communication technologies;
 - (c) select and use modern information and communication technologies for team work and management;
 - (d) select and use modern information and communication technologies for international knowledge transfer.
4. Work in teams.
 - (a) form a team; make a list of roles and a list of activities for each role;
 - (b) work independently in a team, select an appropriate role;
 - (c) prepare formal documentation and reports about teamwork; use different shared resources;
 - (d) use results of teamwork.
5. Communicate orally and in writing.
 - (a) communicate orally and in writing about specific problems, ideas, solutions, results;
 - (b) use presentations, supporting materials, software, videos and other resources in communicate orally and in writing;
 - (c) communicate orally and in writing on topics in the area of knowledge transfer management.

Special competences

1. Problem solving.
 - (a) accurate formulation of problems;
 - (b) search for different solutions of specific problems;
 - (c) form criteria for solutions evaluation, select appropriate solutions.
2. Use explicit and implicit forms of knowledge.
 - (a) independently use explicit forms of knowledge;
 - (b) confidently identify and use implicit knowledge necessary for solving specific tasks.
3. Use software for 3D modeling, 3D printing, 3D scanning.
 - (a) independently select and use appropriate 3D modeling software for different projects;
 - (b) independently select and use appropriate 3D printing software for different projects;
 - (c) independently select and use appropriate 3D scanning software for different projects.
4. Use 3D equipment.
 - (a) independently select and use appropriate 3D printing technologies for different projects;

- (b) independently select and use appropriate 3D printing materials for different projects;
- (c) independently select and use appropriate 3D scanning technologies for different projects.
- 5. Customer focus.
 - (a) perform marketing researches (interviews with customers, analysis of literature, news and web-resources); create a list of products and services required by customers;
 - (b) continuously develop and improve services required by customers;
 - (c) prepare descriptions of innovations required by customers;
 - (d) use individual approach to each customer.
- 6. Critical thinking.
 - (a) independently analyze and systematize tasks, ideas, methods, solutions, tools and results;
 - (b) independently analyze exceptions;
 - (c) independently analyze and correct mistakes.
- 7. Life long learning.
 - (a) evolve with the needs, improve their intellectual level;
 - (b) apply the acquired knowledge for solving new multidisciplinary tasks.

3 Conclusion

Thus, the course "3D PRINTING TECHNOLOGY TRANSFER", developed by the authors, allows the students not only to master knowledge and practical skills in the field of 3D printing (scanning), but also to get acquainted with the basics of entrepreneurship and the principles of 3D technologies transfer. The competences, acquired by the students after getting acquainted with the course materials, will allow them to become highly qualified specialists demanded in a modern labour-market.

Course "3D PRINTING TECHNOLOGY TRANSFER" is effectively implemented not only in the curricula of bachelor's and master's degrees, but is also used for the preparation of post-graduate students, retraining and refresher courses. Thus, the introduction of a course developed by the authors in educational activities will enable higher education institution to develop "learning throughout the life" concept, using the competency approach. In addition, the reading of the "3D PRINTING TECHNOLOGY TRANSFER" course in DSTU and KSMA allows us to develop activities according to the knowledge triangle concept: "education - science innovation".

References

1. Kholodov, I.: 3D printing: past, present and a little about future and the realities of Russia in this area. http://www.ixbt.com/printer/3d/3d_common.shtml (2014). Accessed 12 Dec 2015 [in Russian]

2. History of 3D Printing: The Free Beginner's Guide. <http://3dprintingindustry.com/3d-printing-basics-free-beginners-guide/history/> (2014). Accessed 12 Dec 2015
3. McLellan, C.: The history of 3D printing: A timeline. <http://www.zdnet.com/article/the-history-of-3d-printing-a-timeline/> (2014)
4. History of 3D Printing. <http://3dprintingfromscratch.com/common/history-of-3d-printing/> (2015). Accessed 10 Nov 2015
5. History of 3D printing. <http://www.orgprint.com/wiki/3d-pechat/istorija-3d-pechat/> (2012). Accessed 10 Nov 2015 [in Russian]
6. Applications of 3D printing. <http://www.orgprint.com/wiki/3d-pechat/sfery-primenenija-3d-pechat/> (2013). Accessed 15 Dec 2015 [in Russian]
7. Zhitkova, V.: 3D printing - a new bubble? http://slon.ru/future/3d_pechat_ocherednoy_puzyr-913100.xhtml (2013). Accessed 15 Dec 2015 [in Russian]
8. World of 3D printers. The future has already begun. Part 3 (Ending). [http://www.foto-business.ru/mir-3d-printerov-budushee-uje-nachalos-chast-3-\(zaklyuchenie\).html](http://www.foto-business.ru/mir-3d-printerov-budushee-uje-nachalos-chast-3-(zaklyuchenie).html) (2015). Accessed 12 Dec 2015 [in Russian]
9. Yashyna, K. Sadovoy, O.: Strategy of Dniprodzerzhynsk State Technical University knowledge transfer unit Scientific and technical journal Information processes, technologies and system on the transport. 1, 121-127. (2014) [in Ukrainian]
10. Karpenko, O., Yashyna, K.: Knowledge transfer in higher educational institutions of Ukraine. Recent economic problems 2014: Materials of VIII International Conference. 132-134. (2014) [in Ukrainian]
11. Yashyna, K., Tarasenko S., Sadovoy O.: Dniprodzerzhynsk State Technical University knowledge transfer unit. Intelligent technology of system programming. IV Ukrainian scientific conference of young scientists and students. Collection of Scientific Works. 107. (2015) [in Ukrainian]
12. Karpenko, O., Yashyna, K.: Formation of the knowledge transfer system in Ukrainian higher education institutions. Actual problems of economy. 7 (169), 157-166. (2015) [in Ukrainian]

Calculation Methods of the Prognostication of the Computer Systems State under Different Level of Information Uncertainty

Igor P. Atamanyuk¹, Yuriy P. Kondratenko^{2,3}, Vyacheslav S. Shebanin¹

¹ Mykolaiv National Agrarian University, Commune of Paris str. 9,
54010 Mykolaiv, Ukraine

atamanyuk_igor@mail.ru, rector@mnau.edu.ua

² Petro Mohyla Black Sea State University,
68th Desantnykiv Str. 10, 54003 Mykolaiv, Ukraine

³ Cleveland State University,
2121 Euclid Av., 44115, Cleveland, Ohio, USA

y.kondratenko@csuohio.edu, y_kondrat2002@yahoo.com

Abstract. Calculation methods of the prognostication of the computer systems state under different volume of a priori information and accuracy of the measurement of controlled parameters (under absence and presence of measurement errors) are obtained in the work. Canonical expansions of random sequences of the indices characterizing the state of the investigated systems considered as basic features of the methods. Synthesized methods do not impose any significant limitations on the qualities of the sequence of the change of the forecast parameters (linearity, stationarity, Markov behavior, monotonicity, etc.) and allow to take into account the stochastic peculiarities of the process of functioning of the investigated objects as much as possible. Expressions of the determination of a mean-square extrapolation error are obtained for solving the prognostication problems specifically concerning the state of computer systems under different level of information uncertainty.

Keywords. calculation method, random sequence, canonical decomposition, prognostication of the state

Key Terms. computation, mathematical model

1 Mathematical statement of the problem of the prognostication of a technical condition

One of the most important problems that arises constantly in the process of the operation of computer systems and computerized control systems [1,2,3] is based on quite evident fact that any decision about the permission of the system operation (of the realization of a stated problem) is closely connected with the solving of the prognosis problem. For example, the forecast of the remaining functioning time is a rotating machinery prognosis, results of which can be used also for forecasting of the reliability of machinery components (and additional equipment) as well as for forecasting future operational conditions. This kind of prognosis is based on the output data of multi-sensor monitoring system and current results of data processing. The main goal of such prognosis deals with: (a) reducing downtime of the machinery and corresponding equipment; (b) optimizing spares quantity; (c) decreasing functioning cost; (d) increasing safety of the machinery maintenance. In [4] authors analysis known methods of rotating machinery prognosis and classify the approaches to three groups based on different models: (a) general reliability, (b) environmental conditions, (c) combining prognostication and reliability.

Special attention should be paid to prognostication of manufacturing and industrial systems [5]. The results of such prognosis can help to determine the most rational maintenance modes for long-time functioning of different computer-integrated technological complexes. Modelling the degradation mechanism and dynamical degradation monitoring of the most important components of computer-integrated technological complexes are the base for prognosis in [5] within an e-maintenance architecture.

Different forecasting methods can be used for short-term electric load prognosis [6,7] at the enterprises, plants, cities and regions. The surveys of the prognosis methods based on applying Kalman filter, state space models, linear regression, stochastic time series, various smoothing algorithms, and artificial intelligence methods as well as the analysis of their application for solving prognostication problems is presented in [6,7] with implementation to short-term electric load prognosis.

The uncertainties of functioning conditions, external environment, nonstationary parameters and working modes and problems with their mathematical formalization are the main obstacles in using efficient computer models for forecasting future behavior of complex technological systems. As example, in [8] authors consider a special Bayesian computer model for prognostication of the active hydrocarbon reservoir future functioning.

Last years, such powerful theoretical-applied tool as the theory of neuro-fuzzy systems has been introduced successfully for solving different prognostication tasks in engineering, medicine, investment policy, finance and other fields [9,10,11,12,13].

According to reliability of control systems, it is necessary to note that computers are the main components of embedded controllers (traditional, fuzzy, neuro, etc.). The main critical requirement deals with providing efficient functioning such systems and networks in normal and in failure modes, when any component fails. One of the efficient approach for design process of such control systems is based on the applying redundant-elements-design-method [14].

Two most important indexes can be taken in to account in solving forecasting tasks for the e-business systems: (a) insufficient speed of response and (b) preventing failure of the system. In [15] author consider predictive inputs for the designed prognosis system as intrinsic (component activity levels, system response time, etc.) and extrinsic (time, date, whether, etc.) variables.

The prognostication in computer networks is described in [16] for forecasting the level of computer virus spread based on two models of viral epidemiology (differential equation model and the discrete Markov model).

The problem of forecasting control is especially topical for computer systems which are used for the management of the objects that relate to the class of critical or dangerous and under the threat of accident objects (aircraft, sea mobile objects, nuclear power stations, chemical industry plants etc.) [17,18,19].

Computer systems are exploited in the conditions of continuous influence on the great number of external and internal perturbing factors, the influence on the object of which is random by the moment of origin, duration and intensity. And correspondingly the changes of the system state also turn out to be random and form a random sequence. Thereupon the extrapolation of the realization of the random sequence describing the functioning of the investigated system on a certain interval of time is the mathematical content of the problem of the prognostication of a technical condition.

The most general extrapolation form for the solving of the problem of non-linear extrapolation is a Kolmogorov-Gabor polynomial [20] but it is very difficult and laborious procedure to find its parameters for the great number of known values and used order of non-linear relation. Thereupon during the forming of realizable in practice algorithms of the prognosis different simplifications and restrictions on the qualities of random sequence are used. For example, a range of suboptimal methods of non-linear extrapolation with a limited order of stochastic relation on the basis of approximation of a posteriori density of probabilities of an estimable vector by an orthogonal expansion by Hermite polynomials or in the form of Edgeworth series was offered by V.S. Pugachev [21]. Solution of non-stationary equation of A.N. Kolmogorov [20] (particular case of differential equation of R. L. Startanovich for the description of Markovian process) is obtained provided that the drift coefficient is linear function of state and coefficient of diffusion is equal to constant. Exhaustive solution of the problem of optimal linear extrapolation for different classes of random sequences and different level of informational support of the problem of prognosis (A.N. Kolmogorov equation [22] for stationary random sequences measured without errors; Kalman method [23] for markovian noisy random sequences; Wiener-Hopf filter-extrapolator [24] for noisy stationary sequences; algorithms of optimal linear extrapolation of V. D. Kudritsky [25] on the basis of canonical expansion of V. S. Pugachev etc.) exists. But maximal accuracy of the prognosis with the help of the methods of linear extrapolation can be achieved only for Gaussian random sequences.

Thus the development of the new methods of the prognostication of computer systems state which allow to take into account the information about the investigated object as much as possible is a topical problem.

Let us assume without restricting the generality that the state of a computer system is determined in exhaustive way by scalar parameter X the change of the values of which in discrete range of points $t_i, i = \overline{1, I}$ is described by the discrete sequence

$\{X\} = X(i), i = \overline{1, I}$. It is necessary to get optimal estimations of future values of a random sequence under different volume of a priori and a posteriori information.

2 Prognostication under the absence of the errors of measurement

The most universal from the point of view of the limitations that are imposed on the investigated sequence is the method on the basis of canonical model [26]:

$$X(i) = M[X(i)] + \sum_{\nu=1}^i \sum_{\lambda=1}^N W_{\nu}^{(\lambda)} \beta_{\lambda\nu}^{(\lambda)}(i), \quad i = \overline{1, I}, \quad (1)$$

where elements $W_{\nu}^{(\lambda)}, \beta_{\lambda\nu}^{(\lambda)}(i)$ are determined by recurrent correlations:

$$W_{\nu}^{(\lambda)} = X^{\lambda}(\nu) - M[X^{\lambda}(\nu)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N W_{\mu}^{(j)} \beta_{\lambda\mu}^{(j)}(\nu) - \sum_{j=1}^{\lambda-1} W_{\nu}^{(j)} \beta_{\lambda\nu}^{(j)}(\nu), \quad \nu = \overline{1, I}; \quad (2)$$

$$\beta_{\lambda\nu}^{(\lambda)}(i) = \frac{M[W_{\nu}^{(\lambda)}(X^h(i) - M[X^h(i)])]}{M[\{W_{\nu}^{(\lambda)}\}^2]} = \frac{1}{D_{\lambda}(\nu)} \{M[X^{\lambda}(\nu)X^h(i)] -$$

$$-M[X^{\lambda}(\nu)]M[X^h(i)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N D_j(\mu) \beta_{\lambda\mu}^{(j)}(\nu) \beta_{h\mu}^{(j)}(i) -$$

$$- \sum_{j=1}^{\lambda-1} D_j(\nu) \beta_{\lambda\nu}^{(j)}(\nu) \beta_{h\nu}^{(j)}(i)\}, \quad \lambda = \overline{1, h}, \quad \nu = \overline{1, i}, \quad h = \overline{1, N}, \quad i = \overline{1, I}. \quad (3)$$

$$D_{\lambda}(\nu) = M[\{W_{\nu}^{(\lambda)}\}^2] = M[X^{2\lambda}(\nu)] - M^2[X^{\lambda}(\nu)] -$$

$$- \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N D_j(\mu) \{\beta_{\lambda\mu}^{(j)}(\nu)\}^2 - \sum_{j=1}^{\lambda-1} D_j(\nu) \{\beta_{\lambda\nu}^{(j)}(\nu)\}^2, \quad \lambda = \overline{1, N}, \nu = \overline{1, I}; \quad (4)$$

The method of extrapolation on the basis of mathematical model (1) has two forms of notation [27,28,29]

$$m_x^{(\mu, l)}(h, i) = \begin{cases} M[X^h(i)] & \text{when } \mu = 0; \\ m_x^{(\mu, l-1)}(h, i) + (x^l(\mu) - m_x^{(\mu, l-1)}(l, \mu)) \beta_{h\mu}^{(l)}(i) & \text{when } l \neq 1, \\ m_x^{(\mu-1, N)}(h, i) + (x^l(\mu) - m_x^{(\mu-1, N)}(l, \mu)) \beta_{h\mu}^{(1)}(i) & \text{when } l = 1. \end{cases} \quad (5)$$

or

$$m_x^{(k,N)}(1,i) = M[X(i)] + \sum_{j=1}^k \sum_{v=1}^N (x^v(j) - M[X^v(j)]) S_{((j-1)N+v)}^{(kN)}((i-1)N+1), \quad (6)$$

where

$$S_{\lambda}^{(\alpha)}(\xi) = \begin{cases} S_{\lambda}^{(\alpha-1)}(\xi) - S_{\lambda}^{(\alpha-1)}(\alpha) \gamma_k(i), & \text{if } \lambda \leq \alpha-1; \\ \gamma_{\alpha}(\xi), & \text{for } \lambda = \alpha; \end{cases} \quad (7)$$

$$\gamma_{\alpha}(\xi) = \begin{cases} \beta_{1, [\alpha/N]+1}^{(\text{mod}_N(\alpha))}([\alpha/N]+1), & \text{for } \xi \leq kN; \\ \beta_{1, [\alpha/N]+1}^{(\text{mod}_N(\alpha))}(i), & \text{if } \xi = (i-1)N+1. \end{cases} \quad (8)$$

Mean-square error of extrapolation is determined as

$$M[X(i/x^v(j), \nu=1, \overline{N-1}, j=\overline{1, k}) - m_x^{(k,N-1)}(1,i)] = M[X^2(\nu)] - \\ - M^2[X^{\lambda}(\nu)] - \sum_{j=1}^k \sum_{v=1}^{N-1} M[(W_j^{(\nu)})^2] (\beta_{1,j}^{(\nu)}(i))^2, i = \overline{k+1, I}. \quad (9)$$

Expression $m_x^{(\mu,l)}(h,i) = M[X^h(i)/x^v(j), j = \overline{1, \mu-1}, \nu = \overline{1, N}; x^v(\mu), \nu = \overline{1, l}]$ for $h=1, l=N, \mu=k$ is optimal estimation $m_x^{(k,N)}(1,i)$ of the future value $x(i), i = \overline{k+1, I}$ provided that for the calculation of the given estimation values $x^v(j), \nu = \overline{1, N}, j = \overline{1, k}$ are used that is the results of the measurements of sequence $\{X\}$ in points $t_j, j = \overline{1, k}$ are known.

3 Prognostication on the basis of a priori information about the sequence of measurements with errors

Solution of the problem of prognosis (5),(6) presupposes the usage of true values of random sequence $\{X\}$ in the points of discretization $t_j, j = \overline{1, k}$. But in real situations the assumption about that that measured values $x(j), j = \overline{1, k}$ are known absolutely exactly is never carried out. The errors of the determination of the values of the forecast parameter can appear whether as a result of overlay of hindrances in the communication channel between measuring device and investigated object or as a result of influence of hindrances on the measuring tools.

Let us assume that as a result of measurements random sequence is observed

$$Z(i) = X(i) + Y(i), \quad i = \overline{1, I}, \quad (10)$$

where $Y(i), i = \overline{1, I}$, is a random error of measurement, $X(i), i = \overline{1, I}$, is unobserved component. It is necessary to obtain optimal (in mean-square sense) estimation of future values of random sequence $\{X\}$: $M[X^\lambda(\nu)X^h(i)], \lambda, h = \overline{1, N}, \nu, i = \overline{1, I}$ by the results of measurements $z(j), j = \overline{1, k}$.

Within the limits of such a statement the simplest nonoptimal solution of the problem presupposes the usage of algorithms (5),(6) substituting in it the results of measurements

$$m_{x/z}^{(\mu, l)}(h, i) = \begin{cases} M[X^h(i)], \mu = 0; \\ m_{x/z}^{(\mu, l-1)}(h, i) + (z^l(\mu) - m_{x/z}^{(\mu, l-1)}(l, \mu))\beta_{h\mu}^{(l)}(i), l \neq 1; \\ m_{x/z}^{(\mu-1, N)}(h, i) + (z^l(\mu) - m_{x/z}^{(\mu-1, N)}(l, \mu))\beta_{h\mu}^{(1)}(i), l = 1; \end{cases} \quad (11)$$

$$m_{x/z}^{(k, N)}(1, i) = M[X(i)] + \sum_{j=1}^k \sum_{\nu=1}^N (z^\nu(j) - M[Z^\nu(j)]) S_{((j-1)N+\nu)}^{(kN)}((i-1)N+1). \quad (12)$$

Conditional mathematical expectation remains as before unbiased estimation of future values of true extrapolated realization. At the same time the error of a single extrapolation will be written down as:

$$\Delta_{x/z}^{(k)}(i) = m_{x/z}^{(k)}(1, i) - x^{(k)}(i), \quad i = \overline{k+1, I},$$

where $x^{(k)}(i), i = \overline{k+1, I}$ is a true value of extrapolated realization in the area of forecast. These values aren't known actually and realization $x^{(k)}(i)$ is developing in a random way in the area of forecast. As a result of this the error of a single extrapolation acquires random character:

$$\delta_{x/z}^{(k)}(i) = m_{x/z}^{(k)}(1, i) - m_x^{(k)}(i) - \sum_{\nu=k+1}^i \sum_{\lambda=1}^N W_\nu^{(\lambda)} \beta_{1\nu}^{(\lambda)}(i) \quad (13)$$

The application of the operation of mathematical expectation to the last expression

$$\begin{aligned} S_{x/z}^{(k)}[i/z^\nu(j), \nu = \overline{1, N}, j = \overline{1, k}] &= m_{x/z}^{(k, N)}(1, i) - m_x^{(k, N)}(1, i) = \\ &= \sum_{j=1}^k \sum_{\nu=1}^N (z^\nu(j) - M[Z^\nu(j)]) S_{((j-1)N+\nu)}^{(kN)}((i-1)N+1) - \end{aligned} \quad (14)$$

$$\begin{aligned} & -\sum_{j=1}^k \sum_{v=1}^N \left(x^v(j) - M[X^v(j)] \right) S_{((j-1)N+v)}^{(kN)} ((i-1)N+1) = \\ & = \sum_{j=1}^k \sum_{v=1}^N y^v(j) S_{((j-1)N+v)}^{(kN)} ((i-1)N+1), \quad i = \overline{k+1, I}. \end{aligned}$$

shows that in the given case (as distinct from an ideal case) a single extrapolation is accompanied by conditional systematic error.

Correspondingly the dispersion of the error of a single extrapolation from (13), (14) is determined as

$$M \left[\left(\delta_{x/z}^{(k)}(i) - S_{x/z}^{(k)}(i) \right)^2 \right] = \sum_{j=k+1}^i \sum_{v=1}^{N-1} D_v(j) \left\{ \beta_{1j}^{(v)}(i) \right\}^2, \quad i = \overline{k+1, I}. \quad (15)$$

With the usage of (13), (14) mean-square error of a single extrapolation will be written down in the form

$$E_{x/z}^{(k)}(i/z(\mu), \mu = \overline{1, k}) = \left\{ S_{x/z}^{(k)}(i) \right\}^2 + D_x^{(k)}(i), \quad i = \overline{k+1, I}. \quad (16)$$

As error (16) is conditional averaging (16) by condition that values $z(\mu)$, $\mu = \overline{1, k}$ are random is necessary for complete characteristic of the accuracy of algorithm (11), (12). As a result the expression for mean-square error of prognosis is in the form

$$\begin{aligned} E_{x/z}^{(k)}(i) &= D_{x/z}^{(k, N)}(i) = \sum_{l=1}^k \sum_{\mu=1}^N \sum_{j=1}^k \sum_{v=1}^N M \left[Y^\mu(l) Y^v(j) \right] S_{((l-1)N+\mu)}^{(kN)} ((i-1)N+1) \times \\ & \times S_{((j-1)N+v)}^{(kN)} ((i-1)N+1) + \sum_{j=k+1}^i \sum_{v=1}^{N-1} D_v(j) \left\{ \beta_{1j}^{(v)}(i) \right\}^2, \quad i = \overline{k+1, I}. \end{aligned} \quad (17)$$

4 Prognostication with preliminary filtration of the errors of measurements

Increase of the quality of extrapolation of random sequence $\{X\}$, measured with noises is possible at the expense of transition from the results of measurement $z(\mu)$, $\mu = \overline{1, k}$, $k < I$ to estimation.

$$x^*(\mu) = M[X(\mu)] + (1 - F^{(\mu)}) m_x^{(\mu-1, N)}(1, \mu) + F^{(\mu)} \overset{o}{z}(\mu), \quad \mu = \overline{1, k}. \quad (18)$$

Unbiased estimation of unknown value $x(\mu)$ being studied as a balanced mean value of the result of the forecast at μ -th step $m_x^{(k,N-1)}(1, \mu)$ and result μ - of that measurement $z(\mu)$.

By means of consecutive substitution with the application of estimation (18) the algorithm of extrapolation (5) is brought to the form [30]:

$$m_x^{(\mu,l)}(h,i) = \begin{cases} M[X^h(i)], \mu = 0; \\ m_x^{(\mu,l-1)}(h,i) + F^{(\mu)}(z^l(\mu) - m_x^{(\mu,l-1)}(l, \mu))\beta_{h\mu}^{(l)}(i), l \neq 1; \\ m_x^{(\mu-1,N)}(h,i) + F^{(\mu)}(z^l(\mu) - m_x^{(\mu-1,N)}(l, \mu))\beta_{h\mu}^{(1)}(i), l = 1. \end{cases} \quad (19)$$

Algorithm (19) has equivalent form of notation as following

$$m_x^{(k,N)}(1,i) = M[X(i)] + \sum_{j=1}^k \sum_{v=1}^N \left(\overset{o}{z}(j) \right)^v G_{((j-1)N+v)}^{(kN)}((i-1)N+1), \quad (20)$$

$$G_\lambda^{(\alpha)}(\xi) = \begin{cases} G_\lambda^{(\alpha-1)}(\xi) - G_\lambda^{(\alpha-1)}(\alpha)\gamma_k(i), \lambda \leq \alpha-1; \\ \gamma_\alpha(\xi), \lambda = \alpha; \end{cases} \quad (21)$$

$$\gamma_\alpha(\xi) = \begin{cases} F^{([\alpha/N]+1)}\beta_{1, [\alpha/N]+1}^{(\text{mod}_{N-1}(\alpha))}([\alpha/N]+1), \text{ for } \xi \leq kN; \\ F^{([\alpha/N]+1)}\beta_{1, [\alpha/N]+1}^{(\text{mod}_N(\alpha))}(i), \text{ if } \xi = (i-1)N+1. \end{cases} \quad (22)$$

Optimal values of weight coefficients are determined from the condition of minimum of mean-square error of filtration

$$E_f(k) = M\left[|X^*(k) - X(k)|^2\right] = M\left[\left|(1 - F^{(k)})\sum_{j=1}^k \sum_{v=1}^N \left(\overset{o}{Z}(j)\right)^v \times \right.\right. \\ \left.\left. \times G_{(j-1)(N+v)}^{(kN)}((k-1)N+1) + F^{(k)}\overset{o}{Z}(k) - \overset{o}{X}(k)\right|^2\right]. \quad (23)$$

After differentiation of this expression on $F^{(k)}$ and solution of the corresponding equation the expression for calculation of the optimal value of the coefficient is obtained

$$F^{(k)} = \frac{F_1^{(k)} + F_2^{(k)} - F_3^{(k)}}{F_1^{(k)} + F_2^{(k)} - 2F_3^{(k)} + D_y(k)}, \quad (24)$$

$$\begin{aligned}
 F_1^{(k)} &= D_x(k) - 2 \sum_{j=1}^{k-1} \sum_{v=1}^N M \left[\left(\overset{o}{X}(j) \right)^v \overset{o}{X}(k) \right] G_{((j-1)N+v)}^{((k-1)N)} ((k-1)N+1) + \\
 &+ \sum_{j=1}^{k-1} \sum_{v=1}^N \sum_{l=1}^{k-1} \sum_{\mu=1}^N M \left[\left(\overset{o}{X}(j) \right)^v \left(\overset{o}{X}(l) \right)^\mu \right] G_{((j-1)N+v)}^{((k-1)N)} ((k-1)N+1) \times \\
 &\quad \times G_{((l-1)N+\mu)}^{((k-1)N)} ((k-1)N+1); \\
 F_2^{(k)} &= \sum_{j=1}^{k-1} \sum_{v=1}^N \sum_{l=1}^{k-1} \sum_{\mu=1}^N M \left[Y^v(j) Y^\mu(l) \right] G_{((j-1)N+v)}^{((k-1)N)} ((k-1)N+1) G_{((l-1)N+\mu)}^{((k-1)N)} ((k-1)N+1); \\
 &+ \sum_{j=1}^{k-1} \sum_{v=1}^N \sum_{l=1}^{k-1} \sum_{\mu=1}^N M \left[\left(\overset{o}{X}(j) \right)^v \left(\overset{o}{X}(l) \right)^\mu \right] G_{((j-1)N+v)}^{((k-1)N)} ((k-1)N+1) \times \\
 &\quad \times G_{((l-1)N+\mu)}^{((k-1)N)} ((k-1)N+1); \\
 F_3^{(k)} &= \sum_{j=1}^{k-1} \sum_{v=1}^N M \left[Y^v(j) Y(k) \right] G_{((j-1)N+v)}^{((k-1)N)} ((k-1)N+1).
 \end{aligned}$$

Each element of the formula (24) has evident physical sense. Specifically summand $F_1^{(k)}$ determines contribution to resultant error made by stochastic nature of random sequence $\{X\}$, summands $F_2^{(k)}$ and $F_3^{(k)}$ are connected with the errors of past measurements and summand $D_y(k)$ is dispersion of the last measurement. Algorithm (19),(20) got on the basis of function $M[X^\lambda(\nu)X^h(i)], \lambda, h = \overline{1, N}, \nu, i = \overline{1, I}$ and results of measurements $z(j), j = \overline{1, k}$ provides minimum of mean-square error of the prognosis for the given volume of known information about investigated random sequence as two interconnected consecutive stages (filtration-extrapolation) are fulfilled in optimal way: weight coefficients of estimation (18) are determined from the condition of the minimum of mean-square error of approximation to true values and parameters of extrapolator on the stage of preliminary filtration and further forecast are optimal which was proved earlier in the theorem.

Mean-square error of extrapolation with the use of the algorithm of polynomial filtration (19),(20) is determined as

$$\begin{aligned}
 E_x^{(k)}(i) &= \sum_{j=1}^k \sum_{v=1}^N \sum_{l=1}^k \sum_{\mu=1}^N M \left[\left(\overset{o}{X}(j) \right)^v \left(\overset{o}{X}(l) \right)^\mu \right] \left(G_{((j-1)N+v)}^{(kN)} ((i-1)N+1) - \right. \\
 &\quad \left. - S_{((j-1)N+v)}^{(kN)} ((i-1)N+1) \right) \times \left(G_{((l-1)N+\mu)}^{(kN)} ((i-1)N+1) - S_{((l-1)N+\mu)}^{(kN)} ((i-1)N+1) \right) - \quad (25) \\
 &+ \sum_{j=1}^k \sum_{v=1}^N \sum_{l=1}^k \sum_{\mu=1}^N M \left[Y^v(j) Y^\mu(l) \right] G_{((j-1)N+v)}^{(kN)} ((i-1)N+1) G_{((l-1)N+\mu)}^{(kN)} ((i-1)N+1).
 \end{aligned}$$

$$+D_x^{(k)}(i), i = \overline{k+1, I},$$

5 Prognostication on the basis of complete a priori information about the sequence measured with errors

Application of the operation of filtration in algorithm (19),(20) allows to decrease mean-square error of extrapolation compared with (11),(12) as the estimation $x^*(\mu)$, $\mu=\overline{1, k}$ has better accuracy characteristics compared with $z(\mu)$, $\mu=\overline{1, k}$. But in algorithm (19),(20) as well as in (11),(12) there is a mismatch between stochastic qualities of a posteriori information $x^*(\mu)$, $\mu=\overline{1, k}$ and parameters of extrapolation form (5),(6) on the basis of which the method under study is formed.

For the forming of the method of the prognosis by noisy measurements let's introduce into consideration the mixed random sequence $\{X'\} = \{Z(1), Z(2), \dots, Z(k), X(k+1), \dots, X(I)\}$ combining in itself the results of measurements till $i = k$, as well as the data about the sequence $\{X\}$ for $i = \overline{k+1, I}$.

The canonical expansion for such a sequence is of the form

$$X'(i) = M[X'(i)] + \sum_{\nu=1}^i \sum_{\lambda=1}^N U_{\nu}^{(\lambda)} \gamma_{1\nu}^{(\lambda)}(i), i = \overline{1, I}. \quad (26)$$

Random coefficients of the canonical decomposition (26) defined by the following recurrence formulas:

- for observation interval $[t_1, \dots, t_k]$

$$U_{\nu}^{(\lambda)} = Z^{\lambda}(\nu) - M[Z^{\lambda}(\nu)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N U_{\mu}^{(j)} \gamma_{\lambda\mu}^{(j)}(\nu) - \sum_{j=1}^{\lambda-1} U_{\nu}^{(j)} \gamma_{\lambda\nu}^{(j)}(\nu), \nu = \overline{1, k}; \quad (27)$$

- for forecasting interval $[t_{k+1}, \dots, t_I]$

$$U_{\nu}^{(\lambda)} = X^{\lambda}(\nu) - M[X^{\lambda}(\nu)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N U_{\mu}^{(j)} \gamma_{\lambda\mu}^{(j)}(\nu) - \sum_{j=1}^{\lambda-1} U_{\nu}^{(j)} \gamma_{\lambda\nu}^{(j)}(\nu), \nu = \overline{k+1, I}. \quad (28)$$

Accordingly, the expression for the dispersion of the random coefficients $U_{\nu}^{(\lambda)}$, $\lambda=\overline{1, N}$, $\nu = \overline{1, I}$ are of the form:

- for observation interval $[t_1, \dots, t_k]$

$$D_{\lambda}(\nu) = M[(U_{\nu}^{(\lambda)})^2] = M[Z^{2\lambda}(\nu)] - M^2[Z^{\lambda}(\nu)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N D_j(\mu) (\gamma_{\lambda\mu}^{(j)}(\nu))^2 - \sum_{j=1}^{\lambda-1} D_j(\nu) (\gamma_{\lambda\nu}^{(j)}(\nu))^2, \nu = \overline{1, k}; \quad (29)$$

- for forecasting interval $[t_{k+1}, \dots, t_I]$

$$D_{\lambda}(\nu) = M \left[\left(U_{\nu}^{(\lambda)} \right)^2 \right] = M \left[X^{2\lambda}(\nu) \right] - M^2 \left[X^{\lambda}(\nu) \right] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N D_j(\mu) \left(\gamma_{\lambda\mu}^{(j)}(\nu) \right)^2 - \sum_{j=1}^{\lambda-1} D_j(\nu) \left(\gamma_{\lambda\nu}^{(j)}(\nu) \right)^2, \quad \nu = \overline{k+1, I}. \quad (30)$$

The coordinate functions $\gamma_{h\nu}^{(\lambda)}(i)$ are calculated using the formulas:

- for observation interval $[t_1, \dots, t_k]$ (function $\gamma_{h\nu}^{(\lambda)}(i)$ describes the stochastic relationship between the variables $Z^{\lambda}(\nu)$ and $Z^h(i)$)

$$\gamma_{h\nu}^{(\lambda)}(i) = \frac{1}{D_{\lambda}(\nu)} \left\{ M \left[Z^{\lambda}(\nu) Z^h(i) \right] - M \left[Z^{\lambda}(\nu) \right] M \left[Z^h(i) \right] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N D_j(\mu) \gamma_{\lambda\mu}^{(j)}(\nu) \gamma_{h\mu}^{(j)}(i) - \sum_{j=1}^{\lambda-1} D_j(\nu) \gamma_{\lambda\nu}^{(j)}(\nu) \gamma_{h\nu}^{(j)}(i) \right\}, \quad \lambda = \overline{1, h}, \quad 1 \leq \nu \leq i \leq l \quad (31)$$

- for description in the canonical decomposition of stochastic correlation between intervals $[t_1, \dots, t_k]$ and $[t_{k+1}, \dots, t_I]$ ($\gamma_{h\nu}^{(\lambda)}(i)$ describes the relationship between random variables $Z^{\lambda}(\nu)$ and $X^h(i)$)

$$\gamma_{h\nu}^{(\lambda)}(i) = \frac{1}{D_{\lambda}(\nu)} \left\{ M \left[Z^{\lambda}(\nu) X^h(i) \right] - M \left[Z^{\lambda}(\nu) \right] M \left[X^h(i) \right] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N D_j(\mu) \gamma_{\lambda\mu}^{(j)}(\nu) \gamma_{h\mu}^{(j)}(i) - \sum_{j=1}^{\lambda-1} D_j(\nu) \gamma_{\lambda\nu}^{(j)}(\nu) \gamma_{h\nu}^{(j)}(i) \right\}, \quad \nu = \overline{1, k}, \quad i = \overline{k+1, I}; \quad (32)$$

- for forecasting interval $[t_{k+1}, \dots, t_I]$ (function $\gamma_{h\nu}^{(\lambda)}(i)$ describes the stochastic relationship between the variables $X^{\lambda}(\nu)$ and $X^h(i)$)

$$\gamma_{h\nu}^{(\lambda)}(i) = \frac{1}{D_{\lambda}(\nu)} \left\{ M \left[X^{\lambda}(\nu) X^h(i) \right] - M \left[X^{\lambda}(\nu) \right] M \left[X^h(i) \right] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^N D_j(\mu) \gamma_{\lambda\mu}^{(j)}(\nu) \gamma_{h\mu}^{(j)}(i) - \sum_{j=1}^{\lambda-1} D_j(\nu) \gamma_{\lambda\nu}^{(j)}(\nu) \gamma_{h\nu}^{(j)}(i) \right\}, \quad k \leq \nu \leq i \leq I. \quad (33)$$

The necessity of the two expressions (28) and (29) for the determination of the random coefficients of the canonical expansion (27) is explained with the technology of the forming of random sequence $\{X'\} : \{X'\} = \{Z\}, t_i, i = \overline{1, k}$ and $\{X'\} = \{X\}, t_i, i = \overline{k+1, I}$. The stated peculiarity also results in the increasing compared to the expansion (1) of the number of formulae for the calculation of the dispersions of the random coefficients (29), (30) and coordinate functions (31), (32), (33).

In the canonical expansion (26) the random sequence $\{X'\}$ is presented in the investigated range of points $t_i, i=\overline{1, I}$ with the help of N arrays $\{U^{(\lambda)}\}, \lambda=\overline{1, N}$ of uncorrelated centered random coefficients $U_i^{(\lambda)}, i=\overline{1, I}$. The given coefficients contain information about the values $Z^\lambda(i), \lambda=\overline{1, N}, i=\overline{1, k}$ and $X^\lambda(i), \lambda=\overline{1, N}, i=\overline{k+1, I}$, and coordinate functions $\gamma_{hv}^{(\lambda)}(i), \lambda, h=\overline{1, N}, v, i=\overline{1, I}$ describe probabilistic connections of the order $\lambda+h$ between sections t_v and $t_i, v, i=\overline{1, I}$.

Let us assume that as a result of measurement in the first point of discretization t_1 value $z(1)$ becomes known (additive mixture of unobserved true value $x(1)$ and error $y(1)$). Measurement $z(1)$ concretizes random coefficients $U_1^{(\lambda)}, \lambda=\overline{1, N}$ for section t_1 :

$$u_1^{(\lambda)} = z^\lambda(1) - M[Z^\lambda(1)] - \sum_{j=1}^{\lambda-1} u_1^{(j)} \gamma_{\lambda 1}^{(j)}(1), \lambda=\overline{1, N}. \quad (34)$$

Substitution of values (34) in canonical expansion (26) and further application of the operation of mathematical expectation allow to write down the expression for the estimation of future values $x^h(i)$ with the use of a posteriori information $z^l(1), l=\overline{1, N}$ in the following form

$$m_{x/z}^{(1,l)}(h, i) = m_{x/z}^{(1,l-1)}(h, i) + (z^l(1) - m_{x/z}^{(1,l-1)}(l, 1)) \gamma_{h1}^{(l)}(i) \quad (35)$$

where $m_{x/z}^{(1,l)}(h, i)$ is optimal (in mean-square sense) estimation of value $x^h(i)$ provided that for the prognosis values $z^j(1), j=\overline{1, l}$ are used.

Measurement $z(2)$ leads to the fixation of random coefficients $u_2^{(\lambda)}, \lambda=\overline{1, N}$ for t_2 :

$$u_2^{(\lambda)} = z^\lambda(2) - M[Z^\lambda(2)] - \sum_{j=1}^N u_1^{(j)} \gamma_{\lambda 1}^{(j)}(2) - \sum_{j=1}^{\lambda-1} u_2^{(j)} \gamma_{\lambda 2}^{(j)}(2), \lambda=\overline{1, N}. \quad (36)$$

Use of the values of random coefficients (36) allows to obtain prognosis algorithm taking into consideration $z^l(1), z^l(2) l=\overline{1, N}$:

$$m_{x/z}^{(2,l)}(h, i) = \begin{cases} m_{x/z}^{(2,l-1)}(h, i) + (z^l(2) - m_{x/z}^{(2,l-1)}(l, 2)) \gamma_{h2}^{(l)}(i), & l \neq 1; \\ m_{x/z}^{(1,N)}(h, i) + (z(2) - m_{x/z}^{(1,N)}(1, 2)) \gamma_{h2}^{(1)}(i), & l = 1. \end{cases} \quad (37)$$

For random quantity of measurements $z(\mu)$, $\mu = \overline{1, I}$ the algorithm of optimal extrapolation takes on form:

$$m_{x/z}^{(\mu, l)}(h, i) = \begin{cases} M[X^h(i)], \mu = 0; \\ m_{x/z}^{(\mu, l-1)}(h, i) + (z^l(\mu) - m_{x/z}^{(\mu, l-1)}(l, \mu))\gamma_{h\mu}^{(l)}(i), l \neq 1; \\ m_{x/z}^{(\mu-1, N)}(h, i) + (z^l(\mu) - m_{x/z}^{(\mu-1, N)}(l, \mu))\gamma_{h\mu}^{(l)}(i), l = 1. \end{cases} \quad (38)$$

Expression $m_{x/z}^{(\mu, l)}(h, i) = M[X^h(i)/z^\nu(j), j = \overline{1, \mu-1}, \nu = \overline{1, N}; z^\nu(\mu)]$, for $h = 1, l = N, \mu = k$ is unbiased optimal estimation $m_{x/z}^{(k, N-1)}(1, i)$ of future value $x(i)$, $i = \overline{k+1, I}$ provided that for the calculation of given estimation values $z^\nu(j)$, $\nu = \overline{1, N}, j = \overline{1, k}$ are used that is the results of the measurements of sequence $\{X'\}$ in points $t_j, j = \overline{1, k}$ are known.

In Fig. 1 the diagram is presented that reflects peculiarities of functioning of the method of prognosis (38).

Mean-square error of extrapolation with the help of method (39) is determined by the expression:

$$M\left[\left\{X(i/z^\nu(j), \nu = \overline{1, N}, j = \overline{1, k}) - m_{x/z}^{(k, N)}(1, i)\right\}^2\right] = M[X^2(i)] - M^2[X(i)] - \sum_{j=1}^k \sum_{\nu=1}^N D_\nu(j) (\gamma_{1j}^{(\nu)}(i))^2, i = \overline{k+1, I}. \quad (39)$$

6 Conclusion

In nowadays, the prognostication of the current state of complex computer systems, especially, for the class of critical applications, is an important and actual problem for providing high functioning reliability of the various control objects and decision-making systems. The proposed approach, based on the mathematical formalization of the parametrical changes of computer systems using nonlinear canonic models of the random sequences, allows to take into account the stochastic properties of investigated computer systems. Exhaustive solutions of the prognostication problems are obtained by authors with the aim of the evaluation of the computer system state and analysis of its further operational capability in the situations with different volume of a priori and a posteriori information or with various levels of information uncertainty. Synthesized prognostication methods as well as assumed, as their basis canonical expansions do not impose any significant limitations on random sequences

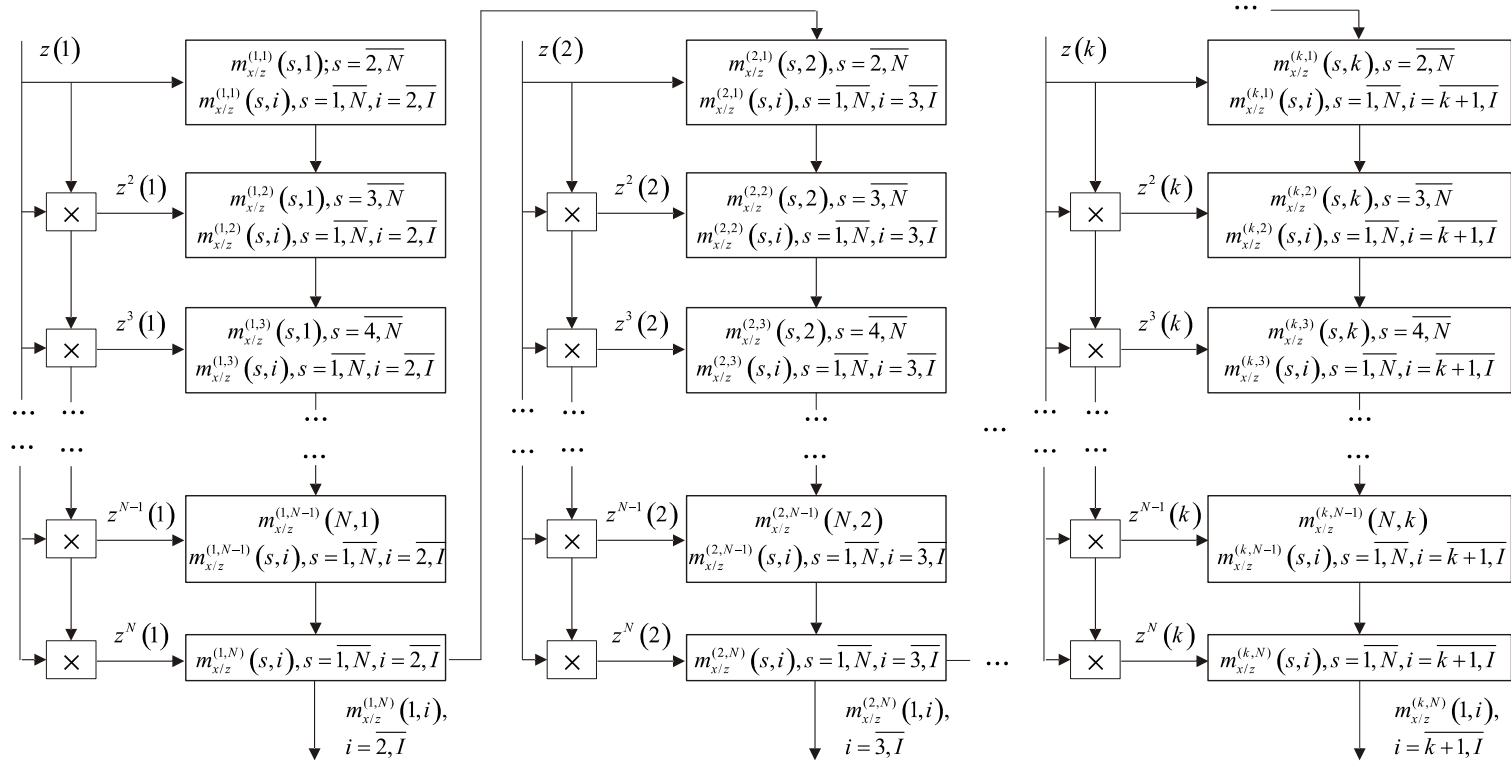


Fig. 1. Diagram of the prognosis of a noise random sequence with the help of extrapolator (38)

of the change of the values of controlled parameters including linearity, stationarity, Markov behavior, monotonicity, etc. Suggested mathematical expressions for the determination of mean-square error of extrapolation allow to make a decision about the choice of the most appropriate method from the totality of the introduced ones for the solution of the prognostication problem of computer system with prescribed accuracy. The specific diagram, presented in the paper, reflects the peculiarities of the synthesized prognostication methods. Proposed methods are fairly simple in computing aspects and may be applied for solving computer system prognostication tasks in real time taking into account that all parameters of the prognostication models can be defined previously.

References

1. Tkachenko, A.N., Brovinskaya, N.M., Kondratenko, Y.P.: Evolutionary adaptation of control processes in robots operating in non-stationary environments. *J. Mechanism and Machine Theory*. Printed in Great Britain, Vol. 18, No. 4, pp. 275-278 (1983)
2. Kondratenko, Y., Korobko, V., Korobko, O.: Distributed Computer System for Monitoring and Control of Thermoacoustic Processes. In: *Proceedings of the 7th IEEE international conference IDAACS'2013, Berlin, Germany, Vol. 1*, pp. 249-253 (2013)
3. Kondratenko, Y.P., Timchenko, V.L.: Increase in Navigation Safety by Developing Distributed Man-Machine Control Systems. In: *Proceedings of the Third Intern. Offshore and Polar Engineering Conference*. Singapore, Vol. 2, pp. 512-519 (1993)
4. Heng, A., Zhang, S., Tan, A.C., Mathew, J.: Rotating machinery prognostics: State of the art, challenges and opportunities. *J. Mechanical Systems and Signal Processing*, 23(3), pp.724-739 (2009)
5. Muller, A., Suhner, M.C., Iung, B.: Formalisation of a new prognosis model for supporting proactive maintenance implementation on industrial system. *J. Reliability Engineering & System Safety*, 93(2), pp. 234-253 (2008)
6. Metaxiotis, K., Kagiannas, A., Askounis, D., Psarras, J.: Artificial intelligence in short term electric load forecasting: a state-of-the-art survey for the researcher. *J. Energy Conversion and Management*, 44(9), pp.1525-1534 (2003)
7. Moghram, I., Rahman, S.: Analysis and evaluation of five short-term load forecasting techniques. *Power Systems, IEEE Transactions on*, 4(4), pp.1484-1491 (1989)
8. Craig, P.S., Goldstein, M., Rougier, J.C., Seheult, A.H.: Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96(454), pp.717-729 (2001)
9. Welstead, S.T.: *Neural Network and Fuzzy Logic Applications in C-C++*. John Wiley & Sons, Inc. (1994)
10. Kondratenko, Y.P., Sidenko, Ie.V.: Decision-Making Based on Fuzzy Estimation of Quality Level for Cargo Delivery. In book: *Recent Developments and New Directions in Soft Computing. Studies in Fuzziness and Soft Computing 317*, Zadeh, L.A. et al. (Eds), Springer International Publishing Switzerland, pp. 331-344 (2014)
11. Kondratenko, G.V., Kondratenko, Y.P., Romanov, D.O.: Fuzzy Models for Capacitive Vehicle Routing Problem in Uncertainty. In: *Proc. 17th International DAAAM Symposium "Intelligent Manufacturing and Automation: Focus on Mechatronics & Robotics"*, Vienna, Austria, pp. 205-206 (2006)
12. Kondratenko, Y., Gordienko, E.: Neural Networks for Adaptive Control System of Caterpillar Turn. In: *Proceeding of the 22th International DAAAM Symposium "Intelligent Manufacturing and Automation"*, Vienna, Austria, pp. 0305-0306 (2011)

13. Kondratenko, Y., Gordienko, E.: Implementation of the neural networks for adaptive control system on FPGA. In.: Proceeding of the 23th International DAAAM Symposium "Intelligent Manufacturing and Automation", Vol. 23, No.1, B. Katalinic (Ed.), Vienna, Austria, EU, pp. 0389-0392 (2011)
14. Shooman, M.L.: Reliability of computer systems and networks: fault tolerance, analysis, and design. John Wiley & Sons (2003)
15. Helsper, David, Clayton Wilkinson, Robert Zack, John T. Tatum, Robert J. Jannarone, Bernd Harzog.: Enhanced computer performance forecasting system. U.S. Patent 6,876,988, issued April 5, (2005)
16. Billings, L., Spears, W.M. and Schwartz, I.B.: A unified prediction of computer virus spread in connected networks. Physics Letters A, 297(3), pp.261-266 (2002)
17. Silva N., Vieira M.: Towards Making Safety-Critical Systems Safer: Learning from Mistakes. In: ISSREW, 2014 IEEE International Symposium on Software Reliability Engineering Workshops (ISSREW), pp. 162-167 (2014)
18. Irrera, I., Duraes J., Vieira M.: On the need for training Failure Prediction algorithms in evolving software systems. In: High-Assurance Systems Engineering (HASE), 2014 IEEE 15th International Symposium on. IEEE, pp. 216-223 (2014)
19. Kharchenko, V.S. (Ed.): Safety of Critical Infrastructures: Mathematical Analysis and Engineering Methods of Analysis and Ensuring, National Aerospace University named after N.E.Zhukovsky, KhAI, Kharkiv (2011)
20. Box, G.E.P., Jenkins, G.M.: Time-series Analysis, Forecasting and Control. Holden-Day, San Francisco (1970)
21. Pugachev, V.S.: The Theory of Random Functions and its Application. Fitmatgiz, Moscow (1962)
22. Kolmogorov, A.N.: Interpolation and Extrapolation of Stationary Random Sequences. J. Proceedings of Academy of Sciences of the USSR. Mathematical Series 5, pp. 3-14 (1941)
23. Kalman, R.E.: A New Approach to Linear Filtering and Prediction Problems. Trans. ASME Series D, J. Basic Engg., Vol. 82 (Series D), pp. 35-45 (1960)
24. Wiener, N.: Extrapolation, Interpolation and Smoothing of Stationary Time Series: With Engineering Applications. MIT Press, New-York (1949)
25. Kudritsky V.D.: Filtering, Extrapolation and Recognition Realizations of Random Functions. FADA Ltd., Kyiv, (2001)
28. Atamanyuk I., Kondratenko Y.: Calculation Method for a Computer's Diagnostics of Cardiovascular Diseases Based on Canonical Decompositions of Random Sequences. ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proceedings of the 11 International Conference ICTERI-2015, S.Batsakis, et al. (Eds.), CEUR-WS, Vol. 1356, pp. 108-120 (2015)
29. Atamanyuk, I.P., Kondratenko, V.Y., Kozlov, O.V., Kondratenko, Y.P.: The algorithm of optimal polynomial extrapolation of random processes, In: Modeling and Simulation in Engineering, Economics and Management, K.J.Engemann, A.M.Gil-Lafuente, J.L.Merigo (Eds.), International Conference MS 2012, New Rochelle, NY, USA, Proceedings. Lecture Notes in Business Information Processing, Volume 115, Springer, pp. 78-87 (2012)
30. Atamanyuk, I. Kondratenko Y.: Computer's Analysis Method and Reliability Assessment of Fault-Tolerance Operation of Information Systems. In: Proceedings of the 11th International Conference ICTERI-2015, S.Batsakis, et al. (Eds.), CEUR-WS, Vol. 1356, pp. 507-522 (2015)
31. Atamanyuk I.P., Kondratenko Y.P.: The method of forecasting technical condition of objects. Patent 73885, Ukraine, G05B 23/02, Bul. № 19 (2012)
32. Atamanyuk, I.P. Optimal Polynomial Extrapolation of Realization of a Random Process with a Filtration of Measurement Errors. J. Automation and Information Sciences. Volume 41, Issue 8, Begell House, USA, pp. 38-48 (2009)

Classification and Research of the Reactor Protection Instrumentation and Control System Functional Safety Markov Models in a Normal Operation Mode

Yevgeniy Bulba^{1,3}, Yuriy Ponochohny², Vladimir Sklyar¹, Aleksandr Ivasiuk³

¹ National Aerospace University KhAI, Kharkiv, Ukraine

evhenb@gmail.com, vvsklyar@ukr.net

² Poltava National Technical University named after Yuriy Kondratyuk, Poltava, Ukraine

pnchl@rambler.ru

³ RPC Radiy, Kirovograd, Ukraine

ivasiuk.radiks@gmail.com

Abstract. This article presents the basic phases of the development and research instrumentation and control system (ICS) functional safety Markov models on the basis of the self-diagnosing programmable platform. The set of the models sets is obtained on the basis of the failure tree development and analysis, which includes the ICS hardware channels detected and undetected failures. There the classification of ICS in a normal operation mode, considering the different majority body modes and diagnostics levels is presented. The application of the models has allowed determining the boundaries of the ICS safety integrity level 3 (SIL3) in two-dimensional space of the input parameters and system operational time.

Keywords: instrumentation and control system, functional safety, Markov model, safety integrity level

Key terms. MathematicalModeling, MathematicalModel, SoftwareSystems

1 Introduction

Instrumentation and control systems (ICS) of the critical objects, which perform the safety-relevant functions, are estimated from the functional safety positions. The functional safety depends on the right operation of the electric, electronic and programmable electronic (E/E/PE) systems, integrated with the technological systems

and equipment safety to reduce the external risk [1]. Functional safety analysis principles are interpreted in [2].

The estimation of the functional safety is a definition of risk level in the field of safety. Its value is the composition of probability of dangerous situations on production and gravity of all consequences which can arise during operation. The estimation of the functional safety for the reactor protection systems (RPS) takes the specific place.

The functional safety estimation models are considered in 6 parts of the IEC-61508 [3] standard in details. This document presents the examples of models: reliability block diagrams, failure trees, Markov and multiphase, Petri nets and Monte-Carlo, the formal languages models. Also, as noted in this standard, the given models are examples for creation of models of real systems only. So, in the papers [4, 5] the models of the functional safety of control systems of nuclear reactors and sensor systems of protection taking into account their constraints and operating conditions are analyzed.

The purpose of this article is the classification and creation of ICS RPS Markov safety models in the mode of normal operation and influence of input parameters of model on a measure value of the functional safety is probed.

The article consists of 7 sections. The actual part is an introduction. The second part represents the analysis of the abnormal protection systems operating conditions in the mode of normal operation. The flowchart of reliability and a complete tree of system failures are constructed on the basis of the carried-out analysis. The third section represents the justifications of the main assumptions allowing using of Markov simulation. The fourth section gives us the signs of classification of Markov models of the functional safety of ICS RPS and also the 6 main models considering the hardware failures are selected here. The fifth section presents the Markov graphs and the description of three models: MSaf1, MSaf2 and MSaf6. The sixth section gives the justification of models input parameters values and the ranges of their change.

The last section presents the simulation results of Markov's models.

2 The analysis of reactor protection system operating conditions in the mode of normal operation

The analysis of the abnormal protection system functional safety is mandatory in case of design of the unit. The RPS of the reactor is one of the most important security arrangements and safety of reactor installation in general in many respects depends on its reliability.

The Abnormal Protection Systems can be realized on the basis of platforms with using of the Field Programmable Gates Arrays (FPGA). The main attention in such platforms shall be paid to self-diagnosing for determination of dangerous and safe system failures.

We will consider the operation of ICS which is part of RPS in the mode of normal operation. Normal operation (normal operation) is understood as operation in the set operational limits and conditions. ICS turns on three independent hardware channels,

each of which is diagnosed on existence of dangerous failures by the diagnostic system. The considered system functions in the mode with low frequency of requests to safety features. Respectively, for an assessment of the functional safety it is necessary to use Average Probability of Failures on Demand (PFD_{avg}) for each of the safety function.

The reliability block diagram of ICS RPS considering the majority voting component and diagnostic system is shown on Fig.1. This work presents the one-version part of ICS which operates in the mode of normal operation is considered (in emergency case two such ICS parts are involved, and each has the separate software version).

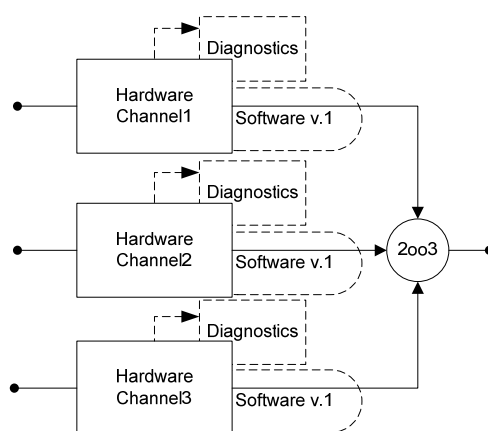


Fig. 1. The reliability block diagram of ICS RPS in the normal operation mode

The analysis of failures is made by the methods of collection and research of information on system failures in general, or elements of system. The majority of methods are based on carrying out inquiries of experts, application of numerical methods, the pilot studies, methods of probability theory and mathematical statistics [6].

The RPS failure tree creation and, as a result, Markov model of the system states can be result of such analysis.

Fig.2 shows the ICS tree of failures and at the same time the graphic notation (+, −, ×, *, #) for display of corresponding states is used:

- (+) – up state of hardware channel;
- (*) – up state of software;
- (−) – reveal of the dangerous failure in hardware channel revealed by the diagnostic system (the detected dangerous failure);
- (×) – reveal of a dangerous failure in hardware channel, undetected by the diagnostic system (the undetected dangerous failure);
- (#) – reveal of a software failure, a software failure is considered as a failure for the general reason.

Note: as in this system repair is made at once after reveal of an explicit failure, the repair state isn't considered, and return to run state with a repairing rate μ is modeled.

For display of change of states (transitions in between) the following notation is used:

- the solid line shows the transitions which will happen in system aren't dependent on operation modes of its elements;
- the dotted line shows the possible transitions which existence depends on a specific operation mode of elements of system.

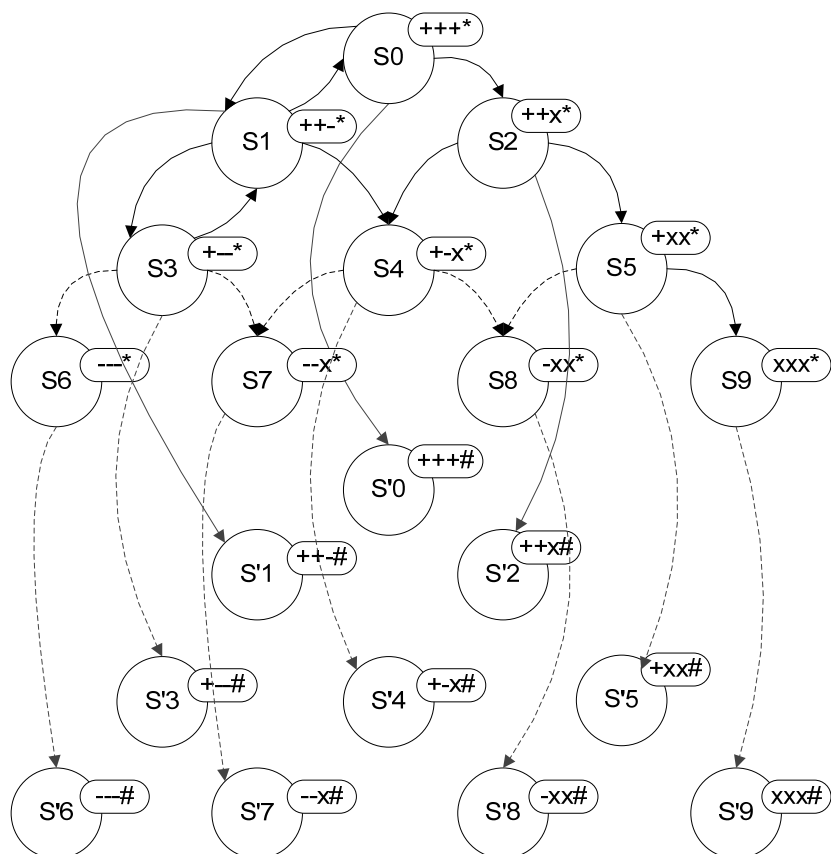


Fig. 2. The general failure tree of ICS RPS in the normal operation mode

Not all states, which are shown on the Fig.2, are admissible for the specific mode of functioning of the ICS elements and restrictions accepted in case of creation of model of estimation of the functional safety.

3 Main assumptions of instrumentation and control systems markov models

The ICS is characterized by the parameter of diagnostic coverage (DC). Unlike the models provided in [3,5] in the considered the diagnostic system is executed continuously (not periodically) and the revealed failures are eliminated immediately after detection. The remaining assumptions in case of model creation are the following:

- the failure and repair events of the hardware channels make the elementary flows (stationary, ordinary and without aftereffect), with constant parameters λ (failure rate) and μ (repair rate);
- the identical hardware channels with identical failures rate are used in the system;
- the majority device and the diagnostic system failure rate are scornfully small;
- in this model only the dangerous ICS hardware channels failures, those failure rate is calculated as $\lambda_D = 0.5 * \lambda$ [3] is considered;
- the common cause failure rate is scornfully small, so it is not considered in this model [2];
- when diagnosing, the part of the dangerous failures is revealed, consequently, the rate of the found dangerous failures is $\lambda_{DD} = \lambda_D * DC$, and the rate of undetected dangerous failures is $\lambda_{DU} = \lambda_D * (1 - DC)$;

The software failures are not within the scope of this work.

4 ICS RPS functional safety Markov models classification features

Work [3] presents the typical models of systems with diagnosing of dangerous failures and majority vote system. The programmed logic-based implementation of ICS RPS allows using the additional functions and operation modes of the diagnostic system and majority device.

The list of ICS function enhanced modes and the appropriate classification features are shown in Table 1.

Table 1. ICS RPS operation modes classification features

#	Classification feature	Feature meaning
1	Majority voting component response on the reveal of the hardware undetected failures	1.0 – single majority device (without response) 1.1 – majority device with the undetected failure channel switching-off function 1.2 – majority device with the function of the enhanced diagnostics in the case of the unidentified failure
2	Undetected failures countermeasure (warfare) methods	2.0 – without response 2.1 – with «migration» of the undetected failure in detected 2.2 – with the enhanced undetected failures diagnostics
3	Software failure response	3.1 – without software failure repair 3.2 – with software failure repair

From the 18 possible classification features combinations the 12 is permissible. The half of 12 permissible combinations has such feature as the software failures repair, which upsets the condition of markov property in the construction of models [7]. Therefore, the set of markov models according to the proposed classification features will be the following:

$M = \{MSaf1, MSaf2, MSaf3, MSaf4, MSaf5, MSaf6\}$,
where

$MSaf1 = \{1.0 \cup 2.0 \cup 3.1\}$ is a typical model with the single majority device, in which the failures, undetected by the diagnostic system, are collected;

$MSaf2 = \{1.0 \cup 2.1 \cup 3.1\}$ is a model of the system with the single majority device, in which the hardware channels do not switch off after the reveal of the undetected failure, and the reveal of the detected failure is possible in them (in this paper it is called as the «migration» of the undetected failure in detected);

$MSaf3 = \{1.0 \cup 2.2 \cup 3.1\}$ is a model of the system with the single majority device, in which the enhanced diagnostic of the failures, undetected by the diagnostic system is carried out periodically;

$MSaf4 = \{1.1 \cup 2.0 \cup 3.1\}$ is a model of the system with the majority device, which switches off the hardware channel, if there is a mismatch with other channels, which are regarded as intact by the diagnostic system;

$MSaf5 = \{1.1 \cup 2.2 \cup 3.1\}$ is a model of the system with the majority device, which switches off the hardware channel, if there is a mismatch with other channels, which are regarded as intact by the diagnostic system; also in the ICS the enhanced diagnostic of the failures, undetected by the diagnostic system is carried out periodically;

$MSaf6 = \{1.2 \cup 2.2 \cup 3.1\}$ – is a model of the system with the majority device, which initiates the hardware channels enhanced diagnostic in the case of the mismatch, and the diagnostic system acknowledges them as intent.

To lower the dimensionality this paper represents the approach to division on the models of hardware and software features, which is used in work [8]. The software models are not within the scope of this work, they are described in [7, 8]. Further, three ICS RPS function safety estimation Markov models, taking into account the hardware dangerous failures will be considered.

5 ICS RPS functional safety markov models

5.1 ICS RPS functional safety model with the single majority device without the responses on the diagnostic system non-identified failures (MSaf1)

The marked graph (digraph) of the model of ICS operation in the conditions of dangerous failures reveal is shown on Fig.3. This graph is constructed on the classical approach, which is described in [3] and contains the absorbing state S_8 with

undetected dangerous failures. The output from a state of an undetected dangerous failure without carrying out additional actions isn't provided in the classical model.

Proceeding from the functioning of the diagnostic system and majority device logic, contains:

a) Operational states (up states): S_0 (all channels are operational), S_1 (in one of channels the dangerous failure appears and is found) and S_3 (in one of channels the dangerous failure appears, but isn't found);

b) non-operation states (down states): S_2 (in two channels dangerous failures appear and are found), S_4 (in one of channels the dangerous failure appears and is found, in another - it appears, but isn't found) and S_5 (in two channels dangerous failures appear and are found, in the third channel it appears, but isn't found);

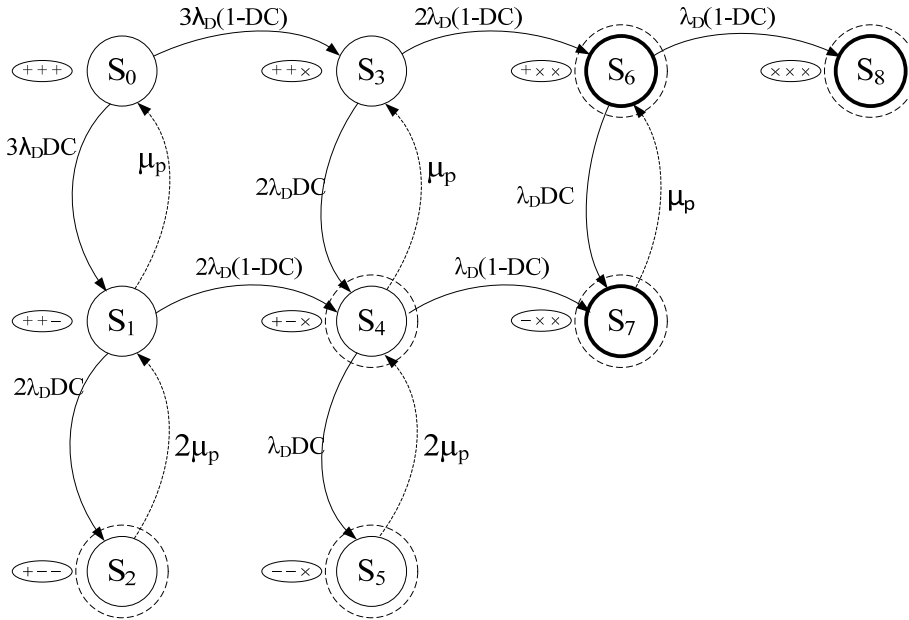


Fig. 3. The marked graph of the model of ICS RPS operation with the absorbing state (MSaf1)

c) states with undetected dangerous failures which majority device is incapable to parry an organ: S_6 (in two channels the dangerous failures appear but aren't found), S_7 (in one of channels the dangerous failure appears and is found, in two channels the dangerous failures appear, but aren't found), S_8 (in three channels the dangerous failures appear, but aren't found).

After detection of a dangerous failure reveal, the non-serviceable channel is disconnected and recovered with intensity μ_p , it is modeled by the appropriate transitions of $S_1 \rightarrow S_0$, $S_2 \rightarrow S_1$, $S_4 \rightarrow S_3$, $S_5 \rightarrow S_4$, $S_7 \rightarrow S_6$. The safety function PFD_{avg} is defined as:

$$PFD_{avg} = 1 - P_0(t) - P_1(t) - P_3(t). \quad (1)$$

5.2 ICS RPS functional safety model with the single majority device without the “migration” of the failures, undetected by the diagnostic system (MSaf2)

The practice of use of the considered systems [4] shows that the hardware channel with the shown undetected dangerous failures continues to be used. In the course of its use the reveal of other defects which can be detected by the diagnostic system is probable. Respectively, the ICS, after the reveal of an undetected dangerous failure and the subsequent reveal of new failures (the found dangerous failure) can pass into the channel repair state.

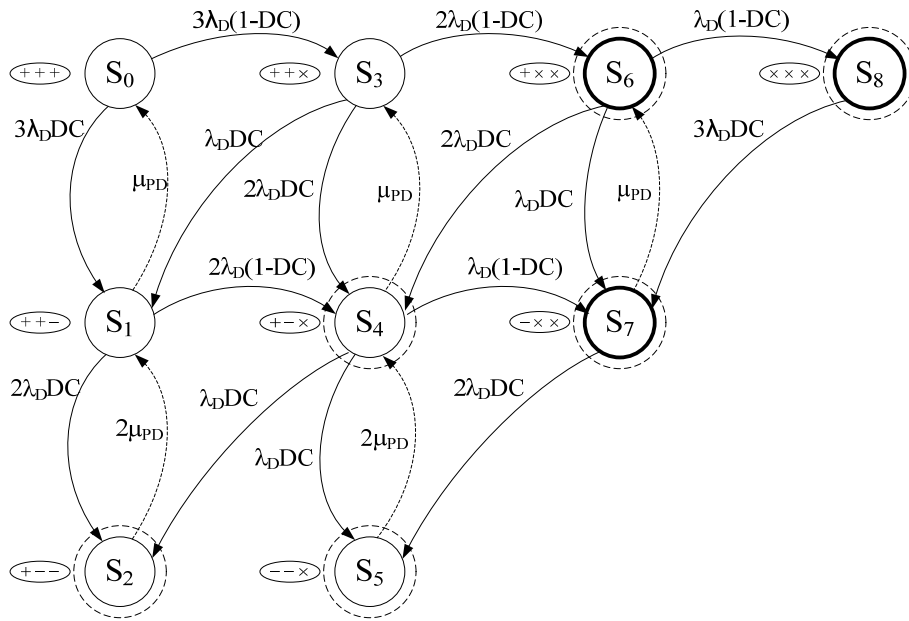


Fig. 4. The marked graph of the ICS RPS operation model without absorbing states (MSaf2)

At the same time, the complete diagnostics of the channel with elimination of all (found and undetected) defects is carried out during the recovery operations.

Also it effects on the duration of reveal duration, and respectively $\mu_{PD} = 1/(MRT + T_D) < \mu_P$. Here MRT (Mean Time to Repair) is the average duration of repair of one ICS channel, T_D is an extra time of diagnosing of undetected failures. The marked graph of such model is provided on Fig.4.

The repeated reveal of the found dangerous failures on the graph is illustrated with transitions of $S_8 \rightarrow S_7$, $S_7 \rightarrow S_5$, $S_6 \rightarrow S_4$, $S_4 \rightarrow S_2$. Thus, a graph on Fig.4 doesn't contain the absorbing states.

The safety function PFD_{avg} is defined on (1), also, as well as in the MSaf1 model.

5.3 ICS RPS functional safety model with the majority device, initiating the enhanced diagnostic of the undetected failures (MSaf6)

The marked graph of such model is provided on Fig.5. The system is initially in operation (all three channels are in operation). In a case of the identified failures reveal the system alternatively passes into states of S_1 and S_2 , with a possibility of resetting after repair.

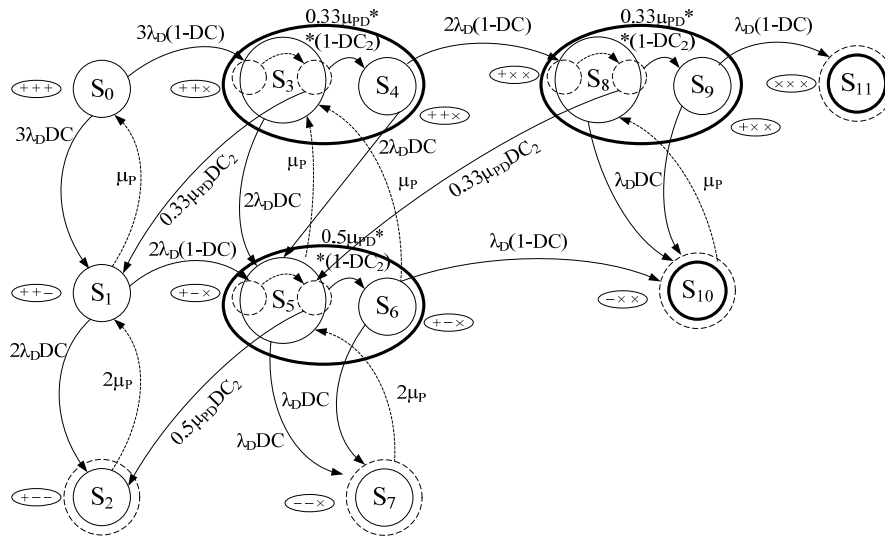


Fig. 5. The marked graph of the model of ICS RPS operation with the enhanced diagnostics on the majority device signal (MSaf6)

In the case of first undetected failure reveal the system passes into S_3 state. Further the majority device informs about the discrepancy of channels operation results, but it is unknown, in what channel is a failure, so the system channels are alternatively passing the enhanced diagnostics with a covering DC_2 . As a result, the undetected failure either reveals (transition to S_1 state) or does not reveal and the system passes into S_4 state. The duration of the enhanced diagnostics of one channel is equal T_{PD} , thus to check the three channels the transition with intensity $\mu_{PD} = 1/3T_{PD}$ is initiated.

As the initiation of the reinforced diagnostics happens practically right after detection of a discrepancy a majority organ, the state of S_3 is "assembly". It integrates the states of the hidden failure reveal and the beginning of carrying out the reinforced diagnostics. And unlike the previous models, the state of S_3 is non-serviceable (in it the reinforced diagnostics is carried out).

If it wasn't succeeded to find the hidden failure, the system passes into S_4 state (it is operable), from which two transitions are possible: reveal of one more hidden failure (transition to S_8 state), or reveal of the found failure – transition in state S_5 .

The S_5 state is characterized by the fact that the system knows about the found failure of one of channels, but the majority organ informs about an undetected failure in one of two remained channels. The enhanced diagnostics of these channels, as a

result of which the failure can be found (transition to S_2 state), is launched, or if it is undetected the transition to S_6 state is made.

Similarly the combination of states of S_8 and S_9 is explained.

As in the course of the reinforced diagnostics reveal as found, and undetected failures is possible, reveal of the found failures is modeled by transitions of $S_3 \rightarrow S_5$, $S_4 \rightarrow S_5$, $S_5 \rightarrow S_7$, $S_6 \rightarrow S_7$, $S_8 \rightarrow S_{10}$, $S_9 \rightarrow S_{10}$; reveal of undetected failures is modeled by transitions of $S_4 \rightarrow S_8$, $S_6 \rightarrow S_{10}$, $S_9 \rightarrow S_{11}$.

Thus, the states of S_{10} and S_{11} are most dangerous, as it is impossible to find failures in them neither with the system of diagnostics, nor in a majority organ. An additional periodic prevention of the hidden failures for their detection is necessary. The safety function PFD_{avg} is defined as:

$$PFD_{avg} = 1 - P_0(t) - P_1(t) - P_4(t). \quad (2)$$

6 Discussion on definition of input parameters values

The values of input parameters were defined as a matter of experience practical operation of the considered class of systems, and also proceeding from the recommendations explained in [3].

As it is required to provide a measure value of the functional safety at the SIL3 level, that is $PFD_{avg} \in]1e-4...1e-3]$, it is necessary to conduct additional researches of models for the purpose of selection of values of input parameters. Values of input parameters concerning which researches are conducted are considered as basic and are shown in a Table 2.

Table 2. ICS RPS functional safety models input parameters base values

#	Name	Base value	Variation range	Unit
1	$\lambda_D = 0.5 * \lambda$	2.5e-5	[0.05 ... 5] * 1e-5	1/hour
2	$\lambda_{DD} = \lambda_D * DC$	2.25e-5		1/hour
3	$\lambda_{DU} = \lambda_D * (1 - DC)$	2.5e-6		1/hour
4	$\mu_P = 1/MRT$	1/8		1/hour
5	$\mu_{PD} = 1/(MRT + T_D)$	1/(8+4)		1/hour
6	DC	0.9 (for MSaf1, MSaf2) 0.5 (for MSaf6)	[0.01...1]	
7	$\mu_{PD*} = 1/(MRT + T_D)$	1/(8+4)=1/12		1/hour
8	$DC2 = 1 - (1 - DC)/10$	0.95	[0.95...1]	

Also the Table 2 demonstrates the options of input parameters λ_{DU} and DC change, and also DC2 for the research of their influence on an function of the functional safety.

7 ICS RPS safety markov models research

To determine the resultant function PFD_{avg} for each graphs, given on Fig.3 – Fig.5, the differential equation system of Kolmogorov-Chapman was generated. Differential equation system solve was executed in the Matlab system by means of the ode15s method for time slot $[0 \dots 500000]$ hours. The results of this modeling are given in Fig.6.

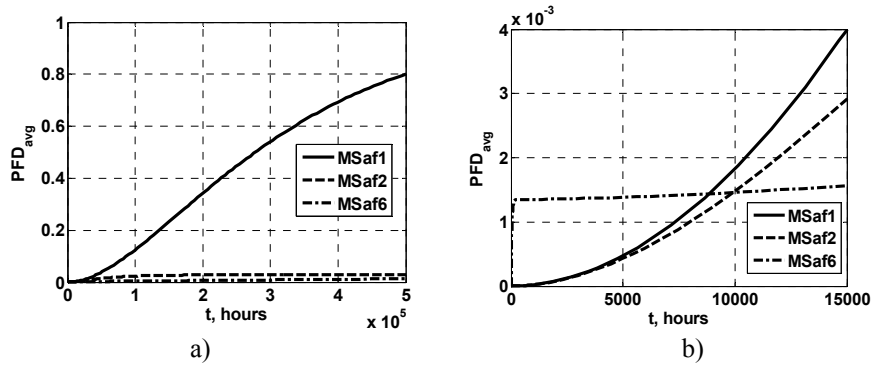


Fig. 6. The dependency of functions PFD_{avg} from the operation time on the interval $[0 \dots 500000]$ hours (a) and $[0 \dots 15000]$ hours (b)

Fig.6 shows that the existence of the absorbing states in the MSaf1 model underlies the continuous growth of function PFD_{avg} to one. On the other hand, the MSaf2 model without the absorbing states illustrates the asymptotic convergence of the functional safety function to stationary of PFD_{avg} value = 0,028 through 16000 working hours. At the same time the SIL3 requirements are provided on time slot till 7200 operation hours (10 operation months) for the MSaf1 model; and on time slot till 8000 operation hours for the MSaf2 model.

At the basic parameter values from the Table 2, the SIL3 requirements in the MSaf6 model aren't satisfied. It is required to define the values of input parameters for this model, in case of which the $PFD_{avg} < 1e-3$ condition will be satisfied.

Fig.7 demonstrates in three-dimensional representation the dependence of the functional safety $PFD_{avg}(t)$ on values of input parameter $DC \in [0 \dots 1]$ for the MSaf1 and MSaf2 models.

Analyzing the diagrams, it is possible to mark that in the absence of dangerous failures diagnostics ($DC = 0$), the both models show the identical behavior of the $PFD_{avg}(t)$ function (diagrams match). In the case of detection of all dangerous failures ($DC=1$) both models show the identical behavior of the $PFD_{avg}(t)$ function, such as the asymptotic convergence to the settled value.

The dynamics of $PFD_{avg}(t)$ functional safety function change shows that value of input parameter of spanning by diagnostics of DC influences duration of the temporal period of execution by system of requirements of SIL3. The diagram of the three-dimensional figure projection to the plane $[t, DC]$ on the level $PFD_{avg}=1e-3$ (Fig.8) demonstrates such influence $[t, DC]$ in more details. For the best visualization of a graphics are shown in different scales concerning an axis DC . For the MSaf6 model in

case of $DC=0.64$ value execution of requirements of SIL3 on all temporal interval of research is provided.

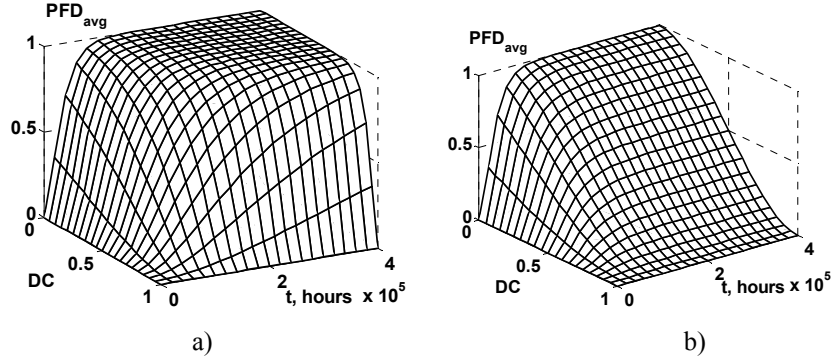


Fig. 7. Function $PFD_{avg}(t)$ behavior dependence on the input parameter DC (diagnostic coverage) for MSaf1 (a) and MSaf2 (b)

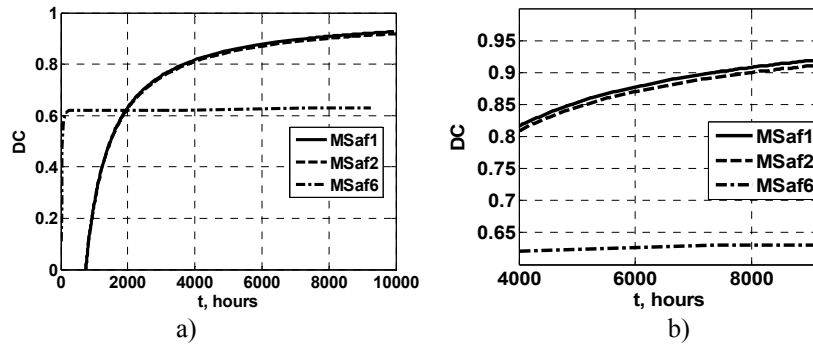


Fig. 8. Projections of function PFD_{avg} on the plane $[t, DC]$ on level $PFD_{avg}=1e-3$ on a scale $t \in [0...10000]$ hours (a) and $t \in [4000...10000]$ hours (b)

Fig.9 represents in three-dimensional representation the functional safety $PFD_{avg}(t)$ dependence on the dangerous failures rate value λ_D for the MSaf1 and MSaf2 models.

At first sight, the MSaf2 model (without the absorbing states) illustrates the best result as a function $PFD_{avg}(t)$ in it aims to the settled PFD_{avg} value = 0,028 (value is caused by a stable combination of parameters $DC = 0.9$ and $\mu_{PD} = 0.0833$).

Model MSaf1 illustrates the convergence of function $PFD_{avg}(t)$ to 1. And than more the dangerous failures rate is, the approximation of $PFD_{avg}(t)$ function to the steady-state mean is faster.

However, if we look at the projection of three-dimensional figures on Fig.9 to the plane $[t, \lambda_D]$ on the upper cutoff of requirements of SIL-3, then the difference between results of simulation of MSaf1 and MSaf2 doesn't exceed $\Delta t=100$ hours in case of $\lambda_D=1e-4$ 1/hour (Fig.10).

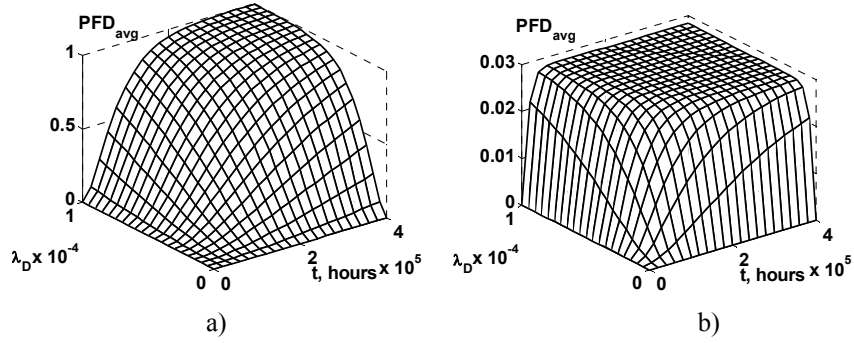


Fig. 9. Dependence of $PFD_{avg}(t)$ function behavior from the input parameter λ_D for MSaf1 (a) and MSaf2 (b)

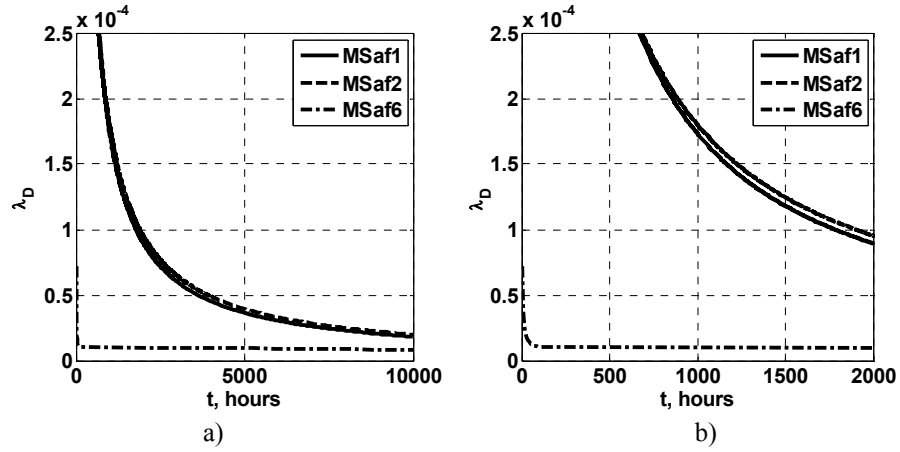


Fig. 10. Projections of function PFD_{avg} to the plane $[t, \lambda_D]$ on the $PFD_{avg}=1e-3$ level on the scale of $t \in [0 \dots 10000]$ (a) and $t \in [0 \dots 2000]$ (b)

For the MSaf6 model for basic values of input parameters (in particular $DC=0.5$) the requirements of SIL3 are fulfilled in case of $\lambda_D < 1e-5$ 1/hr.

8 Conclusions

The analysis of ICS functional safety simulation received results showed that:

a) in case of the accounting of secondary manifestation of dangerous failures and detections their diagnostic system (model MSaf2) for base-line values of input parameters reaches the settled PFD_{avg} value = 0.028 that it isn't enough for safety arrangements of the SIL3 level;

b) in case of value of dangerous failures rate $\lambda_D = 2.5e-5$ (1/hour) the considered system meets requirements of SIL3 during the first 8000 operation hours; for

extension of this period till 10000 hours it is necessary to increase spanning by diagnostics to the $DC=0.92$ level (models MSaf1 and MSaf2);

c) if it is impossible to increase the spanning by diagnostics, then it is necessary to reduce failure density of each channel to $\lambda = 2 * \lambda_D = 4e-5$ 1/hour for the extension of the temporal period of SIL3 requirements support till 10000 hours (models MSaf1 and MSaf2);

d) for the system with majority device which initiates additional diagnostics of the hardware channels (model MSaf6) a sufficient condition of support of requirements of SIL3 is the spanning by diagnostics $DC > 0.65$.

The developed Matlab-programs which can be used in engineering practice are a subject of practical interest.

The essential lack of the developed models is the absence of taking note of software failures in ICS channels. The accounting of manifestation of software defects and their elimination during rescue and recovery operations as it is described in [8], is the direction of further researches and improvement of the developed models. Also, further researches and improvement of the developed models is to analyze software features and interactions with hardware failures in instrumentation and control system channels.

References

1. IEC 61508-1:2010. Functional safety of electrical/electronic/programmable electronic safety-related systems - Part 1: General requirements (2010)
2. Sklyar, V.V. Elements of the information and control systems functional safety analysis methodology. Radioelectronic and computer systems 6(40), 75--79 (2009)
3. IEC 61508-6:2010. Functional safety of electrical/electronic/programmable electronic safety-related systems - Part 6: Guidelines on the application of IEC 61508-2 and IEC 61508-3 (2010)
4. Bahmach, E.S. Siora, A.A. Sklyar, V.V. Tokarev, V.I. Kharchenko, V.S. FPGA-based NPP information and control systems safety accession and provision Radioelectronic and computer systems 7, 75--82 (2007)
5. Langeron, Y. Barros, A. Grall, A. Berenguer, C. Combination of safety integrity levels (SILs): A study of IEC61508 merging rules. Journal of Loss Prevention in the Process Industries 21(4), 437--449 (2008)
6. Al-Kuwaiti, M., Kyriakopoulos, N., Hussein, S.: A comparative analysis of network dependability, fault-tolerance, reliability, safety, and survivability. IEEE Communications Surveys & Tutorials 11(2), 106--124 (2009)
7. Trivedi, K.S., Dong Seong Kim, Roy, A., Medhi, D.: Dependability and safety models. In: Proceedings 7th International Workshop on the Design of Reliable Communication Networks (DRCN 2009), pp. 11-20, IEEE Press, Washington, DC (2009)
8. Ponochozny, Y.L., Siora, A.A., Kharchenko, V.S. Models of dual-channel information management system readiness considering the software updating. Radioelectronic and computer systems 6(70), 135--139 (2014)

Use of Natural LUT Redundancy to Improve Trustworthiness of FPGA Design

Alex Drozd, Miroslav Drozd, Mykola Kuznietsov

Institute of Computer Systems, Odessa National Polytechnic University,
1 Shevchenko St., 65044 Odessa, Ukraine

Drozd@ukr.net, miroslav_dr@mail.ru,
koliaodessa@mail.ru

Abstract. Paper is devoted to a design FPGA problem regarding increase in trustworthiness of the results calculated in digital components of computer systems. The modern CAD solves this problem by implementation of significant hardware redundancy in fault tolerant decisions. The method for increase in trustworthiness of the results calculated on FPGA with the LUT-oriented architecture is offered. This method is directed to improving of the ready project without change of its hardware decision. The method generates 16 versions in programming of LUT memory and suggests selecting the best version by the given criterion. The method is shown on the example of masking of short circuit between LUT inputs. The method allows selecting the decision taking into account risk-value of bits in LUT memory.

Keywords. FPGA design, resource-based approach, LUT-oriented architecture, natural redundancy, trustworthiness of results, masking of short circuit

Key Terms. High Performance Computing, Concurrent Computation, Model, Method, Simulation

1 Introduction

We offer a method for improving of FPGA projects in view of ample opportunities and the perspective directions in design of computer systems and their components on FPGA. External manifestations of these two aspects are watched in a gain of the market in the field of the design of systems for digital processing of signal and images, accelerators, reconfigurable systems and safety-related instrumentation and control systems [1 – 4].

The internal reasons of successful development of FPGA design can be considered by means of resource-based approach which analysis process of integration of the artificial world created by the human into the natural world (NW). It allows not only to explain history of the computer world which repeats in development the NW during a short time but also to offer new methods based on systematization in features and perspectives of further development [5].

Integration process consists in the solution of challenges of the NW, including problems of the computer world. The solution is based on achievement of throughput (for execution of all amount of works for limited time), of result trustworthiness and on attachment of resources: models, methods and means.

The models are human understandings of the NW and its components. Methods serve for conversion and an assessment of resources. Models and methods form an informational part of the resources but means (materials and tools) belong to the technological one. Resources develop from simple to real by structuring under features of the NW. Simple forms are exact and sequential according to initial opportunities and understandings of the human. Real resources reflect features of the NW [6].

Development of computer resources most brightly showed two such features: parallelism and fuzziness [7, 8]. In the resource development occurring under these features, the resource-based approach identifies a number of levels: replication, diversification and autonomy. In the hierarchy they occupy the low, L, the middle, M and the high, H levels, respectively. Each preceding level serves for the next one. In all the levels, from L to H, the goal is surviving, i.e. integration into the NW. It is provided by different methods: increasing throughput, trustworthiness of results and access to resources (self-sufficiency), respectively [9].

The LUT-oriented architecture of FPGA projects has the high level of fuzziness in the form of natural redundancy in the organization and functionality. We offer a method which shows opportunities to use this redundancy for development of a CAD in the modes of FPGA-project optimization for improving of result trustworthiness.

The method allows replicate different versions of the LUT program code for the ready FPGA project with the subsequent choice of the best by the selected criterion. The method is shown on the example of fault masking for the purpose of increase in trustworthiness of the results calculated on FPGA. The problem of FPGA project improvement by criterion of result trustworthiness is defined in section 2. Section 3 describes a model of data transfer between LUT. Versions of LUT programming are generated in section 3. In section 4 the received versions are evaluated from a line item of masking of faults like shorts circuit of adjacent inputs of the LUT. Increase in trustworthiness of the results calculated on FPGA in case of faults like shorts is shown in section 5. Experimental results are considered in section 6. The received results are analyzed in the conclusions.

2 Problem Definition

All levels of resource development are represented in FPGA design. The matrix and pipeline types of parallelism corresponding to levels of replication and diversification are supported by arrays of the configurable blocks and the structure of a logical element as section of the pipeline containing the LUT and the register [10].

Matrix parallelism restricts possibilities in paralleling of computations by data dependency when high orders of numbers will be processed after low orders, and by control dependency with waiting for computation of conditions in branching of an algorithm. Pipeline parallelism removes data dependency processing at the same time both high and low orders in different sections of the pipeline [11].

A method of results preparation belongs to the high level of autonomy. This method is free from both restrictions, calculating a set of possible results, not waiting for receiving all operands [12].

All the libraries, including software and library of IP cores used in CAD are built by the method of result preparing. All of the modern design of digital systems and their components are also based on the method of result preparing. For example, the FPGA chips are preparations for a lot of projects, and the chip programmed for a specific project contains the prepared results in the tables stored in the LUT (Look-Up Table) memory [13].

Mainly due to the preparation of the results, FPGA projects receive features, allowing provide high level a number of characteristics: capacity of calculations and the trustworthiness of their results, versatility in functionality, efficiency of design, manufacturability and flexibility [14].

However, the modern CAD is oriented first of all on matrix parallelism. The FPGA chips contain sets of the prepared iterative array multipliers and the prepared circuits of look-ahead carry for design of digital units with matrix parallelism [15].

As a rule, the CAD supports the optimization modes in energy consumption and throughput, ignoring requirements for the result trustworthiness important for the safe computer systems including autonomous systems relating to levels of diversification and autonomy [16, 17]. Opportunities of such systems completely are defined at design stage, for example, onboard equipment of space crafts for distant flights, the pacemakers implanted in a body of the person.

Possibilities of FPGA design associated with natural redundancy of the LUT-oriented architecture allow improve significantly projects by a method of versions which repeats development of the NW by means of mutations. This method provides replication of the ready decision with deviations, admissible from it, in any parameters with the subsequent choice of decision which is the best by the main criterion. The choice of the decision can be executed with use of analytical estimates or simulation [12].

Use of the method of versions completely comes true within the principle of ALARA (As Low As Reasonability Applicable/Practicable), when there is no alternative of improving of the main characteristics of the decision directed to the greatest possible lowering of risk at the expense of really available limited resources [18].

Thus, the problem, motivation of its decision and resources necessary for this purpose are as follows:

1. As a rule, the modern CAD doesn't support the modes of project optimization in trustworthiness of results.
2. Increase of result trustworthiness is an important condition of maintenance in the functional safety of computer systems in areas of critical application and is traditionally provided with use of fault-tolerant decisions with significant hardware redundancy.
3. FPGA projects have the high level of fuzziness which allows to use effectively the method of versions for increase of result trustworthiness on the basis of natural redundancy of the LUT-oriented architecture without extra hardware.

We offer the method using natural redundancy to replicate different versions of the program LUT code for the given hardware implementation of the FPGA project. Replication is executed with changes, unessential for project parameters. The choice of the version is carried out in the most important parameter.

The method is shown on the example of the best masking of typical faults of the digital circuits for the purpose of the greatest increase of trustworthiness in results of the computation executed on FPGA.

3 The Model of Data Transfer between LUT

We can consider the LUT-oriented architecture on the example of Cyclone FPGA Family Data Sheet of Altera Corporation. Cyclone structure contains arrays of logic elements. Each logic element consists of a four-input LUT and programmable register. The LUT is a function generator that can implement any function of four variables A , B , C and D . Inputs of the LUT are called as well as variables [19].

We offer model of data transfer between LUT which allows change a program code for LUT without changing the hardware structure of FPGA-project.

The program code can be diversely replicated with use of two versions in programming of ordered pair of the LUT units connected among themselves. The output of the first of them is connected directly or via the programmable register (or other elements with implementation of a self-dual function) to an input of the second LUT. Two versions of programming which differ from each other in direct or inverse value of the bit transferred from the first LUT to the second one are shown in Fig. 1.

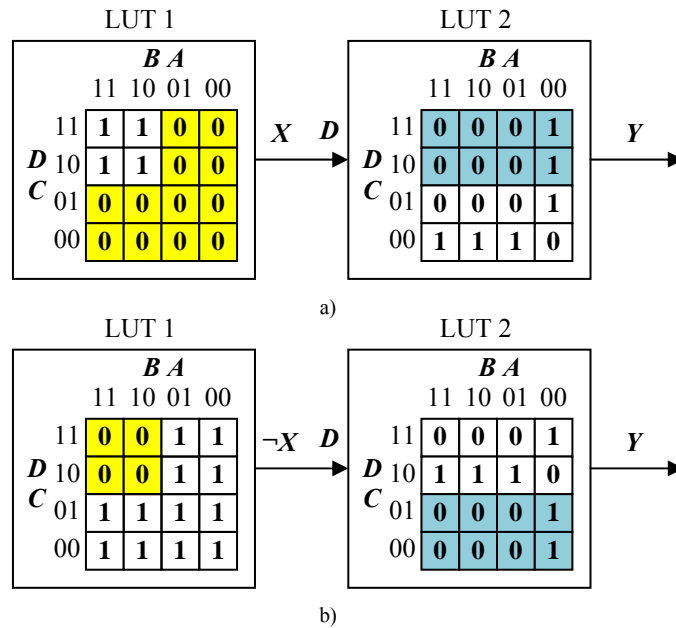


Fig. 1. Two versions of LUT units programming: a – transfer of a direct value of X ;
b – transfer of an inverse value of X

The first LUT 1 and second LUT 2 generate functions $X = D \wedge B$ and $Y = (A \wedge B) \oplus (C \wedge D)$, respectively. The functions are described by tables which rows and columns are determined by values of 2-bit codes DC and BA , respectively. These codes are made of the A, B, C, D variables. Connection between units transfers value X from an output of the LUT 1 to the D input of the LUT 2 (Fig. 1, a).

The second version in programming of units can be executed by transmission of inverse value $\neg X$ between them (Fig. 1, b). In this case, all bits of the LUT 1 will be inverted. For saving value Y , i.e. for compensating of inverse value X , the table of the LUT 2 changes mutual position of values "0" and "1". In case of inverse value of the variable D , positions of two top rows and two bottom ones are interchanged.

4 Versions in Programming of the LUT Units

Direct and inverse values of the variables A, B, C, D generate 16 different versions in programming of the LUT 2. All of them are shown in Fig. 2.

	BA	$B\neg A$	$\neg BA$	$\neg B\neg A$																																																																
D	<table><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>0</td></tr></table>	0	0	0	1	0	0	0	1	0	0	0	1	1	1	1	0	<table><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>1</td><td>0</td><td>1</td></tr></table>	0	0	1	0	0	0	1	0	0	0	1	0	1	1	0	1	<table><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	0	0	0	1	0	0	0	1	0	0	1	0	1	1	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr></table>	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	1
0	0	0	1																																																																	
0	0	0	1																																																																	
0	0	0	1																																																																	
1	1	1	0																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
1	1	0	1																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
1	0	1	1																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
0	1	1	1																																																																	
C	<table><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>0</td></tr></table>	0	0	0	1	0	0	0	1	0	0	0	1	1	1	1	0	<table><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>1</td><td>0</td><td>1</td></tr></table>	0	0	1	0	0	0	1	0	0	0	1	0	1	1	0	1	<table><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	0	0	0	1	0	0	0	1	0	0	1	0	1	1	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr></table>	1	0	0	0	1	0	0	0	1	0	0	0	0	1	1	1
0	0	0	1																																																																	
0	0	0	1																																																																	
0	0	0	1																																																																	
1	1	1	0																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
1	1	0	1																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
1	0	1	1																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
0	1	1	1																																																																	
$\neg D$	<table><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	0	0	0	1	1	1	1	0	0	0	0	1	0	0	0	1	<table><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>1</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr></table>	0	0	1	0	1	1	0	1	0	0	1	0	0	0	1	0	<table><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>1</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr></table>	0	1	0	0	1	0	1	1	0	1	0	0	0	1	0	0	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr></table>	1	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0
0	0	0	1																																																																	
1	1	1	0																																																																	
0	0	0	1																																																																	
0	0	0	1																																																																	
0	0	1	0																																																																	
1	1	0	1																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
0	1	0	0																																																																	
1	0	1	1																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
1	0	0	0																																																																	
0	1	1	1																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
C	<table><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>1</td><td>1</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	0	0	0	1	1	1	1	0	0	0	0	1	0	0	0	1	<table><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>1</td><td>1</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr></table>	0	0	1	0	1	1	0	1	0	0	1	0	0	0	1	0	<table><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>1</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr></table>	0	1	0	0	1	0	1	1	0	1	0	0	0	1	0	0	<table><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr></table>	1	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0
0	0	0	1																																																																	
1	1	1	0																																																																	
0	0	0	1																																																																	
0	0	0	1																																																																	
0	0	1	0																																																																	
1	1	0	1																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
0	1	0	0																																																																	
1	0	1	1																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
1	0	0	0																																																																	
0	1	1	1																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
$\neg D$	<table><tr><td>1</td><td>1</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	<table><tr><td>1</td><td>1</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr></table>	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	<table><tr><td>1</td><td>0</td><td>1</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr></table>	1	0	1	1	0	1	0	0	0	1	0	0	0	1	0	0	<table><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr></table>	0	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0
1	1	1	0																																																																	
0	0	0	1																																																																	
0	0	0	1																																																																	
0	0	0	1																																																																	
1	1	0	1																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
1	0	1	1																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
0	1	1	1																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
$\neg C$	<table><tr><td>1</td><td>1</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	<table><tr><td>1</td><td>1</td><td>0</td><td>1</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr><tr><td>0</td><td>0</td><td>1</td><td>0</td></tr></table>	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	0	<table><tr><td>1</td><td>0</td><td>1</td><td>1</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>0</td><td>0</td></tr></table>	1	0	1	1	0	1	0	0	0	1	0	0	0	1	0	0	<table><tr><td>0</td><td>1</td><td>1</td><td>1</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>1</td><td>0</td><td>0</td><td>0</td></tr></table>	0	1	1	1	1	0	0	0	1	0	0	0	1	0	0	0
1	1	1	0																																																																	
0	0	0	1																																																																	
0	0	0	1																																																																	
0	0	0	1																																																																	
1	1	0	1																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
0	0	1	0																																																																	
1	0	1	1																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
0	1	0	0																																																																	
0	1	1	1																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	
1	0	0	0																																																																	

Fig. 2. Versions in programming of the LUT 2

All versions can be received on condition of connection of all four inputs of the LUT 2 to outputs of four units as LUT 1.

The FPGA-project containing n pair of LUT can be replicated into a set of 2^n decisions which save uniform hardware structure and differ in a program code. For example, the 10 pair of LUT provides 1024 versions of the project. These versions can significantly differ in a number of parameters, including trustworthiness of the calculated results.

5 Increase in Trustworthiness of the Results Calculated on FPGA

The offered method can be shown on the example of increase in trustworthiness of the results calculated on FPGA in case of faults like shorts between adjacent inputs of the LUT [20]. Tables for LUT 2 in cases of the correct functioning and short circuit $A \leftrightarrow B$, $B \leftrightarrow C$, $C \leftrightarrow D$ between inputs A and B , B and C , C and D are shown in Fig. 3.

	$B A$	$A \leftrightarrow B$	$B \leftrightarrow C$	$C \leftrightarrow D$
A	0 0 0 1	0 1 1 1	0 0 0 1	0 0 0 1
B	0 0 0 1	0 1 1 1	0 1 0 1	1 1 1 0
C	0 0 0 1	0 1 1 1	0 0 1 0	1 1 1 0
D	1 1 1 0	1 0 0 0	1 0 1 0	1 1 1 0
a)				
	$B A$	$A \leftrightarrow B$	$B \leftrightarrow C$	$C \leftrightarrow D$
$\neg A$	0 1 1 1	0 1 1 1	0 1 0 0	0 1 1 1
$\neg B$	1 0 0 0	1 0 0 0	0 0 0 0	1 0 0 0
$\neg C$	1 0 0 0	1 0 0 0	1 0 0 0	1 0 0 0
$\neg D$	1 0 0 0	1 0 0 0	0 0 0 0	1 0 0 0
b)				

Fig. 3. Tables for cases of the correct and faulty functioning of the LUT 2: a – the initial programming of the LUT; b – programming of the LUT with inverse values of inputs

Short circuit of two inputs distorts their values "01" and "10", substituting with them by "00" value. It leads to substitution of eight bits determined in the table by "01" and "10" values with four bits addressed on "00" value. Comparing of tables for cases of the correct and faulty functioning of the LUT 2 allows detect the erroneous bits highlighted in the color.

The first row of tables is shown for the initial programming of the LUT 2 (Fig. 3, a). Short circuit $A \leftrightarrow B$ determines 8 erroneous bits in two internal columns of the table. Short circuit $B \leftrightarrow C$ is shown in four erroneous bits. Short circuit $C \leftrightarrow D$ determines 8 erroneous bits in two internal rows of the table. The total quantity of erroneous bits is equal 20. It makes 42% of total quantity of bits in memory of LUT.

The second row of tables is shown for the programming of the LUT 2 with inverse values of inputs A , B , C and D (Fig. 3, b). Short circuits $A \leftrightarrow B$ and $C \leftrightarrow D$ don't lead to distortion of bits in the table. Short circuit $B \leftrightarrow C$ is shown in four erroneous bits. The total quantity of erroneous bits is equal 4. It makes 8% of total quantity of bits in

memory of LUT. Thus, conversion of the version with the initial programming of the LUT units reduces the number of erroneous bits by 5 times, i.e. repeatedly improves masking of short circuits and increases trustworthiness of the results calculated on FPGA. This effect is reached without any changes in the hardware implementation of the circuit decision, i.e. after all types of its optimization. The advantage of the offered method is the independent analysis of the LUT pairs that significantly reduces time necessary for a choice of the best version of the FPGA project.

The method is shown for a case of identical risk-values of erroneous bits for obtaining authentic results. But these values can change depending on the frequency of use of the bit and consequences of its distortion. The accounting of these two factors makes a method as risk-oriented. Such development of the method includes two steps. On the first step, the frequency of the appeal to each bit of each LUT should be evaluated. This step can be executed by simulation of operation of the circuit on characteristic input data. The program model of the circuit decision allows to count the number of appeals to each bit of memory and to calculate the frequency of its use. The second step requires expert estimates of consequences for every error.

6 Experimental Results

The experimental verification of the offered method was executed in process of preparation of a lab classes within the project TEMPUS GREENCO "Green Computing & Communications" (530270-TEMPUS-1-2012-1-UK-TEMPUS-JPCR) [21].

The program model is developed for the analysis of LUT pairs using Delphi 10 Seattle demo-version [22]. The analysis is made in two modes:

- for a specific case of programming of the LUT 2 in view of the risk-value of each bit for trustworthiness of result;
- for all 216 functions generated in the LUT 2 on condition of identical value of bits in relation to trustworthiness of result.

The main panel of the program with setting of the risk-values of bits in LUT 2 memory for function Y is shown in Fig. 4.

The program calculates risk for each of 16 versions in programming of the LUT unit according to the given table "R-values of bits" of bit risk-values and sets of the erroneous bits in versions. For example, in case of table "R-values of bits", the table of risks for 16 versions shows that the risk $Risk_{cur} = 35$ of the initial version matches maximum $Risk_{max} = 35$ and exceeds the minimum risk $Risk_{min} = 5$ by 7 times.

Thus, reprogramming of LUT unit can repeatedly reduce the risks associated with faults in implementation of FPGA projects. Viewing of all functions generated in the LUT 2 unit showed that the average scoring of the best version in comparison with current one and the worst is equal 21% and 53% of the masked errors, respectively.

7 Conclusions

LUT-oriented architecture of FPGA-projects was widely used in various areas of information technologies owing to the high levels of development of its resources:

models, methods and means. At the same time, its opportunities are used not fully, saving reserves for the best application.

The screenshot shows a software window titled 'Form1' with a menu bar containing 'Exit', 'Function', 'Start', and 'Shorts A-B'. The main area displays several tables and risk values:

Table of versions				Table of shorts A - B				R-values of bits							
0	1	2	3	0	0	0	1	1	1	1	1	1	1	2	1
0	0	1	1	1	1	1	1	1	1	1	1	2	3	1	3
1	1	0	0	0	1	0	0	0	0	0	0	3	1	1	1
2	1	0	0	0	0	1	0	0	0	0	0	1	2	1	1
3	1	0	0	0	0	1	0	0	0	0	0				

Table of risks			
35	29	28	22
29	23	22	16
24	18	17	11
18	12	11	5

Risk.cur = 35
 Risk.min = 5
 Risk.max = 35

Fig. 4. Panel of the program with setting of the risk-values of LUT memory bits

All levels of resource development are provided to FPGA chips:

- replication in matrix parallelism which is offered for design of digital units with data processing in parallel codes;
- diversification in the pipeline parallelism supported in structure of logical elements or reconfigurable logic blocks;
- autonomy in preparation of results and circuits of their distribution in adding units, in preparation of iterative array multipliers.

However, the matrix parallelism relating to the low level of resource development dominates. For example, preparation of results is used for improving of circuit decisions with matrix parallelism. The modern CAD offers the optimization modes aimed at enhancement of FPGA-projects in parameters of throughput and energy consumption which are important at the level of replication and autonomy, respectively. At the same time, improving of projects in trustworthiness of the results calculated on FPGA isn't executed. However increase of trustworthiness is important for safety-related computer systems including such autonomous systems.

The offered method allows improve ready FPGA-project without change of the reached throughput and energy consumption. The method saves the hardware decision and executes reprogramming of LUT, each of which can have 16 versions of programming. The method is shown on the example of the best masking of short circuit between LUT inputs. Reprogramming of LUT allows reducing quantity of possible errors by 5 times. Development of the method taking into account risk-value of bits in LUT memory can raise this index considerably.

References

1. Bakhmach, E., Kharchenko, V., Siora, A., Sklyar, V., Tokarev, V.: Design and Qualification of I&C Systems on the Basis of FPGA Technologies. In: 7th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC&HMIT 2010), pp. 916--924. Las Vegas, Nevada (2010)
2. Dick, C. H., Harris, F. J.: Direct Digital Synthesis – Some Options for FPGA Implementation. In: SPIE International Symposium on Voice Video and Data Communication: Reconfigurable Technology: FPGAs for Computing and Applications, pp. 2-10. Stream, Boston, MA, USA (1999)
3. Palagin, A.V., Opanasenko, V.N.: Design and application of the PLD-based reconfigurable devices. Design of Digital Systems and Devices, Springer Verlag, Berlin, Heidelberg, vol. 79, pp. 59--91 (2011)
4. Melnyk, A. O., Melnyk, V. A.: Personal Supercomputers: Architecture, Design, Application: Lvivska Polytechnika, Lviv, Ukraine (2013)
5. Drozd, J., Drozd, A., Antoshchuk, S., Kharchenko, V.: Natural Development of the Resources in Design and Testing of the Computer Systems and their Components. In: Proc. 7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, pp. 233--237. Berlin, Germany (2013)
6. Tanenbaum, A.: Structured Computer Organization, 4th ed. Prentice Hall, Upper Saddle River, NJ (1999)
7. Kondratenko, G., Kondratenko, Y., Romanov, D.: Fuzzy Models for Capacitive Vehicle Routing Problems in Uncertainty. In: Proc. International DAAAM Symposium, pp. 205--206 (2006)
8. Drozd J., Drozd A.: Models, Methods and Means as Resources for Solving Challenges in Co-Design and Testing of Computer Systems and their Components. In: Proc. 9th International Conference on Digital Technologies (DT'2013), pp. 176--180. Zhilina, Slovakia (2013)
9. Drozd, J., Drozd, A., Maevsky, D., Shapa, L.: The Levels of Target Resources Development in Computer Systems. In: Proc. IEEE East-West Design & Test Symposium, pp. 185--189. Kiev, Ukraine (2014)
10. Drozd, J., Drozd, A., Antoshchuk, S., Kushnerov, A., Nikul, V.: Effectiveness of Matrix and Pipeline FPGA-Based Arithmetic Components of Safety-Related Systems. In: Proc. 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. pp. 785--789. Warsaw, Poland (2015)
11. Kharchenko V.S. (ed): Green IT-Engineering. Volume 1. Principles, Models, Components. National Aerospace University "KhAI", Kharkiv, Ukraine (2014)
12. Drozd A., Kharchenko, V. (eds): Online Testing of the Safe Instrumentation and Control Systems. National Aerospace University named after N.E. Zhukovsky "KhAI", Kharkiv, Ukraine (2012)
13. Netlist Optimizations and Physical Synthesis. Qii520072.0. Quartus II Handbook. Vol. 2: Altera Corporation (2004)
14. Kharchenko, V.S., Sklyar, V.V. (eds): FPGA-based NPP I&C Systems: Development and Safety Assessment. RPC Radiy, National Aerospace University "KhAI", SSTC on Nuclear and Radiation Safety, Kharkiv, Ukraine (2008)
15. Xilinx Staff. Virtex-II Pro and Virtex-II Pro X Platform FPGAs: Complete Data Sheet (2005)
16. PowerPlay Power Analysis. Qii53013. Quartus II Handbook: Altera Corporation (2013)
17. Design Optimization for Altera Devices. Qii52005-2.0. Quartus II Handbook. Vol. 2: Altera Corporation (2004)

18. Kharchenko, V.S. (ed): Safety of Critical Infrastructures: Mathematical and Engineering Methods of Analysis and Ensuring: Ministry of Education and Science of Ukraine, National Aerospace University "KhAI", Kharkiv, Ukraine (2013)
19. Cyclone FPGA Family Data Sheet. Altera Corporation. Online: <http://www.altera.com> (2003)
20. Blyzniuk, M., Kazymyra, I., Kuzmich, W.: Probabilistic Analysis of CMOS Physical Defects in VLSI Circuits for Test Coverage Improvement. Microelectronics Reliability, 41(12), 2023--2040 (2001)
21. Drozd, A., Kharchenko, V. (eds): Green Hardware and Programmable Systems: Ministry of Education and Science of Ukraine, National Aerospace University "KhAI", Kharkiv, Ukraine (2015)
22. Delphi 10 Seattle: Embarcadero. Online: <https://www.embarcadero.com/ru/products/delphi> (2015)

Ontological Approach to the Assessment of Information Sufficiency for Software Quality Determination

Tetiana Hovorushchenko, Oksana Pomorova

Khmeltsky National University, 11, Institutskaya St.,
29016, Khmeltsky, Ukraine

tat_yana@ukr.net, o.pomorova@gmail.com

Abstract. The aim of this study is the development of approach to the assessment of information sufficiency for software quality determination (according to ISO 25010: 2011). The proposed approach to assessment of information sufficiency is based on the comparative analysis of fragments of ontology of the subject domain "Software Engineering" and ontologies, that are developed on the basis of software requirements and system specification of the developed software. The approach provides the improvement the specifications for the presence of measures, that are necessary to the determination of software quality sub characteristics and characteristics. The work of developed approach is illustrated by the assessment of information sufficiency for software quality determination of automated system for large-format photo print.

Keywords. Software, software project, software requirements specification (SRS), software quality, ontology, ISO 25010:2011

Key Terms. Model-Based Software System Development, Software Component, Software System, Method

1 Introduction

The software quality is basic factor for its successful implementation and exploitation. According to the standards of ISO 25010 [1], ISO 25030 [2] the software quality is the ability of the software to meet the stated and predicted needs when using under certain conditions. This definition differs from the definition of software quality of ISO 9000 [3] mainly because this definition of software quality provides for needs satisfaction, while the definition of ISO 9000 [3] provides for requirements satisfaction. The development of modern software system is user-oriented [4], namely the software quality is the important characteristic in terms of the stakeholders (especially customers). Obviously, even attracting the best experts for the development of technologies and standards of the software systems quality assurance doesn't guarantee sufficient software quality.

The essential and integral feature of modern software systems is their complexity, so the attempts to describe the software objects with abstraction from their complexity lead to the abstraction from their essence. The constant growth of the software functions complexity inevitably leads to the increasing their volume and laboriousness (effort applied) [5]. One of the most important causes of poor quality of large software projects are the increasing the number of components (subsystems) and the interfaces between them, and uncontrolled complexity of software systems, in opinion of the researchers The Standish Group International [6]. Research [6] shows that statistics of success of small, moderate, medium, large and grand software projects is significantly different - Fig. 1.

CHAOS RESOLUTION BY PROJECT SIZE			
	SUCCESSFUL	CHALLENGED	FAILED
Grand	2%	7%	17%
Large	6%	17%	24%
Medium	9%	26%	31%
Moderate	21%	32%	17%
Small	62%	16%	11%
TOTAL	100%	100%	100%

The resolution of all software projects by size from FY2011–2015 within the new CHAOS database.

Fig. 1. Statistics of success of small, moderate, medium, large and grand software projects in 2011-2015

Analysis of Fig. 1 provides the conclusion that 62% of small projects are successful while both only 6% of large projects and only 2% of grand projects are successful, i.e. small projects are ten times more successful than large and thirty times more successful than grand projects. This conclusion is confirmed by the statistics of software projects, that is presented in [5], on the basis of the function points as the main modern units of software size - Fig. 2.

During the software project, we often can not estimate the share of the informational indeterminacy of the project. Identification of the information, that appears in the process of interaction of "subsystems-interfaces-data-external influences", is especially difficult. Identification of future properties of developed software, that will display this information, is even more difficult task.

The cause of appearance of informational indeterminacy of the project is the low level of knowledge documentation, especially at the system level (Fig. 3 [7]).

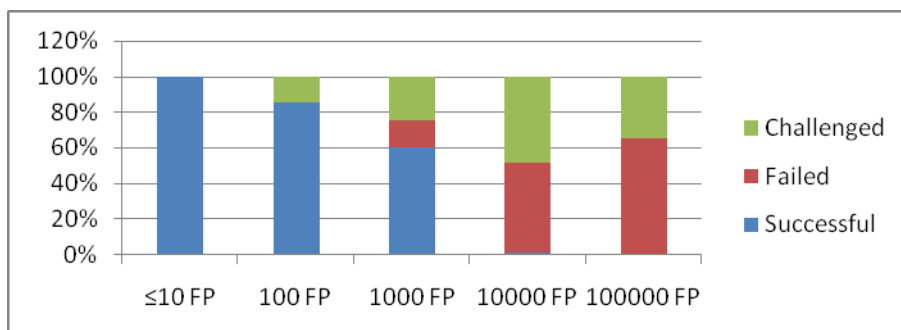


Fig. 2. Statistics of success of software projects with size 10, 100, 1000, 10000, 100000 functional points in 2010-2013

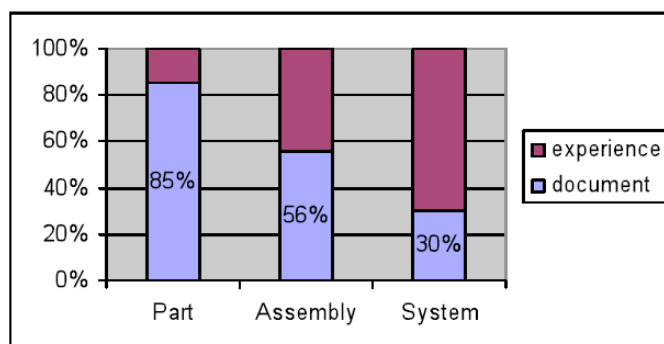


Fig. 3. How well system level knowledge is documented

Fig. 4 depicts the situation, characterized by premature design decisions and their documentation, prior to understanding the design. Fig. 4 shows an area, referred to as the “knowledge gap,” that is the result of the low level of knowledge documentation and the root cause of many engineering failures [8]. The size of knowledge gap is not constant for software project – during the lifecycle it can increase and decrease, since new information appears and it should be taken into account. The presented on Fig. 4 viewpoint on the knowledge gap does not quite correspond to reality. We assume that partial consideration of the subject domain information in the software quality models and impact of this information on only finished product lead to increase of the knowledge gap size during the life cycle (new boundaries of knowledge gap are delineated dotted line on Fig. 5), that can be the cause of the software accidents and disasters [9]. For safe software functioning the knowledge gap size is desirable to reduce. This can be done by the consideration of as much subject domain information in the software quality models and standards. Reducing the knowledge gap size will provide the improvement of the software quality.

Given the above, all the available knowledge and information about the software system can be represented as the diagram, which has the sector that reflects the volume of insufficient (unknown) information (knowledge gap) - Fig. 6. This sector consists of unconsidered subject domain information. The size of this sector is not

determined, because it is unclear what information and how much information is unknown. Sector of the unknown information should be narrow by fully consideration of subject domain information. The smaller size of sector of the unknown information indicates to the higher quality and safer work of software system, i.e. the main task is the reducing the share of unknown information about software system.

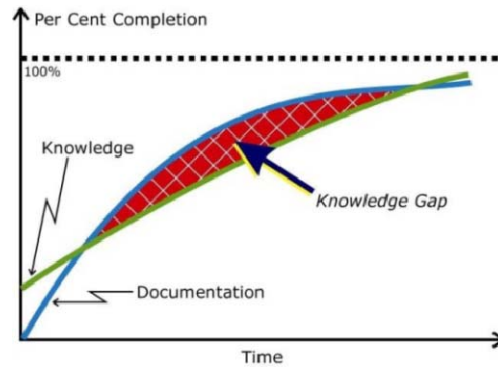


Fig. 4. Knowledge gap

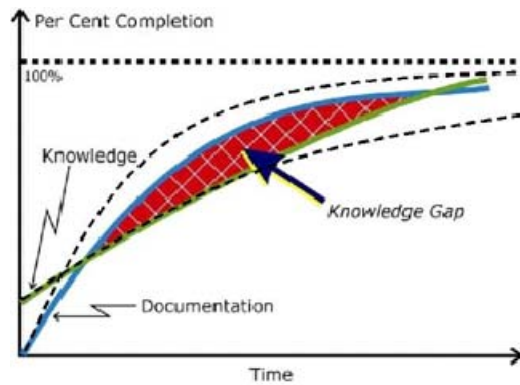


Fig. 5. Real size of Knowledge gap

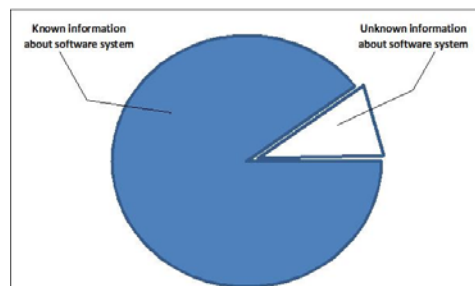


Fig. 6. Field of knowledge about software system with sector of the unknown information

Then the *actual task* is the assessment of information sufficiency as to software (for example, the possibility of obtaining of trustworthy information on the measures for calculation of the values of the software quality characteristics and subcharacteristics), on the basis of which software quality (by ISO 25010 [1]) is determined. Incompleteness and inaccuracy of such information lead to fall of veracity of software quality assessments. So the *purpose of this study* is the development the approach to the assessment of information sufficiency for software quality determination.

2 Ontological Approach to the Assessment of Information Sufficiency for Software Quality Determination

The most used model for software quality assessment is the model ISO 25010 [1]. The idea of this standard is that each of the characteristics is something we can analyse directly at the software product. Model ISO 25010 proposes to assess the software quality as a function of the eight characteristics, each of which is a function of several subcharacteristics (total 31 subcharacteristics). But subcharacteristics, in turn, are the functions of several measures. Analysis of [10-13], ISO 9126-2 [14], ISO 9126-3 [15] and revised on their basis ISO 25023 [16] provide the determination the dependence of quality subcharacteristics from the measures (total 203 measures). The basic idea is that the assessment of quality, its characteristics and subcharacteristics should be comprehensively performed, considering all these characteristics, subcharacteristics and measures accordingly.

Some of the measures are part of several quality subcharacteristics. So, if such measures are inaccurate or missing, then simultaneously use of these subcharacteristics in determining the several quality characteristics will significantly affect to the veracity of the software quality estimates. In such situation the condition of the mitigation of influence of these subcharacteristics cross-correlation is the important when using them in the quality models. Such mitigation is performed by identifying the joint measures, improving the accuracy of their values, or, if possible, limiting the simultaneous using of subcharacteristics that containing the same measures.

The information on determining the software quality characteristics and subcharacteristics is conveniently presented as semantic networks or other structures, which provide the displaying of the causal relationships between concepts. One of these structures is ontology. The advantages of ontology are the systematic approach to the study of the subject domain, the possibility of the holistic filing of known subject domain information, the identification of the overlaps and gaps in knowledge on the basis of the visualization of missing logical relationships.

Researchers *have already used the ontologies in software design*. E. Burov proposed methods and tools for development of software systems based on the ontological models [17, 18]. I. Shostak & I. Butenko developed the ontological models and methods of forming the profile during the software certification [19]. L. Babenko proposed the ontological approach to specifying the features of software systems and their components [20]. We use ontologies for assessment of information sufficiency for software quality determination.

For development and visualization of ontologies the large number of software tools, including universal, that provide the work with different subject domains, are today developed: Ontolingua Server, SMART, Protégé, OntoEdit, WebOnto, ODE (Ontological Design Environment), DOE (Differential Ontology Editor), CONE, OntoEditor +. The authors use a free software Protégé 4.2, which provides the work (creation, edition, vizualization and comparison) with ontologies of the various subject domains (<http://protege.stanford.edu/>).

First and foremost, the base ontology of the subject domain "Software Engineering" was developed. In it there is information about the software quality characteristics, subcharacteristics and measures. For development of this ontology the 8 software quality characteristics by ISO 25010 [1] (Functional Suitability, Reliability, Usability, Security, Performance Efficiency, Maintainability, Compatibility, Portability) were used. For determination of the Functional Suitability ISO 25010 proposed 3 subcharacteristics, which in turn are based on 15 measures. For determination of the Reliability ISO 25010 proposed 4 subcharacteristics, which are based on 30 measures. For determination of the Usability ISO 25010 proposed 6 subcharacteristics, which are based on 49 measures. For determination of the Security ISO 25010 proposed 5 subcharacteristics, which are based on 23 measures. For determination of the Performance Efficiency ISO 25010 proposed 3 subcharacteristics, which are based on 26 measures. For determination of the Maintainability ISO 25010 proposed 5 subcharacteristics, which are based on 33 measures. For determination of the Compatibility ISO 25010 proposed 2 subcharacteristics, which are based on 9 measures. For determination of the Portability ISO 25010 proposed 3 subcharacteristics, which are based on 18 measures. The idea of developed base ontology is shown on Fig. 7.

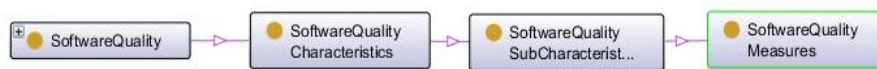


Fig. 7. Base ontology for subject domain "Software engineering" (part "Software quality")

The components of the base ontology are: base ontology for Functional Suitability (Fig. 8), the base ontology for Reliability (Fig. 10), the base ontology for Usability (Fig. 12), the base ontology for Security (Fig. 14), base ontology for Performance Efficiency (Fig. 16), the base ontology for Maintainability (Fig. 18), the base ontology for Compatibility (Fig. 20), the base ontology for Portability (Fig. 22).

The developed base ontology provides the following conclusions: 1) Functional Suitability: subcharacteristics Functional Completeness, Functional Appropriateness have 4 joint measures; Functional Appropriateness, Functional Correctness have 2 joint measures; 2) Reliability: subcharacteristics Maturity, Availability, Recoverability have 1 joint measure; Maturity, Fault Tolerance have 2 joint measures; Fault Tolerance, Recoverability have 1 joint measure; 3) Usability: subcharacteristics Learnability, Operability have 2 joint measures; Appropriateness Recognisability has 1 joint measure with the Learnability, Operability; Operability, User Error Protection have 1 joint measure; Operability, User Interface Aesthetics have 1 joint measure; 4) Security: subcharacteristics Confidentiality, Integrity have 8 joint measures; 5) Performance Efficiency: subcharacteristics Time Behaviour, Resource Utilization

have 2 joint measures; Time Behaviour, Capacity have 1 joint measure; 6) Maintainability: subcharacteristics Modularity, Modifiability have 3 joint measures; Testability has 2 joint measures with Modularity, Modifiability; Modularity, Analysability have 1 joint measure; Analysability, Modifiability have 1 joint measure; 7) Compatibility: subcharacteristics Co-existence, Interoperability have 1 joint measure; 8) Portability: subcharacteristics Adaptability, Replaceability have 2 joint measures; Adaptability, Installability have 1 joint measure.

In addition, there are measures, which are included in the formulas of several subcharacteristics of different characteristics (for example, measure Operation Time is included in subcharacteristics of all 8 quality characteristics). One of the basic properties of the base ontology is precisely the possibility of manifestation of cross-correlation of characteristics and subcharacteristics when using them in quality models. Because the important condition is the mitigation of the cross-correlation of such subcharacteristics when using them in quality models, therefore during assessment of the software quality it is necessary to pay special attention to those measures, which are part of simultaneously several subcharacteristics.

Ontological approach to the assessment of information sufficiency for software quality determination (by ISO 25010:2011 [1]) consists of the next steps: 1) analysis of the software requirements specification for the concrete software project for the presence of measures, that necessary for determining the quality characteristics and subcharacteristics of software project and software; 2) the development of ontology for determining the quality of the concrete software; 3) comparison of the developed ontology with base ontology for software quality determination, components of which are shown on Fig. 8, 10, 12, 14, 16, 18, 20, 22; 4) identification of measures, which are absent in the ontology for determination of the quality of the concrete software; 5) identification of quality characteristics and subcharacteristics, that cannot be calculated on the basis of the existing measures (at the same time should remember about the basic idea of ISO 25010 [1], which says that the quality assessment should be performed comprehensively, considering all quality characteristics; the assessment of quality characteristics also should be performed comprehensively, considering all subcharacteristics; the assessment of quality subcharacteristics, in turn, should be performed comprehensively, considering all measures); 6) the presence of subcharacteristics and characteristics, values of which cannot be determined on the basis of measures, that available in the software requirements specification, indicates the need to complement of this specification by the necessary measures (at this stage adding the necessary information and deleting other relevant information are possible); 7) repeating the steps 2-6 until all quality characteristics and subcharacteristics will be possible to identify or until the conclusion will be formed, that data for software quality determination are insufficient.

3 Experiments: Assessment of Information Sufficiency for Determination of Quality of Software of Automated System for Large-Format Photo Print

During the study the specification of automated system for large-format photo print was analyzed. On the basis of the specification analysis the available measures were

determined, that necessary for determining the quality characteristics and subcharacteristics of software project and software. These measures provides the development of ontology for determination of the quality of this software, consisting of the: ontology for Functional Suitability (Fig. 9), ontology for Reliability (Fig. 11), ontology for Usability (Fig. 13), ontology for Security (Fig. 15), ontology for Performance Efficiency (Fig. 17), ontology for Maintainability (Fig. 19), ontology for Compatibility (Fig. 21), ontology for Portability (Fig. 23) for concrete software project.

The comparison of the developed ontology for software project of automated system with fragments of the base ontology for subject domain "Software engineering" provides to find that in the ontology for project the 4 measures (Number of Functions, Operation Time, Number of Data Items, Number of Test Cases) are missing.

In addition, on the basis of the comparison of the ontology for software project of automated system for large-format photo print with base ontology was found that in the concrete ontology the data for determination of some quality characteristics and sub-characteristics are insufficient due to the absence of the above 4 measures.

Analysis of Fig. 8 and Fig. 9 provides the conclusion that the data for determination of all 3 subcharacteristics of Functional Suitability are insufficient. Therefore, none of subcharacteristics cannot be calculated, so Functional Suitability of software project cannot be determined too, and therefore the quality of the software project cannot be determined.

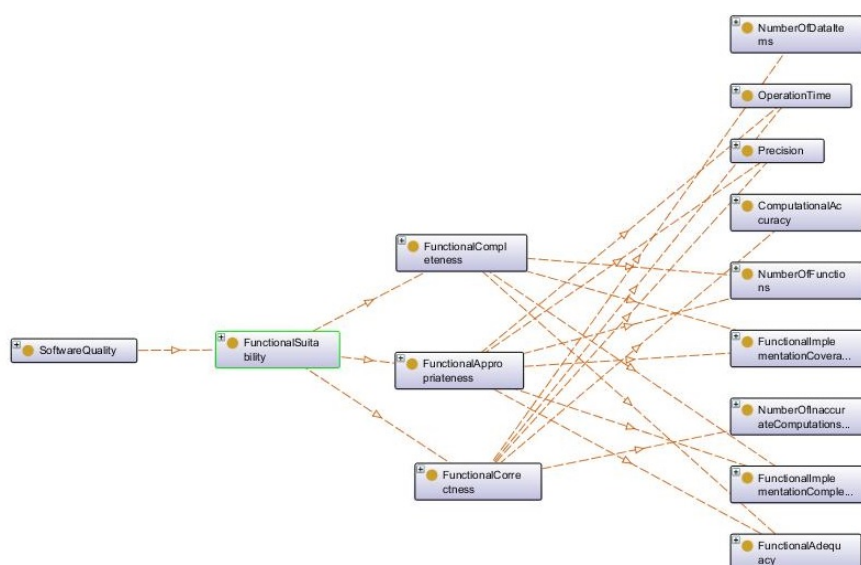


Fig. 8. Base ontology for Functional Suitability

Analysis of Fig. 10 and Fig. 11 provides the conclusion that the data for determination of all 4 subcharacteristics of Reliability are insufficient. Therefore, none of subcharacteristics cannot be calculated, so Reliability of software project cannot be determined, and therefore the quality of the software project cannot be determined.

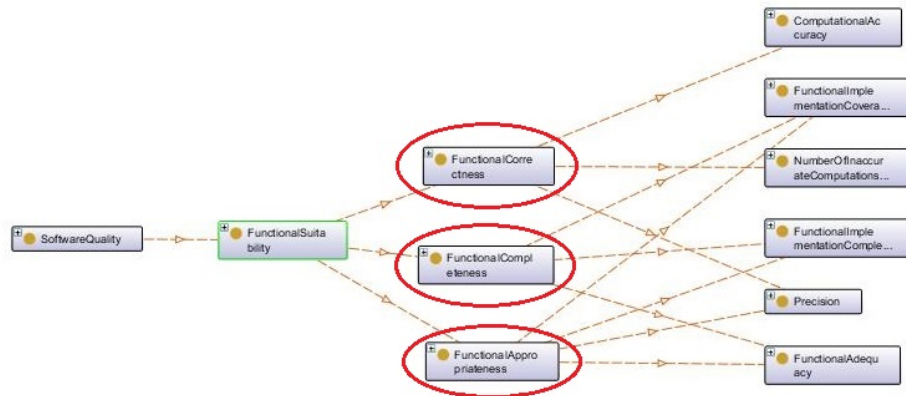


Fig. 9. Ontology for Functional Suitability for automated system for large-format photo print

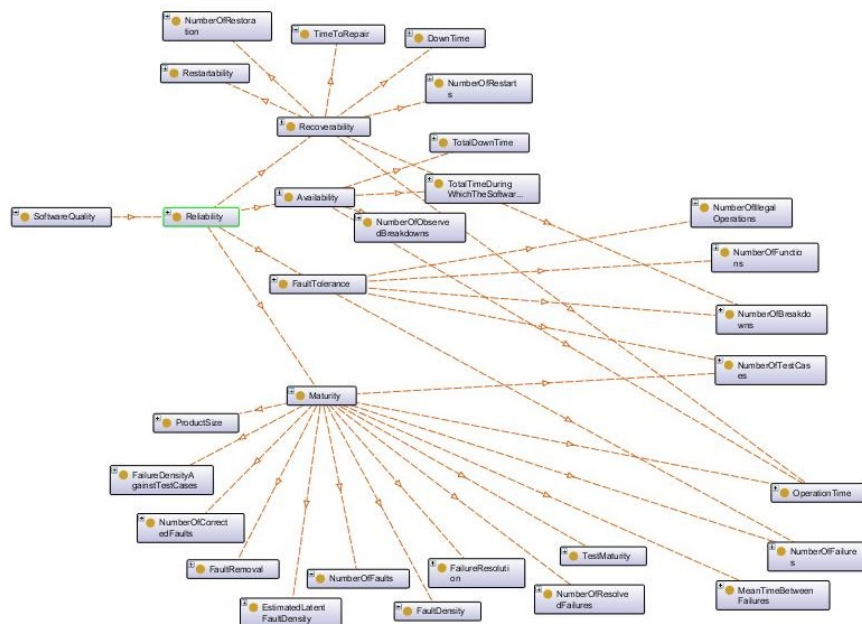


Fig. 10. Base ontology for Reliability

Analysis of Fig. 12 and Fig. 13 provides the conclusion that the data for determination of 3 from 6 subcharacteristics of Usability are insufficient. Therefore, 3 subcharacteristics cannot be calculated, so Usability of software project cannot be determined, and therefore the quality of the software project cannot be determined.

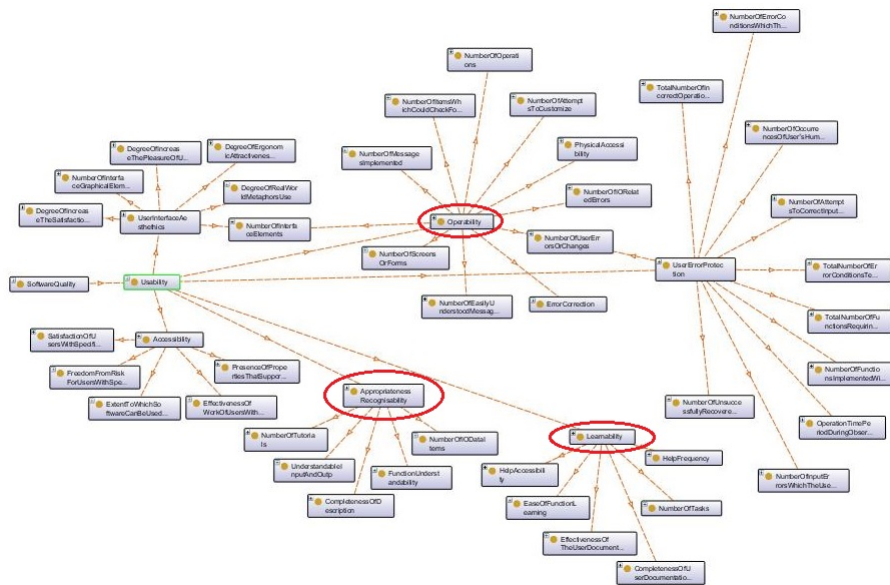


Fig. 13. Ontology for Usability for automated system for large-format photo print

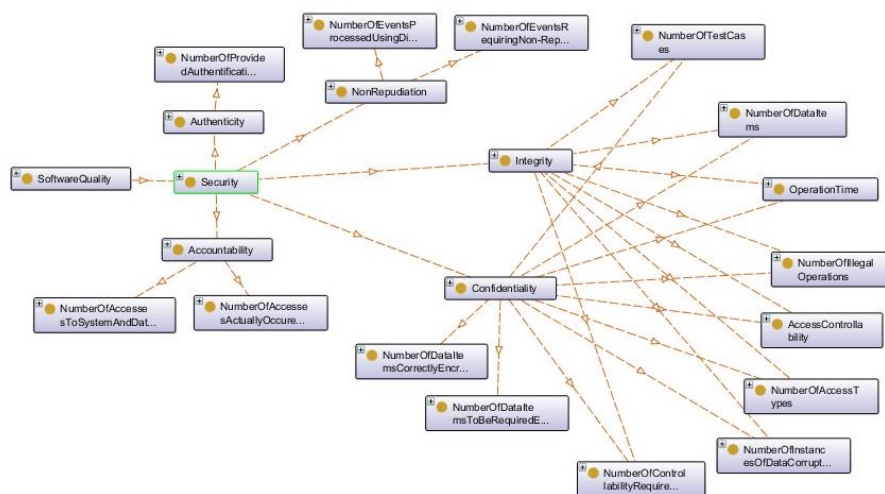


Fig. 14. Base ontology for Security

Analysis of Fig. 16 and Fig. 17 provides the conclusion that the data for determination of all 3 sub characteristics of Performance Efficiency are insufficient. Therefore, none of sub characteristics cannot be calculated, so Performance Efficiency of software project cannot be determined, and therefore the quality of the software project cannot be determined.

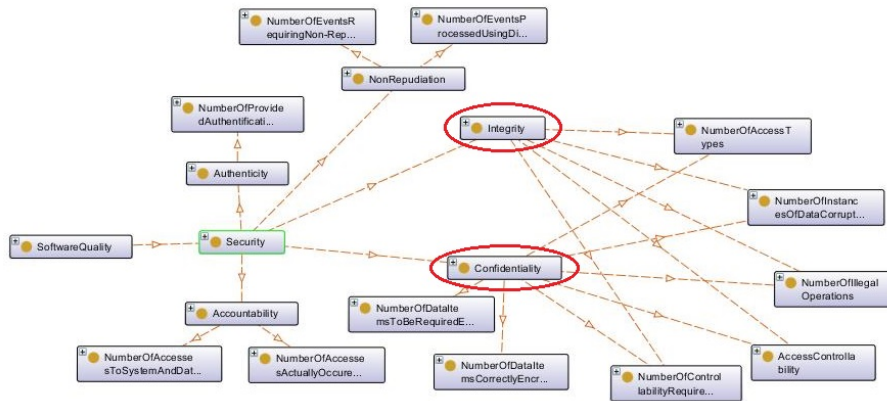


Fig. 15. Ontology for Security for automated system for large-format photo print

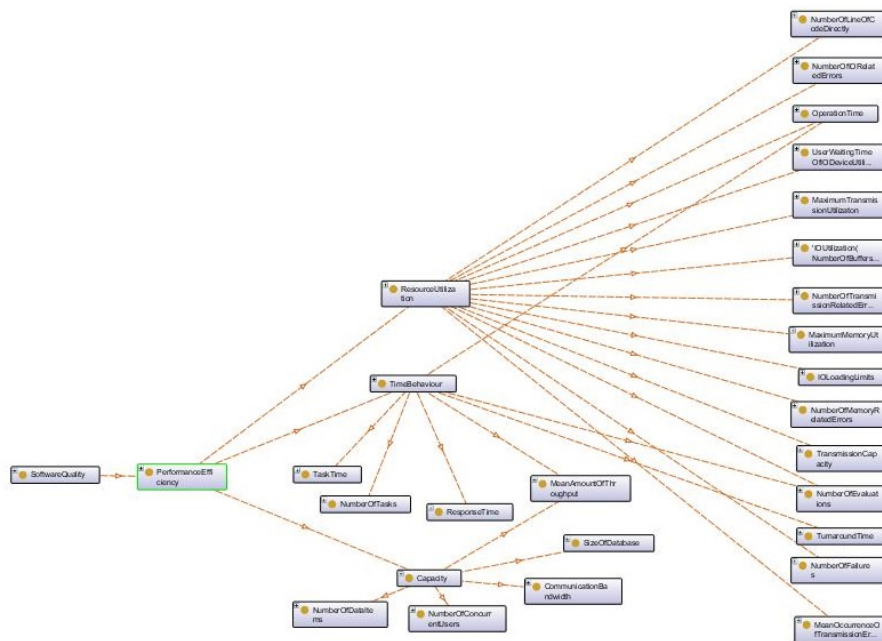


Fig. 16. Base ontology for Performance Efficiency

Analysis of Fig. 18 and Fig. 19 provides the conclusion that the data for determination of 4 from 5 subcharacteristics of Maintainability are insufficient. Therefore, 4 subcharacteristics cannot be calculated, so Maintainability of software project cannot be determined, and therefore the quality of the software project cannot be determined.

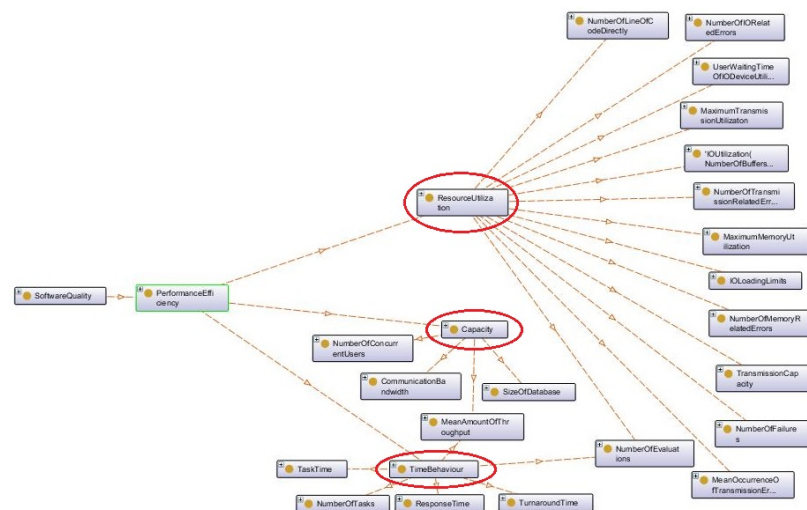


Fig. 17. Ontology for Performance Efficiency for automated system for large-format photo print

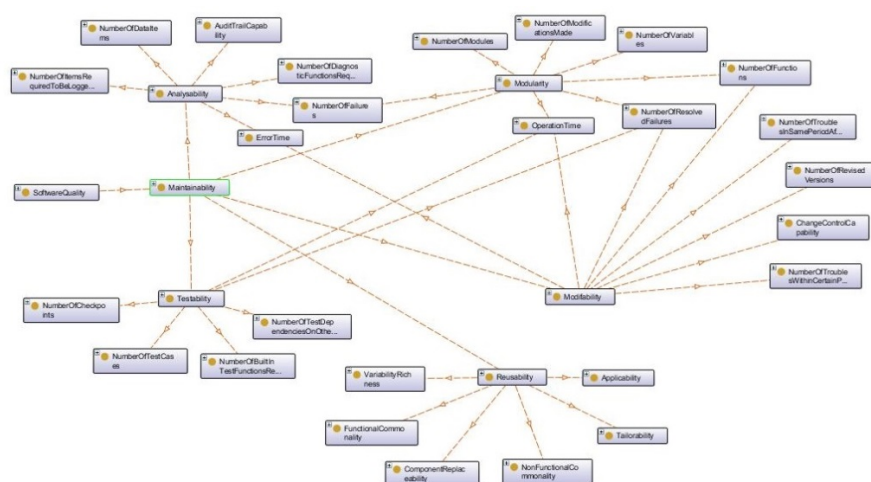


Fig. 18. Base ontology for Maintainability

Analysis of Fig. 20 and Fig. 21 provides the conclusion that the data for determination of all 2 subcharacteristics of Compatibility are insufficient. Therefore, none of subcharacteristics cannot be calculated, so Compatibility of software project cannot be determined, and therefore the quality of the software project cannot be determined.

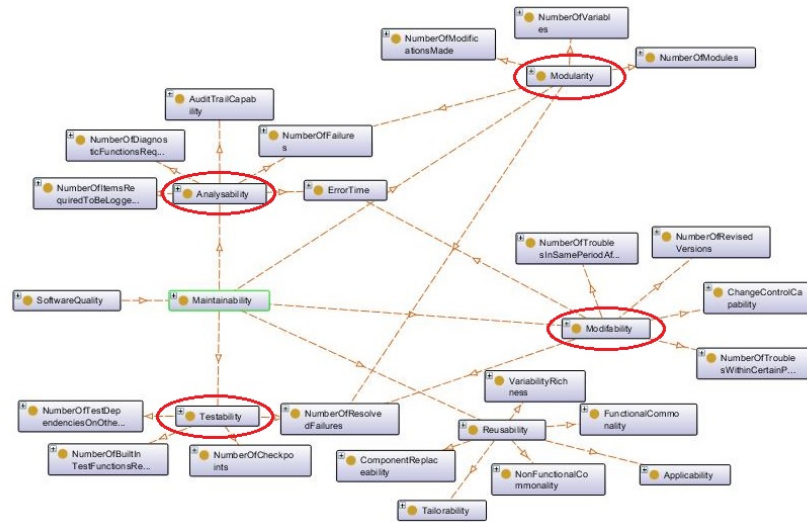


Fig. 19. Ontology for Maintainability for automated system for large-format photo print

Analysis of Fig. 22 and Fig. 23 provides the conclusion that the data for determination of 2 from 3 subcharacteristics of Portability are insufficient. Therefore, 2 subcharacteristics cannot be calculated, so Portability of software project cannot be determined, and therefore the quality of the software project cannot be determined.

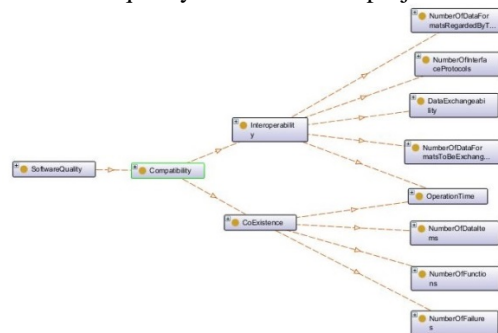


Fig. 20. Base ontology for Compatibility

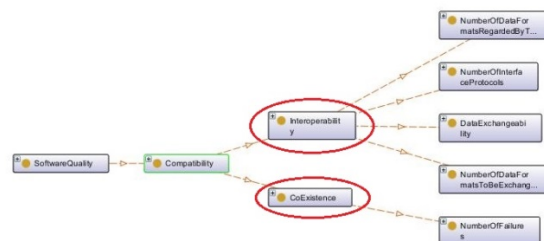


Fig. 21. Ontology for Compatibility for automated system for large-format photo print

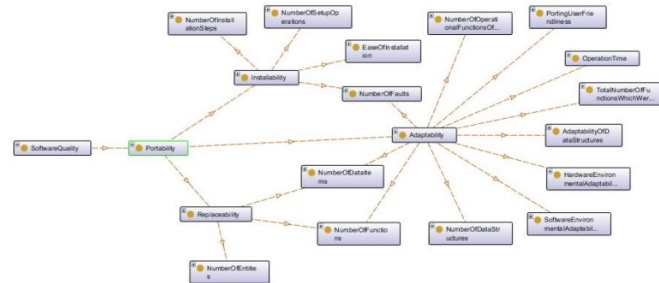


Fig. 22. Base ontology for Portability

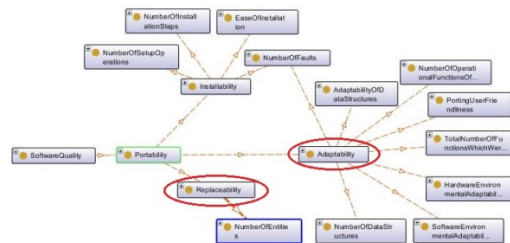


Fig. 23. Ontology for Portability for automated system for large-format photo print

Then the lack of 4 these measures in software requirements specification led to the impossibility of determination of all quality characteristics and the quality of the project and developed software.

For the concrete software project there are characteristics and sub characteristics, that are impossible to define or possible to insufficient define according to available information in the specification. Because the proposed approach to the assessment of information sufficiency for software quality determination is iterative, then the complement of the software requirements specification was conducted. The measures Number of Functions, Number of Data Items were added, and then the new version of the ontology for determination of the quality of the concrete software was created.

Comparative analysis of the new version of ontology with the base ontology showed that changes have occurred in the determination of Functional Completeness, Capacity, Appropriateness Recognisability, Analyzability, and Replaceability of the concrete software project. But the lack other 2 measures in specification (Operation Time, Number of Test Cases) leaves impossible the determination of all software quality characteristics and the quality of the project and developed software (still insufficient information). The process of complement the specification is iterative. But customer of developed automated system has decided that further complement of the specification is economically inexpedient therefore the conclusion about insufficient data for determination of the software quality was formed.

4 Conclusions

The measures analysis is an effective mean of assessing the software quality upon availability of veracity information for it conduct. One of the factors affecting the veracity of such information is sufficiency of the volumes of information about measures in the SRS. Therefore, solving the task of assessment of sufficiency information about measures in the SRS generally enhances the veracity of software quality assessment.

In the analysis of software quality subcharacteristics (as sources of information) the cross-correlation of these subcharacteristics because they have joint measures. Correlation of subcharacteristics, that displayed by base ontology, should be considered because it can reduce the accuracy and veracity of software quality assessment.

Knowledge of experienced professionals on interference and correlation of software quality subcharacteristics are valuable, so they should be stored and used in assessing the software specifications in terms of information sufficiency for software quality characteristics and subcharacteristics.

For displaying of these knowledge we selected ontologies that became the basis of the approach to the assessment of information sufficiency for software quality determination (according to ISO 25010: 2011).

The proposed ontological approach provides the development of recommendations for improvement of the software specification that illustrated by the example of the assessment of information sufficiency for determination of quality for software of automated system of large-format photo print.

References

1. ISO 25010:2011 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models (2011)
2. ISO 25030:2007 Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Quality requirements (2007)
3. ISO 9000:2015 Quality management systems -- Fundamentals and vocabulary (2015)
4. ISO 10002:2014 Quality management – Customer satisfaction – Guidelines for complaints handling in organizations (2014)
5. McConnell, S.: Code complete. Microsoft Press (2013)
6. The Standish Group International: CHAOS Report. Technical report, CHAOS Knowledge Center (2015)
7. Maier, R.: Knowledge Management Systems. Information and Communication Technologies for Knowledge Management (2013)
8. Patterson, Jr F.G.: Life cycles for system acquisition. Encyclopedia of Life Support Systems, Systems Engineering and Management for Sustainable Development, 82--110 (2004)
9. Pomorova, O., Hovorushchenko, T.: The Way to Detection of Software Emergent Properties. In: Proc. 2015 IEEE 8-th Int. Conf. on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. Vol.2, 779--784 (2015)
10. Abran, A., Al-Quitash, R.E., Desharnais, J.-M., Habra, N.: ISO-Based Models to Measure Software Product Quality. Software Quality Measurement: Concepts and Approaches. Chapter 5, 61--96 (2014)

11. Montagud Gregori, S.; Abrahao Gonzales, SM.; Insfrán Pelozo: A systematic review of quality attributes and measures for software product lines. *Software Quality Journal*. 20(3-4), 425--486 (2012)
12. Sun Her, J., Hyeok Kim, J., Hun Oh, S., Yul Rhew, S., Dong Kim, S.: A framework for evaluating reusability of core asset in product line engineering. *Information and Software Technology*. 49, 740--760 (2007)
13. Biscoglio, I., Marchetti, E.: Definition of Software Quality Evaluation and Measurement Plans: A Reported Experience Inside the Audio-Visual Preservation Context. In *Software Technologies: 9th International Joint Conference, ICSOFT 2014, Revised Selected Papers, CCIS 555*, 63--80 (2015)
14. ISO 9126-2:2003 Software engineering -- Product quality -- Part 2: External metrics (2003)
15. ISO 9126-3:2003 Software engineering -- Product quality -- Part 3: Internal metrics (2003)
16. ISO 25023:2015 Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Measurement of system and software product quality (2015)
17. Burov, E.: Complex ontology management using task models. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, 18(2), 111--120 (2014)
18. Burov, E., Pasitchnyk, V., Gritsyk, V.: Modeling software testing processes with task ontologies. *British Journal of Education and Science*, 2(6), 256--263 (2014)
19. Shostak, I., Butenko, I.: Ontology approach to realization of information technology for normative profile forming at critical software certification. *J. Military Institute of Kiev National University named after Taras Shevchenko*, 38, 250--253 (2012)
20. Babenko, L.: Ontological approach to specification of software systems features and components. *Cybernetics and System Analysis*, 1, 180--187 (2009) (in Russian)

Semi-Markov Availability Models for an Infrastructure as a Service Cloud with Multiple Pools

Oleg Ivanchenko¹, Vyacheslav Kharchenko²

¹ University of Customs and Finance,
2/4 Dzerzhinskogo, 217 km, Dnepropetrovsk, Ukraine

vmsul2@gmail.com

² National Aerospace University “Kharkiv Aviation Institute”
17 Chkalova St., 61070, Kharkiv, Ukraine

v_s_kharchenko@ukr.net

Abstract. Solving of tasks for Cloud Computing is impossible without maintaining of high availability level of Infrastructure as a Service (IaaS) Cloud. Several large IaaS Cloud providers try to solve this problem by means of increasing number of physical machines (PMs) in multiple pools. However, migrations of available PMs from one pool to another and also repairs, diagnostic instances of failed physical machines are quite complex task for modeling of availability for an IaaS Cloud. In this paper, we show how we can build Semi-Markov availability models with discrete states and how we can use it in order to determine availability level for the IaaS Cloud with three pools.

Keywords. Infrastructure as a Service Cloud, three pools of physical machines, Semi-Markov availability models

Key Terms. Infrastructure, Mathematical Model, Development, Characteristic

1 Introduction

Nowadays Cloud Computing is one of the most widely used services in an enterprise environment. Therefore availability of cloud infrastructures is of paramount importance to improve quality of service (QoS) and development of cloud user's possibilities. Despite this fact, researchers of cloud infrastructure behavior, including availability and reliability analysis of respective components are still regarded as quite complex scientific direction for different kinds of modeling.

Large Infrastructure as a Service (IaaS) Cloud providers try to use multiple pools of physical machines (PMs) in order to maintain normal operation of cloud's components on quite a long period of time. However, with a larger number of PMs number of states for stochastic model of an IaaS Cloud also increase; the model ought to in-

clude a large number of parameters while still being tractable [1]. Some famous researchers in order to perform availability analysis of large side clouds had built interacting sub-models, before they started to build monolithic model for an IaaS Cloud [2]. They built interacting sub-models on based Markov models and used stochastic reward nets. At the same time these sub-models were built by them with using of their own software package SHARPE [3]. Earlier in paper [4] authors tried to build a continuous time Markov chain (CTMC) availability model, for an example with two PMs in each pool. Another researcher also proposed the availability model with the failure of PMs, repairing process and employment of cold PMs in case of failure in running machines [5]. In this model PMs in each pool are modeled by a three-dimensional CTMC too. In spite of the fact that various authors used stochastic approach based Markov models to describe behavior of the physical machine pools, they couldn't get rigorous analytical expressions.

Focus of this paper is to build Semi-Markov availability models for the IaaS Cloud with three pools and different number of PMs in each pool.

2 Statement of the Researches Results

2.1 Metamodel for Availability Analysis of IaaS Cloud

Note that the architecture of an IaaS Cloud is not tied to a real cloud implementation [6]. Suppose that researchers have used a simple cloud infrastructure with certain number of PMs. To reduce power consumption, cooling and infrastructure costs, PMs are grouped into three pools such as: hot, warm and cold pools. Assume that hot pool consists of turned on and running PMs; warm pool contains turned on, although not ready physical machines; cold pool consists of turned off PMs. Moreover this architecture has certain number of virtual machines (VMs), which are deployed on PMs. Deployment of VMs on base PMs allows to reduce power consumption and to maintain enough high performance of the cloud implementation.

In difference from other, proposed concept for maintaining of availability for a cloud infrastructure bases on use of two additional systems, namely Technical State-Control System (TSCS) and Resource Provisioning Decision Engine (RPDE) [7]. Our IaaS Cloud should be used TSCS, which is working in monitoring and diagnostic modes. In this case, these modes as regarded as an organization form of constant control of the significant parameters that the determinate not only the PMs performability, but also affect cloud infrastructure readiness to make effective intended use [8]. It's obviously, that monitoring and diagnostic sub-systems provide repair facilitated by information which is needed to repair and migration of PMs from one pool to other. As described in [6], RPDE tries to find a PM that can accept the job provisioning.

Figure 1 shows the portions of the taxonomy metamodel for availability analysis of IaaS Cloud. Researchers in order to deal with the complexities of metamodeling should work in the paradigm of four models, such as scalability, performance, flexibility (elasticity), power consumption. Each model ensures the overall metamodel by input parameters, namely initial number of PMs for each pool (scalability model), power consumption for each PM (power consumption model), management metrics

values, search rates (flexibility model) and failure rates, repair rates, migration rates, number of repair facilities (reliability model). In other words output parameters of these models are input parameters for meta model. At the same time values of design and temporal parameters of such models can be experimentally measured. The stages of meta modeling are colorfully shown by this figure. According to the illustrated Fig. 1, we will try to create analytical models with considering states and stochastic changing of all times failures, repairs and migrations of PMs.

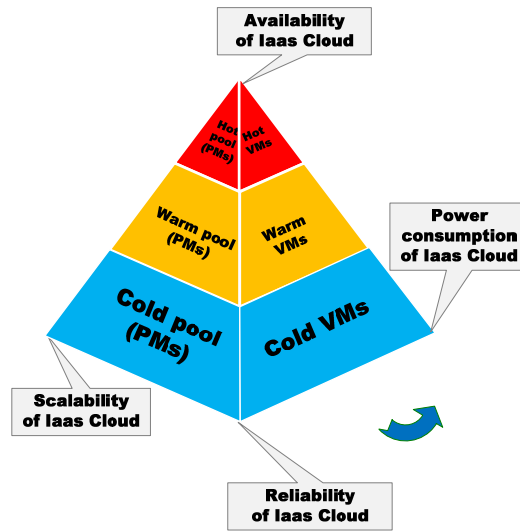


Fig. 1. Taxonomy metamodel for availability analysis of IaaS Cloud

On this basis, we will construct Semi-Markov model for availability analysis of an IaaS Cloud with three pools. Therefore it is proposed to describe various options of interactions of PMs at availability-model level.

2.2 Analytical Availability Models for an IaaS Cloud with Three Pools

Let's consider two interesting analytical models of an IaaS Cloud. Fig. 2 shows a Semi-Markov (SM) model for availability analysis of the IaaS Cloud with three pools (hot, warm and cold) and three PMs in each pool.

In our modeling we use the following assumptions and limitations.

- Hot, warm, and cold pools contain identical PMs [9]. If a hot PM fails the failed PM is replaced by available (non-failed) PM from warm or cold pools, respectively.
- We assume that periodic technical state control (CTS) of hot PMs is operated during a time interval, which lasts τ_c .
- To analyze the availability of the IaaS Cloud we also assume that all times to failure of all PMs are exponentially distributed. Typically, mean time to failure (MTTF) of warm PMs ($1/\lambda_w$) is higher than MTTF of hot PMs ($1/\lambda_h$) by a factor

of two to four [7]. At the same time MTTF of cold PMs is a very lower than $1/\lambda_w$. However, for process of SM modeling we will use only MTTF of hot PMs, considering quite high reliability level of warm and cold PMs.

- Moreover in real situations providers haven't enough time for repair of failed hot PMs, as well as they haven't enough number of repair facilities. Therefore we also assume that all times to repair are not exponentially distributed. In this occasion we have preferred to use Erlang- k distribution, where $k = 2, 3$ [10]. Parameter $1/\mu$ is mean time to repair (MTTR) of a PM.
- Available PMs can migrate from warm and cold pools to hot pool. We also assume that all times to migration (migration delays) of PMs are exponentially distributed. For modeling we have used mean time to migration (MTTM) of PMs from warm ($1/\gamma_{wh}$) and cold ($1/\gamma_{ch}$) pools to hot pool.
- The migrations of PMs to hot pool are implemented when providers can search non-failed warm or cold PMs with mean time to searches (MTTSs) $1/\delta_w$ and $1/\delta_c$.
- We consider that IaaS Cloud becomes unavailable when the SM model enters the state S_{I5} .

Suppose that this infrastructure is operated during a time interval $t \in [0, T]$ and at the initial moment $t = 0$ the IaaS Cloud is ready for using (state S_0). The transition from state S_0 to state S_I occurs at fixed nonrandom time $\tau_c < T$, where parameter T is operation time of IaaS Cloud between two periodic controls of technical state. The state S_I is state of CTS. Note that the periodic CTS includes monitoring and diagnostic operations of hot PMs and conduct by means of using Technical State Control System. If third hot PM is available the SM model returns from state S_I to state S_0 .

Otherwise when the TSCS detects a failure, model goes to state S_2 with rate λ_h . In state failure of the third hot PM, model tries to search non-failed warm PM (transition from state S_2 to state S_3) with rate δ_w or cold PM (transition from state S_2 to state S_4) with rate δ_c . When warm or cold PMs are available, model transforms from state S_3 to state S_0 or from state S_4 to state S_0 respectively.

If the warm and cold pools are empty, repair facility tries to recover the failed hot PM, that is model goes from state S_2 to state S_0 with repair rate μ . When recovery the third failed PM is impossible, model transforms from state S_2 to state S_5 with overall failure rate $3\lambda_h$. It means that next steps of modeling as regards states of $S_5 - S_9$ for second hot PM and states of $S_{10} - S_{12}$ for first hot PM are repeated. Note that in this case we can maintain that transition from state S_7 to state S_{10} and transition from state S_{12} to state S_{15} are implemented with overall failure rates $2\lambda_h$ and λ_h respectively. We also consider that the model will transition from state S_{12} to state S_{15} when the last hot PM fails.

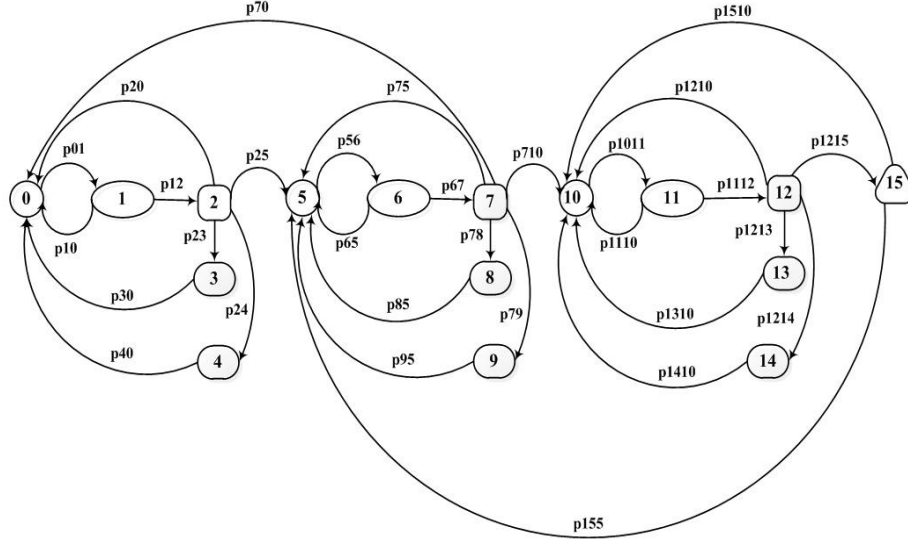


Fig. 2. SM model for the availability analysis of the IaaS Cloud with three PMs in each pool

To solve this task in the following we are inclined to use method of transformation of the SM models into embedded Markov chains [10]. For this type of models the transitions of process from state i to state j occur through unit time. Therefore the transitions of this SM process are interpreted as follows. Since CTS performs within fixed deterministic period of time T , consequently transition from state S_0 to state S_l is given by:

$$Q_{0l}(t) = \begin{cases} 0, & t < T, \\ 1, & t \geq T. \end{cases}$$

The transition from state S_l to state S_0 is then given by:

$$Q_{l0}(t) = \begin{cases} 0, & t < \tau_c, \\ 1, & t \geq \tau_c. \end{cases}$$

The other similar transitions can be got as follows:

$$Q_{56}(t) = Q_{1011}(t) = \begin{cases} 0, & t < T, \\ 1, & t \geq T, \end{cases}$$

$$Q_{65}(t) = Q_{1110}(t) = \begin{cases} 0, & t < \tau_c, \\ 1, & t \geq \tau_c. \end{cases}$$

At the same time, probabilities of sudden failures of hot PMs at random times for transitions from state S_l to state S_2 , from state S_6 to state S_7 and from state S_{11} to state S_{12} are given by:

$$Q_{l2}(t) = Q_{67}(t) = Q_{1112}(t) = 1 - e^{-\lambda_h t}.$$

Implementations of transitions from state S_2 to state S_0 , from state S_7 to state S_5 , from state S_{12} to state S_{10} and from state S_7 to state S_0 , from state S_{12} to state S_5 , from state S_{15} to state S_{10} depend from time to repair of the hot PMs. Therefore in these cases, distribution functions of repair time are given by:

$$\begin{aligned} Q_{20}(t) &= Q_{75}(t) = Q_{1210}(t) = 1 - (1 + \mu t)e^{-\mu t}, \\ Q_{70}(t) &= Q_{125}(t) = Q_{1510}(t) = 1 - \left(1 + \mu t + \frac{(\mu t)^2}{2}\right)e^{-\mu t}. \end{aligned}$$

For our SM availability model, we assume that distribution functions of search time of non-failed warm and cold PMs respectively are given by:

$$\begin{aligned} Q_{23}(t) &= Q_{78}(t) = Q_{1213}(t) = 1 - e^{-\delta_w t}, \\ Q_{24}(t) &= Q_{79}(t) = Q_{1214}(t) = 1 - e^{-\delta_c t}. \end{aligned}$$

Similarly, distribution functions of migration time for warm and cold PMs respectively are given by:

$$\begin{aligned} Q_{30}(t) &= Q_{85}(t) = Q_{1310}(t) = 1 - e^{-\gamma_{wh} t}, \\ Q_{40}(t) &= Q_{95}(t) = Q_{1410}(t) = 1 - e^{-\gamma_{ch} t}. \end{aligned}$$

Then steady-state availability [10] of the cloud can be computed as

$$A = \pi_0 + \pi_5 + \pi_{10}, \quad (1)$$

where π_0, π_5, π_{10} are steady-state probabilities for states S_0, S_5, S_{10} .

On the other hand, the steady-state availability A (1) is given by [11]:

$$A = \lim_{t \rightarrow \infty} A(t),$$

where $A(t)$ – instantaneous availability of the cloud infrastructure.

In the overall case steady-state probabilities of SM availability model are given by:

$$\begin{aligned} \pi_0 &= \frac{\bar{t}_0}{U}, \quad \pi_5 = \beta \frac{\bar{t}_5}{U}, \quad \pi_{10} = \alpha \beta \frac{\bar{t}_{10}}{U}, \\ U &= \bar{t}_0 + \bar{t}_1 + p_{12}(\bar{t}_2 + p_{23}\bar{t}_3 + p_{24}\bar{t}_4) + \beta[\bar{t}_5 + \bar{t}_6 + p_{67}(\bar{t}_7 + p_{78}\bar{t}_8 + p_{79}\bar{t}_9) + \\ &\quad + \alpha(\bar{t}_{10} + \bar{t}_{11})] + \varepsilon(\bar{t}_{12} + p_{1213}\bar{t}_{13} + p_{1214}\bar{t}_{14} + p_{1215}\bar{t}_{15}), \end{aligned}$$

where

$$\begin{aligned} \alpha &= \frac{p_{67}p_{710}}{1 - p_{1110} - \xi p_{1112}}, \quad \beta = \frac{p_{12}p_{25}}{1 - p_{65} - \alpha p_{1112}p_{125} - \nu p_{67}}, \\ \xi &= p_{1210} + p_{1213} + p_{1214} + p_{1215}, \quad \nu = p_{75} - p_{78} - p_{79}, \quad \varepsilon = \alpha \beta p_{1112}, \\ p_{12} &= p_{67} = p_{1112} = 1 - e^{-\lambda_h \tau_c}, \end{aligned}$$

$$\begin{aligned}
 p_{23} &= \delta_w \int_0^\infty (1 + \mu t) e^{-(3\lambda_h + \delta_w + \delta_c + \mu)t} dt, \quad p_{24} = \delta_c \int_0^\infty (1 + \mu t) e^{-(3\lambda_h + \delta_w + \delta_c + \mu)t} dt, \\
 p_{25} &= \lambda_h \int_0^\infty (1 + \mu t) e^{-(3\lambda_h + \delta_w + \delta_c + \mu)t} dt, \\
 p_{75} &= \frac{1}{2} \int_0^\infty \mu^2 t (\mu^2 t^2 + 2\mu t + 2) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt, \\
 p_{78} &= \delta_w \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt, \\
 p_{79} &= \delta_c \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt, \\
 p_{710} &= \lambda_h \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt, \quad p_{65} = p_{1110} = e^{-\lambda_h \tau_c}, \\
 p_{1210} &= \frac{1}{2} \int_0^\infty \mu^2 t (\mu^2 t^2 + 2\mu t + 2) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(\lambda_h + \delta_w + \delta_c + 2\mu)t} dt, \\
 p_{1214} &= \delta_c \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(\lambda_h + \delta_w + \delta_c + 2\mu)t} dt, \\
 p_{1215} &= \lambda_h \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(\lambda_h + \delta_w + \delta_c + 2\mu)t} dt, \\
 p_{125} &= 1 - p_{1210} - p_{1213} - p_{1214} - p_{1215}, \\
 \overline{t_0} = \overline{t_5} = \overline{t_{10}} &= T, \quad \overline{t_2} = \frac{3\lambda_h + \delta_w + \delta_c + 2\mu}{(3\lambda_h + \delta_w + \delta_c + \mu)^2}, \quad \overline{t_3} = \overline{t_8} = \overline{t_{13}} = \frac{1}{\gamma_{wh}}, \quad \overline{t_4} = \overline{t_9} = \overline{t_{14}} = \frac{1}{\gamma_{ch}}, \\
 \overline{t_{15}} &= \frac{3}{\mu}, \quad \overline{t_1} = \overline{t_6} = \overline{t_{11}} = \frac{1}{\lambda_h} (1 - e^{-\lambda_h \tau_c}), \\
 \overline{t_7} &= \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(2\lambda_h + \delta_w + \delta_c + 2\mu)t} dt, \\
 \overline{t_{12}} &= \int_0^\infty (1 + \mu t) \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-(\lambda_h + \delta_w + \delta_c + 2\mu)t} dt.
 \end{aligned}$$

Plots depending of steady-state availability A from failure rates λ_h of hot PMs and operation time T (repair rates μ are constant values) are shown in Fig. 3, Fig. 4. The values of steady-state availability A are greatly increased by means of increasing of repair rate μ and reduction of failure rate λ_h of hot PMs, as depicted in Fig. 3 and Fig. 4.

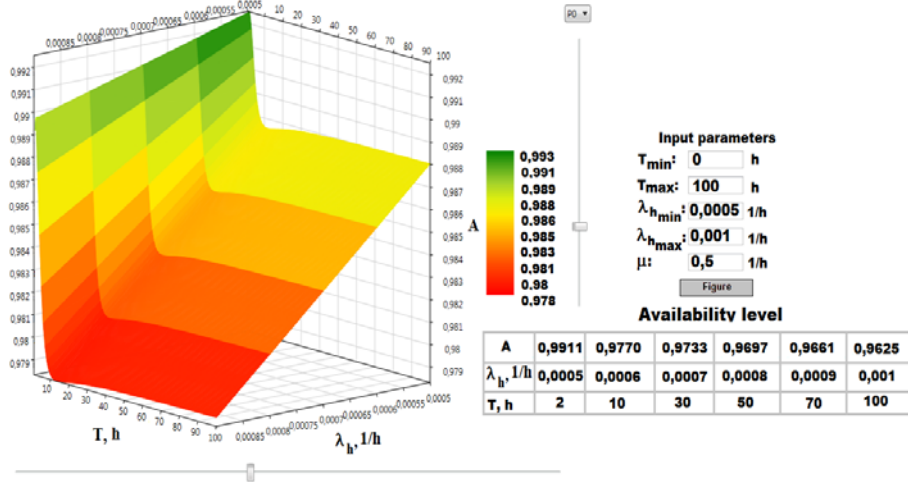


Fig. 3. Depending of steady-state availability $A(\lambda_h, T)$ for $T = 100$ h, $\mu = 0,5$ 1/h

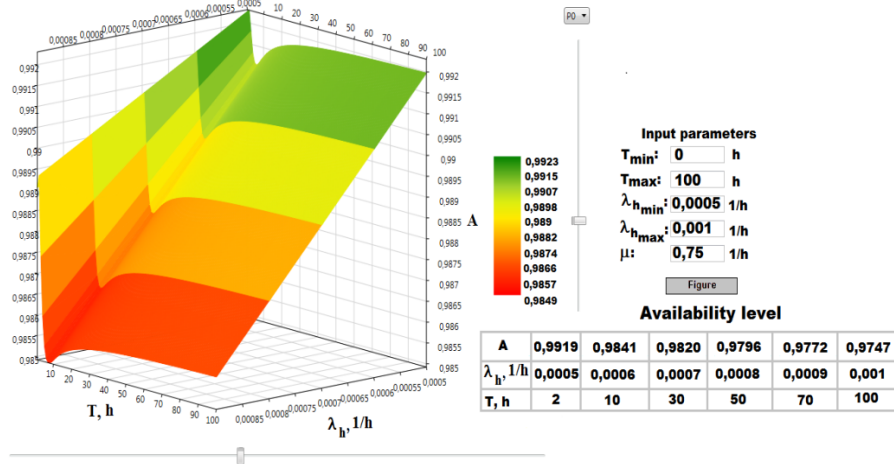


Fig. 4. Depending of steady-state availability $A(\lambda_h, T)$ for $T = 100$ h, $\mu = 0,75$ 1/h

Let's continue our researches by means of creation more scalable stochastic model for IaaS Cloud. Because with a larger number of PMs in a data center, the overall Cloud service availability increases, leading to lower cost of service downtime [7]. Therefore within a unified methodological approach we will try to create an improved SM availability model of infrastructure with a larger number of PMs.

Additional researches have shown that IaaS Cloud providers wish to increase number of PMs in order to minimize downtime cost and damage business reputation [4], [6], [7]. Perhaps inspired by using stochastic approaches for solution various

serious tasks of determining the optimal PM capacity configuration of IaaS Cloud [6], we have been proposing next SM availability model.

Assume that our infrastructure contains similar three pools with ten PMs in each pool. This SM model for availability analysis of the IaaS Cloud is shown in Fig. 5. Also suppose that all times to failure of PMs are exponentially distributed and Erlang- k distribution, where $k = 2$ is general distribution for all times to repair. In spite of the fact, that both models are SMs models, we have to take into consideration some interesting features of their implementation.

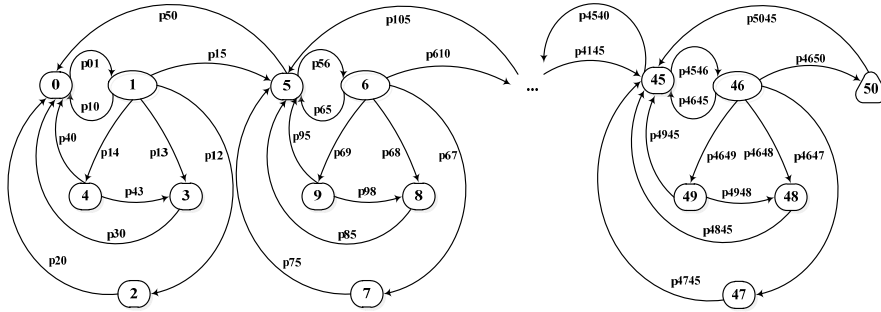


Fig. 5. SM model for the availability analysis of the IaaS Cloud with ten PMs in each pool

Unlike first SM model, second SM availability model of the IaaS Cloud includes modeling kernel from five states. The states S_0 , S_I , S_5 for second model (Fig. 5) are the same as the first model (Fig. 2). But the difference between kernels of first model and second model is that states S_4 , S_9 for first model are states of search of the cold PM, whilst these states for second model are states of failure of the warm PMs. For second model the following group assumptions can take place.

- Model contains hot, warm, and cold pools. Every pool consists of ten identical PMs[9]. If a hot PM fails the failed PM is replaced by available (non-failed) PM from warm or cold pools too.
- Upon failure of the warm PM, the failed PM is replaced by available (non-failed) PM from cold pool.
- We also assume that periodic technical state control (CTS) of hot PMs is operated during a time interval, which lasts τ_c .
- To analyze performance and availability of the IaaS Cloud we also assume that all times to failure of all hot and warm PMs are exponentially distributed.
- We also consider that all times to repair are not exponentially distributed. In this case we have used Erlang- k distribution, where $k = 2, 3$ [10]. Parameter $1/\mu$ is mean time to repair (MTTR) of a PM.
- Cloud infrastructure becomes unavailable when the SM model enters the state S_{50} .

Therefore the transitions of modeling kernel for second SM model can be written as follows:

$$\begin{aligned}
 Q_{01}(t) &= Q_{56}(t) = Q_{1011}(t) = Q_{1516}(t) = Q_{2021}(t) = Q_{2526}(t) = Q_{3031}(t) = Q_{3536}(t) = \\
 &= Q_{4041}(t) = Q_{4546}(t) = \begin{cases} 0, t < T, \\ 1, t \geq T, \end{cases} \\
 Q_{10}(t) &= Q_{65}(t) = Q_{1110}(t) = Q_{1615}(t) = Q_{2120}(t) = Q_{2625}(t) = Q_{3130}(t) = Q_{3635}(t) = \\
 &= Q_{4140}(t) = Q_{4645}(t) = \begin{cases} 0, t < \tau_c, \\ 1, t \geq \tau_c, \end{cases}
 \end{aligned}$$

For other functions we can write the following:

$$\begin{aligned}
 Q_{12}(t) &= Q_{67}(t) = Q_{1112}(t) = Q_{1617}(t) = Q_{2122}(t) = Q_{2627}(t) = Q_{3132}(t) = Q_{3637}(t) = \\
 &= Q_{4142}(t) = Q_{4647}(t) = 1 - e^{-\delta_w t}, \\
 Q_{13}(t) &= Q_{68}(t) = Q_{1113}(t) = Q_{1618}(t) = Q_{2123}(t) = Q_{2628}(t) = Q_{3133}(t) = Q_{3638}(t) = \\
 &= Q_{4143}(t) = Q_{4648}(t) = 1 - e^{-\delta_c t}, \\
 Q_{14}(t) &= Q_{69}(t) = Q_{1114}(t) = Q_{1619}(t) = Q_{2124}(t) = Q_{2629}(t) = Q_{3134}(t) = Q_{3639}(t) = \\
 &= Q_{4144}(t) = Q_{4649}(t) = 1 - e^{-\lambda_w t}, \\
 Q_{20}(t) &= Q_{75}(t) = Q_{1210}(t) = Q_{1715}(t) = Q_{2220}(t) = Q_{2725}(t) = Q_{3230}(t) = Q_{3735}(t) = \\
 &= Q_{4240}(t) = Q_{4745}(t) = 1 - e^{-\gamma_{wh} t}, \\
 Q_{30}(t) &= Q_{85}(t) = Q_{1310}(t) = Q_{1815}(t) = Q_{2320}(t) = Q_{2825}(t) = Q_{3330}(t) = Q_{3835}(t) = \\
 &= Q_{4340}(t) = Q_{4845}(t) = 1 - e^{-\gamma_{ch} t}, \\
 Q_{43}(t) &= Q_{98}(t) = Q_{1413}(t) = Q_{1918}(t) = Q_{2423}(t) = Q_{2928}(t) = Q_{3433}(t) = Q_{3938}(t) = \\
 &= Q_{4443}(t) = Q_{4948}(t) = 1 - e^{-\delta_c t}, \\
 Q_{40}(t) &= Q_{95}(t) = Q_{1410}(t) = Q_{1915}(t) = Q_{2420}(t) = Q_{2925}(t) = Q_{3430}(t) = Q_{3935}(t) = \\
 &= Q_{4440}(t) = Q_{4945}(t) = 1 - (1 + \mu t) e^{-\mu t}, \\
 Q_{50}(t) &= Q_{105}(t) = Q_{1510}(t) = Q_{2015}(t) = Q_{2520}(t) = Q_{3025}(t) = Q_{3530}(t) = Q_{4035}(t) = \\
 &= Q_{4450}(t) = Q_{5045}(t) = 1 - \left(1 + \mu t + \frac{(\mu t)^2}{2} \right) e^{-\mu t}, \\
 Q_{15}(t) &= 1 - e^{-\lambda_1 t}, \tag{2} \\
 Q_{610}(t) &= 1 - e^{-\lambda_2 t}, \tag{3} \\
 Q_{1115}(t) &= 1 - e^{-\lambda_3 t}, \tag{4} \\
 Q_{1620}(t) &= 1 - e^{-\lambda_4 t}, \tag{5} \\
 Q_{2125}(t) &= 1 - e^{-\lambda_5 t}, \tag{6} \\
 Q_{2630}(t) &= 1 - e^{-\lambda_6 t}, \tag{7} \\
 Q_{3135}(t) &= 1 - e^{-\lambda_7 t}, \tag{8} \\
 Q_{3640}(t) &= 1 - e^{-\lambda_8 t}, \tag{9} \\
 Q_{4145}(t) &= 1 - e^{-\lambda_9 t}, \tag{10}
 \end{aligned}$$

$$Q_{4650}(t) = 1 - e^{-\lambda_{10}t}. \quad (11)$$

We define the failure rates for $j = 1, 2, \dots, n_h$ ($n_h = 10$) PMs nodes[9], [10]:

$$\lambda_j = (n_h - i)\lambda_0, i = 0, 1, \dots, k \text{ (for } k = n_h - 1 \text{)}, \quad (12)$$

where λ_0 – basic failure rate value for all PMs.

By replacing the λ_j expression (12) to the λ_s values in the equations (2), (3), ..., (11), we will be finished description of second model.

As can be seen in Fig. 3 and Fig. 4 in case with three PMs in each pool, IaaS Cloud has quite high of availability level. Results of modeling for second SM availability model will get in the near future time. Overall feature for both SM models is identical modeling kernels.

3 Conclusions Statement of the Researches Results

Clearly, proposed stochastic approach based SM models gives opportunity to perform availability analysis of the Cloud Infrastructures with using of different modeling kernels. Thus, the contributions of this paper are the following.

If you wanted to make a deep availability and reliability analysis of the IaaS Cloud, for example, when this infrastructure is one of the most important components of Management System Critical Infrastructure, in particular during the accidents and disasters or other negative events, such as sudden or hidden failures, you would be able to use proposed SMs availability models. An additional advantage of these SMs models is that researches can use rigorous analytical expressions from this paper in order to determine availability and reliability values for the IaaS Cloud. Moreover you can use this stochastic approach in order to choose optimal architectures among the many various Cloud Infrastructures. Several optimization problems, including capacity planning, management of resources of Cloud Infrastructures can be solved using stochastic approach and SM models described in this paper.

References

1. Khazaei, H., Misic, J., Misic, V.B., Mohammadi, N.B.: Availability analysis of cloud computing centers. In: Globecom 2012 – Communications Software, Services and Multimedia Symposium (GC12 CSSM) (2012)
2. Longo, F., Ghosh, R., Naik, V.K., Trivedi, K.S.: A scalable availability model for infrastructure-as-a-service cloud. In: Proc. Int Conf on Dependable Systems and Networks, pp. 335--346 (2011)
3. Trivedi, K.S., and Sahner, R.: SHARPE at the age of twenty two. ACM Sigmetrics Performance Evaluation Review, 36(4), 52--57 (2009)
4. Ghosh, R., Trivedi, K.S., Naik, V.K., Kim, D.S.: End-to-end performability analysis for infrastructure-as-a-service cloud: An Interacting Stochastic Models Approach. In: IEEE PRDC, Tokyo (2010)

5. Khazaei, H.: Performance modeling of cloud computing centers. Diss. The University of Manitoba (2013)
6. Ghosh, R., Longo, F., Xia, R., Naik, V., Trivedi, K.: Stochastic model driven capacity planning for an infrastructure-as-a-service cloud. *IEEE Transactions on Services Computing*, 7(4), 667--680 (2014)
7. Ghosh, R.: Scalable stochastic models for cloud services. Diss. Duke of University (2012)
8. Ivanchenko, O., Kharchenko, V., Skatkov, A.: Management of critical infrastructures Based on Technical Megastate. *Int. J. "Information and Security"*, 28(1), 37--51 (2012)
9. Ghosh, R., Longo, F., Frattini, F., Russo, S., Trivedi, K.S.: Scalable analytics for IaaS cloud availability. *IEEE Transactions on Cloud Computing*, 2(1), 57--70 (2014)
10. Ivanchenko, O., Lovyagin, V., Maschenko, E., Skatkov, A., Shevchenko, V.: Distributed critical systems and infrastructures. National Aerospace University named after N. Zhukovsky "KhAI", Kharkiv, Ukraine (2013)
11. Abbadi, I.M.: Toward Trustworthy Clouds' Internet Scale Critical Infrastructure. In: Bao, F., Weng J. (eds.) *Information Security Practice and Experience*. LNCS-6672, pp. 71--82. Springer Verlag, Heidelberg (2011)

Application of Structure Function in System Reliability Analysis based on Uncertain Data

Vitaly Levashenko, Elena Zaitseva, Miroslav Kvassay

University of Zilina, Department of Infomatics,
Univerzitna 8215/1, 010 26, Zilina, Slovakia

{elena.zaitseva, vitaly.levashenko,
miroslav.kvassay}@fri.uniza.sk

Abstract. An important step in reliability evaluation of any system is selection of an appropriate mathematical representation. One of the possible mathematical representations is structure function that expresses dependency of system state on states of its components. This function must be completely specified for reliability evaluation of the analyzed system. The structure function is constructed based on complete information about the system structure and possible components states. However, there are a lot of practical problems when the complete information is not available because data from which it can be derived cannot be collected. As a rule, other mathematical representations and methods for evaluation of system reliability are used in these situations. In this paper, we propose a new method for construction of the structure function from uncertain or incomplete data. This method is developed based on application of Fuzzy Decision Tree.

Keywords: Fuzzy Decision Tree, Multi-State System, Structure Function, Uncertainty

Key Terms. Reliability, Model, Approach, Methodology, ScientificField

1 Introduction

As has been shown in paper [1], selecting a mathematical representation of an analyzed system is an important step in reliability analysis. Depending on the number of performance levels, two types of models can be recognized. These models are named as *Binary-State Systems* (BSSs) and *Multi-State Systems* (MSSs).

A BSS admits only two states in investigation of the system and its components: perfect functioning and complete failure. However, in practice, many systems can go through different performance levels between these two extreme states [1, 2]. A MSS is a mathematical model that is used to describe such systems since it allows defining more than two levels of performance [2, 3, 4].

There are different types of mathematical representations of a system. In reliability engineering, structure function, fault trees, reliability block diagrams, Markov models

and Petri nets are typically used for the mathematical representation of real systems under study. Historically, mathematical models based on the structure function have been proposed firstly. In this case, a system is modeled as a mapping that assigns system state to all possible combinations of component states. The system performance level is known based on the states of all its components. This interpretation of the structure function supposes the exact definition of all possible states of the system and its components. Therefore, any uncertainty cannot be considered and taken into account. However, this indicates that methods based on the structure function approach have some difficulties in application on real-world problems because, as a rule, data about behavior of such systems are uncertain. Two approaches can be used to solve this problem.

The first of them is development of a new model that takes uncertainties into account [5, 6, 7]. The application of a new mathematical model leads to a development of new mathematical methods for the analysis of this model. The second solution is to use one of the traditional models and develop new methods for construction of the structure function that will take uncertainties of the initial data into account.

Specifics of the uncertainty have to be analyzed before the development of the new method for the structure function construction based on the uncertain data. There are different factors of uncertain data. In our investigation, we will take into account two of them. The first are ambiguity and vagueness of initial data. It means that initial data about the system operation are collected based on (a) measurement that can be inaccurate and with an error or (b) experts that can have different opinions on one situation. Therefore, values of states of the components or system performance level cannot be indicated as exact (integers). Ambiguity and vagueness in a real system have been studied using the probability theory. However, it is worth pointing out that some uncertainties that are not random in nature can play important roles in construction of the structure function [5, 6, 8]. The fuzzy logic makes it possible to define the structure function in a more flexible form for such data than the probabilistic approach. So, non-exact values are the first factor of the uncertainty of initial data, and it can be expressed using fuzzy values.

Secondly, situations in which it is impossible to indicate some values of the system components states or performance level can exist. For example, it can be very expensive, or it needs unacceptable long time. This implies that some information about the system behavior can be absent. Therefore, the data are incomplete.

Based on the previous text, we have a task of construction of exact and completely specified structure function based on uncertain and incomplete data, what is a typical problem of Data Mining [9]. One of the approaches used for solving this problem is application of *Fuzzy Decision Trees* (FDTs), which are widely used in Data Mining for analysis of uncertain data and decision making in ambiguities [10, 11].

In this paper, we propose a method based on the application of an FDT for construction of the structure function. FDTs allow taking into account uncertainties of two types. The first of them is ambiguity of initial data. This can occur when it is expensive to obtain all data about real system behavior, or there are poorly documented data. This type of uncertainty is covered by fuzzy values in an FDT. It means that initial data can be defined and interpreted with some possibility and might not be

exact. The second type of uncertainty agrees with incompletely specified initial data. As a rule, if the exact values of the actual data about the system behavior cannot be determined, we need to rely on more data to get additional information necessary to correct the used theoretical model [6, 12]. An FDT allows reconstructing these data with different levels of the confidence [10, 11].

This paper is structured as follows. Section 2 discusses the concept of the structure function. The principal steps of the proposed method are considered in sections 3 – 5. These steps are Collection of data into a repository (section 3), Representation of the system model in the form of an FDT (section 4), and Construction of the structure function based on the FDT (section 5).

2 Structure function of the system

The *structure function* as a mathematical model was introduced in reliability engineering as one of the firsts [13]. This function captures the relationships between components of the system and the system itself in such a way that the state of the system is known based on the states of its components through the structure function.

Let us suppose that the system can be divided into n components (subsystems). A state of each component can be denoted by a random variable x_i that can be in one of m_i possible values. This variable takes value 0 if the component fails and one of values $1, \dots, m_i - 1$ if the component works satisfactorily.

Let us denote the structure function as $\phi(\mathbf{x})$. Then it agrees with the next map:

$$\phi(\mathbf{x}) = \phi(x_1, \dots, x_n): \{0, \dots, m_1 - 1\} \times \dots \times \{0, \dots, m_n - 1\} \rightarrow \{0, \dots, M - 1\}, \quad (1)$$

where $\phi(\mathbf{x})$ defines system state from complete failure ($\phi(\mathbf{x}) = 0$) to perfect functioning ($\phi(\mathbf{x}) = M - 1$); $\mathbf{x} = (x_1, \dots, x_n)$ is a state vector; x_i is the i -th component state that changes from complete failure ($x_i = 0$) to perfect functioning ($x_i = m_i - 1$).

Next, let us suppose that the system is coherent. This means: (a) the system structure function is monotone: $\phi(x_i, \mathbf{x}) \leq \phi(x_j, \mathbf{x})$ for any $x_i \leq x_j$; and (b) there are no irrelevant components in the system.

Every system component is characterized by the probabilities of individual states:

$$p_{i,s} = \Pr\{x_i = s\}, s = 0, \dots, m_i - 1. \quad (2)$$

Please note that the structure function of MSS (1) is transformed into the structure function of BSS if $m_i = M = 2$.

Many reliability indices and measures can be calculated based on the system structure function. One of them is the probability of the system performance level that is calculated as follows [3]:

$$A_j = \Pr\{\phi(\mathbf{x}) = j\}, j = 0, \dots, m_i - 1. \quad (3)$$

The structure function also allows calculating the boundary system states [14], minimal cut/path sets [15] and importance measures [16]. However, defining structure function as equation (1) for a real application can be a difficult problem.

As a rule, the structure function can be defined as a result of the system structure analysis or based on expert data [12, 17]. In system structure analysis, the system is interpreted as a set of components (subsystems) with correlations. These correlations can be defined by functional relations that are interpreted as the structure function (1). However, there are many structure-complex systems for which correlations and/or connections of components are hidden or uncertain (e.g. power systems, network systems). As a rule, other methods are used in reliability estimation for such systems [5, 18]. Construction of a structure function based on the expert data requires special analysis and transformation of initial data [12, 19]. We suggest the new method for construction of the structure function (1) that is based on the application of an FDT.

In terms of Data Mining, the structure function can be interpreted as a table of decisions [9, 20], where state vector $\mathbf{x} = (x_1, \dots, x_n)$ is interpreted as a set of input attributes and value of the structure function as an output attribute. This table of decisions can be constructed based on an FDT for all combinations of the input attributes. So, values of the structure function can be defined for all combinations of component states using the FDT: component states are interpreted as FDT attributes, and the structure function value agrees with one of M values (classes) representing system performance levels. The FDT is induced based on some samples (not all) of the inputs and output attributes. In case of construction of the structure function, the samples are state vectors with the corresponding function value. These samples have to be collected as initial information about the system.

The method proposed in this paper includes the following steps:

- collection of data into the repository according to requests of FDT induction;
- representation of the system model in the form of an FDT that classifies components states according to the system performance levels;
- construction of the structure function as a decision table that is created by induced FDT.

The structure function is constructed as a decision table that classifies the system performance level for each possible combination of components states. The decision table is formed based on the FDT that provides the mapping for all possible components states (input data) in M performance levels. The FDT is induced using uncertain data that are presented in the form of a specified repository.

3 Data repository construction

Collection of data in the form of a repository is provided by the monitoring of values of system component states and system performance level. This repository can be presented in the form of a table where the columns agree with the input and output attributes. The number of the input attributes is n and the i -th has m_i possible values (the i -th column includes m_i sub-columns). Every row contains a real sample of components states and the corresponding system performance level.

For example, let us consider the offshore electrical power generation system presented in [2]. The purpose of this system (Fig. 1) is to supply two nearby oilrigs with

electric power. The system includes 3 generators: two main generators A_1 and A_3 , and standby generator A_2 . Both main generators are at oilrigs. In addition, oilrig 1 has generator A_2 that is switched into the network in case of outage of A_1 or A_3 . The control unit U continuously supervises the supply from each of the generators with automatic control of the switches. If, for instance, the supply from A_3 to oilrig 2 is not sufficient, whereas the supply from A_1 to oilrig 1 is sufficient, U can activate A_2 to supply oilrig 2 with electric power through the standby subsea cables L . This implies that the system consists of 5 relevant components ($n = 5$): generators A_1 , A_2 , and A_3 , control unit U , and the standby subsea cables L . Furthermore, according to the description of the system activity in [2], we assume that the system and all its components have 3 states/performance levels ($M = 3$ and $m_i = 3$, for $i = 1, \dots, 5$). Next, let us denote variables defining states of the system components in the following way: main generators A_1 and A_3 as x_1 and x_3 respectively, standby generator A_2 as x_2 , and control unit U and standby subsea cables L as x_4 and x_5 respectively.

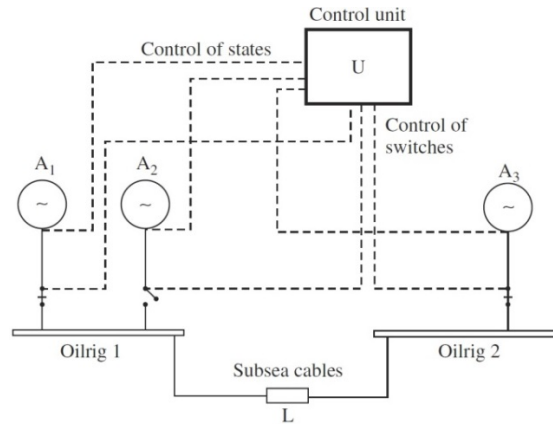


Fig. 1. Outline of the offshore electrical power generation system [2]

Let us suppose monitoring of the offshore power generation system that allowed collecting 108 (from 243 possible) samples of the system behavior. Some of them are shown in Table 1. The monitoring of this system permitted obtaining information about some combinations of component states and the corresponding performance levels of the system. However, this information is not complete because the data from the real monitoring are uncertain. This uncertainty is caused by the ambiguity of classification of component states and system performance levels into classes of exact values [12, 20]. Therefore, special type of data is used to define values of the input and output attributes in the repository. These data can be interpreted as quasi-fuzzy data that describe occurrence of every value of every attribute with some possibility ranging from 0 to 1. For example, the first row in Table 1 indicates the nonworking ($x_1 = 0$) and insufficient ($x_1 = 1$) states of generator A_1 with possibility of 0.8 and 0.2 respectively, while the possibility of the working state ($x_1 = 2$) is 0. In case of stable generator A_2 , the state is indicated as nonworking ($x_2 = 0$) with possibility of 0.8 and as other values ($x_2 = 1$ and $x_2 = 2$) with possibilities of 0.1. State of main generator A_3

is nonworking ($x_3 = 0$) with possibility 0.7 and insufficient ($x_3 = 1$) or working ($x_3 = 2$) with possibilities 0.2 and 0.1 respectively. States of control unit U are defined as $x_4 = 0$ with possibility 0.8, $x_4 = 1$ with possibility 0.2 and $x_4 = 2$ with possibility 0. Only 2 of 3 states of the standby subsea cables L are relevant in this case because possibility of state $x_5 = 2$ is 0. The relevant states have possibilities 0.7 for $x_5 = 0$ and 0.3 for $x_5 = 1$. The system state is interpreted as a failure for this components states with the possibility 0.7 ($\phi(x) = 0$) and as the sufficient state ($\phi(x) = 1$) with the possibility 0.3, while the state of perfect operation ($\phi(x) = 2$) is not indicated since its possibility is 0.

Table 1. Data obtained based on the monitoring of the offshore electrical power generation system

No	x_1			x_2			x_3			x_4			x_5			$\phi(x)$		
	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
1	0.8	0.2	0.0	0.8	0.1	0.1	0.7	0.2	0.1	0.8	0.2	0.0	0.7	0.3	0.0	0.7	0.3	0.0
2	0.8	0.1	0.1	0.7	0.1	0.2	0.6	0.2	0.2	0.8	0.2	0.0	0.0	1.0	0.0	0.8	0.1	0.1
3	1.0	0.0	0.0	0.7	0.3	0.0	0.9	0.1	0.0	0.0	0.9	0.1	0.7	0.2	0.1	1.0	0.0	0.0
4	0.8	0.1	0.1	0.8	0.1	0.1	0.0	0.9	0.1	0.1	0.9	0.0	0.8	0.1	0.1	0.2	0.7	0.1
5	0.7	0.2	0.1	1.0	0.0	0.0	0.2	0.7	0.1	0.1	0.6	0.3	0.1	0.9	0.0	0.0	0.7	0.3
6	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.7	0.2	0.1	0.5	0.3	0.2	0.7	0.2	0.1
7	0.8	0.1	0.1	0.2	0.6	0.2	0.8	0.1	0.1	0.0	0.0	1.0	0.8	0.1	0.1	0.6	0.2	0.2
8	1.0	0.0	0.0	0.0	0.9	0.1	0.0	0.9	0.1	0.0	0.9	0.1	0.7	0.3	0.0	0.0	0.6	0.4
9	0.8	0.1	0.1	0.0	0.9	0.1	0.1	0.8	0.1	0.0	0.9	0.1	0.2	0.7	0.1	0.1	0.6	0.3
10	0.7	0.3	0.0	0.1	0.8	0.1	0.0	0.1	0.9	0.8	0.1	0.1	0.8	0.2	0.0	0.7	0.1	0.2
11	0.7	0.2	0.1	0.0	0.9	0.1	0.0	0.1	0.9	0.0	0.1	0.9	0.7	0.2	0.1	0.0	1.0	0.0
12	0.1	0.6	0.3	0.8	0.2	0.0	0.2	0.8	0.0	0.1	0.7	0.3	0.0	1.0	0.0	0.1	0.6	0.5
13	0.2	0.8	0.0	0.7	0.3	0.0	0.1	0.1	0.8	0.1	0.6	0.3	1.0	0.0	0.0	0.0	0.7	0.3
14	0.1	0.8	0.1	0.2	0.7	0.1	0.0	1.0	0.0	0.3	0.6	0.4	0.9	0.1	0.0	0.2	0.8	0.0
15	0.0	0.2	0.8	0.9	0.1	0.0	0.2	0.8	0.0	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.6	0.4
16	0.0	0.1	0.9	1.0	0.0	0.0	0.0	0.1	0.9	0.1	0.6	0.3	0.0	0.2	0.8	0.2	0.5	0.3
17	0.0	0.2	0.8	0.1	0.6	0.3	0.2	0.5	0.3	0.2	0.7	0.1	0.0	0.3	0.7	0.1	0.1	0.8
18	0.0	0.2	0.8	0.2	0.7	0.1	0.0	0.2	0.8	1.0	0.0	0.0	1.0	0.0	0.0	0.7	0.0	0.3
19	0.0	0.1	0.9	0.0	0.2	0.8	0.2	0.7	0.1	0.0	0.1	0.9	0.2	0.7	0.1	0.0	0.1	1.9
...
108	0.0	0.0	1.0	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.8	0.2	0.0	0.0	1.0

The data obtained based on the monitoring and presented in Table 1 can be interpreted as fuzzy data [21]. The possibilities of individual states of the system components and of the system correspond to membership functions of fuzzy data.

The data obtained based on the monitoring of the offshore electrical power generation system are incompletely specified because we have 108 of all 243 combinations of components states. Traditional mathematical approach for system reliability analysis based on the structure function cannot be used in this case. Therefore, construction of structure function (1) based on incomplete data requires a special transformation and development of new methods. In this paper, we suggest the new method for construction of the structure function based on an FDT. This method allows reducing indeterminate values and obtaining a completely specified structure function.

4 Construction of FDT for representation of system

A decision tree is a formalism for expressing mappings of input attributes (components states) to output attribute/attributes (system performance level), consisting of an analysis of attribute nodes (input attributes) linked to two or more sub-trees and leafs or decision nodes labeled with classes of the output attribute (in our case, a class agrees with a system performance level) [21]. An FDT is one of the possible types of decision trees that permit operating with fuzzy data (attributes) and that use methods of fuzzy logic. Construction of a structure function assumes manipulation with real data, but the analysis of these data is implemented based on the methods of fuzzy logic the data are uncertain [22, 23]. The uncertainty may be present in obtaining numeric values of the attributes (system components states) or in obtaining the exact class (system performance level) where the instance belongs to.

There are different methods for inducing an FDT [10, 22, 24]. An FDT induction is implemented by the definition of the correlation between n input attributes $\{A_1, \dots, A_n\}$ and an output attribute B . The construction of the system structure function supposes that the system performance level is the output attribute and component states defined by a state vector are input attributes. Each input attribute (component state) A_i ($1 \leq i \leq n$) is measured by a group of discrete values ranging from 0 to $m_i - 1$, which agree with the values of states of the i -th component: $\{A_{i,0}, \dots, A_{i,j}, \dots, A_{i,m_i-1}\}$. An FDT assumes that the input set $\mathbf{A} = \{A_1, \dots, A_n\}$ is classified as one of the values of output attribute B . Value B_w of output attribute B agrees with one of the system performance levels and is defined as M values ranging from 0 to $M - 1$ ($w = 0, \dots, M - 1$). The correlation between the terminologies and basic concepts of FDTs and reliability analysis are shown in Table 2.

Table 2. Correlation between the terminologies of FDTs and reliability analysis

FDT	System reliability
Number of input attributes: n	Number of the system components: n
Attribute A_i ($i = 1, \dots, n$)	System component x_i ($i = 1, \dots, n$)
Values of attribute A_i : $\{A_{i,0}, \dots, A_{i,j}, \dots, A_{i,m_i-1}\}$	State of component i : $\{0, \dots, m_i - 1\}$
Output attribute B	System performance level $\phi(\mathbf{x})$
Values of output attribute B : $\{B_0, \dots, B_{M-1}\}$	Values of system performance level: $\{0, \dots, M - 1\}$
Decision table	Structure function

A fuzzy set A with respect to a universe U is characterized by a *membership function* $\mu_A : U \rightarrow [0,1]$, which assign an A -membership degree, $\mu_A(u)$, to each element u in U . $\mu_A(u)$ gives us an estimation that u belongs to A . The *cardinality measure* of the fuzzy set A is defined by $M(A) = \sum_{u \in U} \mu_A(u)$, and it is measure of size of set A . For

$u \in U$, $\mu_A(u) = 1$ means that u is definitely a member of A and $\mu_A(u) = 0$ means that u is definitely not a member of A , while $0 < \mu_A(u) < 1$ means that u is a partial member of A . If either $\mu_A(u) = 0$ or $\mu_A(u) = 1$ for all $u \in U$, A is a crisp set. The set of input attributes A is crisp if $\mu_A(u) = 0$ or $\mu_A(u) = 1$.

For example, let us consider input attributes $A = \{A_1, A_2, A_3, A_4, A_5\}$ and the output attribute B for the offshore electrical power generation system in Fig. 1. This system is represented by 5 input attributes. Each input attribute is defined as: $A_i = \{A_{i,0}, A_{i,1}, A_{i,2}\}$, for $i = 1, \dots, 5$, and the output attribute is $B = \{B_0, B_1, B_2\}$. The values of the input attributes and the output attribute are defined in Table 3. These values are obtained based on the data from Table 1 and are used for the FDT construction as a training test. The principal difference of Table 1 and 3 is interpretation of initial data as attributes. The induction of the FDT based on this training test can be implemented using some of methods for FDT induction [10, 22, 24]. We propose to induct the FDT using the method based on the *cumulative information estimates* proposed in [25]. These estimations allow inducting FDTs with various properties. Criteria for building non-ordered, ordered or stable FDTs, as well as, development of this method have been considered in [26].

Table 3. A training set for the FDT induction

No	A ₁			A ₂			A ₃			A ₄			A ₅			B		
	A _{1,0}	A _{1,1}	A _{1,2}	A _{2,0}	A _{2,1}	A _{2,2}	A _{3,0}	A _{3,1}	A _{3,2}	A _{4,0}	A _{4,1}	A _{4,2}	A _{5,0}	A _{5,1}	A _{5,2}	B ₀	B ₁	B ₂
1	0.8	0.2	0.0	0.8	0.1	0.1	0.7	0.2	0.1	0.8	0.2	0.0	0.7	0.3	0.0	0.7	0.3	0.0
2	0.8	0.1	0.1	0.7	0.1	0.2	0.6	0.2	0.2	0.8	0.2	0.0	0.0	1.0	0.0	0.8	0.1	0.1
3	1.0	0.0	0.0	0.7	0.3	0.0	0.9	0.1	0.0	0.0	0.9	0.1	0.7	0.2	0.1	1.0	0.0	0.0
4	0.8	0.1	0.1	0.8	0.1	0.1	0.0	0.9	0.1	0.1	0.9	0.0	0.8	0.1	0.1	0.2	0.7	0.1
5	0.7	0.2	0.1	1.0	0.0	0.0	0.2	0.7	0.1	0.1	0.6	0.3	0.1	0.9	0.0	0.0	0.7	0.3
6	1.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	0.0	0.7	0.2	0.1	0.5	0.3	0.2	0.7	0.2	0.1
7	0.8	0.1	0.1	0.2	0.6	0.2	0.8	0.1	0.1	0.0	0.0	1.0	0.8	0.1	0.1	0.6	0.2	0.2
8	1.0	0.0	0.0	0.0	0.9	0.1	0.0	0.9	0.1	0.0	0.9	0.1	0.7	0.3	0.0	0.0	0.6	0.4
9	0.8	0.1	0.1	0.0	0.9	0.1	0.1	0.8	0.1	0.0	0.9	0.1	0.2	0.7	0.1	0.1	0.6	0.3
10	0.7	0.3	0.0	0.1	0.8	0.1	0.0	0.1	0.9	0.8	0.1	0.1	0.8	0.2	0.0	0.7	0.1	0.2
11	0.7	0.2	0.1	0.0	0.9	0.1	0.0	0.1	0.9	0.0	0.1	0.9	0.7	0.2	0.1	0.0	1.0	0.0
12	0.1	0.6	0.3	0.8	0.2	0.0	0.2	0.8	0.0	0.1	0.7	0.3	0.0	1.0	0.0	0.1	0.6	0.5
13	0.2	0.8	0.0	0.7	0.3	0.0	0.1	0.1	0.8	0.1	0.6	0.3	1.0	0.0	0.0	0.0	0.7	0.3
14	0.1	0.8	0.1	0.2	0.7	0.1	0.0	1.0	0.0	0.3	0.6	0.4	0.9	0.1	0.0	0.2	0.8	0.0
15	0.0	0.2	0.8	0.9	0.1	0.0	0.2	0.8	0.0	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.6	0.4
16	0.0	0.1	0.9	1.0	0.0	0.0	0.0	0.1	0.9	0.1	0.6	0.3	0.0	0.2	0.8	0.2	0.5	0.3
17	0.0	0.2	0.8	0.1	0.6	0.3	0.2	0.5	0.3	0.2	0.7	0.1	0.0	0.3	0.7	0.1	0.1	0.8
18	0.0	0.2	0.8	0.2	0.7	0.1	0.0	0.2	0.8	1.0	0.0	0.0	1.0	0.0	0.0	0.7	0.0	0.3
19	0.0	0.1	0.9	0.0	0.2	0.8	0.2	0.7	0.1	0.0	0.1	0.9	0.2	0.7	0.1	0.0	0.1	1.9
...
108	0.0	0.0	1.0	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.1	0.9	0.0	0.8	0.2	0.0	0.0	1.0

The FDT resulted from the training set presented in Table 3 has been inducted by application of the cumulative information estimates using the method in [24]. This FDT is presented in Fig. 2. The nodes of this FDT agree with the input attributes. Every node has 3 branches according to the values of the corresponding input attrib-

ute from the training test (Table 3). Every branch correlates with some values of the output attribute. The set of output attribute values in a branch is named as a leaf if the analysis finish and one of the values of the output attribute can be chosen according to algorithms proposed in [25, 26].

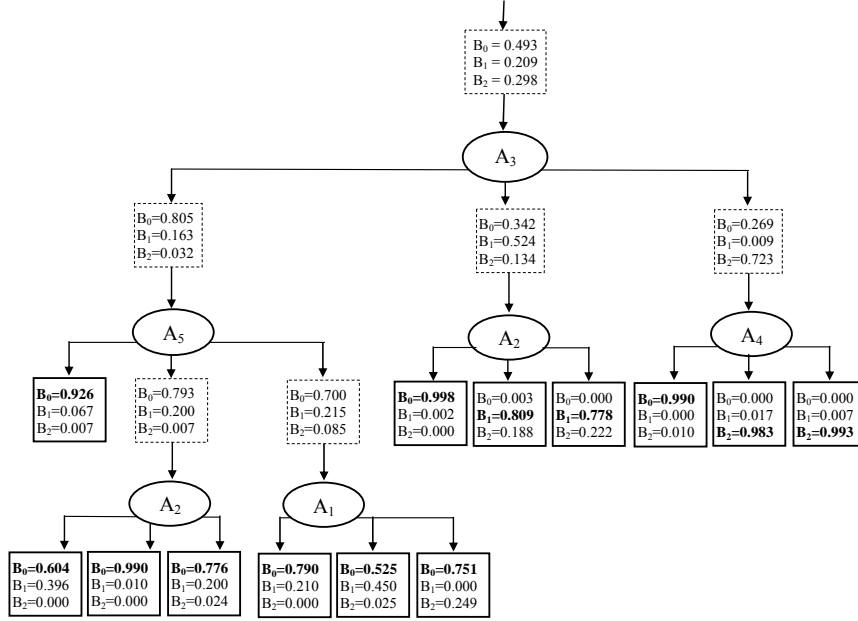


Fig. 2. Non-ordered FDT constructed based on the data obtained by the monitoring of the off-shore electrical power generation system from Fig. 1

This FDT can be used for the analysis of all possible states of system components to construct the structure function of the offshore electrical power generation system. This process is considered below.

5 Construction of structure function based on FDT

According to [20], FDTs allow developing fuzzy decision rules or a decision table. A decision table contains all possible values of input attributes and the corresponding values of the output attribute that is calculated using the FDT. Such decision table agrees with the structure function. This implies that all possible combinations of values of the component states (all state vectors) have to be analyzed by the FDT to classify state vectors into M classes of the system performance levels.

Each non-leaf node is associated with an attribute $A_i \in \mathbf{A}$, or in terms of reliability analysis: each non-leaf node is associated with a component. The non-leaf node agreeing with attribute A_i has m_i outgoing branches. The s -th outgoing branch ($s = 0, \dots, m_i - 1$) from the non-leaf node corresponding to attribute A_i agrees with state s of

the i -th component ($x_i = s$). A path from the root to a leaf defines one or more state vectors (according to the values of the input attributes (component states) occurred in the path) for which the structure function takes value determined by the value of the output attribute. If any input attribute is absent in the path, all possible states have to be considered for the associated component.

Let us consider construction of the structure function of the offshore electrical power generation system from Fig. 1 using the FDT depicted in Fig. 2. All possible component states (all state vectors) have to be used for calculation of the system performance level by the FDT to form the decision table (structure function). Let us explain this idea for the first level of the FDT in more detail.

Preliminary analysis of the data obtained based on the monitoring (see Table 3) shows that possible values of the output attribute B are distributed as follows: value 0 – with confidence 0.493, value 1 – with confidence 0.209 and value 2 – with confidence 0.298. These values are implied by frequency of every output value in the training test (Table 3). Attribute A_3 is associated with the FDT root. So, analysis of the data starts from this attribute. This attribute has the following possible values: $A_{3,0}$, $A_{3,1}$, and $A_{3,2}$. Value $A_{3,0}$ of this attribute makes the output attribute B to be B_0 (the system is non-operational) with the confidence of 0.805. Other variants, B_1 and B_2 , of output attribute B can be chosen with the confidence of 0.163 and 0.012 respectively. If the attribute A_3 has other values, i.e. $A_{3,1}$ or $A_{3,2}$, then the analysis is done similarly.

If the value of attribute A_3 is $A_{3,0}$, then the next attribute in the analysis is A_5 , which have values $A_{5,0}$, $A_{5,1}$, and $A_{5,2}$. Value $A_{5,0}$ of this attribute agrees with a leaf representing the output attribute. Therefore, in this situation, the analysis is stopped and value of the output attribute is defined: value B_0 of attribute B should be chosen with the confidence of 0.926, and values B_1 and B_2 with confidences 0.067 and 0.007 respectively. Similarly, the process of the analysis of the non-ordered FDT continues for the other input attributes and their values.

Next, let us consider state vector $\mathbf{x} = (0,0,0,0,0)$. The analysis based on the FDT starts with the attribute A_3 (Fig. 2) that is associated with the 3-rd component. State of this component is 0 ($x_3 = 0$) for the specified state vector. Therefore, the branch for value $A_{3,0}$ of attribute A_3 value is taken into account. According to this value, the identification of the output attribute value (system performance level) has to continue through attribute A_5 . According to the state vector, $x_5 = 0$, therefore, attribute A_5 has value $A_{5,0}$. Now, value of the output attribute can be indicated because the branch has a leaf. So, value of the output attribute is defined as 0 with the confidence of 0.926 without analysis of other attributes. Analysis of other state vectors is similar and allows obtaining all possible values of the system performance level in the form of the structure function. The analysis of all possible state vectors from $\mathbf{x} = (0,0,0,0,0)$ to $\mathbf{x} = (2,2,2,2,2)$ allows us to construct the complete structure function of the offshore electrical power generation system depicted in Fig. 1.

It is important to note that this method of construction of the structure function based on FDTs permits to compute (restore) data missing from the monitoring.

A representation of the system using the structure function allows calculating different indices and measures for estimation of system reliability. Probabilities of system performance levels (3) are one of them. Suppose that probabilities of the compo-

nents states of the offshore electrical power generation system have values shown in Table 4. In this case, the probabilities of system performance levels are: $A_2 = 0.73$, $A_1 = 0.20$ and $A_0 = 0.07$. Other measures can be computed using the structure function too. For example, importance measures for this system can be calculated using the algorithms considered in [15, 27, 28].

Table 4. Components states probabilities

Component state, s	Probabilities				
	$p_{1,s}$	$p_{2,s}$	$p_{3,s}$	$p_{4,s}$	$p_{5,s}$
0	0.1	0.2	0.1	0.1	0.1
1	0.4	0.4	0.4	0.2	0.1
2	0.5	0.4	0.5	0.6	0.8

Let us present a simple case study to verify the modelling approach described in the previous sections. We use structure function of the offshore electrical power generation system to examine efficiency and accuracy of the proposed method for construction of the structure function based on uncertain data. Therefore, the structure function must be transformed to ambiguous and incompletely specified form. In the proposed methods, two types of uncertainty are included. The first is ambiguity of data values. Therefore, all integer values of components states and performance level have to be transformed to values with some possibilities. We can use algorithm from [29] to transform data from numeric to fuzzy cases in this case. The second type of the considered uncertainty in the proposed method is incompletely specification of initial data. This incompleteness is modeled by random deleting of some state vectors and the corresponding values of system performance levels. The range of deleted states was changed from 5% to 90%. Each transformed structure function can be interpreted as uncertain data obtained by the aforementioned monitoring. We used these data to construct the structure function based on the proposed methods using FDT induction. The FDTs were induced based on the method presented in [22, 25]. The structure function construction was implemented according to the concept introduced in section 5. As a result, a single or a small group of state vectors might be misclassified. Therefore, we had to estimate this misclassification by the error rate. The constructed structure functions were compared with the exact specified function (it was defined at the beginning of the experiments), and the error rate was calculated as a ratio of wrong values of the structure function to the dimension of unspecified part of the function. The experiments were repeated 1000 times for every version of incompletely specified offshore electrical power generation system. The unspecified state vectors were selected randomly in proportion to dimension of the structure function from 5% to 85%. The results for the investigated system are shown in Table 5. The error rate depends on unspecified proportion of the initial data. This error increases essentially, if the unspecified part is most than 80%. This indicates that the proposed method is acceptable for large range of incompletely defined initial data and can be used for construction of the structure function based on incompletely specified data.

This method can be considered for special cases and types of initial data with application of other algorithms from Data Mining to improve the obtained results. For example, initial data can be obtained for similar samples and, in this case, special algorithms for pattern recognition and intelligent diagnosis can be used [30].

Table 5. The error rate for the construction of the structure function of the offshore electrical power generation system

Unspecified state vectors, in %	The error
5	0.0661
10	0.0637
15	0.0682
20	0.0661
25	0.0670
30	0.0662
35	0.0648
40	0.0663
45	0.0657
50	0.0659
55	0.0671
60	0.0673
65	0.0679
70	0.0700
75	0.0743
80	0.0813
85	0.0995
90	0.1465

6 Conclusion

The new method for constructing the structure function is proposed in this paper. This method allows obtaining a structure function based on incompletely specified data (for example, data obtained from some monitoring). The term “incompletely specified” assumes uncertainties of two types.

The first type of uncertainty deals with some state vectors missing from the initial data. In practical application, it can be caused by the impossibility to obtain or indicate all possible combinations of system component states.

These uncertainties are considered and taken into account in the interpretation of the initial data as quasi-fuzzy data. This interpretation requires use of mathematical methods of fuzzy logic for the analysis. In this paper, an FDT is used for system behavior modeling and construction of the system structure function. This mathematical method transforms incomplete and uncertain initial data to a correct decision [10,24]. The induction of FDT is implemented based on cumulative information estimates [25] that take into account mathematical concept of entropy. These estimates are then adopted for the analysis of uncertain data. Therefore, the system structure function can be constructed using an FDT based on uncertain data, and the FDT transforms

incompletely specified data about system reliability/availability into a completely specified mathematical model that is known as the system structure function.

The second type of uncertainty results from ambiguity of initial data. In this case, the system performance level and components states can be defined with some possibilities. According to the typical definition of the structure function (1), performance level can have only one value for every state vector from set $\{0, \dots, M-1\}$. However, the boundary between two neighboring values can be diffused in real applications. Both such values can be therefore indicated with some possibility. The proposed method takes such ambiguity into account and permits indicating performance level using some values ranging from 0 to $M-1$ with a possibility that is considered in the next steps of the method and is not disregarded. The component states are indicated in a similar manner and the state of the i -th component is considered as a value ranging from 0 to m_i-1 with possibilities. For example, the data in Table 1 are presented with consideration of such ambiguity: every value is indicated with some possibility.

7 Acknowledgment

This work is supported by the grant grants VEGA 1/0498/14 and VEGA 1/0038/16.

References

- 8 Zio E., Reliability engineering: Old problems and new challenges, *Reliability Engineering and System Safety*, 94, 125–141 (2009)
- 9 Natvig B., *Multistate Systems Reliability Theory with Applications*, Wiley, New York (2011)
- 10 Aven T., Heide B. (1993), On performance measures for multistate monotone system, *Reliability Engineering and System Safety*, Vol. 41, pp.259–266
- 11 Lisnianski A., Levitin G., *Multi-state System Reliability. Assessment, Optimization and Applications*. World Scientific, Singapore (2003)
- 12 Cutello V., Montero J., Yanez J., Structure functions with fuzzy states, *Fuzzy Sets and Systems*, 83, 189-202 (1996)
- 13 Huang H.-Z., Structural reliability analysis using fuzzy sets theory. *Eksplotacja i Niezawodność – Maintenance and Reliability*, 14 (4), 284–294 (2012).
- 14 Liu Y., Huang H.-Z., Ling D., Reliability prediction for evolutionary product in the conceptual design phase using neural network-based fuzzy synthetic assessment, *International Journal of Systems Science*, 44 (3), 545–555 (2013)
- 15 Aven T., Heide B., Reliability and validity of risk analysis, *Reliability Engineering and System Safety*, 94 (11), 1862–1868 (2009)
- 16 Witten I.H., Frank E., Hall M. A., *Data mining : practical machine learning tools and techniques*, Elsevier (2011)
- 17 Mitra S., Konwar K.M., and Pal S.K., Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network: Generation and Evaluation. *Journal of IEEE Trans. on Syst., Man Cybernetics — Part C: Applications and Reviews*, 32, 328-339 (2002)
- 18 Tsipouras M. G., Exarchos T.P., Fotiadis D.I., A methodology for automated fuzzy model generation, *Fuzzy Sets and Systems*, 159(23), 3201-3220 (2008)

- 19 Aven T., Zio E., Baraldi P., Flage R., *Uncertainty in Risk Assessment: The Representation and Treatment of Uncertainties by Probabilistic and Non-Probabilistic Methods*, Wiley (2014)
- 20 Barlow R.E. Proschan B. *Statistical Theory of Reliability and Life Testing, Probability Models*. Holt, Rinehart, and Winston, New York (1975),
- 21 Zaitseva, E., Levashenko, V., Kostolny J., Kvassay, M., Direct partial logic derivatives in analysis of boundary states of multi-state system, 11th International Conference on ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer, pp. 535-549 (2015)
- 22 Kvassay M., Zaitseva E., Levashenko V. Minimal cut sets and direct partial logic derivatives in reliability analysis. In: *Safety and Reliability: Methodology and Applications - Proceedings of the European Safety and Reliability Conference*, pp. 241-248, CRC press (2015)
- 23 Kuo W., Zhu X., *Importance Measures in Reliability, Risk, and Optimization: Principles and Applications*. Chichester, Wiley (2012)
- 24 Wang Y. and Li L., Effects of Uncertainty in Both Component Reliability and Load Demand on Multistate System Reliability, *IEEE Trans. on System, Man and Cybernetic – Part A: Systems and Humans*, 42(4), 958-969 (2012)
- 25 Li M., Liu J., Li J., Kim B.U., Bayesian modeling of multi-state hierarchical systems with multi-level information aggregation, *Reliability Engineering & System Safety*, 124(4), 158–164 (2014)
- 26 Stanton N.A., Baber C., Error by design: methods for predicting device usability, *Design Studies*, 23(4), 363-384 (2002)
- 27 Levashenko V., Zaitseva E., Kovalik S., *Decision Marking based on the fuzzy data. Background, methods and algorithms*, LAP-Lambert Academic Publisher (2014)
- 28 Quinlan J.R. (1987), Simplifying decision trees. *International Journal of Manmachine Studies*, 27, pp. 221-234.
- 29 Levashenko V., Zaitseva E., Puuronen S. (2007), Fuzzy Classified Based on Fuzzy Decision Tree., *Proc. IEEE Int. Conf. on Computer as a tool (EUROCON 2007)*, Warsaw, Poland, pp.823 – 827.
- 30 Dong M., Kothari R. (2001), Look-Ahead Based Fuzzy Decision Tree Induction. *IEEE Trans. on Fuzzy Systems*, vol. 9, no.3, pp. 461-468.
- 31 Olaru C., Whenkel L. (2003), A Complete Fuzzy Decision Tree Technique. *Fuzzy Sets and Systems*, , pp.221-254
- 32 Levashenko V., Zaitseva E. (2002), Usage of New Information Estimations for Induction of Fuzzy Decision Trees. H. Yin et al. (Eds.): *IDEAL 2002, Lecture Notes in Computer Science LCNS2412*, Springer-Verlag, pp. 493-499.
- 33 Levashenko V., Zaitseva E. (2012), Fuzzy decision trees in medical decision making support system, *Proc. of the IEEE Federated Conf. on Computer Science & Information Systems*, (IEEE FedCSIS'12), September 9-12, Wrocław, Poland, pp.213-219.
- 34 Zaitseva E. (2012), Importance analysis of a multi-state system based on multiple-valued logic methods. In: Lisnianski A. and Frenkel I. (eds) *Recent Advances in System Reliability: Signatures, Multi-state Systems and Statistical Inference*, London: Springer, pp. 113–134
- 35 Artsiomchyk, Y. , Zhivitskaya, H. (2015), Improving the supply chain efficiency based on reliability analysis, *Proc. Int. Conf. on Information and Digital Technologies, IDT 2015; Zilina; Slovakia*, 2015, pp. 16-21
- 36 Lee H.M., Chen C.-M. et al. An efficient fuzzy classifier with feature selection based on fuzzy entropy, *IEEE Trans on Sys, Man, and Cyber, Part B: Cybernetics*, vol.31, 2001, pp. 426 – 432
- 37 Subbotin S.A.(2014), The sample properties evaluation for pattern recognition and intelligent diagnosis, *Proc. Int. Conf. on Digital Technologies, Zilina, Slovakia*, pp. 321-332.

Metamorphic Viruses Detection Technique Based on the Modified Emulators

Oksana Pomorova, Oleg Savenko, Sergii Lysenko, Andrii Nicheporuk

Khmeltsky National University, 11, Institutskaya St.,
29016, Khmeltsky, Ukraine

o.pomorova@gmail.com, savenko_oleg_st@ukr.net,
sirogyk@ukr.net, andrey.nicheporuk@gmail.com

Abstract. An article presents a new technique for metamorphic viruses detection using modified emulators, placed in the hosts of the network. Proposed technique provides the classification of the metamorphic virus in classes with the usage of the fuzzy logic. Technique makes it possible to detect the metamorphic viruses, which use obfuscation techniques. The results of experimental studies showed the effectiveness of the proposed method of detection metamorphic virus copies at 85%.

Keywords. Malware detection, metamorphic viruses, polymorphic viruses, obfuscation, modified network emulators

Key Terms. Machine Intelligence, KE, and KM for ICT, Software Component, Software System

1 Introduction

The detection of computer viruses is one of the main challenges of information security for today. Viruses execute harmful activity on infected hosts, such as stealing system information, accessing private information, corrupting data, spamming, logging their keystrokes. Among computer viruses one of the leaders are metamorphic viruses. Thus, metamorphic viruses cause damage to cost millions of dollars. According to Symantec in 2011 metamorphic virus Sality infected about 3 million computers in the world [1], and at the end of 2015 it is in the top five of the most spread viruses (1.43% of the total number of detected threats) [2].

Detection of the metamorphic viruses is a very complicated task due to its usage of the obfuscation techniques for program code. It allows virus to create multiple copies of the same virus and its detection become a very difficult task. In addition, in order to complicate the process of reverse engineering and the data protection obfuscation technique is often used in the trusted applications by software developers [3].

That is why the actual problem is to develop a new technique for metamorphic viruses detection with the involvement of fuzzy logic that will allow the suspicious

programs classification to one of the metamorphic viruses classes using the modified emulators placed on each host of the network.

2 Related Works

Known techniques for virus detection, based on signature analysis are not able to detect the altered copies of metamorphic virus [4-7]. In order to detect this type of viruses most antivirus scanners use the heuristic's method, which deploys the sequence of API-functions calls, the control flow graph of the program, the structural features of the PE .EXE files, opcode instructions and their combinations.

In [7] a method for metamorphic malware detection is presented. A malware signature is described by the set of control flow graphs the malware contains. Technique uses a distance metric based on the distance between feature vectors. The drawback of the approach is that it is computationally inefficient.

In [4] a statistical technique based on comparison of the similarity between two files infected by two morphed versions of a given metamorphic virus is used. The proposed solution based on static analysis and it uses the histogram of machine instructions frequency in various offspring of obfuscated viruses. The disadvantage of the approach is that it is inefficient for detection of viruses, which use code transposition techniques.

A malware detection system based on API call graph is proposed in [5]. Each malware sample is represented as data dependent API call graph. Graph matching algorithm is used to calculate similarity between the input sample and malware API call graph samples stored in a database. The main drawback that most malware samples are generated from previous existing samples, therefore sequences API calls are similar.

In [8] the technique of botnet detection which bots use polymorphic code is proposed. Performed detection is based on the multi-agent system by means of antiviral agents that contain sensors. Developed technique makes it possible to perform provocative actions against probably infected file. The disadvantage of this technique is large computational complexity of the behavior analysis.

The state-of-art demonstrates a need of development new and improve existing techniques for metamorphic viruses' detection.

3 Metamorphic Virus Detection Method-based on the Modified Emulators

A new metamorphic viruses detection technique based on the modified emulators is proposed. It uses the emulation on each host in the network. The main function of the hosts is to implement a single-time emulation and execution of the unknown potentially malicious program and sending the results to the server.

The server is used for processing the results of the emulation process obtained from the hosts. In order to complicate the reverse engineering process and data protection, the obfuscation techniques are often used for trusted applications by software developers. Therefore, the main task of the server is classification of the feature vec-

tors based on the comparison of the metamorphic viruses' copies, which were obtained from network hosts.

The proposed technique specifies three classes of the metamorphic viruses, which are to be classified. In addition, there is the fourth class of programs, which have similarity to metamorphic viruses by its behavior, but these programs are not malicious. Fig.1 shows a scheme of the metamorphic viruses detection.

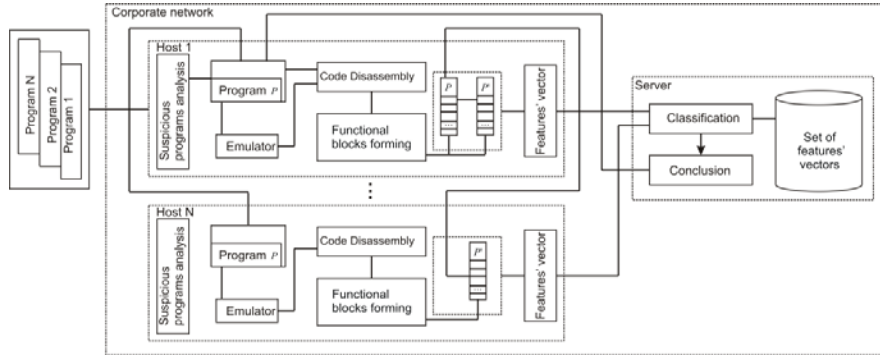


Fig. 1. The scheme of the metamorphic viruses detection

Let us consider the proposed technique. In order to detect suspicious program activity on each host of the corporate network, the analyzer of the suspicion programs is used. Every single operation that is performed by suspicious program is not dangerous one. However, the execution of the sequence of such operations may indicate a possible risk of infection with the virus.

Each application that comes into the system is marked as suspicious or un suspicious. Let us present the feature vector, which defines the program membership to one of two classes as follows:

$$\overline{U} = (M, Q, J, Y, L, N, H),$$

where, M —attempt of the program to get the system admin is trator rights, Q - attempt to open or close the system port, J - attempt to delete the file, Y - create a file or process, L —key logging attempt, N – sending messages to the network, H -creation a key or an entry to the registry.

Each feature is able to posses a value 0 or 1, where 1 indicates the activation of the featuretrait,0 -vice versa. Program consider suspicious if:

$$P = suspicious, if \forall u \in \overline{U}, (u_i = 1 \wedge u_j = 1)$$

So, if $P = suspicious$ for some program, it comes to the metamorphic viruses detection system. In order to get the modified code sample F_s , the emulation of program P is carried out. The emulation process consists of the instruction execution in a virtual environment, and the extraction instructions from the software package.

The usage of the one-type emulators in all the network hosts does not guarantee the metamorphic virus detection with high efficiency, because their usage will produce only the same code samples. In order to detect the metamorphic viruses properties and features, different conditions for malicious code execution are needed.

Therefore, the modified emulators on each host are created. The structure of the emulator includes the virtual processor. It is able to execute the set of instruction such as MMX, SSE, SSE2, etc, and it includes a set of virtual registers. Also, the emulator consists of RAM and virtual stack, virtual network controller; the operating system (it supports API functions, registry and ports).

To avoid the anti-emulation technologies, which are used by metamorphic viruses, the emulator includes a heuristics module. For each operation performed by virtual CPU the fixed, the processing time is determined and the checking of repeating for some operation is executed.

In order to obtain the original sample code F_P , the disassembly of program P is carried out. The result of disassembly is a set assembler instructions x86/x64. In order to construct the feature vector only opcodes are used and operands are discarded.

The resulting listing of the disassembly instructions is partitioned into the functional blocks (FB). One of the techniques that are used to perform the instruction obfuscation is moving of program blocks. It is carried out by using the conditional state transition instructions (jz, jnz, jmpetc).

The main mechanisms creation of the metamorphic viruses copies are the insertion, deletion and transposition of their instructions. For the purpose of finding the similarities between the two FB code samples F_P and F_S the Damerau–Levenshtein distance was used.

In order to evaluate the Damerau–Levenshtein distance the polynomial algorithm complexity of Wagner-Fisher was used. It made it possible to create a short transformation chain in order to transform the set of opcodes of the program after the emulation into the opcode set of the program before the emulation.

Consider a program P , which consists of a set of assembler commands p_i , $P = \{p_1, p_2, \dots, p_k\}$. Let us partition the program P into the functional blocks of an arbitrary length. Such blocks start and end by the instructions of the conditional state transitions, such as jmp, jzetc, that is $P = \{B_1, B_2, \dots, B_l\}$. Then we can write: $P = \{B_1 = \{p_1, p_2, \dots, p_{i-1}\}, \dots, B_l = \{p_i, p_{i+1}, \dots, p_k\}\}$.

Let us denote program before the emulation F_P , and program after the emulation F_S .

Let us present the functional unit B , which consists of a set of opcodes of length $|B| = m$, as p_1, p_2, \dots, p_m . So, the subset of opcodes x_i, x_{i+1}, \dots, x_j of the functional unit B will be specified as $B(i, j)$.

Let us denote the transformation weight of the opcode a into b as $w(a, b)$. Thus, $w(a, b)$ is the weight of the replacement of one opcode into another one, when $a \neq b$, $w(b, a)$ -weight of the transposition, $w(a, \varepsilon)$ -weight of the deletion, and $w(\varepsilon, b)$ -weight of the insertion for opcode b .

Let us assume that B_g and B_h - two FB, which consist of the opcodes sequence (of n and m length respectively) defined by a finite alphabet of the assembler instruc-

tions $A = (a_1, a_2, \dots, a_k)$. Then B_g of the FB program F_p we will denote as $B_g^{F_p}$, and B_h of the same FB program F_s after the emulation we will denote as $B_h^{F_s}$. Then the Damerau-Levenshte in distance $dL(B_g^{F_p}, B_h^{F_s})$ is calculated as $dL(B_g^{F_p}, B_h^{F_s}) = OPT(N, M)$, where

$$OPT = \begin{cases} 0, & i = 0, j = 0 \\ i, & j = 0, i > 0 \\ j & i = 0, j > 0 \\ \min \begin{cases} OPT(i, j - 1) + w(a, \varepsilon) \\ OPT(i, -1j) + w(\varepsilon, b) \\ OPT(i - 1, j - 1) + w(a, b) \\ OPT(i - 2, j - 2) + w(b, a) \end{cases} & j > 0, i > 0 \end{cases} \quad (1)$$

After the Damerau-Levenshte in distance for two bocks B_g and B_h is evaluated, the weighted averages of the corresponding parameter of the feature vector for all code blocks is to be formed. In order to obtain such weighted averages of the parameters, the index of the weighted arithmetic mean is used (2).

$$dL = \left[\frac{\sum_{i=1}^n dL_i * f_i}{\sum_{i=1}^n f_i} \right], \quad (2)$$

where, dL_i – the Levenshte in distance for FB B_i , f_i - number of FB with the value dL_i .

TheDamerau-Levenshtein distance estimates the minimum value for the required operations of the replacement, insertion, deletion and transposition, and is an integer value. In this case, for the finding of the lowest difference between the metamorphic viruses' copies the obtained values are rounded down.

For the rest of the features the normalization is performed in the same way.

Thus, the feature vector of similarity for metamorphic viruses' copies based on theDamerau-Levenshte in metrics will be presented as follows:

$$\bar{S} = \langle dL, T, D, I, R, M \rangle, \quad (3)$$

where dL – the Damerau-Levenshte in distance for functional unit between programs F_p and F_s ; T –number of required operations of the opcodes exchange for the program block's transformation F_p into F_s ($F_p = F_s$); D - number of operations required for the opcode deletion; I - number of operations required for the opcode inser-

tion; R - number of operations required for the opcode replacement; M - number of matches between opcodes of the functional units of programs F_p and F_s .

In order to make a conclusion about the infection by metamorphic virus, constructed feature vectors of the similarity are sent to server, where they are analyzed by the fuzzy inference system for the purpose of its classification [9].

The input linguistic variables for fuzzy inference systems are the feature vectors of similarity for the copies of the metamorphic viruses (3). The terms of the linguistic variable are Low, Medium and High.

As the membership function for input variables the trapezoidal one was chosen, and for output –the triangular. For feature dL we can present equations as follows:

$$\mu_{low}(x) = \begin{cases} 0, & 72 < x \\ \frac{72-x}{64}, & 8 < x \leq 72 \\ 1, & 0 \leq x \leq 8 \end{cases} \quad \mu_{medium}(x) = \begin{cases} 0, & x < 16 \\ \frac{x-16}{49}, & 16 \leq x < 65 \\ 1, & 65 \leq x < 96 \\ \frac{145-x}{49}, & 96 \leq x < 145 \\ 0, & 145 \leq x \end{cases}$$

$$\mu_{high}(x) = \begin{cases} 0, & 96 < x \\ \frac{x-96}{38}, & 96 \leq x < 134 \\ 1, & 134 \leq x \leq 161 \end{cases}$$

Fuzzy inference system uses 38 rules for making the conclusion about belonging of the metamorphic virus to one of the class:

- 1 *if (dL is Low) and (T is Low) and (D is Medium) and (I is Hight)*
and (R is Low) and (M is Medium) then class1
- 2 *if (dL is Low) and (T is Medium) and (D is Medium) and (I is Hight)*
and (R is Low) and (M is Medium) then class1
- 38 *if (dL is Hight) and (T is Low) and (D is Hight) and (I is Hight)*
and (R is Medium) and (M is Low) then class3

A result of the fuzzy inference system is a determination of the member ship degree of each virus copy to one of the class of the metamorphic viruses.

4 Experiments

In order to determine the efficiency of the proposed method several experiments were held. For this purpose the set of metamorphic viruses was generated, and metamorphic generators Next Generation Virus Creation Kits, Second Generation Virus Generator and Virus Creation Lab for Win32 were used [10]. They are able to infect .EXE and .DLL files and perform the obfuscation operations with files.

For experiments, the university network with 80 users' hosts and one server was used. Each host us edits modified emulator. The settings of the emulators are presented in Table 1.

Table 1. The settings of the modified emulators, presented on each host of the network

Host's number	The initial network settings of modified emulators							
	A set of instructions	OS type	The size of virtual RAM	System date	Virtual network controller	Runtime emulation instruction,	Start address of emulating	Architecture of CPU
1	SSE2	W7	8	25.11	Yes	1	0x001A4810	x86
2	Hyper-threading	W8.1	8	06.02	No	2	0x001A4620	x86
...
80	MMX	W10	4	10.04	No	1	0x001A5214	x64

Fuzzy inference system was built using Fuzzy Logic Toolbox package in Matlab system. Built system has the following parameters: algorithm –Mamdani, aggregation method, accumulation method, the method defuzzification. The system consists of six inputs and one output.

The result of fuzzy inference system is a conclusion about membership of the unknown program to one of three metamorphic viruses' class or program is trusted. If the resulting value of the membership degree for unknown object belongs to the range from 0 to 0.25 - it is classified as a trusted application; if the membership degree of is in the range of 0.26 to 1, then the unknown object belongs to one of the metamorphic viruses' classes. Values from 0.26 to 0.49 determine the first class of the metamorphic viruses, values from 0.5 to 0.74 –the second class, the value of 0.75 to 1 –the third class.

Table 2 and figure 2 demonstratethe results of fuzzy inference system for suspicious file. As a result,15% of the copies have not changed, 5% were classified asthe first class,11.25%-the third one, and 68, 75% as the second class.

Table 2. The result of fuzzy inference system

№ of host						Result of the fuzzy inference system					
I	II	I	II	I	II	I	II	I	II	I	II
1	0,65	15	0,57	29	0,81	42	0,72	55	–	68	–
2	–	16	0,86	30	0,70	43	0,51	56	0,61	69	0,72
3	0,19	17	0,57	31	0,55	44	0,53	57	0,63	70	0,83
4	0,64	18	0,63	32	–	45	0,63	58	0,84	71	–
5	0,45	19	0,56	33	0,62	46	0,78	59	0,67	72	0,53
6	0,59	20	0,68	34	0,23	47	0,62	60	0,69	73	–
7	–	21	0,72	35	0,59	48	–	61	0,67	74	0,59
8	0,52	22	0,41	36	0,54	49	0,70	62	–	75	0,47
9	–	23	0,69	37	0,68	50	0,26	63	0,51	76	0,74
10	0,61	24	0,79	38	–	51	0,56	64	0,82	77	–
11	0,14	25	0,68	39	0,39	52	0,81	65	0,69	78	0,58
12	0,69	26	0,24	40	0,56	53	0,74	66	0,36	79	0,57
13	0,78	27	0,56	41	0,54	54	0,56	67	0,52	80	0,73
14	0,35	28	0,69								

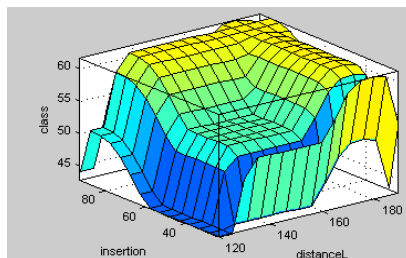


Fig. 2. The result of fuzzy inference about membership of the unknown program to one of metamorphic viruses' class

5 Conclusions

The analysis of the subject area revealed the need to improve existing techniques for metamorphic viruses' detection. In the article, the new technique for metamorphic viruses' detection using modified emulators in network hosts is proposed. The classification of viruses into classes of metamorphic viruses is based on a usage of the fuzzy inference system. Proposed technique makes it possible to detect metamorphic viruses that use obfuscation techniques of the program code. Such approach enables the increase of the efficiency of the metamorphic viruses detection. The results of experimental studies have demonstrated the efficiency technique for metamorphic viruses' detection at about 85%.

References

1. Falliere, N.: Salty: Story of a Peer-to-Peer ViralNetwork. Technical report, Symantec Labs (2011)
2. Virus statistics. Overview cyber November 2014. online: <https://www.esetnod32.ru/company/viruslab/statistics/?id=896397> [in Russian]
3. Attaluri, S., McGhee, S., Stamp, M.: Profile Hidden Markov Models and Metamorphic Virus Detection. *Journal on Computer Virology*, 5, 151--169 (2009)
4. Rad, B. B., Masrom, M.: Metamorphic Virus Variants Classification Using Opcode Frequency Histogram. In: *Proc. 14th WSEAS Int Conf on Computers*, pp. 147--155, WSEAS (2010)
5. Elhadi, A. A. E., Maarof, M. A., Barry, B. I. A.: Improving the Detection of Malware Behaviour Using Simplified Data Dependent API Call Graph. *Int. J. of Security and Its Applications*, 7(5), 29--42 (2013)
6. Kaushal, K., Swadas, P., Prajapati, N.: Metamorphic Malware Detection Using Statistical Analysis. *Int. J. of Soft Computing and Engineering*, 2, 49--53 (2012)
7. Cesare, S., Xiang, Y.: Malware variant detection using similarity search over sets of control flow graphs. In: *Proc. 10th Int. Conf. on Trust, Security and Privacy in Computing and Communications*, Washington, DC, USA, pp. 181-189 (2011)
8. Pomorova, O., Savenko, O., Lysenko, S., Kryshchuk, A., Nicheporuk, A.: A technique for detection of bots which are using polymorphic code. In: *Proc. 21st Int. Conf, CN, Springer, Brunów, Poland*, pp. 265--276 (2014)

9. Shtovba, S, Pankevich, O., Nagorna, A.: Analyzing the criteria for fuzzy classifier learning. Automatic control and computer sciences, 49(3), 123--132 (2015)
10. VX Heavens Computer virus collection. online: <http://vx.netlux.org>

Mobile Post-Emergency Monitoring System for Nuclear Power Plants

Anatoliy Sachenko¹, Volodymyr Kochan¹, Vyacheslav Kharchenko², Hubert Roth³,
Vasyl Yatskiv¹, Mykhaylo Chernyshov⁴, Pavlo Bykovyy¹, Oleksiy Roshchupkin⁵,
Vasyl Koval¹, Herman Fesenko⁶

¹ Department of Information Computer Systems and Control,
Ternopil National Economic University
as@tneu.edu.ua, jazkiv@ukr.net

² National Aerospace University, «Kharkiv Aviation Institute» (KhAI)
v_s_kharchenko@ukr.net

³ Institute of Automatic Control Engineering,
Department of Electrical Engineering and Computer Science, University of Siegen
hubert.roth@uni-siegen.de

⁴ «WESTRON», Kharkiv, Ukraine
office@westron.kharkov.ua

⁵ Yuriy Fedkovych Chernivtsi National University
oleksiy.roshchupkin@gmail.com

⁶ O.M. Beketov National University of Urban Economy in Kharkiv
fesennko@mail.ru

Abstract. In this paper, a concept of designing the post-emergency system for monitoring the equipment and territory of nuclear power plant after a severe accident was developed. Power and communications network lines are found out as the most vulnerable ones during the accident monitoring, and self-descriptiveness and survivability and veracity are recognized as system basic parameters. To ensure the self-descriptiveness it's proposed to equip measurement and control modules with backup wireless communication channels and deploy the repeaters network based on drones. To provide the survivability modules possess the backup power battery, and repeaters appear in the appropriate places after the accident. Moreover an optimization of drone's location is proposed according to the minimum energy consumption criterion. To ensure the veracity it's expected to design the noise-immune protocol for message exchange and archiving and self-diagnostics of all system components. Formulas for estimating the reliability level of the post-emergency monitoring system were obtained.

Keywords. Infrastructure, Network, Controller, NPP, Reliability Model

Key Terms. Data, Integration, Infrastructure, Object, Technology

1 Introduction

Nowadays nuclear power is one of the most powerful and clean energy sources [1]. Constant control and monitoring of the power unit equipment parameters and components within premises and adjacent areas is held to ensure its stable work. The number of sensors reached up to several thousand per power unit [1, 2]. Requirements for reliability and survivability of monitoring tools increased in post-Fukushima period [3]. Paradigm of systems and services providing localization, minimization and crash recovery is changing. Proper information support is needed for successful and critical management during and after the accident. It is impossible to provide optimal decisions before and after the accident without timely and reliable information about the current state and change dynamics.

Almost all power unit sensors and associated equipment are combined into several networks using measurement modules [1, 4, 5]. Often these networks are based on the widespread wired serial interfaces [1, 4]. It should be noted that data flow in measuring channels of Nuclear Power Plant (NPP) is relatively small and sufficient to control and make decisions during normal operation. However, the data flow increases dramatically in pre-emergency and post-emergency conditions. It is caused by a significant increase of sensor data acquisition frequency in obtaining information about the dynamics of the situation. Furthermore, additional data acquisition systems are activated, the ones that are not involved in NNP regular operation. Therefore, emergency mode channel capacity is used to design such kind of networks that have duplication and mechanical protection of data channels to improve the communication reliability.

During the accident a part of data channels will inevitably fail. The probability of the total communication loss (with all data acquisition systems) is low, even in case of a serious accident. However during the damage of a part of communication channels the information flows should be redirected to not damage one. Thus the frequency of collisions increases and the probability of the local networks overload is high [4, 6]. It creates a lack of information when it is most needed. To overcome this situation, the additional wireless channels may be used [5, 7]. It will provide a necessary bandwidth and restore the connection with the data acquisition modules that are damaged.

The aim of this work is to design a concept of NPP integrated system for both post-emergency monitoring and decision support. It is expected that such integrated system (infrastructure as a system of systems), despite the injury during the accident, can ensure a high reliability of data exchange with measurement (and under certain conditions - control) channels to make correct decisions. Within this study we believe that such system can monitor the appropriate equipment as well as buildings and territories nearby.

2 Basic Requirements for Communication Systems of Post-Emergency Monitoring

The main characteristics of the post-emergency NPP monitoring system have to be self-descriptiveness, survivability and veracity [8]. Thus self-descriptiveness means that the data flow from data channels and control channels must be timely, sufficient and stable. The system has to provide necessary bandwidth in the post-emergency

conditions. It is obviously that a high survivability can only provide the required self-descriptiveness. Obtained data must be reliable. So in the post-emergency conditions, when a part of measurement and control modules are damaged, a required veracity can be provided by self-testing the data acquisition modules. A last one will define the current metrological characteristics of measurement and control channels.

To ensure the mentioned characteristics above in post-emergency conditions using wireless communication channels is a difficult problem, because measurement and control modules can be located indoors or under the rubble, which significantly reduces the signal level. Thus, the level of electromagnetic interference in post-emergency conditions increases dramatically at NPP. At the same time the emergency service for localization and minimization of the consequences after the accidents, should be located in a distance which excludes the emergency service from damage during the accident. Thus, providing a direct wireless connection between the measurement and control modules and control center requires significant power transmitters. However, a probability of the supplying network lines for measurement and control modules damage is very high in the accident, as well, so those modules and wireless communication components should be able to use additional emergency battery. In this case, the usage of powerful transmitters can reduce survivability of the post-emergency monitoring system by limiting its working time without battery recovery.

3 The Basic Principles of Self-Descriptiveness, Survivability and Veracity

To support the uninterrupted long functioning of the post-emergency monitoring system on in base of the wireless network within the noise high level environment we are required to reduce significantly a distance of the wireless communication as well as increase essentially the noise resistance of such communication. The first requirement can be satisfied through the usage of intermediate repeater modules. These modules have to be mobile in order to: (i) be located in the distance before the accident, such distance practically excludes them from the damage during the accident; (ii) fairly quickly (within minutes) occupy a proper place according to the accident's nature, as well as, noise level and the distance to the serviced measurement and control modules; (iii) move (change the dislocation place) according to the environment changes, specified in point (ii).

The mentioned above requirements can be easily implemented, when repeater modules are moved by air. Recently drones became to be used widely [9-11]. Nowadays, they are the most profitable platform where repeater modules can be placed. Such solution provides the maximum access to the required location as well as high-speed movement. Network of repeaters based on drones, which are located at the considerable distance from NPP and fly to locations after the accident, is able to provide the necessary data flow in minutes after the accident occurred.

To increase the wireless communication resistibility to noises, it is necessary to implement error correction and detection codes [12, 13]. However, for such environments it is required to provide noise immunity for data and instructions, as well as, for all communication systems. Existing protocols are designed for the normal noise level [14] only, and they may not function properly during the accident at NPP. It is neces-

sary therefore to develop a communication protocol able to function at the allowable noise level that corresponds to the most severe post-emergency conditions in NPP. On other hand, the high noise immunity requires using the redundant codes, therefore it is necessary to archive essentially the sensor data before coding that enables to reduce a traffic [15].

To ensure the sufficient bandwidth of communication channels it should be possible to provide a parallelization of data streams [16] i.e. a distribution of measurement and control modules between different repeater modules. However, to maintain the self-descriptiveness of the post-emergency monitoring system at a sufficient level it's necessary to provide the dynamic adaptation of repeater modules and the software-technical reconfiguration of the structure according to system functioning conditions. Some repeaters may be damaged in the post-emergency conditions or may return to the place of deployment for the battery recharging. In such cases their traffic must be taken by other repeaters.

Survivability of the post-emergency monitoring system is ensured, firstly, by providing the additional battery for a consumption of measurement and control channels. Thus, it is necessary to save the battery charge for the longer system operation. For this purpose, besides of using the energy-efficient hardware, the wireless transmitter's power should be reduced as much as possible. This can be achieved by reducing the distance between repeaters from one side, and served by them measurement and control modules from other side. It is possible to do that by optimizing the distribution of the serviced measurement and control modules between different repeaters. However, to minimize the energy consumption, it is possible to use a dynamic evaluation of errors level during the data exchange as well as adaptation of transmitter's capacity to such level.

Secondly, a high survivability of the post-emergency monitoring system is provided by different features of network repeaters. They appear on the place of accident after its occurrence (they cannot be damaged during the accident) and must be able to reallocate data streams dynamically, and optimize the own position regarding the serviced measurement and control modules, as well as a configuration of the territory and occurrence of mechanical obstacles. Drone-repeaters have to restore the battery charge during the temporary return to the service base. It should be possible to have additional drones replacing the damaged ones.

The high veracity of the data flow should be ensured by the constant self-testing of channel repeaters. Thus, it is not necessary to provide by default a frequent self-testing procedure. The usage of noise immunity codes with errors correction will allow running a current control of communication channels for errors level. At the same time, the errors level reflects the level of noise and generates information for a subsystem of drones' optimal placement. Thus, it is possible to choose the location and transmitting power within the high level of errors. However, such level has to be acceptable for a given system of noise-immune coding.

Ensuring the self-testing procedure for the measurement and control modules should be considered as a separate problem. Those modules can be damaged during the accident or after it, for example, due to the penetrating radiation. Hence data reliability can be achieved by introducing the metrological support autonomous subsystems for the measurement and control modules [17-19]. These subsystems must be

resistant to errors of its components which can increase rapidly in the post-emergency conditions, for example, due to the penetrating radiation.

Thus, the following principles of the system functioning are proposed:

- A communication network of the system for the NPP post-emergency condition monitoring is put in the drones group (fleet), that located permanently at a considerable distance from the NPP. The communication network is deployed after the accident, when drones are flying into the accident zone.
- Drones fleet is divided by the role and equipment into: repeaters, observers (equipped with a TV camera) and additional sensors (they can be located in drones or be dropped down in certain places). Drones should be able to change their role by upgrading equipment at the location base.
- Drone-repeaters work together on a principle of “one leader”. This principle ensures a maximum reliability of the wireless communication system (a minimum of collisions). If the “leading drone-repeater” (Master) is damaged then other drone-repeater takes Master functions, for example, drone with the smallest working time(among all involved ones) at the accident place.
- The Master drone-repeater determines the location zone per each drone-repeater and measurement modules which will interact with Master, or some other task.
- Each drone-repeater independently selects a location with the minimum noise (as given areas by Master), and the necessary transmitter power for measurement modules (in terms of errors in transmitted data), and the possibility of landing (with the permission of the Accident Liquidation Centre, according to the accident assessment using drones observers).
- Drone-observers enable to run the continuous visual monitoring of the accident location for: actions assessing the drones of other purpose, selecting the safe places for drone-repeater landing, assessing the trajectory of drone-sensors and their location.
- Measurement and control modules are equipped with backup batteries, blocks of wireless communication, as well as, self-testing and self-diagnostic systems.
- To meet the system requirements the self-adaptability, self-testing and self-healing procedures are used.

4 Structure of the System

The constantly active wire network of the measurement and control modules and its wireless extension [20] includes sensors and actuators from 1 to n which are connected to m traditional measurement and control modules (Fig. 1). Each of them includes the multichannel analog to digital converter ADC or digital to analog converter DAC, microcontroller of the traditional data processing and an adapter of the wired interface. Through this interface the measurement and control modules interact with the control and decision making center receiving commands and sending measurement results. To provide the work of those modules within the system of post-emergency monitoring each module is equipped with an additional wireless microcontroller which receives data from the wireless network, or prepares data for the transmission through wireless network. This microcontroller operates only in pre-

emergency and post-emergency modes. It receives measurement results (from the measuring module microcontroller or adapter of wired interface), then compresses, encodes and transmits it through the wireless interface. Measurement and control modules, with the absence of network power, are charged from the independent accumulator unit (it is not presented on Fig 1).

In the normal exploitation mode the data and commands exchange is running through the wired network. If it is damaged during an accident, another wireless network is created on the basis of drones. Due to “Master’s” commands drones are situated in the air in a way to run following functions: to cover all measuring modules which are equipped with the wireless connection; to distribute data streams through drones as evenly as possible; to secure the highest possible veracity of the transmission for sensor data and controlling commands; to avoid obstacles and making no obstacles per each other.

In the independent power supply of the measurement and control modules (from the backup accumulators only) it is very important to minimize their power consumption. For this purpose all possibilities have to be explored including a limitation of the wireless interface power, and drones must be placed in the appropriate zones close enough to the signal sources. Error level during the message exchange can be considered as one of the important criteria for the effective energy-saving. If the error level is acceptable for the selected coding method then it is enabled to try decreasing the transmitter’s power of the wireless interface both as a part of measuring modules and a part of drones.

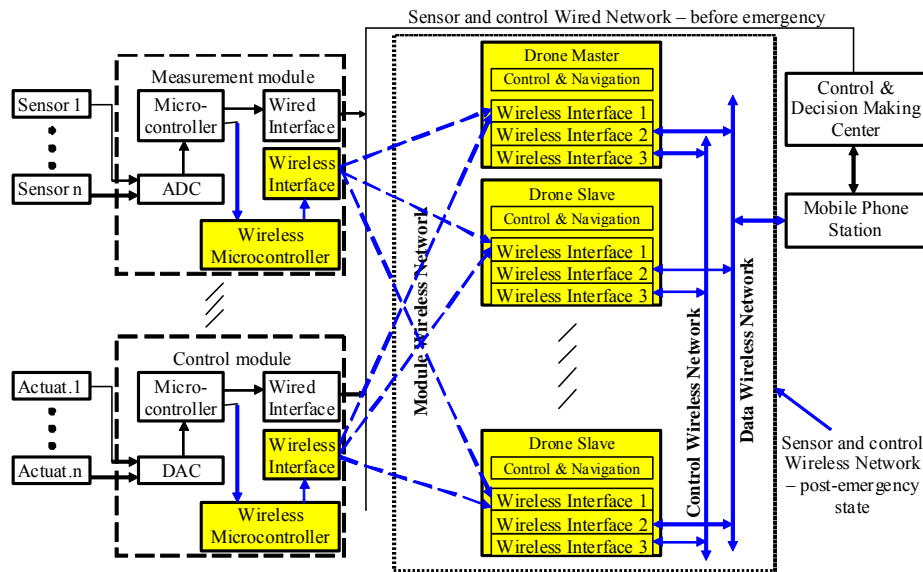


Fig. 1. Structural scheme of wired network for measurement and control modules and its wireless extension

In its turn the request to draw closer to the signal sources is connected closely with the running the principles 4 and 5 above. Hence all drones have to be equipped with

rather the high quality navigation system. Such system must provide different types of navigation below:

- Using the existing system of the global navigation (GPS) [21]. Note the level of noises is increasing rapidly in post-emergency conditions at the NPP. At the same time the signal from the GPS satellite is rather weak. So a probability is very high that such signal cannot be able for a reliable navigation;
- Using the local system of beacons similar to the GPS [22 – 25]. A distance to the local beacons (sources of the individual signals which location is tied to the local terrain map) may be even thousand times smaller. Ensuring the required signal level, even if it being powered from the accumulators, won't be a problem;
- Protection from collisions with the local obstacles is better performed using the ultrasound location [25, 26]. But the ultrasound probing impulses must have the unique individual coding. Otherwise the impulses from the other drones will be accepted as native ones, hence the defense system against the collision with the obstacles will be out of an order;
- The previous analysis showed the ultrasound locating system is extremely inaccurate, and it is too slow in dynamic conditions. It will not be able therefore to prevent a collision of the drones between each other. The assessment of the protection variants against such collisions showed that the better variant would be considered when positions of all drones are taking into account by the "leading" one (Master) with its following delivering commands about the positions and movement vectors per each subordinate (Master –Slave principle). However, when the Master is damaged, the system becomes inoperable. Hence it is necessary to try predicting a possibility with commands absence from the Master when its function may be replaced by each (or assigned one) drone [27, 28]. To avoid the collisions, such kind of replacement should be organised according to the hierarchical principle.

Note that sensor data collection and actuators control (exchange in the network of measurement and control modules), and retransmitting of these data (message exchange with the center of decision making and control), and drones control (following the "Master's" commands) are different tasks which have very little in common. Except while running the exchange task in the network of the measurement and control modules, the errors level may be defined and this information should be included when selecting the place for drone's dislocation. That's why to increase the reliability of the post-emergency monitoring system's functioning it is reasonable to divide the solution of these tasks above on the hardware level. Those tasks must be run by different microcontrollers equipped with their own peripheral devices. During this it is expedient to form the three independent wireless networks of data exchange (measurement and control modules, retransmitted data and drones control networks) with will not conflict with each other, create queues, and etc.

5 Reliability Models

According to the proposed concept the three systems of post-emergency monitoring systems (S1, S2, S3) and reliability block diagrams (correspondingly: RBD1 (Fig. 2), RBD2 (Fig. 3) and RBD3 (Fig. 4)) have been developed.

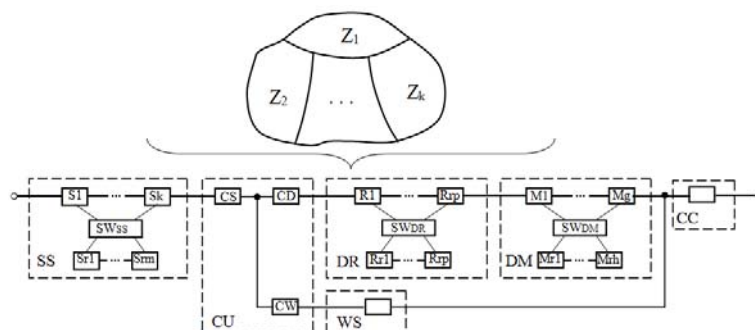


Fig. 2. Reliability block diagram of the system with general sensors, where S_i (S_{rv}) – sensors (redundant sensors), SS – sensor system, SW – switching units, CS – sensor controller, CD – drone system interface controller, CW – wire system interface controller, CU – controller unit, DR – drone transmission system, DM – drone monitoring system, R_j – transmission drones, M_b – monitoring drones, WS – wired system, CC – crisis centre

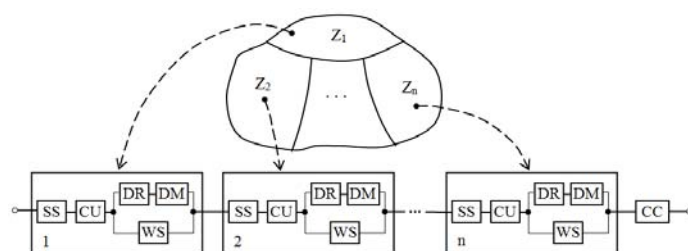


Fig. 3. Reliability block diagram of the system with separated zones of sensors and drones

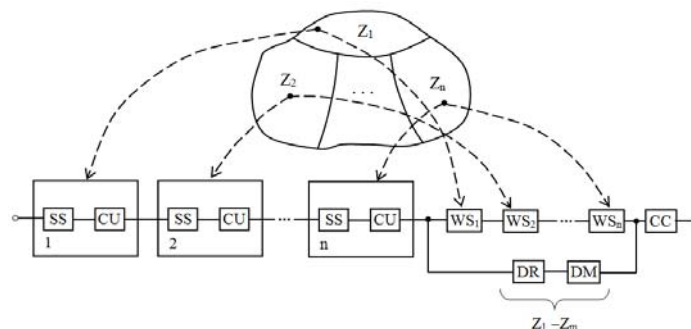


Fig. 4. Reliability block diagram of the system with separated zones of sensors and general drone fleet

Each system has a general way for increasing the system reliability, which includes sliding redundancy in SS, DR and DM – any failed element of the main chain ((S_1 - S_2 -...- S_k) for SS, (R_1 - R_2 -...- R_q) for DR and (M_1 - M_2 -...- M_g) for DM) can be replaced by means of any element of the redundancy chain ((S_{r1} - S_{r2} -...- S_{rm}) for SS, (R_{r1} - R_{r2} -...- R_{rp}) for DR and (M_{r1} - M_{r2} -...- M_{rp}) for DM). Moreover, each system

has a possibility to replace the failed main chain by means of the redundancy chain: (DR-DM) by means of (CW-WS) for S1, (CD-DR-DM) by means of (CW-WS) for S2, (WS₁-WS₂-...-WS_n) by means of ((DR-DM)₁-(DR-DM)₂-...-(DR-DM)_m) for S3.

Based on the proposed reliability block diagrams we can obtain the following formulas for calculating the reliability function (RF) per each of these systems:

$$P_{S1}(t) = p_{SS}(t) \cdot p_{CS}(t) \cdot [1 - (1 - p_{CD}(t) \cdot p_{DR}(t) \cdot p_{DM}(t)) \cdot (1 - p_{CW}(t) \cdot p_{WS}(t))] \cdot p_{CC}(t), \quad (1)$$

$$\begin{aligned} \text{where } p_{SS}(t) &= e^{-k \cdot \lambda_S \cdot t} \cdot \sum_{i=0}^m \frac{k \cdot \lambda_S \cdot t}{i!}; & p_{CS}(t) &= e^{-\lambda_{CS} \cdot t}; & p_{CD}(t) &= e^{-\lambda_{CD} \cdot t}; \\ p_{DR}(t) &= e^{-q \cdot \lambda_R \cdot t} \cdot \sum_{j=0}^p \frac{q \cdot \lambda_R \cdot t}{j!}; & p_{DM}(t) &= e^{-g \cdot \lambda_M \cdot t} \cdot \sum_{l=0}^h \frac{g \cdot \lambda_M \cdot t}{l!}; & p_{CW}(t) &= e^{-\lambda_{CW} \cdot t}; \\ p_{WS}(t) &= e^{-\lambda_{WS} \cdot t}; & p_{CC}(t) &= e^{-\lambda_{CC} \cdot t}. \end{aligned}$$

$$P_{S2}(t) = \{p_{SS}(t) \cdot p_{CU}(t) \cdot [1 - (1 - p_{DR}(t) \cdot p_{DM}(t)) \cdot (1 - p_{WS}(t))]\}^n \cdot p_{CC}(t), \quad (2)$$

$$\text{where } p_{CU}(t) = e^{-\lambda_{CU} \cdot t}.$$

$$P_{S3}(t) = (p_{SS}(t) \cdot p_{CU}(t))^n \cdot \left[1 - (1 - \prod_{i=1}^n p_{WS_i}(t)) \cdot (1 - (p_{DR}(t) \cdot p_{DM}(t))^m) \right] \cdot p_{CC}(t) \quad (3)$$

The fact, that each of the systems has devices with redundancy elements, enables to consider both those devices separately as well as systems with a multi-level degradation [29, 30]. For example, in the Table 1-3 the degradation levels with corresponding characteristics for the systems SS, DR, DM are shown

Table 1. Degradation levels characteristics of sensor system

Number of the degradation level	Condition that determines the degradation level	Formula for calculating the RF
m+1	All elements of the main chain (S1- S2-...-Sk) are functioning, all elements of the redundancy chain (Sr1- Sr2-...-Srm) are functioning	$p_{SS}(t) = e^{-k \cdot \lambda_S \cdot t} \cdot \sum_{i=0}^m \frac{k \cdot \lambda_S \cdot t}{i!}$
m	All elements of the main chain (S1- S2-...-Sk) are functioning, one element of the redundancy chain (Sr1- Sr2-...-Srm) is failed, or it is functioning instead of a failed element for the main chain	$p_{SS}(t) = e^{-k \cdot \lambda_S \cdot t} \cdot \sum_{i=0}^{m-1} \frac{k \cdot \lambda_S \cdot t}{i!}$
...

Number of the degradation level	Condition that determines the degradation level	Formula for calculating the RF
1	All elements of the main chain (S1- S2-...-Sk) are functioning, m elements of the redundancy chain (Sr1- Sr2-...-Srm) are failed, or they are functioning instead of failed elements for the main chain	$p_{SS}(t) = e^{-k \cdot \lambda_S \cdot t} \cdot \sum_{i=0}^{m-m} \frac{k \cdot \lambda_S \cdot t}{i!} =$ $e^{-k \cdot \lambda_S \cdot t} \cdot \sum_{i=0}^0 \frac{k \cdot \lambda_S \cdot t}{i!} = e^{-k \cdot \lambda_S \cdot t}$
0	At least one of the elements of the main chain (S1- S2-...-Sk) is failed, m elements of the redundancy chain (Sr1- Sr2-...-Srm) are failed or they are functioning instead of failed elements for the main chain	0

Table 2. Degradation levels characteristics of drone transmission system

Number of the degradation level	Condition that determines the degradation level	Formula for calculating the RF
p+1	All elements of the main chain (R1- R2-...-Rq) are functioning, all elements of the redundancy chain (Rr1- Rr2-...-Rrp) are functioning	$p_{DR(d=p+1)}(t) = e^{-q \cdot \lambda_R \cdot t} \cdot \sum_{j=0}^p \frac{q \cdot \lambda_R \cdot t}{j!}$
p	All elements of the main chain (R1- R2-...-Rq) are functioning, one element of the redundancy chain (Rr1- Rr2-...-Rrp) is failed or it's functioning instead of a failed element for the main chain	$p_{DR(d=p)}(t) = e^{-q \cdot \lambda_R \cdot t} \cdot \sum_{j=0}^{p-1} \frac{q \cdot \lambda_R \cdot t}{j!}$
...
1	All elements of the main chain (R1- R2-...-Rq) are functioning, p elements of the redundancy chain (Rr1- Rr2-...-Rrp) are failed or they are functioning instead of failed elements for the main chain	$p_{DR(d=1)}(t) = e^{-q \cdot \lambda_R \cdot t} \cdot \sum_{j=0}^{p-p} \frac{q \cdot \lambda_R \cdot t}{j!} =$ $e^{-q \cdot \lambda_R \cdot t} \cdot \sum_{j=0}^0 \frac{q \cdot \lambda_R \cdot t}{j!} = e^{-q \cdot \lambda_R \cdot t}$
0	At least one of the elements for the main chain (R1- R2-...-Rq) is failed, p elements of the redundancy chain (Rr1- Rr2-...-Rrp) are failed, or they are functioning instead of failed elements for the main chain	$p_{DR(d=0)}(t) = 0$

Table 3. Degradation levels characteristics of drone monitoring system

Number of the degradation level	Condition that determines the degradation level	Formula for calculating the RF
$h+1$	All elements of the main chain (M1- M2-...-Mg) are functioning, all elements of the redundancy chain (Mr1- Mr2-...-Mrp) are functioning	$P_{DM(d=h+1)}(t) = e^{-g \cdot \lambda_M \cdot t} \cdot \sum_{l=0}^h \frac{g \cdot \lambda_M \cdot t}{l!}$
h	All elements of the main chain (M1- M2-...-Mg) are functioning, one element of the redundancy chain (Mr1- Mr2-...-Mrp) is failed, or it's functioning instead of a failed element for the main chain	$P_{DM(d=h)}(t) = e^{-g \cdot \lambda_M \cdot t} \cdot \sum_{l=0}^{h-1} \frac{g \cdot \lambda_M \cdot t}{l!}$
...
1	All elements of the main chain (M1- M2-...-Mg) are functioning, h elements of the redundancy chain (Mr1- Mr2-...-Mrp) are failed, or they are functioning instead of failed elements for the main chain	$P_{DM(d=1)}(t) = e^{-g \cdot \lambda_M \cdot t} \cdot \sum_{l=0}^{h-h} \frac{g \cdot \lambda_M \cdot t}{l!} =$ $e^{-g \cdot \lambda_M \cdot t} \cdot \sum_{l=0}^0 \frac{g \cdot \lambda_M \cdot t}{l!} = e^{-g \cdot \lambda_M \cdot t}$
0	At least one of the elements of the main chain (M1- M2-...-Mg) is failed, h elements of the redundancy chain (Mr1- Mr2-...-Mrp) are failed, or they are functioning instead of failed elements for the main chain	$P_{DM(d=0)}(t) = 0$

Using the data from Tables 1-3 and assuming that devices CS, CD, CW, WS, CC are characterized by a two-level degradation (these devices have only 1 and 0 degradation level), we can determine levels and provide characteristics per each of them per each of the systems S1, S2, S3 correspondingly.

For example, in the table 4 the characteristics of the given degradation level a for the system with common sensors S1 are shown, and in Table 5 the characteristics of the lowest degradation level for the system S1 are shown too.

Table 4. Characteristics of the given degradation level a for system S1 according to degradation levels of devices, which this system includes

Degradation level of the system		Shorthand names for the devices							
		SS	DR	DM	CS	CD	SW	WS	CC
a	Degradation level of the device	m (Tab.1)	p+1 (Tab.2)	h+1 (Tab.3)	1	1	0	0	1
	Formula for calculating the RF	$P_{SI(d=a)}(t) = e^{-k \cdot \lambda_S \cdot t} \cdot \sum_{i=0}^{m-1} \frac{k \cdot \lambda_S \cdot t}{i!} \cdot e^{-q \cdot \lambda_R \cdot t} \cdot$ $\cdot \sum_{j=0}^p \frac{q \cdot \lambda_R \cdot t}{j!} \cdot e^{-g \cdot \lambda_M \cdot t} \cdot \sum_{l=0}^h \frac{g \cdot \lambda_M \cdot t}{l!} \cdot$ $\cdot e^{-\lambda_{CS} \cdot t} \cdot e^{-\lambda_{CD} \cdot t} \cdot e^{-\lambda_{CC} \cdot t}$							

Table 5. Characteristics of the lowest degradation level for the system S1

Degradation level of the system		Shorthand names for the devices							
		SS	DR	DM	CS	CD	SW	WS	CC
1	Degradation level of the device	1 (Tab.1)	0 (Tab.2)	0 (Tab.3)	1	1	1	1	1
	Formula for calculating the RF	$P_{SI(d=1)}(t) = e^{-k \cdot \lambda_S \cdot t} \cdot e^{-\lambda_{CS} \cdot t} \cdot e^{-\lambda_{CD} \cdot t} \cdot e^{-\lambda_{SW} \cdot t} \cdot e^{-\lambda_{WS} \cdot t} \cdot e^{-\lambda_{CC} \cdot t}$							

Similar tables are available for systems S2 and S3 too. For example, the characteristics of their lowest degradation levels are given in Tables 6 and 7 correspondingly.

Table 6. Characteristics of the lowest degradation level for the system S2

Degradation level of the system		Shorthand names for the devices					
		SSi	DRi	DMi	CUi	WSi	CC
1	Degradation level of the device	1 (Tab.1)	0 (Tab.2)	0 (Tab.3)	1	1	1
	Formula for calculating the RF	$P_{S2(d=1)}(t) = (e^{-k \cdot \lambda_S \cdot t} \cdot e^{-\lambda_{CU} \cdot t} \cdot e^{-\lambda_{WS} \cdot t})^n \cdot e^{-\lambda_{CC} \cdot t}$					

Table 7. Characteristics of the lowest degradation level for the system S3

Degradation level of the system		Shorthand names for the devices					
		SSi	CUi	WSi	DRj	DMj	CC
1	Degradation level of the device	1 (Tab.1)	1 (Tab.2)	0 (Tab.3)	1	1	1
	Formula for calculating the RF	$PS3(d=1)(t) = (e^{-k \cdot \lambda_S \cdot t} \cdot e^{-\lambda_{CU} \cdot t})^n \cdot (e^{-q \cdot \lambda_S \cdot t} \cdot e^{-g \cdot \lambda_S \cdot t})^m \cdot e^{-\lambda_{CC} \cdot t}$					

6 Conclusions

A proposed concept of NPP post-emergency monitoring based on drones satisfies requirements to self-descriptiveness and survivability and veracity. Such approach enables: (i) to avoid the unacceptable damage and fatal failure of post-emergency monitoring system during the accident (ii) to ensure the minimal time of system deployment (iii) to provide the sufficient bandwidth of communication channels with possible (if needed) extension (iv) to employ the recovery operation if components are damaged(v) to ensure the ability of flexible usage for other problems solving (delivery of sensors and other needed equipment to the certain location).

Proposed framework models of assessing the reliability and survivability enable to compare the output options for its further selection. Future studies should be related to the specification of output parameters as well as detailed analysis of their values for fixed solutions of systems design. Moreover it is necessary to specify the models of failures due to accidents for survivability computing.

References

1. Morichi, M.; Toubon, H.; Bronson, F.; Venkataraman, R.; Beaujoin, J.; Dubart, P.: Nuclear Measurement Technologies & Solutions Implemented During Nuclear Accident at Fukushima. In: Advancements in Nuclear Instrumentation Measurement Methods and their Applications (ANIMMA), 2013 3rd International Conference. doi: 10.1109/ANIMMA.2013.6727908, pp.1-8 (2013)
2. Hashemian, H. M.: Maintenance of Process Instrumentation in Nuclear Power Plants . Springer, p. 304. (2006)
3. Bandstra, Mark S.; Vetter, K.; Chivers, D.H.; Aucott, T.; Bates, C.; Coffey, A.; Curtis, J.; Hogan, D.; Iyengar, A.; Looker, Q.; Miller, J.; Negut, V.; Plimley, B.; Satterlee, N.; Supic, L.; Yee, B.: Measurements of Fukushima fallout by the Berkeley Radiological Air and Water Monitoring project. In: Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), DOI: 10.1109/NSSMIC.2011.6154316. IEEE, pp.18-24. (2011)
4. Sachenko, A., Kulakov, Yu., Kochan, V., Turchenko, V., Bykovyy, P., Maykiv, I.: Computer Networks. P. 476, Ternopil. (2012)
5. Hashemian, H.M.: Wireless Sensors for Predictive Maintenance of Rotating Equipment in Research Reactors. In: Annals of Nuclear Energy, Volume 38, Issues 2–3, ISSN 0306-4549, <http://dx.doi.org/10.1016/j.anucene.2010.09.012>, pp. 665-680. (2011)

6. Ryan, A.; Tisdale, J.; Godwin, M.; Coatta, D.; Nguyen, D.; Spry, S.; Sengupta, Raja; Hedrick, J.K.: Decentralized Control of Unmanned Aerial Vehicle Collaborative Sensing Missions. In: American Control Conference, 2007. ACC '07, DOI: 10.1109/ACC.2007.4282397, pp.4672-4677. (2007)
7. Daniel, K.; Rohde, S.; Wietfeld, C.: Leveraging Public Wireless Communication Infrastructures for UAV-based sensor networks. In: Technologies for Homeland Security (HST), 2010 IEEE International Conference, DOI: 10.1109/THS.2010.5655064, pp.179-184. (2010)
8. Yastrebenetsky, M., Kharchenko, V. (editors): Nuclear Power Plant Instrumentation and Control Systems for Safety and Security, IGI-Global, USA, 472p. (2014)
9. Nikolakopoulos, G., Alexis, K.: Switching Networked Attitude Control of an Unmanned Quadrotor. In: International Journal of Control, Automation, and Systems. 11(2). DOI 10.1007/s12555-011-0132-4. pp. 389-397. (2013)
10. Ryan, A.; Hedrick, J.K.: A Mode-Switching Path Planner for UAV-assisted Search and Rescue. In: Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference. DOI: 10.1109/CDC.2005.1582366. pp.1471-1476, (2005)
11. McCune, R.; Purta, R.; Dobski, M.; Jaworski, A.; Madey, G.; Madey, A.; Yi Wei; Blake, M.B.: Investigations of DDDAS for Command and Control of UAV Swarms with Agent-based modeling. In: Simulation Conference (WSC), DOI: 10.1109/WSC.2013.6721531, pp.1467-1478. (2013)
12. Zhengbing, Hu, Yatskiv, V., Sachenko, A.: Increasing the Data Transmission Robustness in WSN Using the Modified Error Correction Codes on Residue Number System. In: Elektronika ir Elektrotechnika, Vol 21. № 1, pp. 76-81. (2015)
13. Oualha, N.; Olivereau, A.: Sensor and Data Privacy in Industrial Wireless Sensor Networks. In: Network and Information Systems Security (SAR-SSI), Conference, DOI: 10.1109/SAR-SSI.2011.5931357, pp.1-8. (2011)
14. Gomaa, R.; Adly, I.; Sharshar, K.; Safwat, A.; Ragai, H.: ZigBee wireless sensor network for radiation monitoring at nuclear facilities. In: Wireless and Mobile Networking Conference (WMNC), 2013 6th Joint IFIP, DOI: 10.1109/WMNC.2013.6548954, pp.1-4. (2013)
15. Jun, Su, Yatskiv, V.: Method and Device for Image Coding & Transferring Based on Residue Number System. In: Sensors & Transducers Journal, Vol.18, Special Issue. pp.60-65, (2013)
16. Jun, Su, Yatskiv, N., Sachenko, A., Yatskiv, V.: Improved Method of Ant Colonies to Search Independent Data Transmission Routes in WSN. In: Proceedings 2012 IEEE 1st International Symposium on Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems. IDAACS-SWS-2012. Offenburg, Germany, pp.52-57. (2012)
17. Kochan, R., Kochan, O.: Method of ADC's Nonlinearity Identification. In: Proc. of the IEEE International Workshop on Intelligent Data Acquisition and Advancing Computing Systems (IDAACS'2007), Dortmund, Germany, pp. 14 – 17. (2007)
18. Kochan, R., Kochan, O. Klym, H., Osolinskiy, O., Bykovyy, P.: Integral Nonlinearity Correction of ADC Using Multi-Resistors Voltage Divider. In: Proc. of the IEEE International Workshop on Intelligent Data Acquisition and Advancing Computing Systems (IDAACS'2015), Warsaw, Poland, pp. 14-17. (2015)
19. Hashemian, H. M., Kerlin, T. W., & Upadhyaya, B. R.: Apparatus for Measuring the Degradation of a Sensor Time Constant. U.S. Patent No. 4,295,128. Washington, DC: U.S. Patent and Trademark Office. (1981)

20. Hiromoto, R.E., Sachenko, A., Kochan, V., Koval, V., Turchenko, V., Roshchupkin, O. V. Yatskiv, O., Kovalok, K.: Mobile Ad Hoc Wireless Network for Pre- and Post-Emergency Situations in Nuclear Power Plant. In: The 2nd IEEE International Symposium on Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems, Offenburg, Germany, pp.92-96. (2014)
21. Ryan, A.; Zennaro, Marco; Howell, A., Sengupta, Raja; Hedrick, J.K.: An Overview of Emerging Results in Cooperative UAV Control. In: Decision and Control, 2004. CDC. 43rd IEEE Conference, vol.1. DOI: 10.1109/CDC.2004.1428700, pp.602-607. (2004)
22. Rudol, P.; Wzorek, M.; Doherty, P.: Vision-Based Pose Estimation for Autonomous Indoor Navigation of Micro-scale Unmanned Aircraft Systems. In: Robotics and Automation (ICRA), 2010 IEEE International Conference. DOI: 10.1109/ROBOT.2010.5509203. pp.1913-1920. (2010)
23. Kim, J.; Sukkarieh, S.: Autonomous Airborne Navigation in Unknown Terrain Environments. In: Aerospace and Electronic Systems, IEEE Transactions, vol.40, no.3. DOI: 10.1109/TAES.2004.1337472, pp.1031-1045. (2004)
24. Tisdale, J.; Ryan, A.; Zu Kim; Tornqvist, D.; Hedrick, J.K.: A Multiple UAV System for Vision-based Search and Localization. In: American Control Conference. DOI: 10.1109/ACC.2008.4586784, pp.1985-1990. (2008)
25. Girard, A.R.; Howell, A.S.; Hedrick, J.K.: Border Patrol and Surveillance Missions using Multiple Unmanned Air Vehicles. In: Decision and Control, 2004. CDC. 43rd IEEE Conference, vol.1, DOI: 10.1109/CDC.2004.1428713, pp.620-625. (2004)
26. Jong-Hyuk Kim; Sukkarieh, S.: Airborne Simultaneous Localisation and Map Building. In: Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference, vol.1. DOI: 10.1109/ROBOT.2003.1241629, pp.406-411. (2003)
27. Rogers, A.; Corkill, Daniel D.; Jennings, N.R.: Agent Technologies for Sensor Networks. In: Intelligent Systems, IEEE, vol.24, no.2, DOI: 10.1109/MIS.2009.22, pp.13-17. (2009)
28. Delle Fave, F.M.; Rogers, A.; Xu, Z.; Sukkarieh, S.; Jennings, N.R.: Deploying the Maximum Algorithm for Decentralized Coordination and Task Allocation of Unmanned Aerial Vehicles for Live Aerial Imagery Collection. In: Robotics and Automation (ICRA), IEEE International Conference. DOI: 10.1109/ICRA.2012.6225053, pp.469-476. (2012)
29. Kharchenko, V. (ed.): Fundamentals of Digital Systems Reliability. National Aerospace University KhAI, 421p. (2003)
30. Kharchenko, V.: The Probabilistic Assessment of Survivability and Safety of an Unmanned Control Systems with Multistage Degradation by Use of QD-diagrams. In: Proceeding of International Conference PSAM'5, vol.1, Osaka, Japan, pp. 525-531. (2000)

Availability Model of Critical NPP I&C Systems Considering Software Reliability Indices

Bogdan Volochiy, Vitaliy Yakovyna, Oleksandr Mulyak

National University Lviv Polytechnic, 12 Bandera St., 79013, Lviv, Ukraine

bvolochiy@ukr.net, yakovyna@polynet.lviv.ua,
mulyak.oleksandr@gmail.com

Abstract. Providing the high availability level for the Instrumentation and Control (I&C) Systems in Nuclear Power Plants (NPP) is highly important. The availability of the critical NPP I&C systems depends on the hardware and software reliability behavior. The high availability of the I&C systems is ensured by the following measures: structural redundancy with choice of the I&C system configurations (two comparable sub-systems in the I&C system, majority voting "2oo3", "2oo4", etc.); maintenance of the I&C system, which implies the repair (changing) of no operational modules; using the N-version programming; software updates; automatic software restart after temporary interrupts caused by the hardware fault. This paper proposes solution of the following case: the configuration of the fault-tolerant I&C system with known reliability indexes of hardware (failure rate and temporary failure rate) is chosen, the maintenance strategy of hardware (mean time to repair, numbers of repair), methods to forecast the number of software failures and the failure rate is specified. To solve this issue, the availability model of the fault-tolerant I&C system was developed in the discrete-continuous stochastic system form. We have estimated the influence of the I&C system on the operational software parameters. Two configurations of I&C systems are presented in this paper: two comparable sub-systems in I&C system, and I&C system with majority voting "2oo3".

Keywords. Instrumentation and Control (I&C) System, Discrete-Continuous Stochastic System, Reliability Behavior, Structural-Automated Model, Markovian Chains, Software Reliability

Key Terms. Mathematical Modeling, Method, Software Systems

1 Introduction

1.1 Motivation

Nowadays the development of fault-tolerant computer-based systems (FTCSs) is a part of weaponry components, space, aviation, energy and other critical systems. One of the main tasks is to provide requirements of reliability, availability and functional safety. Thus the two types of possible risks relate to the assessment of risk, and to ensuring their safety and security.

Reliability (dependability) related design (RRD) [1-6] is a main part of development of complex fault-tolerant systems based on computers, software (SW) and hardware (HW) components. The goal of RRD is to develop the structure of FTCS tolerating HW physical failure and SW designs faults and assure required values of reliability, availability and other dependability attributes. To ensure fault-tolerance software, two or more versions of software (developed by different developers, using other languages and technologies, etc) are used [7]. Therefore use of structural redundancy for FTCS with multiple versions of software is mandatory. When commissioning software some bugs (design faults) remain in its code [8], this leads to the shut-down of the FTCS. After detection the bugs, a software update is carried out. These factors have influence on the availability of the FTCS and should be taken into account in the availability indexes. During the operation of FTCS it is also possible that the HW will fail leading to failure of the software. To recover the software operability, an automatic restart procedure, which is time consuming, is performed. The efficiency of fault-tolerant hardware of FTCS is provided by maintenance and repair.

Insufficient level of adequacy of the availability models of FTCS leads either to additional costs (while underestimating of the indexes), or to the risk of total failure (when inflating their values), namely accidents, material damage and even loss of life. Reliability and safety are assured by using (selection and development) fault-tolerant structures at RRD of the FTCS, and identifying and implementing strategies for maintenance. Adoption of wrong decisions at this stage leads to similar risks.

1.2 Related Works Analysis

Research papers, which focus on RRD, consider models of the FTCS. Most models are primarily developed to identify the impact of one the above-listed factors on reliability indexes. The rest of the factors are overlooked. Papers [4, 5] describe the reliability model of FTCS which illustrates separate HW and SW failures. Paper [6] offer reliability model of a fault-tolerant system, in which HW and SW failures are differentiated and after corrections in the program code the software failure rate is accounted for. Paper [8] describes the reliability model of the FTCS, which accounts for the software updates. In paper [10] the author outlines the relevance of the estimation of the reliability indexes of FTCS considering the failure of SW and recommends a method for their determination. Such reliability models of the FTCS produce analysis of its conditions under the failure of SW. This research suggests that $MTTF_{system} = MTTF_{software}$. Thus, it is possible to conclude that the author considers the HW of the FTCS as absolutely reliable. Such condition reduces the credibility of the

result, especially when the reliability of the HW is commensurable to the reliability of the SW. Paper [11] presents the assessment of reliability parameters of FTCS through modeling behavior using Markovian chains, which account for multiple software updates. Nevertheless there was no evidence of the quantitative assessments of the reliability measures of presented FTCS.

In paper [12], the authors propose a model of FTCS using Macro-Markovian chains, where the software failure rate, duration of software verification, failure rate and repair rate of HW are accounted for. The presented method of Macro-Markovian chains modeling [12, 13] is based on logical analysis and cannot be used for profound configurations of FTCS due to their complexity and high probability of the occurrence of mistakes. Also there is a discussion around the definition of requirements for operational verification of software of the space system, together with the research model of the object for availability evaluation and scenarios preference. It is noted that over the last ten years out of 27% of space devices failures, which were fatal or such that restricted their use, 6% were associated with HW failure and 21% with SW failure.

Software updates are necessary due to the fact that at the point of SW commissioning they may contain a number of undetected faults, which can lead to critical failures of the FTCS. Presence of HW faults relates to the complexity of the system, and failure to conduct overall testing, as such testing is time consuming and needs substation financial support. To predict the number of SW faults at the time of its commissioning various models can be used, one for example is Jelinski-Moranda [14].

A goal of the paper is to suggest a technique to develop a Markovian chain for critical NPP I&C system with different redundancy types (first of all, structure and version) using the proposed formal procedure and tool. The main idea is to decrease risks of errors during development of Markovian chain (MC) for systems with very large (tens and hundreds) number of states. We propose a special notation which allows supporting development chain step by step and designing final MC using software tools. The paper is structured in the following way. The aim of this research is calculating the availability function of critical NPP I&C system with version-structural redundancy and double software updates.

To achieve this goal we propose a newly designed reliability model of critical NPP I&C system. As an example a special critical NPP I&C system is researched (Fig.1). The following factors are accounted for in this model: overall reserve of critical NPP I&C system and joint cold redundancy of modules of main and diverse systems of critical NPP I&C system; the existence of three software versions; SW double update; physical fault.

Structure of the paper is the following. Researched critical NPP I&C system is described in the second section. An approach to developing mathematical model based on Markovian chain and detailed procedure for the critical NPP I&C system are suggested in the third and fourth sections correspondingly. Simulation results for researched Markov's model are analyzed in the section 4. Last section concludes the paper and presents some directions of future researches and developments.

2 Researched Typical NPP Instrumentation and Control Systems Based on Digital Platform

Here we provide the structure (Fig.1) of researched typical NPP Instrumentation and Control system (I&C) based on the digital platform [15]. This platform consists of main and diverse systems which are based on the Field Programmable Gates Arrays (FPGA) chips [16]. Main and Diverse systems based on the FPGA safety controller (FSC) with three parallel channels on voting logic “2-out-of-3”.

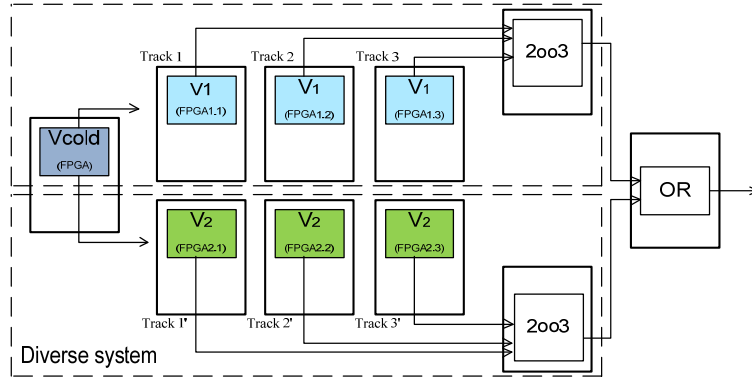


Fig. 1. Configuration of critical NPP I&C systems

This architecture consists of two system (main, diverse) each of them consists three channels connected in parallel with majority voting arrangement for the output signals, such that the output state is not charged if only one channel gives a different result, which disagrees with the other two channels.

The signals from Main and Diverse systems are comparing by element OR.

3 Methods to Forecast the Number of Software Failures and the Software Failure Rate

The papers [18, 19] describe methods of predicting numbers of undetected SW design faults. This method is based on the SW reliability model with index of complexity [20, 21]. The SW reliability model [20] describes the behavior of SW failures in non homogeneous Poisson process forms. The cumulative number of SW failures up to time t is calculated based on formula (1):

$$m(t) = \alpha \left(-\beta^s t^s e^{-\beta t} + s G_{\beta}(s) \right), \quad (1)$$

where $G_z(p) = \int_0^z t^{p-1} e^{-t} dt$ – an incomplete gamma function, α – the coefficient describes the total number of SW failure, β – the factor that represents the rate of

detection of SW failures, s – an index of SW complexity.

Work [21] researches and specifies the intervals of value of the complexity index of SW s . This circumstance has allowed for the elaboration of a formal selection rule for SW reliability models with different complexity indexes. The total number of SW failures (and, consequently, the total number of SW design faults N_{def} , on condition that one SW failure is caused by one SW design fault) is determined by the value of the function of the cumulative number of SW failures (1) at $t \rightarrow \infty$:

$$N_{def} = m(\infty) = \alpha s G(s), \quad (2)$$

where $G(s)$ – the Gamma function.

To estimate the undetected numbers of SW design faults, the following steps [22] should be performed:

- carry out SW testing and represent the result as the number of SW failures in defined interval. The input range of statistical sampling is divided into equal interval $l \leq 5 \lg(n)$ (where n – the total number of SW failures obtained during testing);
- define the point estimates of the reliability SW model parameters α , β , and define parameter s by using the method of maximum likelihood [20];
- carry out the Kolmogorov – Smirnov test for quality, reviewing the experimental reliability model described;
- use the point estimates of the reliability SW model parameters according to (2) the defined total number of SW design faults N_{def} . The forecast for the number of undetected SW design faults is obtained by subtracting the total number of SW design faults N_{def} and defined SW design faults.

Using regression analysis [18, 19], it is possible to:

- increase the accuracy of the forecast of the total number of SW design faults using formula (2); or
- decrease the time required to forecast the number of SW design faults.

The number SW faults depends on the duration of SW testing, which provides information about SW failure behavior. The variable N_{def} from formula (2) was estimated using a nonlinear regression with explanatory variables T_i – time of SW testing. The following equation (3) was used as the regression equation.

$$N_{def}(T_i) = A \left(1 - \exp \left(- (k(T_i - T_c))^d \right) \right), \quad (3)$$

where A , k , d , T_c – parameters of the regression equation.

It is then possible to determine the adjusted forecast of the total number of SW failures N_{def}^* from equation (3) on condition of the time of SW testing being unlimited ($T_i \rightarrow \infty$). Based on the equation (3) the total number of SW failure is equal to the value of regression parameter A .

To estimate the adjusted forecast for the total number of undetected SW failures N_{def}^* , the following steps should be performed:

- during the SW testing procedure, it is necessary to calculate the point estimates of the reliability SW model parameters α , β and s [20] by using the methods of maximum likelihood on the interval $(0; T_i)$, where T_i - the current moment of SW testing. It is also necessary to calculate $N_{def}(T_i)$ according the equation (2);
- estimate the parameter of regression equation (3) by using the least squares method for set of $N_{def}(T_i)$; $N_{def}^* = A$;
- the forecast number of undetected SW faults is determined by subtracting the number of detected and fixed SW faults from N_{def}^* ;
- in the case where a continuation of the process for SW testing is necessary, go back to step 1 and continue adding the new value to set $N_{def}(T_i)$.

An example of dependence $N_{def}(T_i)$ [19] which was obtained during the SW testing procedure is presented in figure 2.

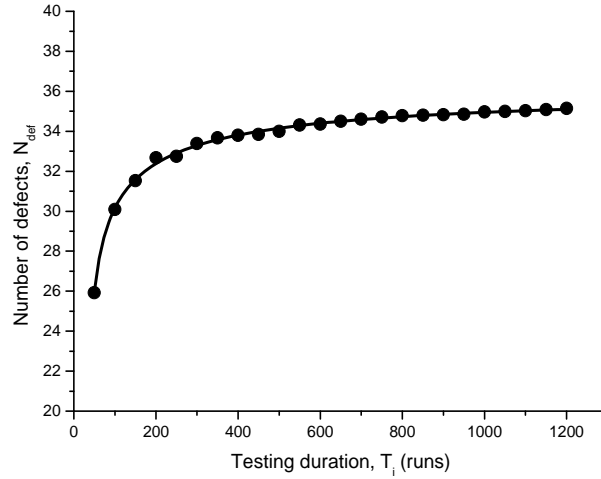


Fig. 2. Dependents of forecasting the numbers of SW faults N_{def} (points), which was calculated according equation (2) from the SW testing durations and appropriated regression equation (line)

In this case, using the methods of forecasting the SW failure numbers and equation (3) increases the accuracy of forecasting by 2-3%. Also, this method decreases the time required to forecast the number of SW failures [19].

An advantage of the SW reliability model [20] is that it is possible to estimate the SW failure rate based on SW testing results at the appropriate level of the life cycle. The SW failure rate depends on the time of SW testing (this dependence is caused by correction of the SW faults on the appropriate live cycle). The relationship takes the form (4):

$$\lambda(t) = \frac{dm(t)}{dt} = \alpha \beta^{s+1} t^s e^{-\beta t}, \quad (4)$$

As a result of using equation (4), the point value of the model parameters and the duration of SW, it is possible to calculate the SW failure rate λ_{SW} – which is constant in time. It is necessary to estimate the value of λ_{SW} , the availability of the I&C NPP system based on Markovian analysis.

The authenticity of the estimate of the undetected SW faults [23, 24] is provided by forecasting the SW failure numbers (as result SW faults) based on artificial neural networks (NN) with radial basis function (RBF). The NN RBF is a nonparametric model of behavior of SW reliability which does not require a priori knowledge and assumptions about the behavior of SW failure. In this research, input data about SW failures were presented in cumulative time series form. The cumulative time series is used for learning about the neural networks RBF and for forecasting the value of SW failure on subsequent time series.

The most reasonable results of forecasting SW failure are obtained by using NN RBF with an Inverse Multi-quadratic function (10 neurons in input layer and 30 neurons hidden layer) [24]. In this configuration, the mean square error of approximation is 1,0%. The coefficient of determination between the forecasting and controlled series is 0,9965. Although the accuracy of forecasting decreases by 1,7%, it is possible to reduce the duration of learning time of the neural network by 3-6 times by using a Gaussian function (15 neurons in the input layer and 10 neurons in the hidden layer) [23, 24].

As a result of the different SW systems analysed, a configuration of neural network RBF was conducted that could be used for time series forecast with homogeneous failure process represented by a cumulative time series.

Figure 3 presents an example of forecasting t, specifically, the total number of SW failures of the web-browser Chromium forecast using the neural network RBF with parameters listed above.

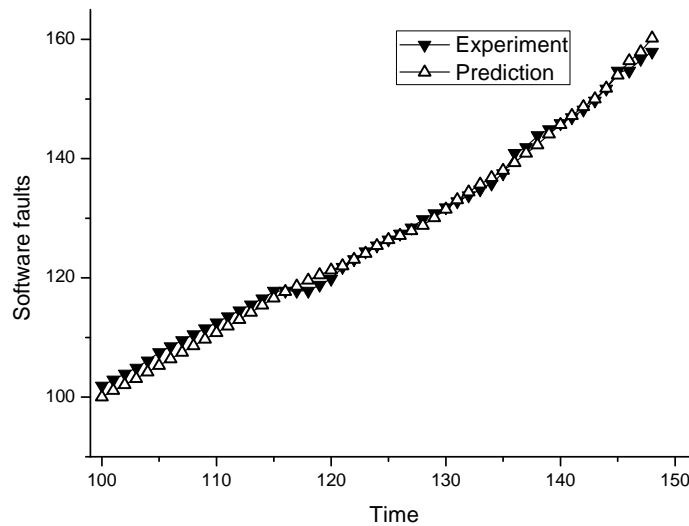


Fig. 3. An example of forecasting t, the total number of SW failures of web-browser Chromium, using the neural network RBF

This parts of paper outlines the estimated numbers of undetected SW faults using two methods based on regressions analysis and neural networks. This is used for reliably estimating the number of undetected SW faults and ensures the requirements of standard [25] are satisfied. It is considered acceptable when number of SW faults calculated by two methods is equal to or less than the standards requirement.

4 Markov's Model for Critical NPP I&C Systems with Software Updates

The method of automated development the Markovian chain of the researched critical NPP I&C systems is described in the works [9, 26]. It involves a formalized representation of the object of study as a "structural-automated model". To develop this availability model of the critical NPP I&C systems one needs to perform the following tasks: develop a verbal description of the research object (fig. 1); define the basic events; define the components of vector states, which can be described as a state of random time; define the parameters for the object of research, which should be in the model; and shape the tree of the modification of the rules and component of the vector of states.

4.1 The Procedures to Describe Behavior of the Critical NPP I&C Systems

The critical NPP I&C systems behavior is described by the following procedures:

- *Procedure 1.* Detection the failure in the critical NPP I&C systems (hardware failure, software failure). Failure can occur in the Main system (MS) and Diverse system (DS).
- *Procedure 2.* Detection of failure in the MS or in the DS of the critical NPP I&C systems.
- *Procedure 3.* Connection of the module from cold standby to faulty systems.
- *Procedure 4.* Loading the software on the module with connections from cold standby to faulty systems.
- *Procedure 5.* Software updating.
- *Procedure 6.* Repair (replacement) of the HW of the faulty systems.

4.2 A Set of the Events for the Critical NPP I&C Systems

According to described procedures which determine the behavior of critical NPP I&C systems, a list of events is composed. Events are presented in pairs corresponding to the start and the end of time intervals to perform each procedure. From this list of events for "structural-automated model" basic events are selected [9].

As a result of analysis, seven basic events in particular were determined: Event 1 - "Hardware failure of the MS module"; Event 2 - "Software failure of the MS module"; Event 3 - "Hardware failure of the DS module"; Event 4 - "Software failure of the DS module"; Event 5 - "Completing of the module switching from cold standby to non-operational systems"; Event 6 - "Completing of the software updates procedure"; Event 7 - "Completing of the procedure of the hardware repair"

4.3 Components of Vector States for the Critical NPP I&C Systems

Components of the vector state that can also be described as a state of random time. To describe the state of the system, eleven components are used: V1 – displays the current number of modules in the MS (the initial value of components V1 equal to n); V2 – displays the current number of modules in the DS (the initial value of components V2 equal to k); V3 – displays the current number of modules in cold standby (the initial value of components V3 equal to m_c); V4 – displays which software version is operated by the MS (V4=0 – first version, V4=1 – second version, V4=2 – third version); V5 – displays which software version operated by DS (V5=0 – first version, V5=1 – second version, V5=2 – third version); V6 – displays the SW faults in the MS; V7 – displays the SW faults in the DS; V8 – displays the SW failure in the MS; V9 – displays the SW failure in the DS; V10 – displays the number of non-operational module, due to HW failure.

4.4 The Parameters of the Critical NPP I&C Systems Markov's Model

Developing Markov's model of the critical NPP I&C systems, its composition and separate components should be set to relevant parameters in particular: n – number of modules that are the part of the MS; k – number of modules that are the part of the DS; m_c – number of the modules in the cold standby; λ_{hw} – the failure rate that is in MS or DS and in the hot standby; λ_{sw11} , λ_{sw12} – the failure rate of first and second software versions; T_{up1} , T_{up2} – mean time of the first and second software updates; T_{switch} – mean time of the module connections from standby; T_{rep} – mean time of hardware repair.

4.5 Structural-Automated Model of the Critical NPP I&C System for the Automated Development the Markovian Chain with Software Updates

According to the technology of a modeling, the discrete-continuous stochastic systems [9] based on certain events using the component vector state and the parameters that describe critical NPP I&C systems, and model of the critical NPP I&C systems for automated development of the Markovian chains are presented on the table 1. Below is describes the procedures of structural-automated model development:

Table 1. Structural-Automated Model of the critical NPP I&C systems for the automated development of the Markovian chains

Terms and conditions	Formula used for the intensity of the events	Rule of modification component for the state vector
Event 1. Hardware failure of the MS module		
$(V1 \geq (n-1)) \text{ AND } (V6=0)$	$V1 \cdot \lambda_{hw}$	$V1:=V1-1; V8:=V8+1$
Event 2. Software failure of the MS module		
$(V1 \geq (n-1)) \text{ AND } (V4=0) \text{ AND } (V6=0)$	$V1 \cdot \lambda_{sw11}$	$V1:=V1-1; V4:=0; V6:=1$

Terms and conditions	Formula used for the intensity of the events	Rule of modification component for the state vector
$(V1 \geq (n-1)) \text{ AND } (V4=1) \text{ AND } (V6=0)$	$V1 \cdot \lambda_{sw12}$	$V1:=V1-1; V4:=1; V6:=1$
Event 3. Hardware failure of the DS module		
$(V2 \geq (k-1)) \text{ AND } (V7=0)$	$V2 \cdot \lambda_{hw}$	$V2:=V2-1; V8:=V8+1$
Event 4. Software failure of the DS module		
$(V2 \geq (k-1)) \text{ AND } (V5=0) \text{ AND } (V7=0)$	$V2 \cdot \lambda_{sw11}$	$V2:=V2-1; V5:=0; V7:=1$
$(V2 \geq (k-1)) \text{ AND } (V5=1) \text{ AND } (V7=0)$	$V2 \cdot \lambda_{sw12}$	$V2:=V2-1; V5:=1; V7:=1$
Event 5. Completing of the module switching procedure from cold standby to non-operational systems		
$(V1 < (n-1)) \text{ AND } (V3 > 0) \text{ AND } (V8 > 0)$	$1/T_{switch}$	$V1:=V1+1; V3:=V3-1$
$(V2 < (n-1)) \text{ AND } (V3 > 0) \text{ AND } (V8 > 0)$	$1/T_{switch}$	$V2:=V2+1; V3:=V3-1$
Event 6. Completing of the software updates procedure		
$(V1 < n) \text{ AND } (V4=0) \text{ AND } (V6=1)$	$1/T_{up1}$	$V1:=n; V4:=1; V6:=0$
$(V1 < n) \text{ AND } (V4=1) \text{ AND } (V6=1)$	$1/T_{up2}$	$V1:=n; V4:=2; V6:=0$
$(V2 < k) \text{ AND } (V5=0) \text{ AND } (V7=1)$	$1/T_{up1}$	$V2:=k; V5:=1; V7:=0$
$(V2 < k) \text{ AND } (V5=1) \text{ AND } (V7=1)$	$1/T_{up2}$	$V2:=k; V5:=2; V7:=0$
Event 7. Completing of the procedure of the hardware repair		
$(V1 < n) \text{ AND } (V2=k) \text{ AND } (V6=0) \text{ AND } (V8 > 0)$	$1/T_{rep}$	$V1:=n; V6:=0; V8:=0$
$(V1=n) \text{ AND } (V2 < k) \text{ AND } (V7=0) \text{ AND } (V8 > 0)$	$1/T_{rep}$	$V2:=k; V7:=0; V8:=0$
$(V1 < n) \text{ AND } (V2 < k) \text{ AND } (V6=0) \text{ AND } (V7=0) \text{ AND } (V8 > 0)$	$1/T_{rep}$	$V1:=n; V2:=k; V8:=0$

The number of software updates can be also changed. It is necessary to change vectors V4 and V5 the *event 6*, that are responsible for the number of updates. For

example, if there are three software updates, the entry component of the event will be as follows:

(V1<n) AND (V4=2) AND (V6=1)	1/T _{up3}	V1:=n; V4:=3; V6:=0
(V2<k) AND (V5=2) AND (V7=1)	1/T _{up3}	V2:=k; V5:=3; V7:=0

4.6 Automated Development of the Markovian Chain and Determining of Availability Function

The developed availability model of the critical NPP I&C system gives the possibilities according to technology [9] for automated construct of the Markovian chains. This construction provides a software module ASNA [17]. The Markovian chains which take into account the following settings critical NPP I&C system: $n=3$; $k=3$; $m_c=0$; λ_{hw} ; λ_{sw11} , λ_{sw12} ; T_{up1} , T_{up2} ; T_{switch} ; T_{rep} , are consists of 169 state and 436 transitions. Information is available on the status of each software module ASNA we have on file "vector.vs", which is written in the form:

State 1: V1=3; V2=3; V3=0; V4=0; V5=0; V6=0; V7=0; V8=0
 State 2: V1=2; V2=3; V3=0; V4=0; V5=0; V6=0; V7=0; V8=1
 State 3: V1=1; V2=3; V3=0; V4=0; V5=0; V6=0; V7=0; V8=2

 State 169: V1=1; V2=1; V3=0; V4=2; V5=2; V6=0; V7=0; V8=4

As the configurations of researched critical NPP I&C system changes the dimension of graphs increases. Therefore for the configuration of critical NPP I&C sys (Fig. 1) with one module in a cold standby graph has 506 states and 1434 transitions.

The proposed availability model of critical NPP I&C system can be easily transformed for other features of the object of study. It is enough to: add / remove basic event; attach / remove components of the state vector; and include / exclude parameters that describe the studied system. Based on information about the work of critical NPP I&C system an appropriate change in the model could be made (Fig. 1).

Basing on the Markovian chains formulas for designing of availability critical NPP I&C system can be assembled. One measure of the availability of recovered critical NPP I&C system reveals it is an availability function. Availability functions of critical NPP I&C system is calculated as the sum of the probability functions staying in operable states of chains. Basing on these states the critical NPP I&C system availability function with parameters of critical NPP I&C is determined by the formula (5):

$$\begin{aligned}
 A(t) = & \sum_{i=1}^{10} P_i(t) + \sum_{i=12}^{14} P_i(t) + \sum_{i=16}^{18} P_i(t) + \sum_{i=21}^{22} P_i(t) + \sum_{i=26}^{51} P_i(t) + P_{53}(t) + \sum_{i=55}^{61} P_i(t) + \\
 & + \sum_{i=67}^{72} P_i(t) + P_{74}(t) + \sum_{i=76}^{77} P_i(t) + P_{79}(t) + \sum_{i=81}^{83} P_i(t) + \sum_{i=86}^{87} P_i(t) + \sum_{i=91}^{95} P_i(t) + \sum_{i=101}^{146} P_i(t) + \\
 & + P_{148}(t) + \sum_{i=150}^{151} P_i(t) + P_{153}(t) + \sum_{i=155}^{158} P_i(t) + P_{160}(t) + \sum_{i=162}^{163} P_i(t) + P_{165}(t) + \sum_{i=167}^{168} P_i(t)
 \end{aligned} \quad (5)$$

Based on the Markovian chains ("vector.vs") a system of differential equations (6) was formed. Its solution allows us to estimate the function availability value of re-searched critical NPP I&C system.

$$\begin{aligned}
 \frac{dP_1(t)}{dt} = & -6 \cdot (\lambda_{hw} + \lambda_{sw11}) \cdot P_2(t) + \frac{1}{T_{repl}} \cdot (P_2(t) + P_3(t) + P_6(t) + P_7(t) + \\
 & + P_8(t) + P_9(t) + P_{11}(t) + P_{16}(t)) \\
 \frac{dP_2(t)}{dt} = & -\frac{1}{T_{repl}} \cdot P_2(t) - 2 \cdot \lambda_{sw11} P_2(t) - 2 \cdot \lambda_{hw} \cdot P_2(t) - 3 \cdot \lambda_{hw} \cdot P_2(t) + \\
 & + 2 \cdot \lambda_{hw} P_1(t) \\
 \frac{dP_3(t)}{dt} = & -\frac{1}{T_{repl}} \cdot P_3(t) - 3 \cdot (\lambda_{hw} + \lambda_{sw11}) \cdot P_3(t) + 2 \cdot \lambda_{hw} P_2(t) \\
 & \vdots \\
 \frac{dP_{169}(t)}{dt} = & -\frac{1}{T_{repl}} \cdot P_{169}(t) + 2 \cdot \lambda_{hw} P_{156}(t) + 2 \cdot \lambda_{sw12} P_{168}(t)
 \end{aligned} \quad (6)$$

Initial conditions for the system (2) are: $P_1(t) = 1; P_2(t) = \dots = P_{169}(t) = 0$.

5 Simulation Results

5.1 Research of Influence of Software Updates Duration on the Availability Function

With the assistance of the proposed model, the following questions can be answered: What are the duration values of the first and the second software update (ensuring the values of the availability function of critical NPP I&C system of the initial phase of its operation do not reach below the specified level)? What are the allowed duration values of the first and the second SW updates? How does the correlation between the first and the second SW updates influence on the availability function?

The experiment is conducted for the condition where the duration of the first software update is significantly shorter than the duration of the second update. The duration of the first update is given within 10 - 50 hours, and the duration of the second update - 200 hours. The experiment is conducted with the following parameters critical NPP I&C system: $\lambda_{hw} = 1 \cdot 10^{-5} \text{ hour}^{-1}$; $\lambda_{sw11} = 2 \cdot 10^{-3} \text{ hour}^{-1}$; $\lambda_{sw12} = 1 \cdot 10^{-3} \text{ hour}^{-1}$; $T_{switch} = 6 \text{ min}$; $T_{rep} = 200 \text{ hour}$; $T_{up2} = 200 \text{ hour}$; (*line 1* - $T_{up1} = 10 \text{ hour}$; *line 2* - $T_{up1} = 20 \text{ hour}$; *line 3* - $T_{up1} = 30 \text{ hour}$; *line 4* - $T_{up1} = 40 \text{ hour}$; *line 5* - $T_{up1} = 50 \text{ hour}$).

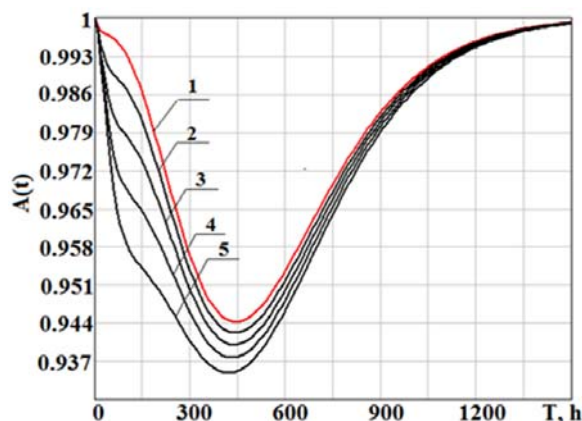


Fig. 4. Dependencies of availability function of the critical NPP I&C system on values of the software update durations (duration of the first software update for 10 to 50 hours; the duration of the second firmware update - 200 hours)

The following results are produced by the proposed experiments:

- The minimal decrease level of the availability function of the readiness of critical NPP I&C system in the first and the second experiments is the different. Hence could be argued that the first and second software updates has different influence on the reliability behavior of the critical NPP I&C system.
- With the assistance of the proposed model it is possible to choose the duration of software updates that helps to ensure a minimum allowed level of the decrease of the availability function of the critical NPP I&C system.

6 Conclusion

This research presents a model of critical NPP I&C system with double software updates to illustrate automated development of Markovian chains using a special technology and tool ASNA. Also this research presents two methods of forecasting the number of software failure with indexes of complexity and software failure rates.

The presented model can be easily adapted to different configurations of critical NPP I&C system, which envisages the use different majority voting, standby of the hardware part and as a consequence in the majority of software versions from different developers. In fact, this model can be adopted for an arbitrary number of software updates.

Future research has the potential to supplement this model with further factors:

- Erlang distribution for durations of software updates;
- Unsuccessful restarting; unreliable commutation of elements and so on.

References

1. Mudry, P.A., Vannel, F., Tempesti, G., Mange, D.: A reconfigurable hardware platform for prototyping cellular architectures. In: International Parallel and Distributed Processing Symposium. IEEE International, pp. 96--103 (2007)
2. Viktorov, O.: Reconfigurable Multiprocessor System Reliability Estimation. Asian Journal of Information Technology 6 (9), pp. 958--960 (2007)
3. Rajesh, S., Vinoth Kumar C., Srivatsan, R., Harini, S., Shanthi, A.: Fault Tolerance in Multicore Processors With Reconfigurable Hardware Unit. In: 15th International conference on high performance computing. Bangalore, INDIA, pp. 166--171 (2008)
4. Amerijckx, C., Legat, J.-D.: A Low-Power Multiprocessor Architecture For Embedded Reconfigurable Systems. In: Power and Timing Modeling, Optimization and Simulation, International Workshop, pp. 83--93 (2008)
5. Zhu, C., Gu, Z., Dick, R., Shang, L.: Reliable multiprocessor system-on-chip synthesis. In: Proc. International Conference Hardware/Software Co-design and System Synthesis, pp. 239--244 (2007)
6. Gostelow, K. P.: The design of a fault-tolerant, realtime, multi-core computer system. In: Proc. Aerospace Conference, IEEE, pp. 1--8 (2011)
7. Lyu, M.R. (ed.): Software Fault Tolerance, John Wiley & Sons, New York, NY, USA (1995)
8. Korotun, T.M.: Models and methods for testing software systems. Programming problems, 2, 76--84 (2007) (in Russian)
9. Volochii, B.: Technology of modeling the information systems. Publishing NU "Lviv Polytechnic", Lviv, Ukraine (2004) (in Ukrainian)
10. Xiong, L., Tan, Q., Xu, J.: Effects of Soft Error to System Reliability. In: Proc. Workshops of International Conference on Advanced Information Networking and Applications. pp. 204--209 (2011)
11. Ponochohnyi, J.L., Odarushchenko, E.B.: The reliability modeling non-redundant information and control systems with software updated. Radio-electronic and Computer Systems, 4(8), 93--97 (2004) (in Russian)
12. Kharchenko, V., Odarushchenko, O., Odarushchenko, V., Popov, P.: Selecting Mathematical Software for Dependability Assessment of Computer Systems Described by Stiff Markov Chains. In: Ermolayev, V. et al. (Eds.): Proc. Int. Conf. ICTERI 2013, pp. 146--162 (2013)
13. Kharchenko, V., Ponochohnyi, Y., Boyarchuk, A.: Availability Assessment of Information and Control Systems with Online Software Update and Verification. In: Information and Communication Technologies in Education, Research, and Industrial Applications Communications in Computer and Information Science, Vol. 469, , pp. 300--324, Springer International Publishing, Switzerland (2014)
14. Moranda, P. B.: An error detection model for application during software development. IEEE Trans. Reliability, N. 4, 309--312 (1981)
15. Kharchenko, V., Sklyar, V., Volkoviy, A.: Development and Verification of Dependable Multi-Version Systems on the Basic of IP-Cores. In: Proc. Int. Conf. Dependability of Computer Systems (2008)
16. Review Guidelines for Field-Programmable Gate Arrays in Nuclear Power Plant Safety Systems. NUREG/CR-7006, U.S. Nuclear Regulatory Commission, Washington, D.C., USA (2010)
17. Bobalo, J., Volochiy, B., Lozynskyi, O., Mandzii, B., Ozirkovskyi, L., Fedasuk, D., Shcherbovskyh, S., Jakovyna, V.: Mathematical models and methods for reliability

- analysis of electronic, electrical and software systems, Lviv Polytechnic Press, Lviv, Ukraine (2013)
18. Yakovyna, V.S., Seniv, M.M., Harangda, I. J.: Improved procedure for determining the number of software defects in the early stages of testing. In: International scientific conference "Intelligent Decision Support Systems and Computational Intelligence problems" (ISDMCI'2012), pp. 238, Ukraine (2012)
 19. Yakovyna, V.S., Fedasuk, D.V.: Improved procedure forecasting of software failures based on the reliability model with index of complexity. *Software Engineering*, 2, 5--13 (2012)
 20. Chabanjuk, J.M., Yakovyna, V.S., Fedasuk, D.V.: Building and research the software reliability model with index value of the project. *Software Engineering*, 1, 24--29 (2010)
 21. Yakovyna, V.S.: Modeling of software failure flow parameter and determine the series of complexity index. *Proceedings of the National University "Lviv Polytechnic"*, 806, 296--302 (2014)
 22. Yakovyna, V.S., Fedasuk, D.V., Seniv, M.M., Chabanjuk, J.M.: The method of estimation and forecasting software reliability based model with a dynamic index value of the project. *Computing*, 2(10), 97--107.
 23. Yakovyna, V. S.: Influence of RBF neural network input layer parameters on software reliability prediction. In: *Proc. 4-th International Conference Inductive Modelling (ICIM'2013)*. pp. 344--347, Kyiv (2013)
 24. Yakovyna, V. S.: Software failures prediction using RBF neural network. *Proceedings of Odessa Politechnic University*, 2(46), 111--118 (2015)
 25. Nuclear power plants - Instrumentation and control important to safety - General requirements for systems. IEC 61513. (2011)
 26. Volochiy, B., Mulyak, O., Ozirkovskyi, L., Kharchenko, V.: Automation of Quantitative Requirements Determination to Software Reliability of Safety Critical NPP I&C systems. In: *Proc. 2nd Int. Symp. on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO'16)*, Beer Sheva, Israel, February 15-18, pp. 337--346, IEEE CPS, 978-1-4673-9941-8/16, DOI 10.1109/SMRLO.2016.65 (2016)

Analysis of Methods for Providing Availability and Accessibility of Cloud Services

Max Yanovsky¹, Olga Yanovskaya¹, Vyacheslav Kharchenko^{1,2}

¹National Aerospace University named after N.E. Zhukovsky "KhAI"
17, Chkalova St., 61070 Kharkiv, Ukraine

²Centre for Safety Infrastructure-Oriented Research and Analysis
17, Chkalova St., 61070 Kharkiv, Ukraine

(M.Yanovsky, O.Yanovskaya}@csn.khai.edu, V.Kharchenko@khai.edu

Abstract. The article describes methods for dealing with reliability and fault tolerance issues of cloud datacenters. These methods are mainly focused on the elimination of single point of failure within any component of the cloud infrastructure, including the availability of infrastructure and accessibility of cloud services. The methods for providing the availability of hardware, software and network components are also presented. The analysis of the actual accessibility of the cloud services and the matching of cloud datacenter infrastructure with the level of reliability according to the Tier Classification System is described. Non-compliance of the actual accessibility with the level of High Availability for cloud web services was found.

Keywords. Availability, Accessibility, Cloud Datacenter, Service Reliability

Key Terms. ICTInfrastructure, ICTComponent, WebService, FormalMethod

1 Introduction

High availability is a critical issue for the cloud datacenter. Thus, estimating the financial loss due to the failure of datacenter components, which would result in unavailability of services, is a major part of an economic development plan for any cloud datacenter customer. The complexity of the architecture, meaning the large number of components and diversity approaches in designing the structure of the network infrastructure, causes issues in obtaining accurate evaluation of reliability, availability and accessibility of such systems.

In order to ensure end-user quality of service, the required cloud system should have the appropriate characteristics of reliability and performance. The term reliability refers to the property of an object or system to maintain, over time and within the prescribed limits, the ability to perform the required functions in set modes and conditions of use, maintenance, repair, storage and transportation according to ISO 2382-14:1978 [1]. A major property is failure-free operation, a property of an object that refers to permanent operability during some period of time. Time to failure – time to

the first failure: This property is characterized by the probability of failure-free operation – likelihood of absence of failure within a given operating time.

The main internal property of reliability is availability. Availability reflects the system's ability to perform its functions continuously. The availability coefficient is defined as the probability that at any given time t the object is in working state s , except for maintenance periods during which there is no intended use of the system. However, according to the concept of cloud-based architecture, the concept of availability as the main internal property of reliability refers to the entire infrastructure: the value of the availability factor will be determined based on how efficient the functional state of the system is considered and under what conditions the state of the system can be considered workable. The article considers methods for providing the availability of cloud infrastructure and accessibility of cloud services, as well as analyzes the effectiveness of their application based on studies for actual availability of cloud providers' services.

2 State of the Art

Studies' analysis [2, 3] leads to the conclusion that currently there is ambiguous interpretation of cloud datacenter operability conditions as it depends on the number of available and unavailable services in relation to the total amount of services, at the current time. Seeing as the end user interacts with a specific service of the cloud data-center, the term of cloud service accessibility is suggested to be used.

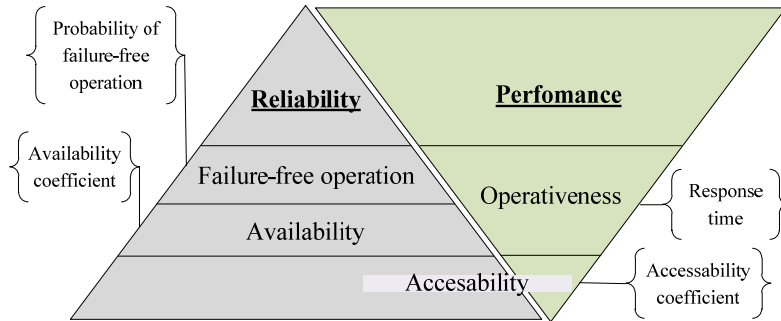


Fig. 1. Correlation between availability, reliability and accessibility indicators

Analysis of the sources [4, 5] shows that the most common availability indicator is determined by the following formula:

$$Ka = MTTF / (MTTF + MTTR), \quad (1)$$

where Ka – availability coefficient,

$MTTF$ – mean time to failure,

$MTTR$ – mean time to recover.

Property of accessibility determines the probability that at any time a certain cloud service will be available to the end user with a satisfactory response time. The main

factor is the accessibility coefficient, which includes not only the availability, but also the functional properties of the system.

With increasing demands on the quality of services in the IT-infrastructure, any kind of failures in the network are unacceptable. Even a relatively small packet loss can have a negative impact on the end-users' quality of service, especially for critical and business-critical processes, so the failure of the main switching node, link or interface may have serious consequences for the provider. The design of the cloud data-center should help minimize network failures and the severity of the consequences of potential accidents.

Advances in technology and the pace of construction of virtual data centers and cloud infrastructures have caused the development of requirements for the distribution functions of management control across multiple geographically dispersed nodes, the division of responsibility between the teams of technical personnel, the extension of monitoring and diagnostics functions support high availability and disaster recovery. According to [6], the datacenter design should include redundant components and distributed platforms, so that the physical connection and access to resources remain constant, regardless of the location and value of the current availability and performance indicators. Furthermore, to protect the competitiveness of enterprises and organizations that are customers of cloud providers, critical business applications need to be available 24/7. In case of environmental or technological disasters, the data must be restored with minimal disruption, calling for an emergency backup and recovery of business applications and the virtual machine in a different availability zone will ensure that user data is protected and accessible from anywhere.

Typically, network architects predict a 4 or 5 "nines" system availability [6]. However, each additional digit = "9" can significantly increase the cost of deployment. To achieve near-zero downtime per year of the cloud data center, one must consider not only the reliability of the hardware and network infrastructure, but also part of software.

3 Classification of Cloud Data Centers Based on Tier Standard

Cloud datacenter reliability levels have been identified in the documents "Data Center Site Infrastructure Tier Standard: Topology" [7] and "Data Center Site Infrastructure Tier Standard: Operational Sustainability" [8] of the world organization Uptime Institute [9], which are engaged in the development and verification of detailed requirements for a fault-tolerant datacenter infrastructure, certification and issuance of recommendations and expert advice on data center infrastructure according to the level of reliability.

Uptime Institute Certification on levels of reliability meets the standard ANSI/TIA-942-A [6]. Classification of data center infrastructure by levels of reliability is carried out on the basis of these basic criteria, the degree of redundancy of equipment and communication channels, the meeting of performance characteristics, functionality, efficiency and expected availability level. The requirements and recommendations apply to the following systems and components:

- architecture and topology;

- power supply system;
- cooling system;
- security;
- fire alarm system;
- structured cabling system;
- maintenance.

There are 4 standard levels of reliability:

1. TIER I: Basic Site Infrastructure;
2. TIER II: Redundant Site Infrastructure Capacity Components;
3. TIER III: Concurrently Maintainable Site Infrastructure;
4. TIER IV: Fault Tolerant Site Infrastructure.

Datacenter infrastructure and operating expenses increase in accordance with the reliability level, which gives grounds for datacenter owners to choose the class of reliability at the designing stage and in accordance with their business needs. The results of the comparative analysis of the datacenter infrastructure reliability levels are shown in the table below.

Table 1. Results of the comparative analysis of the datacenter's infrastructure reliability levels

Properties	TIER I	TIER II	TIER III	TIER IV
Level of active equipment redundancy	N	N+1	N+1	2(N+1)
Redundant channels	1	1	1 active + 1 passive	2 active
Possibility of maintenance without downtime	No	No	Yes	Yes
Fault-free operation	No	No	No	Yes
Continuous cooling system	No	No	No	Yes
Availability coefficient, %	99,671	99,749	99,982	99,995
Downtime per year, h	28,8	22	1,6	0,4

Based on the results of this analysis, the following conclusions are made:

1. The existing Tier Classification System for the reliability assessment of the datacenter infrastructure in terms of business requirements for system performance does not consider the reliability of software components.
2. The classification system of datacenter reliability on Tier levels does not explicitly consider the characteristics of the equipment, such as mean time to failure, which is not correct in assessing the availability of the system.
3. Deploying cloud and business-critical applications requires the highest level of availability TIER IV datacenter infrastructure.
4. During operation of the cloud datacenter and appending servers and equipment within a constant engineering infrastructure needs may change in the required resources, which may lead to a change in the datacenter reliability. Thus, it is necessary

to review and confirm the level of reliability of the datacenter, as well as to ensure the effective operation by highly trained personnel and administration.

In order to meet the levels of reliability and maintenance of the set level of availability and accessibility of cloud infrastructure services, a variety of methods are used to ensure fault tolerance in the core, aggregation and access layers. The main purpose of the application of methods is to ensure availability, which means to eliminate points of single failure of any component of the cloud infrastructure (hardware, software, network) at any layer (core, aggregation, access). Since hardware failure and software faults may appear in components at any layer, there are methods of fault tolerance for each of them. The methods are used to ensure the availability of cloud services at different levels of the architecture will be considered.

4 Methods for Providing Availability of Hardware Components

The objective of this group of methods is to maintain the availability of cloud services and applications in case a particular server becomes unavailable. The method can operate at multiple levels within the datacenter infrastructure. Hardware component accessibility methods are used at the physical layer ISO/OSI model. These include the following.

- Grouping of network adapters and communication channels.

In order to eliminate single points of failure at the level of communication with network, access layer servers have multiple (two or more) network interfaces (Fig. 1). This method is named NIC-Teaming and it involves the grouping of multiple physical connections into one logical channel - LAG (Link Aggregation). The logical connection may be in active-active mode that combines multiple channels into a single logical load sharing or active-passive mode, wherein the second interface is idle as long as the first interface operates as usual.

- Using hot-swappable interfaces.

This method requires the ability to install or remove the interface card on the router or switch without having to power off the device. The controller dynamically recognizes the new interface and begins the data exchange. As a result, new components can be inserted and removed without interrupting the system's operation.

- Use of highly reliable server access layer.

Hardware components of highly reliable servers have the highest values of MTTF.

5 Methods for Providing Availability of Software Components

An analysis [10, 11] found that the main methods of resilience at the application level are the use of pooling resources, protected applications and resources of critical applications as well as and the migration of virtual machines and the use of Unified In-Service Software Upgrades.

- Using application pools of resources.

This method is based on the fact that multiple instances of applications are combined to the pool of resources that are distributed throughout the network (Fig. 2).

According to [11], the use of resource pools is an effective solution for resiliency, but the main disadvantage of this approach is the problem of synchronization. This solution requires more effective methods of planning, synchronization and load balancing on the coordination sites.

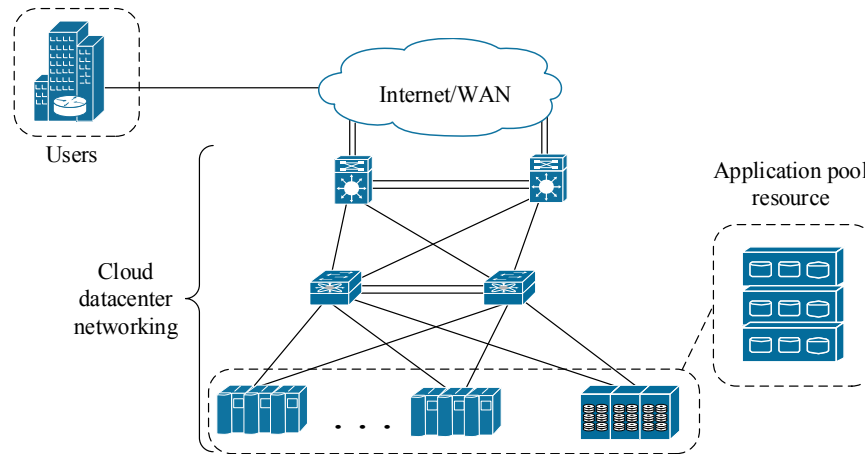


Fig. 2. Use of application pools of resources

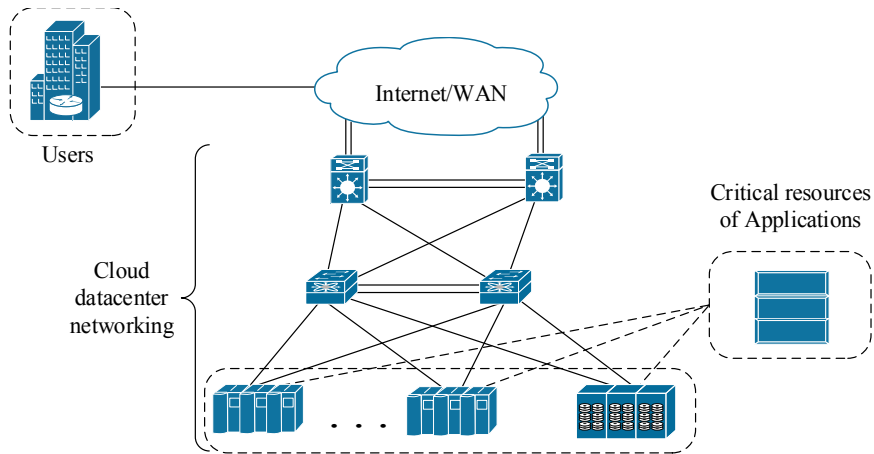


Fig. 3. Transfer of critical resources of applications

- Transfer of critical application resources.

Some applications have critical resources, that for various reasons are not possible or desirable to replicate. They can either work on the basis of high-performance servers, that makes replication too expensive, or they can include critical resources that makes replication not possible due to security threats or exploitation reasons.

Under these conditions, one application server is a single point of failure. In order to minimize the risk of failure for critical resources of applications, the execution of these applications are made on several powerful servers, failover and high availability provided by the active / standby configuration mode for disaster recovery. Connection problems are solved by multisession network connections between the server and clients, and multiple network routes (Fig. 3).

Conditions of effective application of this method are the presence of redundant network links and backup systems as well as continuous monitoring of the status of servers and data replication, in order to maintain synchronization of the active and standby systems.

- Migration of virtual machines.

Virtual machine migration is an effective method for providing fault-tolerance and for maintaining service availability in the event of a failure of the physical server on which it is running. This method assumes that the virtual machine has its own running copy on a server located in another rack or in another datacenter. In this case, services that are deployed on the initial virtual machine are replicated on another virtual machine.

- Using a single integrated service system updates.

The ability to provide unified system ISSU (Unified In-Service Software Upgrades) updates of an operating system without shutting down network devices that are scheduled for preliminary verification of compatibility, is supported by some versions of operating systems within a number of network equipment manufacturers [11], thus avoiding the risks associated with downtime and failed updates of network operating systems.

6 Methods for Providing Availability of Network Components

- Redundant network devices

The analysis of the examined standards and guidelines for the design of the data-center allows us to determine that the redundancy of network devices as a method of fault tolerance, involves the duplication of the core level routers, access layer and distribution switches.

Additional mechanisms for balancing the load between them increase network performance and reduce latency.

Apart from that, in order to minimize the effects of a single point of failure, the network device may also be used in methods such as hot-swappable interface, Unified In-Service Software Upgrades, redundant switching and routing mechanisms.

- Redundant switching and routing mechanisms

The main purpose of this method is to create redundant switching for network devices.

Along with a redundant configuration, switching fabric with two switch modules is used to increase the capacity and performance the switch. The third module, if present, provides an additional precision ($2 + 1$) for switching functions, so that if one of the two functional modules becomes inoperable, a third module can take over the function of the failed module. Redundant routing mechanisms provide simultaneous operation of multiple routing protocols as well as protection from the routing and switching loops based on the following protocols, technology and standards:

- L3 dynamic routing protocols in the core level (OSPF, RIP, or static routing);
- Multiple Spanning Tree Protocol (MSTP);
- MPLS in the core level;
- 802.3ad LAG;
- 802.1q Virtual LANs;
- RTG (Redundant trunk groups);
- VRRP;
- MPLS in the aggregation level.

Based on the analysis, the presented methods highlight a number of common disadvantages in their application: complexity of the architecture, the processes of its maintenance and operation due to demand for high priced resources, additional overhead costs on excess equipment and permanent high-quality maintenance. In order to determine the effectiveness of the methods and architecture considered, an analysis of the accessibility of services from known cloud providers should be conducted.

7 Analysis of Actual Services Accessibility of Cloud Providers

Figures 4 - 5 are bar graphs that illustrate the results of statistical data processing regarding the services accessibility of cloud providers.

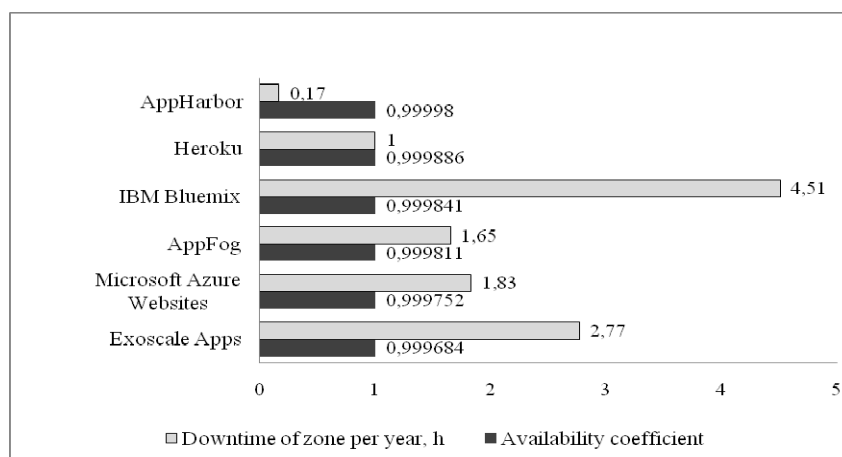


Fig. 4. Actual services accessibility of cloud providers: PaaS service model

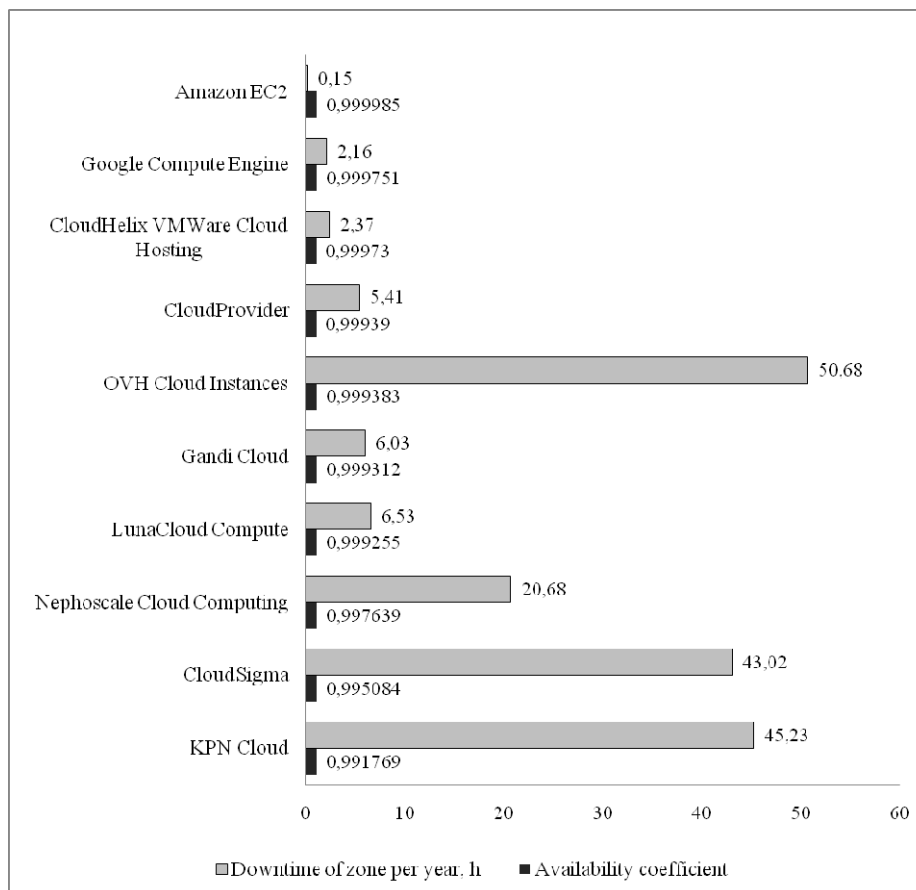


Fig. 5. Actual services accessibility of cloud providers: IaaS service model

The values shown in the above histograms are obtained by analyzing the actual values of downtime for a time period of 1 year, as published by Cloud Harmony [12] for the service models of PaaS and IaaS [13].

According to the analysis of the actual accessibility of cloud providers' services, and as shown by the values represented in the histograms, we can conclude that the average accessibility of a datacenter's cloud services corresponds to a value of 0.999.

In order to determine whether the claimed quality of the service is in compliance with the actual quality of service, an analysis of the service-level agreement (SLA) of known cloud providers was performed. The results of the analysis are summarized in the table 2.

In order to verify compliance of the datacenter infrastructure cloud providers with levels of reliability according to the Tier Classification System, made analysis of the data with characteristics of the datacenter provided by cloud providers. The results of the analysis are presented in the table 3.

Table 2. Results of the comparative analysis of the datacenter infrastructure reliability levels

Cloud provider / Service type	Claimed levels of service accessibility (replication services provided at least two availability zones)
Microsoft Azure / Microsoft Online Services [14]	99.95 %
Microsoft Azure / Virtual Machines [15]	99.95 %
Microsoft Azure / Cloud Services [16]	99.95 %
Google Cloud Platform / Google Compute Engine [17]	99.95 %
Google Cloud Platform / Google App Engine [18]	99.95 %
Amazon EC2 [19]	99.95 %
Rackspace Cloud Servers [20]	99.9 %

Table 3. Reliability levels of cloud providers' datacenters

Cloud provider	Tier Reliability Level
Amazon [21]	IV
Microsoft Azure [22]	IV
Rackspace Cloud [23]	IV

Analysis of the results leads to the conclusion: despite the fact that the cloud data-center infrastructure matches the fourth level of reliability with an availability coefficient of 0,99995, the actual average access-infrastructure of cloud service providers, on average, corresponds to a value of 0.999. Therefore, it is necessary to improve the models and methods of assessing the availability and accessibility for services of client-server cloud infrastructure to obtain more accurate estimates of the reliability indices.

8 Case study

This section provides the results of a case study on accessibility, based on the simulation model of the cloud server that is running 3 virtual machines.

The results of the statistical analysis of time characteristics of the WEB-applications servers' performance [24] confirm that, based on a mathematical model of computing systems, it can be assumed that the random variables: time of the requests towards the server has an exponential distribution, and the input query is a Poisson series (QS M/M/1). Based on this assumption, a model was found. Requests arrive to the physical server network card (Physical NIC), then are distributed among the virtual network interfaces and are processed there. In this model, another element to the input stream applications was added: that of lost requests, due to loss of server performance (hardware failure, operating system or hypervisor refusal). It can be

assumed that hardware failures occur on average once every 300,000 hours, with mean time between OS failures is 1440 hours, and the hypervisor is 2880 hours.

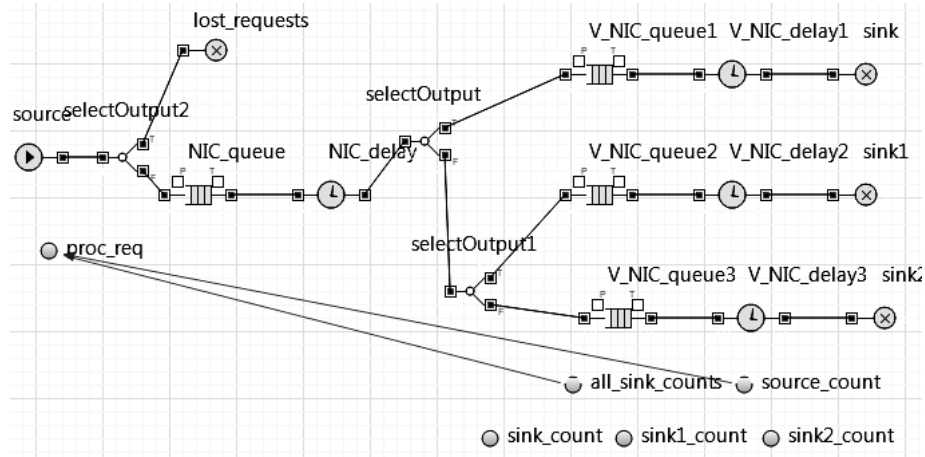


Fig. 6. Simulation model of the cloud server presented in AnyLogic environment

The resulting value, obtained as a percentage of processed requests on the time line, is as follows:

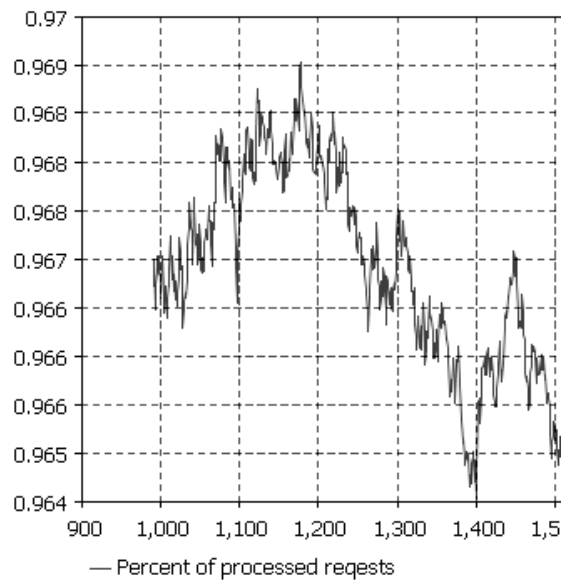


Fig. 7. Simulation results

According to the results shown in the graph, it is evident that during the day (24-hour model) each unit of the system's configuration, that has its test parameters on the average percentage of time, processes user requests of about 96%.

Almost all of the input parameters depend on the specific hardware and software implementation of the system. The exception is the MTTR of hardware failure, the operating system and the hypervisor as well as the corresponding recovery rate. The experiments conducted showed the recovery time takes an average of 1 to 2 hours. The simulation results, when changing the data in this range, show a decrease in the recovery times of up to 1 hour while it is possible to get an increase in availability features in 4 digits.

It is possible to improve this figure in several ways:

- 1) increase the average time between failures of hardware and software server;
- 2) decrease the time of exchange between physical and virtual network adapters, which can be varied from a few milliseconds to tens of milliseconds, depending on the specific application platform and hypervisor which implements virtual server infrastructure.

9 Conclusion

The analysis methods for fault tolerance and availability of client-server cloud infrastructure services were presented. These methods are more focused on the points of single failure elimination for each component of the cloud infrastructure (hardware, software, network) in every level (core, distribution, access), as well as models and methods of estimating the availability and accessibility of cloud-based architectures. The results of the analysis of the actual datacenter services accessibility of cloud service provider for service models PaaS and IaaS, as well as compliance of datacenter infrastructure of cloud providers with levels of reliability according to the Tier classification were presented. Despite the use of different methods of fault tolerance in cloud infrastructures, there is the problem of inconsistency of the actual system availability level of "High Availability" for critical and business-critical web applications. Thus, the direction of future research would be towards the improvement of the models and methods so as to ensure accessibility of cloud services. Furthermore, according to analysis results, it is important for the direction of future research results to find an effective combination of the discussed methods for providing the required level of availability and accessibility.

References

1. ISO/IEC 2382-14: Information technology. Vocabulary. Part 14: Reliability, maintenance and availability (1997)
2. R. S. Couto, M. E. M. Campista and L. H. M. K. Costa: A reliability analysis of datacenter topologies, Global Communications Conference (GLOBECOM), IEEE, Anaheim, CA, pp. 1890--1895. doi: 10.1109/GLOCOM.2012.6503391 (2012)
3. K. V. Vishwanath, N. Nagappan: Characterizing cloud computing hardware reliability, in Proceedings of the 1st ACM symposium on Cloud computing (SoCC '10). ACM, New York, NY, USA, pp. 193--204. doi: 10.1145/1807128.1807161 (2010)

4. Fernandes, S., Tavares, E., Santos, M., Lira, V., & Maciel, P.: Dependability assessment of virtualized networks. IEEE International Conference on Communications (ICC), pp. 2711-2716, Ottawa (2012)
5. Downtime Statistics of Current Cloud Solution,
6. <http://iwgcr.org/wp-content/uploads/2014/03/downtime-statistics-current-1.3.pdf>
7. TIA/EIA-942: Telecommunications Infrastructure Standard for Data Centers (2005)
8. Data Center Site Infrastructure Tier Standart: Topology,
9. <https://uptimeinstitute.com/publications/asset/tier-standard-topology>
10. Data Center Site Infrastructure Tier Standart: Topology,
11. <https://uptimeinstitute.com/publications/asset/tier-standard-operational-sustainability>
12. About Uptime Institute, <https://uptimeinstitute.com/about-ui>
13. Cisco Data Center Infrastructure 2.5 Design Guide,
14. http://www.scn.rain.com/~neighorn/PDF/Cisco_Data_Center_Infrastructure_Design_Guide.pdf
15. Juniper Networks, "Cloud Ready Data Center Network Design Guide",
16. <http://www.juniper.net/us/en/local/pdf/design-guides/8020014-en.pdf>
17. Research and compare cloud providers and services, <https://cloudharmony.com/status>
18. Cloud Computing Vendors Taxonomy, <http://clountaxonomy.opencrowd.com/taxonomy>
19. SLA for App Service, <https://azure.microsoft.com/en-us/support/legal/sla/app-service>
20. SLA for Virtual Machines, <https://azure.microsoft.com/en-us/support/legal/sla/virtual-machines>
21. SLA for Cloud Services, <https://azure.microsoft.com/en-us/support/legal/sla/cloud-services>
22. Google Compute Engine Service Level Agreement, <https://cloud.google.com/compute/sla>
23. Google App Engine Service Level Agreement, <https://cloud.google.com/appengine/sla>
24. Amazon EC2 Service Level Agreement, <https://aws.amazon.com/ru/ec2/sla>
25. Cloud Service Level Agreement, <https://www.rackspace.com/information/legal/cloud/sla>
26. Amazon called out over cloud security,
27. http://www.techworld.com.au/article/326287/amazon_called_over_cloud_security_secrecy
28. Microsoft Cloud Services,
29. <https://assets.digitalmarketplace.service.gov.uk/documents/93064/4504230568132608-pricing-document.pdf>
30. About Rackspace. Global Infrastructure and Uptime Guarantee,
31. <http://www.rackspace.com/about/datacenters>
32. Gorbenko, A., Romanovsky, A.: Time-Outing Internet Services. Security & Privacy, IEEE, 11(2), pp. 68--71. doi: 10.1109/MSP.2013.43 (2013)

Complexity-based Prediction of Faults Number for Software Modules Ranking Before Testing: Technique and Case Study

Svitlana Yaremchuk¹, Vyacheslav Kharchenko²

¹Danube Institute of National University "Odessa Maritime Academy",
8 Didrikhson St., 65029, Odessa, Ukraine

svetlana397@yandex.ru

²National Airspace University of N. E. Zhukovsky "The Kharkiv Aviation Institute",
17 Department, Chkalova St., 61070, Kharkiv, Ukraine

v_s_kharchenko@ukr.net

Abstract. The proposed method is based on estimating software modules' complexity by means of metrics. Indices of source code complexity are the input data for the method. Ranged selection software modules with faults form the output data for the method. The verification of the method has been performed on the basis of representative selection of experimental data for five systems. An appropriate coefficient is proposed and calculated for assessment of the method's efficiency. It shows the correlation between a part of revealed faults to a part of verified modules. The efficiency coefficient value increases on condition of increase in quantity of applied metrics. The proposed method enables to reach the required reliability by means of a choice of certain part of modules for identification as it is possible for the most part of faults. The method provides a software reliability increase in limited resources conditions.

Keywords. Reliability, efficiency, complexity metrics, software modules with faults

Key Terms. Software System, Metric, Software Engineering Process

1 Introduction

1.1 Motivation

Software system reliability (hereinafter referred to as the System) substantially depends on its complicatedness. Complicatedness of multiple and mutually related requirements leads to formation of erroneous, incomplete, contradicting specifications. System's architecture complicatedness causes fatal errors (critical faults) generations,

which may result to project ruination. Complicated system implementation leads to errors' increase in the initial code. Various sources show that testing process reveals 10 – 20 errors per each 1,000 code lines as an average. Changes in initial code are required at the stage of system operation contributing to newer errors generation. Integrated approach is therefore essential to access and ensure reliability basing on system complicatedness evaluation throughout its lifetime. Actual software indicators/metrics should be measured thoroughly and continuously to ensure reliability and handle development and verification procedures. Increasing systems' complicatedness, from the one side and limited resources, from the other side, generate contradictions between required systems' reliability and that achieved once development is completed. It is therefore critically important to develop and to research reliability assessment methods basing on system's complicatedness analysis, both in general and its components taking into note limited testing resources.

1.2 Related Works

Systems' reliability should be continuously controlled throughout their entire lifecycle. Therefore the process of required reliability progress should be controlled at all stages. Any process management demands indicative values evaluation. Multiple number of models, methods and tools are developed to evaluate various reliability indicators at various lifecycle stages.

Works [1-5] propose various models and methods to increase precision of evaluation of reliability indicators. In work [1] the authors analysed the impact of debugging time upon reliability evaluation and forecast. The debugging time causes an overestimation of the perceived software quality up to 15% in studied dataset; similarly, it causes the underestimation of testing time required to obtain a given quality for a software product.

Authors of [2] suggested a method based on matrix allowing known reliability models. Applying matrix enables to 1) form an allowance vector for software system under development to take into account development process peculiarities; 2) to choice appropriate model basing on allowances vector; and 3) calculate model parameters. However, the allowances matrix requires adding newer models and methods.

The most of expenditures are associated with the testing stage. Here the bulk of defects are also revealed. Authors of [3] proved that applying combined operation and debugging tests (OP-D) ensures, in general, addressing both faults occurring with high frequency at the operation time and those occurring with lower frequency. The objective of proposed OP-D technique lies in reliability improving only by means of operation testing revealing, at the same time, as many bugs, as possible, as may be achieved via debugging application. However, the proposed OP-D technique is unable to take into account rigid restrictions in testing resources.

Work [4] has been devoted to the basic theory of the of software systems' dynamics and established the theoretical basis for the reliability assessment of and proposed a new universal method for such assessment. This method takes into account effects of secondary faults, and improves the accuracy of software reliability indices more than twice. Proposed models and methods require experimentally obtained data of faults revealing time in the course of testing. The more data becomes available, the

more precise are reliability evaluating indicators, shorter is the testing period, and the less are resources required to achieve required reliability level.

Work [5] describes software reliability model with complexity index based on the non-homogenous Poisson process (NHPP). The method of software reliability assessment has been developed on the basis of generalized NHPP model and the testing sufficiency criteria. Software application for software failures prediction using artificial neural networks has been developed. However, reliability forecasting by means of neural networks binds developers to use software known not so well and to study network characteristics. Complicatedness, reduced studying rate and high level of operation margin prevent this method from implementation into routine engineering practice.

It is statistically established that revealing and elimination of faults at the earlier stages of development are 10 – 100 times cheaper, than same actions performed at the ready product testing stage. Reliability indicators should be evaluated prior testing commencement to save expenditures. Developers should be aware of faults quantity in the software system. This knowledge enables to draw up testing process and to link it with available resources. However, this data is insufficient. It is more important to know which modules contain the highest quantity of faults. It enables to minimize testing group efforts and to maximize faults revealing which plays an essential part in restricted resources situation. Works [6] through [13] propose certain methods of code complicatedness evaluation by means of metrics.

Work [6] describes analytic model to establish relation between faults quantity in the initial code and complicatedness indexes basing upon developed model. Application and statistical analysis of the proposed method showed that discrepancy between actual faults quantity and estimated one amounted to 11%. The work researches methods of faults localization in software modules. As a result, 9% discrepancy between obtained indicators and actual quantity of faults has been found. However, the problems associated with choice and ranging of the fault prone software modules still has not been solved.

The method proposed in [7] allows combining techniques so as to maximize the number of faults revealed for the tested software from those expected to be available. As for fault types the method refers to well-known orthogonal defects classification (ODC). As for testing techniques, the method applies techniques of functional, statistical, robustness and stress testing. The final result is a forecast of faults quantity of each ODC category each technique is capable to detect for a particular application. However, it still remains a question which modules should be tested if limited monetary and time resources make it impossible to carry on total testing.

Solutions for the problem in question are proposed in works [8] through [13]. Authors of work [8] developed three individual models forecasting fault-proneness in data set: one with ascending stepwise logistic regression, one with descending stepwise logistic regression and one without stepwise selection in logistic regression. The authors concluded that descending stepwise regression provides the best model. The level of false alarm rate is too high in all the models, while it should be contrary.

Complexity metrics in predicting fault-prone software modules have been intensively studied in the work [9]. The binary logistic regression method is applied in studying using as an example commonly available data on five commercial products. The study shows that (1) models generated using more data sets can improve the pre-

diction accuracy but not the recall rate; (2) reducing the cut-off value may improve the recall rate, but the number of false positives will increase, resulting in higher maintenance efforts.

The authors of work [10] studied, whether metrics available in the early lifecycle (i.e. requirement metrics), combined with metrics available in the late lifecycle (i.e. code metrics), may be used to identify fault prone modules using genetic algorithm-based technique. However, applying multiple various metrics increases complicatedness of the method. It also requires running software, which is not well known.

Authors of research [11] have empirically evaluated performance of Hierarchical Clustering Technique (HCT) in predicting fault-prone modules using open source software and metrics. The proposed technique has shown 85% accuracy. However, developers should study thoroughly MatLab hierarchy clusterization algorithms to use this method in practice.

Research [12] addresses the problem of predicting fault prone modules using data mining techniques. In this study the authors applied different data mining rule-based classification techniques on several commonly available datasets. The newly proposed algorithm is an enhanced existing algorithm in terms of effectiveness (i.e. generating less number of rules) and accuracy (i.e. improving the results). Despite of rules reduction the method itself and its automatization possibilities are too complicated to enable its application.

The authors of work [13] have studied various metrics (requirement metrics, design metrics and code metrics) and techniques to identify fault prone modules. The proposed metrics are aimed to provide higher prediction results than existing techniques. However, various metrics combined with their application algorithms substantially increase complicatedness of this method.

The testing stage, aimed to improve reliability of a software system, is the most expensive and time-consuming one. Moreover, since dimensions of software systems have increased significantly during the past decades, effective utilization of limited testing resource has become even more important than before. A software system is typically composed by a number of modules. Each of them needs to be assigned with some testing resource before the testing stage commences. Hence, a natural question is how to allocate the testing resource to modules so that the reliability of a software system is could be maximized. Such a problem was formally defined by Ohtera and Yamada as the Optimal Testing Resource Allocation Problems (OTRAPs). In the works [14-16] it was proved that an optimal allocation scheme may lead to significant improvement in terms of the reliability of a software system. The available resource should be allocated among modules in a way enabling maximum number of faults to be removed from each module to achieve higher software reliability.

The optimization problems are formulated in work [14] as nonlinear programming problems (NLPP), which are solved by means of software reliability improving model based on a non-homogenous Poisson process which incorporates Log-logistic testing-effort function. Work [15] suggests solving the OTRAPs by means of Multi-Objective Algorithms known as Hierarchy Particle Swarm Optimization Algorithm. Experimental results show that the proposed algorithm has overcome the drawbacks of the existing algorithm, and is more efficient. The main goal of the article [16] is to examine the resource allocation plan for fault detection and correction process of the software to save costs during testing and operational phases. The authors developed a model

for fault detection and correction process in pursue of the said aim. Methods proposed in works [14-16] are complicated enough. Their application requires highly qualified personnel.

Analysis of described methods shows situation, as follows. Methods [1] through [5] don't permit to plan and evaluate testing resources, since they require testing data. Methods [6] through [16] require data on projects developed previously. Differences between systems being currently developed and those already existing are inevitably accompanied by substantial margin in evaluation. Allowances methods in realistic development processes don't work, thus reducing precision in reliability indicators' evaluation. Methods are complicated in application. Input data processing requires specific not commonly known software and efforts of skilled highly qualified personnel. All these items increase systems' development costs and reduce popularity of the methods. Methods don't take into account rigid restrictions of the testing resources. Analysis of described defects and drawbacks enables to offer general approaches to reliability evaluation and management throughout systems' lifetime taking into consideration their complicatedness and limited resources assigned for development basing on comparatively simple and more precise methods.

The rest of the work is composed, as follows. Chapter 2 describes general approaches to reliability assessment and management. Chapter 3 presents the check assumption and the description of the complexity-based prediction technique. Chapter 4 describes the case study in terms of the experimental verification methods, results' analysis and discussion. Conclusions are represented in chapter 5.

2 Approaches to Reliability Assessment and Management

There are two key principles of reliability evaluation and reliability management. The first one is technology-oriented. It means that proposed evaluation methods and aids are based on complicatedness recording and may be applicable with various systems development methodologies. Adjustable software may ensure their flexibility and simplify their implementation into business – software corporate processes. The second aspect of reliability management is based on resource considerations and enables to calculate and put foundation under resources allocation for reliability evaluation and ensuring. It comprises determining costs associated with achieving required reliability and comparative analysis of expenditures and incomes derived from faults detection in all the created products (specifications demands, project, initial code, etc.) throughout the entire development stage and risks and losses reduction at the operation stage. The technological- and resources-oriented approach to reliability evaluation and reliability management supposes following methods to be applied.

Complicatedness evaluation method with appropriate metrics enables to identify the most complicated and the most fault-prone requirements. These activities may contribute to design faults identification and elimination at the earliest stages. Their analysis will contribute to earlier faults identification and elimination in specifications. The project complicatedness should be also evaluated using metrics at the development stage. Such an approach enables to analyse project, to identify most complicated subsystems, components, interfaces, carry on their decomposition and im-

prove links. Such activities may contribute to earlier identification and elimination of design faults.

One of proposed ways to reduce complicatedness and to improve program code reliability lies in its optimization (also known as refactoring). A large number of authors of publications in software systems development state that the initial complicatedness cannot be reduced. From their point of view, simplifying any software constructions leads to other ones becoming more complicated. However, the problem of code complicatedness quantification both at pre-optimization and post-optimization stages remains still unsolved.

Once the initial code is written, the testing procedure should be scheduled and resources should be allocated. Here total faults number in the system should be evaluated basing on code metrics. Since lack of resources is a common situation they should be spent in a most efficient manner. Selection and ranking modules with the highest fault number method should be developed. Testing of such modules will enable to reveal maximum faults with minimum number of modules subject to testing.

Conditions analysis for development, testing and allowance for reliability models enables to select a suitable model for evaluating achieved reliability indicators. Profile of system application should be evaluated at the operation stage by means of multiple metrics. Such a measure enables to improve efficient resources allocation to monitor the system to improve its reliability.

The proposed approach is directed to achieve required reliability in the most cost-saving manner. With testing being the most expensive and extended development stage initially, within the framework of general approach, method of improving testing efficiency should be developed. It determines the *researches goal – complexity-based fault number forecasting technique development and verification for ranking software modules prior testing*.

3 The Complexity-Based Prediction Technique of Fault Number for Ranking of Software Modules

3.1 Check Assumption of the Prediction Technique

The proposed method is based on an assumption that the most complicated modules (classes, components) contain the bulk quantity of faults. This assumption is logical and is often applied in scientific publications, requiring, however, practical confirmation. Five sets of experiment data have been used for verification [17]. The mentioned data was placed with Internet resource of the PROMISE (PRedictOr Models In Software Engineering) depository. The data was collected by numerous researchers and experts in software faults forecasting and published as freeware. The data contains evaluations based on various metrics criteria for various software systems for certain periods of time. Selected software systems are developed for various purposes, are implemented by means of various programming languages and methodologies, have various initial code dimensions, various number of active developers, development period, versions quantity, etc. The sole essential factor common for all reviewed software system consists in their being non-commercial projects with freeware initial code. The data in question represent complicatedness indicators by metrics and num-

ber of faults revealed in the course of modules' testing for five various object-oriented systems. Systems characteristics are shown in Table 1. Total volume of explored data exceeds 1.000.000 initial code lines, contains 4.330 modules and 4.449 faults. Comparative analysis of data grouped in the Table 1 shows that the researched system very substantially in their characteristics. Table 1 contains also complicatedness measures per one module in metrics. Multiple metric of choice $M = \{RFC, WMC, LCOM, LOC, NPM, CE, CBO\}$ is explained in the work [6]. Table 1 data analysis shows that complicatedness of software systems differs substantially in a number of metrics. Significant total volume and specified differences in the abovementioned Systems enable to regard this sample as a representative sample. Actions described below are proposed to identify probable correlations between complicatedness and faults quantity in an individual module. Total modules' multitude SET_{mod} for each system has been indexed seven (7) times (according to applied metrics number) in decreasing sequence of complicatedness numerical indicator per each metric. A certain part has been derived from each multitude (10 %, ... 50 %) of the most complicated modules with highest complicatedness level under each metric. As a result seven basic sub-collections have been obtained $SUBSET_{mod}^{RFC}$, $SUBSET_{mod}^{WMC}$, $SUBSET_{mod}^{LCOM}$, $SUBSET_{mod}^{LOC}$, $SUBSET_{mod}^{NPM}$, $SUBSET_{mod}^{CE}$, $SUBSET_{mod}^{CBO}$.

Table 1. Data on Systems being under Research

Systems characteristics	Abbreviated systems names and versions				
	Luc 2.4	Xer 1.4	Ant 1.7	Xal 2.7	Cam 1.4
Modules quantity	340	588	746	910	1 746
Faults quantity	632	1 596	338	1213	670
Fault modules ratio	40%	26%	78%	1%	83%
Number of faults in a module	1.86	2.71	0.45	1.33	0.38
Faults density per 1,000 lines	6.14	11.30	1.62	2.83	3.42
RFC Metric	25	19	34	29	21
WMC Metric	10	10	11	11	9
LCOM Metric	69	75	89	126	73
LOC Metric	303	240	280	471	112
NPM Metric	7	8	8	9	7
CE Metric	5	3	6	7	6
CBO Metric	11	6	11	12	11

Intersection of the sets by seven metrics formed newer modules' sub-collections $SUBSET_{mod}^7$, by six metrics $SUBSET_{mod}^6$, by five metrics $SUBSET_{mod}^5$, by four metrics $SUBSET_{mod}^4$, by three metrics $SUBSET_{mod}^3$, by two metrics $SUBSET_{mod}^2$,

by one metric $SUBSET_{mod}^1$. Actual quantity of faults per one module has been counted for each enlisted sub-collection. The results are displayed in Table 2.

Table 2. Average actual faults quantity per module

Modules complicatedness	Abbreviated systems names and versions				
	Luc 2.4	Xer 1.4	Ant 1.7	Xal 2.7	Cam 1.4
$SUBSET_{mod}^7$	13,5	20,9	3,1	3,0	4,6
$SUBSET_{mod}^6$	5,4	13,2	1,9	2,3	2,0
$SUBSET_{mod}^5$	3,1	14,2	1,1	2,0	1,1
$SUBSET_{mod}^4$	2,1	5,8	1,4	1,7	1,2
$SUBSET_{mod}^3$	2,0	4,3	1,0	1,7	1,9
$SUBSET_{mod}^2$	2,3	3,8	1,0	1,5	1,6
$SUBSET_{mod}^1$	1,5	2,8	0,4	1,7	0,4

Data contained in Table 2 shows that the most complicated software system modules have the greatest number of faults. As modules complicatedness decreases, the average fault number in module reduces confirming thus the validity of allowance method.

The proposed method should be run once the initial code is written, prior its testing commences. Complicatedness indicators of individual modules in metrics form the method's input data. Should the developers lack the corporative and practically checked set of metrics commonly available and most informative for faults forecasting object-oriented metrics considered in the work [6] may be applied as starters. The set of applicable metrics is indicated as $M = \{M_1, \dots, M_n\}$. The method supposed proceeding in five steps, as described below.

Step 1. Complicatedness indicators calculation split into metrics for individual modules in a system being developed using standard or corporate software. ...

Step 2. Total modules set SET_{mod} should be indexed n times as the complicatedness indicator descends with each metric $SET_{mod}^{M_1}, \dots, SET_{mod}^{M_n}$.

Step 3. A certain part of the most complicated modules by a specific metric should be drawn from each indexed set. Dimensions of such a part should be determined referring to testing resources' restrictions. The less are these resources, the less is the part of drawn modules. As a result n basic sub-collections have been obtained for n metrics $SUBSET_{mod}^{M_1}, \dots, SUBSET_{mod}^{M_n}$.

Step 4. Once intersection for basic sub-collections $SUBSET_{mod}^{M_1}, \dots, SUBSET_{mod}^{M_n}$ is determined, intermediate sub-collections should be formed with top complicatedness indicators by the metrics $SUBSET_{mod}^n = SUBSET_{mod}^{M_1} \cap \dots \cap SUBSET_{mod}^{M_n}$. All the

modules within this sub-collections should be ranked as n . Choosing all and any $n-1$ metrics intersections for subset $\text{SUBSET}_{\text{mod}}^{n-1}$ should be determined with all the modules to be ranked as $n-1$. Similar procedure should be applied until subset for any single metric is built as $\text{SUBSET}_{\text{mod}}^1$ with all the modules in it ranked as 1 . Number of ranks should be the same as metrics number.

Step 5. Determining the sum of subsets for $\text{SUBSET}_{\text{mod}}^n, \text{SUBSET}_{\text{mod}}^{n-1}, \dots, \text{SUBSET}_{\text{mod}}^1$ the resulting ranked modules sample should be formed as $\text{SUBSET}_{\text{mod}}^R = \text{SUBSET}_{\text{mod}}^n \cup \text{SUBSET}_{\text{mod}}^{n-1} \cup \dots \cup \text{SUBSET}_{\text{mod}}^1$.

The resulting sample should include modules with top complicatedness indicators simultaneously by n metrics (*rank 7* in our example), $n-1$ metrics (*rank 6*), ... and, finally, one metric (*rank 1*). Modules ranking in resulting sample is necessary to establish sequence of their testing. Top rank modules are the most complicated, supposed to contain the most of faults, and are subject to be tested at the first turn. The proposed method should be applied once the initial code is written and prior its testing starts. Complicatedness indicators of individual modules within the developed system form the input data with output data being the resulting ranked sample of modules with faults. Basic sample dimensions should be determined by restrictions imposed by testing resources.

4 Case Study

The method has been verified by means of commonly accessible experimental data [17] and specially developed software. Number of selected modules has been calculated as well, as faults number in these modules for each system. As this data has been being calculated the quantity of involved metrics has been varied as well, as the modules number with top complicatedness indicators corresponding to these metrics. Obtained data has been analysed. The first subtask of the analysis was to define relations between faults number and modules' complicatedness and was estimated simultaneously by a number of metrics. The second subtask was to define relation between faults number and that of selected modules. Selected ratio values encompassed 10%, ..., 50% including the most complicated modules. The parts dimensions were governed by probable restrictions in testing resources. Since the researched systems substantially varied in their characteristics, relative percentage indicators have been calculated.

4.1 The Experiment Performance Technique

99% modules within one system under research contained faults making it remarkably distinguished among others. Data processing and analysis showed that modules selection based on their complicatedness had not given increase in a quantity of faults in them. Average data for four systems is represented in Table 3. Data is allocated in two lines. The first line displays modules ratio from their total quantity. The lower ratio shows faults ratio contained in these modules in relation to their total quantity.

Values highlighted as examples in the last column in Table 1 should be interpreted, as follows. Base sample 20% of the most complicated modules per each metric forms a resulting ranked modules sample 41,8% ratio of their total number, containing 74,8% ratio of total faults' quantity.

The data in question should be applied in a manner, as follows. Tests are developed for each selected module. These tests take a certain amount of labour expressed in person-hours. Time consumed by running all such tests for all selected modules enables to calculate testing expenditures required to reveal 74,8% of the total faults quantity. At the further stage, restrictions imposed by labour, time, financial, hardware and software resources should be considered. Should there be a lack in resources required to test selected modules, basic sample dimensions should be reduced.

Factor $k = \frac{d_m}{m}$, with m being modules ratio in their total quantity and faults ratio of their total number contained in appropriate modules is proposed to reduce the number of analysed indicators and to facilitate the method's efficiency analysis. The higher is the k value, the more is the number of faults within the selected modules.

Table 3. Average calculated data on four systems under research

Mod- ules' ratio	Indica- tors	Rank 7	Rank 6	Rank 5	Rank 4	Rank 3	Rank 2	Rank 1	Σ
10 %	mod., %	2,0	1,3	2,5	2,5	3,8	3,3	8,0	23,3
	faults, %	15,0	6,5	7,0	8,0	6,5	4,8	7,8	55,5
13 %	mod., %	2,3	1,8	2,8	3,5	4,0	3,8	8,8	26,8
	faults, %	17,3	8,8	8,3	7,0	6,8	5,8	8,5	62,3
17 %	mod., %	2,5	3,0	3,5	5,3	5,0	4,8	11,0	35,5
	faults, %	19,8	13,0	7,3	9,5	6,3	5,8	9,0	70,5
20 %	mod., %	4,5	4,0	4,5	5,5	6,0	5,5	11,8	41,8
	faults, %	23,5	13,3	8,3	7,8	6,5	6,8	8,8	74,8
25 %	mod., %	6,3	5,0	5,5	7,0	7,0	6,3	13,5	50,5
	faults, %	30,0	12,0	9,0	9,0	5,8	5,3	9,0	80,0
33 %	mod., %	9,0	7,5	8,0	8,8	6,5	8,3	14,5	62,5
	faults, %	36,0	14,8	9,3	8,5	5,3	6,5	7,0	87,3
50 %	mod., %	17,8	13,3	10,5	9,8	9,0	9,3	11,5	81,0
	faults, %	47,0	19,3	10,8	5,8	4,3	4,8	4,5	96,3

Basing on the k application, Table 3 data was transformed and represented in Table 4 format. The k values in Table 4 show how the selected and ranked modules share exceeds the part of faults they contain.

Table 4. Method's average efficiency factors k for four systems

Modules' ratio	k for rank 7	k for rank 6	k for rank 5	k for rank 4	k for rank 3	k for rank 2	k for rank 1
10 %	7,50	5,20	2,80	3,20	1,73	1,46	0,97
13 %	7,67	5,00	3,00	2,00	1,69	1,53	0,97
17 %	6,58	4,33	2,07	1,81	1,25	1,21	0,82
20 %	5,22	3,31	1,83	1,41	1,08	1,23	0,74
25 %	4,80	2,40	1,64	1,29	0,82	0,84	0,67
33 %	4,00	1,97	1,16	0,97	0,81	0,79	0,48
50 %	2,65	1,45	1,02	0,59	0,47	0,51	0,39

For example, $k = 7,5$ means that testing of 1% of the selected modules enables to reveal 7,5% of faults. The following factor interpretation is proposed. If $k < 1$ the modules' selection is inefficient from the point of view of their complicatedness evaluated by metrics. If $1 \leq k < 1,5$ modules are selected with minor efficiency. Finally, if $k \geq 1,5$ the modules are selected efficiently. E.g. if 13 % modules are selected their testing may be efficient if they are characterized by high complicatedness levels simultaneously by seven, six, five, four, three, and two metrics. If 20 % modules are selected, testing of modules with high complicatedness values by seven, six, and five metrics will be efficient.

4.2 Results of Analysis

Analysis of Table 4 data enabled to identify two tendencies. The first tendency may be followed in each horizontal data line. Efficient modules selection will be maximum for modules with top complicatedness indications involving simultaneously all the involved metrics. The k decreases as modules' complicatedness reduces. The k is less than 1 in the Table 4, last column. Thus, modules selection by a single metric is not advantageous for faults detection. The second tendency may be followed in each data column. Selection of the least part of the most complicated modules is the most efficient. The k value decreases as the selected module number increases. Consequently, the less is selected modules number, the higher is effect or modules complicatedness evaluation by metrics.

Thus the method's efficiency depends on metrics' quantity determining simultaneously high complicatedness and on modules sample dimensions. The coefficient value increases with the metrics' number increase and sample dimensions decrease. Reverse statement is valid, too. Maximum coefficient value amounted to 7,5 with average value 2,34 and minimum value 0,39.

Selecting modules and tests for them enables to estimate testing expenditures to detect a certain quantity of faults. Since restricted resources prevent testing all the modules, the unrevealed faults may cause risk of losses. Risks calculation and comparing them with testing expenditures may enable managers to find adequate and economically grounded solutions. Tables 3 and 4 data may be helpful not only for efficient

testing efforts allocations, but also for efficient application of other code verification methods, e.g. inspections, statistical analysis, etc.

The proposed set of known metrics for code complicatedness evaluation is an initial, or starting, one. Software corporations' experts may apply their unique corporate metrics set, proven at practice. Discussable aspect is using obtained results not only for code testing but its applicability for other verification methods, such as survey, control, inspections, audits, statistical analysis. It is not still clarified, whether selected modules contain fault-free modules, what is their quantity, how it depends on complicatedness level.

5 Conclusions

Elements are developed for technological and resources oriented approach to reliability management at all the lifecycle stages in restricted resources environment. The technique of complexity-based prediction of faults number for ranking software modules is proposed within the framework of general approach. Supposed statement that the most complicated modules contain the bulk quantity of faults is proved experimentally. The method is applicable after initial code is written prior its testing commencement. The method's input data consists of complicatedness indicators for individual modules of the system being in development by metrics. The output data represents a sample of ranked modules of certain dimensions with a certain number of faults. Sample dimensions are only restricted by testing resources.

The proposed method is aimed to efficient testing with restricted resources. Verification of the method using a representative sample of experimental data demonstrated its efficient operability. Proposed efficiency factor depends on number of metrics by which modules have simultaneous high complicatedness rates and on modules sample dimensions. The achieved results enable to control process of achieving required reliability with restricted resources by means of allocating the testing efforts to a certain number of modules with the highest faults quantity.

Sampling such modules and selecting tests for them enables to estimate *a priori* testing costs by means of summing up testing time for all the selected modules. Maximization of revealed faults number and improved systems' operating reliability may reduce expenditures and increase developers' revenues. Risks of faults triggering at the operation stage may be mitigated. Testing expenditures will be efficient investments into the systems' reliability.

The proposed method is simple enough to be applied in practice. The method does not demand any additional data except the code complicatedness evaluation for the system being in development. The method may be completely automated. Revealed application restriction concerns systems with 99% faulty modules ratio. Only in such a case method application may be inefficient. Prospective direction of further researching may be implementation of the proposed method into business processes involving various methodologies of software systems developments. Testing expenditures calculation model and method of comparing expenditures with unrevealed faults risk evaluation should be developed. Software implementation of the proposed method should be described.

References

1. Cinque, M., Gaiani, C., Stradis, D., Pecchia, A., Pietrantuono R., Russo, S.: On the Impact of Debugging on Software Reliability Growth Analysis: A Case Study. *Computational Science and Its Applications (ICCSA)*, pp 461—475, LNCS-8583 (2014)
2. Kharchenko, V. S., Tarasyuk, O. M., Sklyar, V.V.: The Method of Software Reliability y Growth Models Choice Using Assumptions Matrix. *Proceedings of 26-th Annual Int. Computer Software and Applications Conference, COMPSAC*, Oxford, England, pp. 541--546 (2002)
3. Cotroneo, D., Pietrantuono, R., Russo, S.: Combining Operational and Debug Testing for Improving Reliability. *The Journal of Systems and Software*, vol. 62, no. 2, pp. 408-- 423 (2013)
4. Mayevsky, D. A.: A New Approach to Software Reliability. *Lecture Notes in Computer Science. Software Engineering for Resilient Systems*. Berlin, Springer, no. 8166, pp. 156--168 (2013)
5. Yakovyna, V. S.: Influence of RBF neural network input layer parameters on software reliability prediction. *4-th International Conference on Inductive Modelling ICIM'2013*, Kyiv, pp. 344--347 (2013)
6. Yaremchuk, S. O.: Models, methods and technology of a priori estimation reliability indices of accounting and analytical information systems. *Dissertation for scientific degree of candidate of technical sciences in specialty Information technology*. – Odessa National Polytechnic University, Odessa, Ukraine, 210 p. (2015) (In Ukrainian)
7. Cotroneo, D., Pietrantuono, R., Russo, S.: Testing techniques selection based on ODC fault types and software metrics. *The Journal of Systems and Software*, no. 86, pp. 1613--1637 (2013)
8. Mause, G., Grbac, T. G., Basic, B. D.: Multivariate Logistic Regression Prediction of Fault-Proneness in Software Modules. *MIPRO 2012/CTI*, pp. 813--818 (2012)
9. Yul, L., Mishra, A.: Experience in Predicting Fault-Prone Software Modules Using Complexity Metrics. *Quality Technology & Quantitative Management*, vol. 9, no. 4, pp. 421--433 (2012)
10. Sandhu, P. S., Khullar, S., Singh, S., Bains, S. K., Kaur, M., Singh, G.: A Study on Early Prediction of Fault Proneness in Software Modules using Genetic Algorithm. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 4, no. 12, pp. 1891--1896 (2010)
11. Kaur, S., Mahajan, M., Sandhu, P. S.: Identification of Fault Prone Modules in Open Source Software Systems using Hierarchical based Clustering. *ISEMS*, Bangkok, pp. 238--244 (2011)
12. Najadat, H., Alsmadi, I.: Enhance Rule Based Detection for Software Fault Prone Modules. *International Journal of Software Engineering and Its Applications*, vol. 6, no. 1, pp. 75--86 (2012)
13. Kaur, I.: A Compound Metric for Identification of Fault Prone Modules. *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 17, Issue 6, Ver. V, pp. 31--35 (2015)
14. Ahmad, N., Khan, M., Islam, S.: Optimal Allocation of Testing Resource for Modular Software based on Testing-Effort Dependent Software Reliability Growth. *ICCCNT-12*, Coimbatore, India (2012)
15. Pavithra, M.: Optimal Testing Resource Allocation Problems in Software System using Heuristic Algorithm. *Bonfring International Journal of Software Engineering and Soft Computing*, vol. 2, no. 4, pp. 1--9 (2012)

16. Nasar, M., Johri, P.: Testing and Debugging Resource Allocation for Fault Detection and Removal Process. International Journal of New Computer Architectures and their Applications, no. 4, pp. 193--200 (2014)
17. The PROMISE Repository of empirical software engineering data, <http://openscience.us/repo/defect/ck/>

Importance Analysis of k -out-of- n Multi-State Systems based on Direct Partial Logic Derivatives

Elena Zaitseva, Miroslav Kvassay, Vitaly Levashenko, Jozef Kostolny

University of Zilina, Faculty of Management Science and Informatics, Department of
Informatics, Univerzitna 8215/1, 01026 Zilina, Slovakia

{elena.zaitseva, miroslav.kvassay, vitaly.levashenko,
jozef.kostolny}@fri.uniza.sk

Abstract. In this paper, analysis of k -out-of- n Multi-State Systems (MSSs) is considered. This type of systems consists of n components, and it can be in state j if and only if at least k components are in state j or greater. Investigation of such systems has been considered in several works. However, most of them have dealt only with an efficient computation of several global characteristics, such as system state probability or system availability. In this paper, we deal with importance analysis for such systems. Particularly, we focus on two commonly used importance measures – structural importance and Birnbaum's importance. Using logical differential calculus, we propose an efficient way of how to calculate these measures for a k -out-of- n MSS. The obtained results are then used to analyze an oil supply system.

Keywords. Multi-State System, Structure Function, Structural Importance, Birnbaum's Importance, Logical Differential Calculus

Key Terms. Reliability, Model, Approach, Methodology, Scientific Field

1 Introduction

Reliability has been considered as an important characteristic of many systems [1], [2], [3], [4]. Most of the systems, whose reliability has to be investigated, are composed of more than one element (component) and, therefore, one of the principal tasks of reliability analysis is investigation of influence of individual system components on the proper work of the system [4]. Such investigation requires creation of a mathematical model of the system. As a rule, two approaches are used in reliability analysis. The first one is based on the assumption that the system and all its components (system elements that are assumed to be indivisible into smaller parts) can be in one of only two possible states – functioning (represented by number 1) and failed (presented as number 0). These systems are known as Binary-State Systems (BSSs) [1], [4]. Models based on this approach are suitable for the analysis of consequences of system failure, but they are not very appropriate for the investigation of processes that

result in system failure. For this purpose, the approach based on the idea that the system and all its components can be in one of more than two states is more suitable. In this case, we say about Multi-State Systems (MSSs) [2], [3], [4].

One of the current issues of reliability engineering is evaluation of complex systems. Such systems are composed of many components with very various natures [5]. Typical instances of such systems are healthcare systems [6] containing hardware and software components, human factor, and organizational elements, or complex distribution networks composing of many different hardware elements [7]. This heterogeneity indicates that it can be quite difficult to model such systems as BSSs and, therefore, MSSs are more appropriate.

Reliability analysis of MSSs is a complex problem that includes a lot of tasks. This paper focuses on two specific tasks: identification of situations in which component or its state is critical for system activity, i.e. situations in which a degradation of a component results in system degradation, and quantification of importance of individual system components, i.e. finding components with the greatest influence on system activity. One of the possible ways of how to perform this analysis is application of logical differential calculus.

Logical differential calculus has originally been developed for analysis of dynamic properties of Multiple-Valued Logic (MVL) functions [8]. This tool can also be applied in reliability analysis to identify circumstances under which a change of a state of a system component results in a change of system state. So, it allows us to find situations in which a degradation of a given component or its state is critical for system degradation [9], [10]. In this paper, we consider its application in importance analysis of k -out-of- n MSSs.

A k -out-of- n system is composed of n components. Based on [11], behavior of this system can be described as follows:

- if at least k components are in state $m-1$, then the system is in state $m-1$,
- else if at least k components are in state $m-2$ or better, then it is in state $m-2$,
- ...
- else if at least k components are in state 1 or better, then it is in state 1,
- else the system is in state 0.

Efficient ways of how to calculate some global characteristics, e.g. system state probability or system availability, for this kind of systems have been considered, for example, in [12], [13]. However, those papers have not considered investigation of importance of individual system components (or their states) on system activity. This problem is taken into account on the next pages.

2 Reliability Analysis of Multi-State Systems

A MSS is a mathematical representation of a system under consideration. It allows us to define m levels at which the system or its components can operate. These levels are known as states of the system/component and they take values from the set $\{0, 1, \dots, m-1\}$. State 0 means that the system/component is completely failed, while state $m-1$ implies that it is perfectly functioning. A mapping that defines the dependency of

system state on the states of its components is known as structure function. For a MSS composed of n components, this function has the following form [3], [9]:

$$\phi(\mathbf{x}): \{0,1,\dots, m-1\}^n \rightarrow \{0,1,\dots, m-1\} , \quad (1)$$

where x_i is a variable defining state of the i -th system component for $i = 1, 2, \dots, n$, and $\mathbf{x} = (x_1, x_2, \dots, x_n)$ is a vector of components states (state vector). Specially, if $m = 2$, then definition (1) agrees with the structure function of a BSS. Please note that the structure function of a MSS can also be viewed as a MVL function. In this case, x_i is known as a MVL variable and vector \mathbf{x} can be named as a MVL vector.

Based on the properties of the structure function, two classes of MSSs can be defined – coherent and incoherent. A MSS is coherent if its structure function is non-decreasing in all its arguments, i.e. there exist no circumstances under which degradation (improvement) of a system component can result in improvement (deterioration) of system state. In what follows, only coherent systems are considered.

The structure function defines system topology. However, if we want to investigate not only system topology but also some others characteristics, such as system state probability, system availability, or importance of individual system components, the state probabilities of the system components have to be known. For the i -th system component, they will be denoted as follows:

$$p_{i,s} = \Pr\{x_i = s\}, s = 0, 1, \dots, m-1 . \quad (2)$$

Using these probabilities and the system structure function, we can compute two basic characteristics of MSSs – system state probability [3], [9]:

$$\Pr\{\phi(\mathbf{x}) = j\}, \text{ for } j \in \{0, 1, \dots, m-1\} , \quad (3)$$

and system availability/unavailability with respect to state j of the system [3], [9]:

$$A^{\geq j} = \Pr\{\phi(\mathbf{x}) \geq j\}, U^{\geq j} = \Pr\{\phi(\mathbf{x}) < j\}, \text{ for } j \in \{1, 2, \dots, m-1\} . \quad (4)$$

This definition implies that system availability (unavailability) for system state j agrees with the probability that the system is in such state that its performance can (cannot) satisfy a demand corresponding to state j . For illustration, let us consider a power supply unit that can generate 30 MW, 10 MW, or 0 MW of electricity. Clearly, the system has 3 performance levels from which level 30 MW corresponds to state 2, level 10 MW to state 1 and level 0 MW to state 0. If there is a demand of at least 5 MW of electricity, then the unit is working if it is at least in state 1. This implies that it is available if it is at least in state 1 and, therefore, its availability (unavailability) should be computed with respect to state 1 for this situation.

2.1 Importance Analysis of Multi-State Systems

System state probability and availability are important characteristics of a system. They give us a global view on the system. On the other hand, they carry no information about the system structure, i.e. they do not allow investigating influence of individual system components or their states on the system. For this purpose, other indi-

ces are used. These indices are known as Importance Measures (IMs), and some of the most commonly known are Structural Importance (SI) and Birnbaum's Importance (BI). The SI investigates only system topology while the BI takes into account also state probabilities of the system components. These two indices play a key role in importance analysis because a lot of other measures are defined based on them [4].

Importance analysis of MSSs based on SI and BI has been considered in several papers, e.g., [9], [14], [15], [16]. In those papers, several versions of these measures have been proposed depending on whether we want to:

- investigate influence of a given component state on a given system state/ availability level [9], [14],
- analyze the total influence of a given component state on the system (not only on a specific system state) [15],
- inspect the total importance of a given component [16].

The approaches presented in the aforementioned works have been combined into one complex framework in [10]. According to that work, the SI agrees with a relative number of situations in which a given component (state) is critical for degradation of (a given state/availability level of) the system, while the BI corresponds to the probability that such situation occurs. (The criticality means that degradation of a given component results in system degradation.) These definitions indicate that identification of situations in which a given component (state) is critical for degradation of (a given state/availability level of) the system represents the main issue in the computation of these measures. In the considered paper, this task has been solved using a special tool of MVL that is known as logical differential calculus [8].

Logical differential calculus has been developed for analysis of dynamic properties of MVL functions. Logic derivative is a key term of this tool. There exist several types of logic derivatives but, for the purpose of this paper, Direct Partial Logic Derivatives (DPLDs) are the most important.

A DPLD reveals circumstances under which a considered change of a MVL variable results in the studied change of the analyzed MVL function. Since the formal definition of the structure function of a MSS agrees with the definition of a MVL function, this derivative can also be used in the analysis of MSSs. In this case, it allows us to detect situations in which a given change of a given component state results in the studied change of the system state. More formally, a DPLD with respect to variable x_i is defined as follows [8], [9]:

$$\frac{\partial \phi(j \rightarrow h)}{\partial x_i(s \rightarrow r)} = \begin{cases} 1, & \text{if } \phi(s_i, \mathbf{x}) = j \text{ and } \phi(r_i, \mathbf{x}) = h \\ 0, & \text{otherwise} \end{cases}, \quad (5)$$

for $s, r, j, h \in \{0, 1, \dots, m-1\}, s \neq r, j \neq h$,

where $(a_i, \mathbf{x}) = (x_1, x_2, \dots, x_{i-1}, a, x_{i+1}, \dots, x_n)$ for $a \in \{s, r\}$.

Depending on the relations between s and r and j and h in (5), four kinds of DPLDs with different physical meaning can be used in reliability analysis of MSSs:

- if $s > r$ and $j > h$, then the DPLD identifies situations in which degradation of component i from state s to r results in degradation of system from state j to h ,

- if $s < r$ and $j < h$, then the DPLD detects circumstances under which improvement of component i from state s to r causes improvement of system state from value j to h ,
- if $s > r$ and $j < h$, then the DPLD finds circumstances under which degradation of component i from state s to r results in improvement of system state from value j to h ,
- if $s < r$ and $j > h$, then the DPLD reveals situations in which degradation of the system from state j to h is caused by improvement of component i from state s to r .

Clearly, situations identified by the last two kinds of DPLDs cannot occur in case of a coherent system and, therefore, only the first two DPLDs are meaningful in the analysis of coherent MSSs. Furthermore, in what follows, we will primarily deal with investigation of consequences of component degradation on system activity. This implies that only DPLDs in which $s > r$ and $j > h$ will be taken into account.

Based on the meaning of DPLDs, it is clear that they allow us to find state vectors of the form of (s_i, \mathbf{x}) at which deterioration of state s of component i to state r results in degradation of system state j to h . These state vectors are known as critical state vectors and, clearly, they describe circumstances under which a given component state is critical for a given degradation of system state j .

One of the assumptions that are often used in importance analysis of MSSs is that the system components degrade gradually state by state. This assumption is not unrealistic because even if a component deteriorates from state $m-1$ to state 0, we can assume that it stays in every state from set $\{1, 2, \dots, m-2\}$ for very short time [9]. This implies that only DPLDs of the form of $\partial\phi(j \rightarrow h)/\partial x_i(s \rightarrow s-1)$ have to be taken into account if we want to investigate importance of individual system components.

DPLDs give us a detailed view on the dependency between component degradation and system degradation. However, they are not very appropriate for importance analysis of a general MSS, i.e. a MSS in which a minor degradation (degradation by one state) of any system component can result in degradation of the system by more than one state. This inadequacy results from the fact that a lot of DPLDs have to be computed in such situations, e.g., if we want to investigate consequences of a minor degradation of state s of component i , then we have to compute DPLDs of the form of $\partial\phi(j \rightarrow h)/\partial x_i(s \rightarrow s-1)$ for all $h < j$, i.e. $j-1$ DPLDs. The similar fact can also be observed if we want to use DPLDs to investigate the coincidence between component degradation and decrease in system availability level. To avoid this problem, new types of logic derivatives have been introduced in [10]. These derivatives were named as Integrated Direct Partial Logic Derivatives (IDPLDs) because they combine several types of DPLDs together. Depending on the combined DPLDs, three types of IDPLDs have been defined. In this paper, only IDPLDs of type I and III are used.

An IDPLD of type I is defined as follows:

$$\frac{\partial\phi(j \downarrow)}{\partial x_i(s \rightarrow r)} = \bigcup_{h=0}^{j-1} \frac{\partial\phi(j \rightarrow h)}{\partial x_i(s \rightarrow r)} = \begin{cases} 1, & \text{if } \phi(s_i, \mathbf{x}) = j \text{ and } \phi(r_i, \mathbf{x}) < j, \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

for $s, r \in \{0, 1, \dots, m-1\}, s \neq r, j \in \{1, 2, \dots, m-1\}$,

and it allows us to find situations in which a given degradation of state s of system component i results in a deterioration of system state j . Quantification of these situations allows us to estimate influence of the considered component degradation on system state j .

An IDPLD of type III has the following form:

$$\frac{\partial \phi(h_{\geq j} \rightarrow h_{< j})}{\partial x_i(s \rightarrow r)} = \bigcup_{h_u=j}^{m-1} \bigcup_{h_d=0}^{j-1} \frac{\partial \phi(h_u \rightarrow h_d)}{\partial x_i(s \rightarrow r)} = \begin{cases} 1, & \text{if } \phi(s_i, \mathbf{x}) \geq j \text{ and } \phi(r_i, \mathbf{x}) < j \\ 0, & \text{otherwise} \end{cases}, \quad (7)$$

for $s, r \in \{0, 1, \dots, m-1\}, s \neq r, j \in \{1, 2, \dots, m-1\}$,

where notation $h_{\geq j}$ ($h_{< j}$) means that all system states that are greater than or equal to (less than) j are taken into account. Please note, this definition implies that IDPLDs of type III can be used to find state vectors at which a degradation of a given component state causes degradation of a given level of system availability and, therefore, they can be used to quantify consequences of a given deterioration of state s of component i on level j of system availability.

It has been mentioned in the previous paragraphs that quantification of situations in which an IDPLD of type I or III takes nonzero value allows us to estimate influence of degradation a given component state on system state/availability level. This quantification can be done in two ways [10]. Firstly, we can compute truth density (a relative count of situations in which a function with a Boolean-valued output takes non-zero value) of the considered IDPLD. Result of this computation agrees with the relative number of situations in which a considered degradation of a given component state results in degradation of a given system state/availability level. If we assume that the system components degrade gradually state by state, then this number corresponds to SI of a given component state for a given system state/availability level [10]. This measure does not take the components states probabilities into account and, therefore, it investigates only topological importance of a given component state.

Another possibility is to calculate the probability that the considered IDPLD is nonzero. This agrees with the probability that the studied degradation of a given component state causes decrease in a given state/availability level of the system. If we assume that only minor degradations of the system components can occur, then this number agrees with BI of a given component state for a given system state/availability level [10]. Unlike the SI, the BI provides more information because it considers not only system topology but also state probabilities of the components.

The previously mentioned versions of SI and BI deals with importance of a given component state for a given state/availability level of the system. It has been shown in [10] that these measures can also be used to investigate:

- the total importance of a given component state,
- the total importance of a component for a given system state/availability level,
- the total importance of a given component.

The SI measures that can be used for these purposes are presented in Table 1. The similar table can be shown for the BI measures, but the only difference will be in replacement of the truth density notation with the probability that the IDPLD takes nonzero value. Please note that complex importance analysis of a MSS can be per-

formed by computation of all types of SI or BI measures. Based on the formulae presented in Table 1, the SI (BI) measures investigating all possible dependencies between component degradation and system deterioration can be expressed in the form of Table 2 (IMs concerning with system state) or Table 3 (IMs focusing on system availability level).

Table 1. Summary of structural importance measures investigating topological properties of the system based on component degradation

Structural importance	Definition	Interpretation
The SI of a given component state and for a given system state	$SI_{i,s}^{\downarrow} = TD\left(\frac{\partial\phi(j\downarrow)}{\partial x_i(s \rightarrow s-1)}\right)$	A relative number of situations in which degradation of state s of component i results in degradation of state j of the system.
The SI of a given component state and for a given level of system availability	$SI_{i,s}^{\geq j} = TD\left(\frac{\partial\phi(h_{\geq j} \rightarrow h_{< j})}{\partial x_i(s \rightarrow s-1)}\right)$	A relative number of situations in which degradation of state s of component i results in degradation of level j of system availability.
The SI of a given component state	$SI_{i,s}^{\downarrow} = \sum_{j=1}^{m-1} SI_{i,s}^{j\downarrow}$	A relative number of situations in which state s of component i results in system degradation.
The SI of a given component for a given system state	$SI_i^{j\downarrow} = \frac{1}{m-1} \sum_{s=1}^{m-1} SI_{i,s}^{j\downarrow}$	A relative number of situations in which degradation of component i causes degradation of state j of the system.
The SI of a given component for a given system availability level	$SI_i^{\geq j} = \frac{1}{m-1} \sum_{s=1}^{m-1} SI_{i,s}^{\geq j}$	A relative number of situations in which degradation of component i causes degradation of level j of system availability.
The total SI of a given component	$SI_i^{\downarrow} = \frac{1}{m-1} \sum_{s=1}^{m-1} SI_{i,s}^{\downarrow}$	A relative number of situations in which degradation of component i results in system degradation.

*note: TD(.) – truth density of the argument interpreted as a function with a Boolean-valued output

3 Importance Analysis of k -out-of- n Multi-State Systems

Let us consider a k -out-of- n MSS. According to Table 1 – Table 3, the most important thing in computation of the IMs considered above is efficient identification of non-zero elements of IDPLDs. For example, in case of computing $SI_{i,s}^{j\downarrow}$, this agrees with finding all state vectors $(\cdot, \mathbf{x}) = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ for which IDPLD $\partial\phi(j\downarrow)/\partial x_i(s \rightarrow s-1)$ takes nonzero value. Now, let us find IDPLDs of which form can be nonzero. Firstly, let us assume that $j > s$. Such derivative cannot be nonzero because the k -out-of- n MSS can be in state j if and only if at least k components are in state j or greater. This implies that degradation of a component that is in a state less than j cannot result in degradation of system state j because system state is not

Table 2. Structural importance measures investigating topological properties of the system with respect to system state

		Component state				Average
		1	2	...	$m-1$	
System state	1	$SI_{i,1}^{1\downarrow}$	$SI_{i,2}^{1\downarrow}$...	$SI_{i,m-1}^{1\downarrow}$	$SI_i^{1\downarrow}$
	2	$SI_{i,1}^{2\downarrow}$	$SI_{i,2}^{2\downarrow}$...	$SI_{i,m-1}^{2\downarrow}$	$SI_i^{2\downarrow}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	$m-1$	$SI_{i,1}^{(m-1)\downarrow}$	$SI_{i,2}^{(m-1)\downarrow}$...	$SI_{i,m-1}^{(m-1)\downarrow}$	$SI_i^{(m-1)\downarrow}$
Sum		$SI_{i,1}^\downarrow$	$SI_{i,2}^\downarrow$...	$SI_{i,m-1}^\downarrow$	SI_i^\downarrow

Table 3. Structural importance measures investigating topological properties of the system with respect to system availability level

		Component state				Average
		1	2	...	$m-1$	
System availability level	1	$SI_{i,1}^{\geq 1}$	$SI_{i,2}^{\geq 1}$...	$SI_{i,m-1}^{\geq 1}$	$SI_i^{\geq 1}$
	2	$SI_{i,1}^{\geq 2}$	$SI_{i,2}^{\geq 2}$...	$SI_{i,m-1}^{\geq 2}$	$SI_i^{\geq 2}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	$m-1$	$SI_{i,1}^{\geq (m-1)}$	$SI_{i,2}^{\geq (m-1)}$...	$SI_{i,m-1}^{\geq (m-1)}$	$SI_i^{\geq (m-1)}$

determined by this component. Secondly, let us assume that $j < s$. This derivative cannot also take nonzero values because the system can be in state j if and only if not more than $k-1$ components are in a state greater than j . It follows that a minor degradation of a component that is in a state greater than system state j does not result in violation of this condition and, therefore, there exist no circumstances under which degradation of this component from state s to state $s-1$ can result in degradation of the system if the system is in state j such that $j < s$. Finally, let us consider an IDPLD of the form of $\partial\phi(j\downarrow)/\partial x_i(j \rightarrow j-1)$. This derivative identifies situations in which a minor degradation of state j of component i results in degradation of system state j . Such situations can occur if and only if component i is in state j and exactly $k-1$ from the remaining components are in state j or greater than j . These situations correspond to state vectors of the form of $(j_i, r_{u_1}^{\geq j, (1)}, r_{u_2}^{\geq j, (2)}, \dots, r_{u_{k-1}}^{\geq j, (k-1)}, r_{v_1}^{< j, (1)}, r_{v_2}^{< j, (2)}, \dots, r_{v_{n-k}}^{< j, (n-k)})$ where u_1, u_2, \dots, u_{k-1} are components that are in states greater than or equal to j ; v_1, v_2, \dots, v_{n-k} are components that are in states less than j ; $r_{u_t}^{\geq j, (t)}$, for $t = 1, 2, \dots, k-1$, de-

notes state of component u_t (this state is greater than or equal to j); and $r_{v_t}^{<j,(t)}$, for $t = 1, 2, \dots, n - k$, means that component v_t is in a state less than j . It can be simply shown that $\binom{n-1}{k-1}(m-j)^{k-1}j^{n-k}$ such state vectors exist. It follows that integrated derivatives $\partial\phi(j \downarrow)/\partial x_i(j \rightarrow j-1)$ has $\binom{n-1}{k-1}(m-j)^{k-1}j^{n-k}$ nonzero elements. Since every IDPLD and DPLD has m^{n-1} elements [9], [10], the truth density of integrated derivative $\partial\phi(j \downarrow)/\partial x_i(j \rightarrow j-1)$ can be computed in the following way:

$$\text{TD}\left(\frac{\partial\phi(j \downarrow)}{\partial x_i(j \rightarrow j-1)}\right) = \binom{n-1}{k-1} \frac{(m-j)^{k-1}j^{n-k}}{m^{n-1}}. \quad (8)$$

Based on the results obtained in the previous paragraph and information presented in Table 1, $\text{SI}_{i,s}^{j \downarrow}$ can be computed for a k -out-of- n MSS using the following formula:

$$\text{SI}_{i,s}^{j \downarrow} = \text{TD}\left(\frac{\partial\phi(j \downarrow)}{\partial x_i(s \rightarrow s-1)}\right) = \begin{cases} \binom{n-1}{k-1} \frac{(m-j)^{k-1}j^{n-k}}{m^{n-1}}, & \text{if } s = j \\ 0, & \text{otherwise} \end{cases}. \quad (9)$$

Above, we have shown that only IDPLDs of the form of $\partial\phi(j \downarrow)/\partial x_i(j \rightarrow j-1)$ are nonzero in case of a k -out-of- n MSS. Since this IDPLD is nonzero if and only if states of the system components are characterized by state vectors of the form of $(j_i, r_{u_1}^{\geq j,(1)}, r_{u_2}^{\geq j,(2)}, \dots, r_{u_{k-1}}^{\geq j,(k-1)}, r_{v_1}^{<j,(1)}, r_{v_2}^{<j,(2)}, \dots, r_{v_{n-k}}^{<j,(n-k)})$, then degradation of component i from state j to $j-1$ has to result in degradation of system state from j to $j-1$. This implies that the nonzero elements of this derivative agrees with the nonzero elements of DPLD $\partial\phi(j \rightarrow j-1)/\partial x_i(j \rightarrow j-1)$. It follows that only such DPLDs are nonzero in case of a k -out-of- n MSS. Using this fact and definitions (6) and (7) of IDPLDs, the next formula can be proved simply for k -out-of- n MSSs:

$$\frac{\partial\phi(j \downarrow)}{\partial x_i(s \rightarrow s-1)} = \frac{\partial\phi(h_{\geq j} \rightarrow h_{< j})}{\partial x_i(s \rightarrow s-1)} = \frac{\partial\phi(j \rightarrow j-1)}{\partial x_i(s \rightarrow s-1)} = \begin{cases} \frac{\partial\phi(j \rightarrow j-1)}{\partial x_i(j \rightarrow j-1)}, & \text{if } s = j \\ 0, & \text{otherwise} \end{cases}. \quad (10)$$

This formula implies that the k -out-of- n MSS represents a special type of MSSs in which a minor degradation of a system component can result only in a minor degradation of system state. Furthermore, it follows that the following relationships exist between SI measures presented in Table 1:

$$\begin{aligned}
 SI_{i,s}^{j\downarrow} &= SI_{i,s}^{\geq j} = \begin{cases} \binom{n-1}{k-1} \frac{(m-j)^{k-1} j^{n-k}}{m^{n-1}}, & \text{if } s = j, \\ 0, & \text{otherwise} \end{cases} \\
 SI_i^{j\downarrow} &= SI_i^{\geq j} = \frac{1}{m-1} SI_{i,j}^{j\downarrow} = \binom{n-1}{k-1} \frac{(m-j)^{k-1} j^{n-k}}{(m-1)m^{n-1}}, \\
 SI_{i,s}^{\downarrow} &= SI_{i,s}^{s\downarrow} = \binom{n-1}{k-1} \frac{(m-s)^{k-1} s^{n-k}}{m^{n-1}}.
 \end{aligned} \tag{11}$$

Finally, the total topological importance of component i for the activity of the k -out-of- n MSS can be computed using the following formula:

$$SI_i^{\downarrow} = \frac{1}{m-1} \sum_{j=1}^{m-1} SI_{i,j}^{j\downarrow} = \binom{n-1}{k-1} \frac{\sum_{j=1}^{m-1} (m-j)^{k-1} j^{n-k}}{(m-1)m^{n-1}}. \tag{12}$$

All these formulae imply that there is no sense to distinguish between the SI measures investigating topological properties of the k -out-of- n MSS with respect to system state (Table 2) and with respect to system availability level (Table 3):

Now, let us consider the BI measures. These measures can also be computed using IDPLDs I or III. However, as has been shown in (10), these IDPLDs computed with respect to system state/availability level j agree with $\partial\phi(j \rightarrow j-1)/\partial x_i (s \rightarrow s-1)$ in case of k -out-of- n MSSs and, therefore, the following relationships will hold between BI measures calculated with respect to system state and with respect to system availability level:

$$\begin{aligned}
 BI_{i,s}^{j\downarrow} &= BI_{i,s}^{\geq j} = \begin{cases} \Pr\left\{\frac{\partial\phi(j \rightarrow j-1)}{\partial x_i(j \rightarrow j-1)} = 1\right\}, & \text{if } s = j, \\ 0, & \text{otherwise} \end{cases} \\
 BI_i^{j\downarrow} &= BI_i^{\geq j} = \frac{1}{m-1} BI_{i,j}^{j\downarrow}, \\
 BI_{i,s}^{\downarrow} &= BI_{i,s}^{s\downarrow}, \\
 BI_i^{\downarrow} &= \frac{1}{m-1} \sum_{j=1}^{m-1} BI_{i,j}^{j\downarrow}.
 \end{aligned} \tag{13}$$

This implies that only problem in computation of the BI measures is calculation of the probability that DPLD $\partial\phi(j \rightarrow j-1)/\partial x_i(j \rightarrow j-1)$ takes nonzero value. According to the results presented in the previous paragraphs, this DPLD takes nonzero values for all state vectors of the structure function that have the form of $(j_i, r_{u_1}^{\geq j, (1)}, r_{u_2}^{\geq j, (2)}, \dots, r_{u_{k-1}}^{\geq j, (k-1)}, r_{v_1}^{< j, (1)}, r_{v_2}^{< j, (2)}, \dots, r_{v_{n-k}}^{< j, (n-k)})$. Since a DPLD computed with respect to variable x_i does not depend on this variable [8], [9], the nonzero elements of the DPLD agree with the state vectors that have the following form: $(\cdot, r_{u_1}^{\geq j, (1)}, r_{u_2}^{\geq j, (2)}, \dots, r_{u_{k-1}}^{\geq j, (k-1)}, r_{v_1}^{< j, (1)}, r_{v_2}^{< j, (2)}, \dots, r_{v_{n-k}}^{< j, (n-k)})$. Therefore, the probability that

the DPLD is nonzero can be computed simply as the probability that the system components are in states corresponding to state vectors of this form.

4 Case Study: Oil Supply System

For illustration of our approach, let us consider a modified oil supply system proposed in [12] and considered in [13]. This system is depicted in Fig. 1. There are 4 pipelines that deliver oil from the oil source to 3 oil stations. The system and every pipeline have 4 possible states. The system state is defined by the number of oil stations to which oil can be delivered through the pipelines (Table 4). The state of a pipeline identifies which oil stations can be supplied through the pipeline (Table 4). Next, let us assume that the oil source is perfectly functioning and an oil station is working if at least k pipelines are able to deliver oil to it. This description implies that only relevant components that determine system state are 4 pipelines. So, we obtain a k -out-of-4 MSS in which $m = 4$ and $k \in \{1, 2, 3, 4\}$.

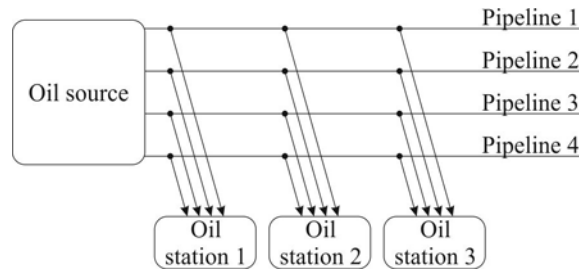


Fig. 1. Oil supply system considered in [12] and [13]

Table 4. Interpretation of system and component states for the oil supply system

State	System	Component (pipeline) i
0	No oil station is supplied.	Pipeline i is not able to deliver oil to any station.
1	Only oil station 1 is supplied.	Pipeline i is able to deliver oil only to station 1.
2	Only oil stations 1 and 2 are supplied.	Pipeline i is able to deliver oil only to oil stations 1 and 2.
3	All oil stations are supplied.	Pipeline i is able to deliver oil to all oil stations.

Firstly, let us investigate topological properties of this system. This can be done simply using SI measures shown in Table 1 and formulae (11) and (12). For pipeline 1 and individual values of k , these results are presented in Table 5. According to the results presented in the form of formulae (11) and (12), the SI measures computed in these tables investigate topological properties of the system not only with respect to system state but also with respect to system availability level. As we can see (the lower right corner of sub-tables), degradation of component 1 has the greatest influence on system activity if $k \in \{2, 3\}$ and the least if $k \in \{1, 4\}$. Next, in the bottom parts of the sub-tables, we can see that, in case of $k \in \{1, 2\}$, the most important state

of the component is state 3, while the least important is state 1. On the other hand, if $k \in \{3,4\}$, then the situation is completely different, i.e. the state with the greatest topological influence on system degradation is state 1, while state 3 has the least influence. Similarly, using the information presented in the right columns of the subtables, we can state if $k \in \{1,2\}$, then pipeline 1 has the greatest influence on system state/availability level 3 and the least on system state/availability level 1; while if $k \in \{3,4\}$, then pipeline 1 has the greatest influence on system state/availability level 1 and the least on system state/availability level 3. Clearly, the same results can be obtained for the remaining pipelines since formulae (11) and (12) imply that all components (or components states) have the same topological influence in case of k -out-of- n systems and fixed values of k and n .

Table 5. Structural importance measures investigating pipelines 1,2,3,4 for $k = 1, 2, 3, 4$

		$k = 1$				$k = 2$			
		Component state			Average	Component state			Average
		1	2	3		1	2	3	
System state	1	0.0156	0	0	0.0052	0.1406	0	0	0.0469
	2	0	0.1250	0	0.0417	0	0.3750	0	0.1250
	3	0	0	0.4219	0.1406	0	0	0.4219	0.1406
Sum		0.0156	0.1250	0.4219	0.1875	0.1406	0.3750	0.4219	0.3125
		$k = 3$				$k = 4$			
		Component state			Average	Component state			Average
		1	2	3		1	2	3	
System state	1	0.4219	0	0	0.1406	0.4219	0	0	0.1406
	2	0	0.3750	0	0.1250	0	0.1250	0	0.0417
	3	0	0	0.1406	0.0469	0	0	0.0156	0.0052
Sum		0.4219	0.3750	0.1406	0.3125	0.4219	0.1250	0.0156	0.1875

Now, let us calculate the BI measures for this system. These measures take into account not only system topology but also state probabilities of the pipelines. In this case, we use numbers presented in [13] (Table 6).

The BI measures for the oil supply system can be computed using formulae (13). According to these formulae, the most important part in computation of these measures is calculating the probability that $DPLD \partial\phi(j \rightarrow j-1)/\partial x_i(j \rightarrow j-1)$, for $j = 1,2,3$ and $i = 1,2,3,4$, takes nonzero value. This can be done simply by identifying the state vectors of the form of $(\cdot, r_{u_1}^{\geq j, (1)}, r_{u_2}^{\geq j, (2)}, \dots, r_{u_{k-1}}^{\geq j, (k-1)}, r_{v_1}^{< j, (1)}, r_{v_2}^{< j, (2)}, \dots, r_{v_{n-k}}^{< j, (n-k)})$ for specific values of j and k . For example, if we want to compute the probability that $DPLD \partial\phi(1 \rightarrow 0)/\partial x_1(1 \rightarrow 0)$ is nonzero for $k = 1$, then its nonzero elements agree

Table 6. State probabilities of the pipelines of the oil supply system

Component	Components state			
	0	1	2	3
1	0.0500	0.0950	0.0684	0.7866
2	0.0500	0.0950	0.0684	0.7866
3	0.0300	0.0776	0.0446	0.8478
4	0.0300	0.0776	0.0446	0.8478

with the state vectors of the form of $(., r^{<1}, r^{<1}, r^{<1})$. Only one state vector has this form, i.e. state vector $(., 0, 0, 0)$; therefore, the probability that the DPLD takes nonzero value is computed as follows:

$$\Pr\left\{\frac{\partial\phi(2 \rightarrow 1)}{\partial x_1(2 \rightarrow 1)} = 1\right\} = \Pr\{(. , \mathbf{x}) = (., 0, 0, 0)\} = p_{2,0} p_{3,0} p_{4,0} \quad (14)$$

$$= 0.05 * 0.03 * 0.03 = 0.000045.$$

If we want to compute the probability that DPLD $\partial\phi(2 \rightarrow 1)/\partial x_1(2 \rightarrow 1)$ takes nonzero value for $k = 1$, then we should calculate the probability that state vector $(., \mathbf{x})$ has the form of $(., r^{<2}, r^{<2}, r^{<2})$. This condition is met if $(., \mathbf{x}) \leq (., 1, 1, 1)$, where relation “ \leq ” between state vectors $(., \mathbf{x}) = (x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$ and $(., \mathbf{y}) = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$ means that $x_k \leq y_k$ for $k = 1, 2, \dots, i-1, i+1, \dots, n$. So, the DPLD is nonzero with the following probability:

$$\Pr\left\{\frac{\partial\phi(2 \rightarrow 1)}{\partial x_1(2 \rightarrow 1)} = 1\right\} = \Pr\{(. , \mathbf{x}) \leq (., 1, 1, 1)\} = (p_{2,0} + p_{2,1})(p_{3,0} + p_{3,1})(p_{4,0} + p_{4,1}) \quad (15)$$

$$\approx 0.001679.$$

Finally, the probability of DPLD $\partial\phi(3 \rightarrow 2)/\partial x_1(3 \rightarrow 2)$ being nonzero for $k = 1$ agrees with the probability that state vector $(., \mathbf{x})$ has the form of $(., r^{<3}, r^{<3}, r^{<3})$. Using the notation “ \leq ”, this can be calculated as follows:

$$\Pr\left\{\frac{\partial\phi(3 \rightarrow 2)}{\partial x_1(3 \rightarrow 2)} = 1\right\} = \Pr\{(. , \mathbf{x}) \leq (., 2, 2, 2)\} \quad (16)$$

$$= (p_{2,0} + p_{2,1} + p_{2,2})(p_{3,0} + p_{3,1} + p_{3,2})(p_{4,0} + p_{4,1} + p_{4,2})$$

$$\approx 0.004943.$$

Equations (14) – (16) identify the probability that DPLDs of the form of $\partial\phi(j \rightarrow j-1)/\partial x_1(j \rightarrow j-1)$ takes nonzero value. Since this probability agrees with values of $BI_{1,j}^{j\downarrow}$ and $BI_{1,j}^{j\geq}$, it allows us to investigate importance of degradation of state j of pipeline 1 on state/availability level j of the oil supply system. Since all other BI measures investigating importance of a degradation of a component on system

state/availability level are equal to 0 (formulae (13)), these values can be used to investigate importance of a given component state for the system, or total importance of pipeline 1 for a specific system state/availability level, or total importance of pipeline 1 for the entire system. All these numbers are presented in the upper left sub-table of Table 7. Clearly, the same results can be obtained if we investigate importance of component 2, since state probabilities for these components are same.

Table 7. Birnbaum's importance measures investigating pipelines 1 and 2 for $k = 1, 2, 3, 4$

		$k = 1$				$k = 2$			
		Component state			Average	Component state			Average
		1	2	3		1	2	3	
System state	1	0.00005	0	0	0.00002	0.0038	0	0	0.0013
	2	0	0.0017	0	0.0006	0	0.0377	0	0.0126
	3	0	0	0.0049	0.0016	0	0	0.0733	0.0244
Sum		0.00004	0.0017	0.0049	0.0022	0.0038	0.0377	0.0733	0.0382
		$k = 3$				$k = 4$			
		Component state			Average	Component state			Average
		1	2	3		1	2	3	
System state	1	0.1023	0	0	0.0341	0.8939	0	0	0.2980
	2	0	0.2797	0	0.0932	0	0.6809	0	0.2270
	3	0	0	0.3563	0.1188	0	0	0.5654	0.1885
Sum		0.1023	0.2797	0.3563	0.2461	0.8939	0.6809	0.5654	0.7134

Using the similar procedure as has been presented above, the BI measures of all system components can be obtained (upper left sub-tables of Table 7 and Table 8). Furthermore, if we repeat this procedure for other values of k , i.e. for $k = 2, 3, 4$, we can investigate importance of individual pipelines for all versions of the oil supply system (the remaining parts of Table 7 and Table 8). (Please note that we obtain the same results for components 1 and 2 and for components 3 and 4 since their state probabilities have the same values.) According to the data presented in Table 7 and Table 8, we can state that pipelines 1 and 2 have less influence on the activity of the oil supply system than pipelines 3 and 4 if $k = 1, 2, 3$ but greater if $k = 4$. Another interesting fact that can be noticed based on Table 7 and Table 8 is that all system components have greater influence on greater states of the system in case of $k \in \{1, 2, 3\}$, (e.g., if $k = 1$, then degradation of component 1 results in degradation of system state 3 with the probability 0.0016 while in degradation of system state 1 with the probability 0.00002. However, in case of $k = 4$, all pipelines are more important for lower states of the system (e.g., a degradation of pipeline 1 causes degradation of system state 1 with the probability 0.2980 and degradation of system state 3 with the probability 0.1885). The similar facts can be observed for the BI measures investigating

the total importance of individual states of the pipelines for the oil supply system (the bottom rows in sub-tables of Table 7 and Table 8). All these results imply that importance of individual system components in case of k -out-of- n systems largely depends on mutual values of k and n .

Table 8. Birnbaum's importance measures investigating pipelines 3 and 4 for $k = 1, 2, 3, 4$

		$k = 1$				$k = 2$			
		Component state			Average	Component state			Average
		1	2	3		1	2	3	
System state	1	0.00008	0	0	0.00003	0.0053	0	0	0.0018
	2	0	0.0023	0	0.0008	0	0.0454	0	0.0151
	3	0	0	0.0069	0.0023	0	0	0.0897	0.0299
Sum		0.00008	0.0023	0.0069	0.0031	0.0053	0.0454	0.0897	0.0468
		$k = 3$				$k = 4$			
		Component state			Average	Component state			Average
		1	2	3		1	2	3	
System state	1	0.1192	0	0	0.0397	0.8755	0	0	0.2918
	2	0	0.2999	0	0.1000	0	0.6524	0	0.2175
	3	0	0	0.3788	0.1263	0	0	0.5246	0.1786
Sum		0.1192	0.2999	0.3788	0.2660	0.8754	0.6524	0.5246	0.6841

5 Conclusion

In this paper, importance analysis of a k -out-of- n MSS was considered. We summarized some results from the qualitative and quantitative analysis of MSSs and proposed the method for calculation of all range of the SI and BI measures. Furthermore, we showed that a k -out-of- n MSS is a special type of MSSs in which a minor degradation of any system component can result only in a minor degradation of system state. Because of that, it is not important to distinguish between IMs focusing on system state and IMs dealing with system availability level. Next, using logical differential calculus, we found closed-form expressions for calculation of the SI measures for a k -out-of- n MSS. All these results were used in the analysis of the oil supply system considered in [12] and [13]. Based on our approach, we identified topological importance of individual system components for different values of k and identified which components of the oil supply system were the most important if the state probabilities of individual system components were known. Finally, we would like to mention that the results presented in this paper could also be applied in the analysis of other types of systems such as medical and temporal database systems studied in [17] and [18].

Furthermore, they could also be used in other research fields, such as data mining, where they can be used to find key attributes in a dataset [19].

Acknowledgment

This work was supported by grants VEGA 1/0498/14 and VEGA 1/0038/16.

References

1. Rausand, M., Høyland, A.: System Reliability Theory. John Wiley & Sons, Inc., Hoboken, NJ (2004)
2. Lisnianski, A., Frenkel, I., Ding, Y.: Multi-state System Reliability Analysis and Optimization for Engineers and Industrial Managers. Springer-Verlag London Ltd., London, UK (2010)
3. Natvig, B.: Multistate Systems Reliability Theory with Applications. John Wiley & Sons, Ltd, Chichester, UK (2011)
4. Kuo, W., Zhu, X.: Importance Measures in Reliability, Risk, and Optimization: Principles and Applications. Wiley, Chichester, UK (2012)
5. Zio, E.: Reliability engineering: Old problems and new challenges. Reliability Engineering & System Safety. 94, 125–141 (2009)
6. Zaitseva, E., Levashenko, V., Rusin, M.: Reliability analysis of healthcare system. In: 2011 Federated Conference on Computer Science and Information Systems, FedCSIS 2011. pp. 169–175. IEEE (2011)
7. Praks, P., Kopustinskas, V., Masera, M.: Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. Reliability Engineering & System Safety. 144, 254–264 (2015)
8. Yanushkevich, S.N., Miller, D.M., Shmerko, V.P., Stankovic, R.S.: Decision Diagram Techniques for Micro- and Nanoelectronic Design Handbook. CRC Press, Boca Raton, FL (2005)
9. Zaitseva, E., Levashenko, V.: Multiple-valued logic mathematical approaches for multi-state system reliability analysis. Journal of Applied Logic. 11, 350–362 (2013)
10. Kvassay, M., Zaitseva, E., Kostolny, J., Levashenko, V.: Importance analysis of multi-state systems based on integrated direct partial logic derivatives. In: 2015 International Conference on Information and Digital Technologies. pp. 183–195. IEEE (2015)
11. Boedigheimer, R.A., Kapur, K.C.: Customer-driven reliability models for multistate coherent systems. IEEE Transactions on Reliability. 43, 46–50 (1994)
12. Tian, Z., Zuo M.J., Yam R.C.M.: The multi-state k-out-of-n systems and their performance evaluation. IIE Trans. 41, 32–44 (2009)
13. Mo, Y., Xing, L., Amari, S. V., Bechta Dugan, J.: Efficient analysis of multi-state k-out-of-n systems. Reliability Engineering and System Safety. 133, 95–105 (2014)
14. Butler, D.A.: A complete importance ranking for components of binary coherent systems, with extensions to multi-state systems. Naval Research Logistics Quarterly. 26, 565–578 (1979)
15. Griffith, W.S.: Multistate reliability models. Journal of Applied Probability. 17, 735–744 (1980)
16. Wu, S.: Joint importance of multistate systems. Computers & Industrial Engineering. 49, 63–75 (2005)
17. Kvet, M., Vajsova, M., Matiaško, K.: Complex Data Management in MRI Results Processing. In: Bris, R., Majernik, J., Pancarz, K., and Zaitseva, E. (eds.) Applications of Computa-

- tional Intelligence in Biomedical Technology. pp. 119–141. Springer International Publishing (2016)
18. Kvet, M., Matiasco, K., Vajsova, M.: Sensor based transaction temporal database architecture. In: 2015 IEEE World Conference on Factory Communication Systems (WFCS). pp. 1–8. IEEE (2015)
 19. Zaitseva, E., Kvassay, M., Levashenko, V., Kostolny, J.: Introduction to knowledge discovery in medical databases and use of reliability analysis in data mining. In: 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015. pp. 311–320. IEEE (2015)

Forecasting Economic Indices of Agricultural Enterprises Based on Vector Polynomial Canonical Expansion of Random Sequences

Igor P. Atamanyuk¹, Yuriy P. Kondratenko^{2,3}, Natalia N. Sirenko¹

¹ Mykolaiv National Agrarian University, Commune of Paris str. 9,
54010 Mykolaiv, Ukraine

atamanyuk_igor@mail.ru, sirenko@mdau.mk.ua

² Petro Mohyla Black Sea State University, 68th Desantnykiv Str. 10,
54003 Mykolaiv, Ukraine

³ Cleveland State University, 2121 Euclid Av.,
44115, Cleveland, Ohio, USA

y.kondratenko@csuohio.edu, y_kondrat2002@yahoo.com

Abstract. Calculating method for forecasting economic indices of agricultural enterprises on the basis of vector polynomial exponential algorithm of extrapolation of the realizations of random sequences is worked out. The model of prognosis allows estimate the results of enterprise functioning (to estimate future gross profit, gross production) after its reorganization (change of land resources, manpower resources, fixed assets). Prognostic model does not impose any restrictions on the forecast random sequence (linearity, stationarity, Markov behavior, monotonicity, etc.) and thus allows fully take into consideration stochastic peculiarities of functioning of agricultural enterprises. The simulation results confirm high efficiency of introduced calculating method. The scheme reflecting the peculiarities of functioning of the forecast model are also introduced in the work. The method can be realized in the decision support systems for agricultural and non-agricultural enterprises with various sets of economic indices.

Keywords. calculation method, random sequence, canonical decomposition, forecasting economic indices

Key Terms. computation, mathematical model

1 Introduction

Many decision-making processes in different areas of economy (management of enterprises, transport logistics, finance forecasting, investment under uncertainty and so on) are based on the different mathematical models, experienced theoretical

methods and modern intelligent algorithms [1-6]. For guaranteeing efficient performance of an enterprise on the market, it is necessary to form the strategy and tactics of enterprise development correctly, to ground the plans and management decisions. To do this is possible only based on effective diagnostics and prognostication of current and future economical situation at the enterprise. Western specialists have the priority in the investigation of the possibilities of the management on the basis of the forecasting of enterprise economic state. Bever (the USA) started theoretical development and building of prognostic models, then it was continued in the works of Altman (the USA) [7-8], Alberichi (Italy), Misha (France) and others [9-10]. More contemporary trend in the building of the algorithms of economic indices forecasting is the usage of stochastic methods of extrapolation. The relevance of such approach is explained with the influence of great number of accidental factors on the results of enterprise functioning (weather conditions, accidental variations of demand and supply, inflation etc.), under the influence of which the change of economic state indices obtains accidental character. It is especially important to take into account stochastic peculiarities of economic indices during the solving of the problems of prognostication of the state of agricultural enterprises.

But the existing models of prognosis impose considerable limitations on the accidental sequence describing the change of economic indices [11-16] (Markovian property, stationarity, monotony, scalarity etc.). Thereupon the problem of the building of the forecast model under the most general assumptions about the stochastic properties of the accidental process of the change of the indices of enterprise economic state arises.

2 Aim and the Raising of Problem

The aim of this work is the development of the efficient and robust method for forecasting agricultural enterprise indices. The main requirement to the forecasting method is the absence of any essential limitations on the stochastic properties of the accidental process of economic indices change.

3 Theoretical Conception of the Proposed Forecasting Method

The most universal method (from the point of view of the requirements to the investigated accidental sequence) is a method that based on the mechanism of canonical expansions [17-18]. The main primary indices of the economic state of agricultural enterprises are the gross profit, gross output, land resources, labour resources, fixed assets that is why the object of the investigation is the vector accidental sequence with five dependant constituents (if necessary the number of figures and their qualitative composition may be changed). Preliminary investigations (the check of dependence of accidental values on the basis of statistical data about the work of agricultural enterprises in Nikolaev region) showed that the accidental sequences describing the change of the economic state of the enterprises which relate to the intensive [19] type of the development during the interval of eleven years that

corresponds to the processing of twelve annual indices for the great number of the enterprises of the mentioned type have the most stable and significant stochastic relations. For such vector accidental sequence the canonical expansion has the following look [20]:

$$X_h(i) = M[X_h(i)] + \sum_{\nu=1}^i \sum_{\lambda=1}^5 V_\nu^{(\lambda)} \varphi_{h\nu}^{(\lambda)}(i), \quad i = \overline{1,12}, \quad h = \overline{1,5}, \quad (1)$$

where $X_1(i)$, $i = \overline{1,12}$ - gross profit;
 $X_2(i)$, $i = \overline{1,12}$ - gross output;
 $X_3(i)$, $i = \overline{1,12}$ - land resources;
 $X_4(i)$, $i = \overline{1,12}$ - labour resources;
 $X_5(i)$, $i = \overline{1,12}$ - fixed assets.

The elements of canonical expansion are the accidental coefficients $V_\nu^{(\lambda)}$, $\nu = \overline{1,12}$, $\lambda = \overline{1,5}$ and nonrandom coordinate functions $\varphi_{h\nu}^{(\lambda)}(i)$, $\nu = \overline{1,12}$, $\lambda = \overline{1,5}$:

$$V_\nu^{(\lambda)} = X_\lambda(\nu) - M[X_\lambda(\nu)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^H V_\mu^{(j)} \varphi_{\lambda\mu}^{(j)}(\nu) - \sum_{j=1}^{\lambda-1} V_\nu^{(j)} \varphi_{\lambda\nu}^{(j)}(\nu), \quad \nu = \overline{1,12}; \quad (2)$$

$$D_\lambda(\nu) = M\left[\left\{V_\nu^{(\lambda)}\right\}^2\right] = M\left[\left\{X_\lambda(\nu)\right\}^2\right] - M^2[X_\lambda(\nu)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^H D_j(\mu) \left\{\varphi_{\lambda\mu}^{(j)}(\nu)\right\}^2 - \sum_{j=1}^{\lambda-1} D_j(\nu) \left\{\varphi_{\lambda\nu}^{(j)}(\nu)\right\}^2, \quad \nu = \overline{1,12}; \quad (3)$$

$$\begin{aligned} \varphi_{h\nu}^{(\lambda)}(i) &= \frac{M[V_\nu^{(\lambda)}(X_h(i) - M[X_h(i)])]}{M\left[\left\{V_\nu^{(\lambda)}\right\}^2\right]} = \frac{1}{D_\lambda(\nu)} (M[X_\lambda(\nu)X_h(i)] - \\ &- M[X_\lambda(\nu)]M[X_h(i)] - \sum_{\mu=1}^{\nu-1} \sum_{j=1}^H D_j(\mu) \varphi_{\lambda\mu}^{(j)}(\nu) \varphi_{h\mu}^{(j)}(i) - \\ &- \sum_{j=1}^{\lambda-1} D_j(\nu) \varphi_{\lambda\nu}^{(j)}(\nu) \varphi_{h\nu}^{(j)}(i)), \quad \lambda = \overline{1,5}, \nu = \overline{1,i}. \end{aligned} \quad (4)$$

Coordinate functions $\varphi_{h\nu}^{(\lambda)}(i)$, $h, \lambda = \overline{1,5}$, $\nu, i = \overline{1,12}$ have the following properties:

$$\varphi_{h\nu}^{(\lambda)}(i) = \begin{cases} 1, & h = \lambda \quad \& \quad \nu = i; \\ 0, & i < \nu \text{ or } h < \lambda \quad \& \quad \nu = i. \end{cases} \quad (5)$$

The algorithm of extrapolation on the basis of canonical expansion has the look [20]:

$$m_h^{(\mu,l)}(i) = \begin{cases} M[X_h(i)], \mu = 0, \\ m_h^{(\mu,l-1)}(i) + [x_l(\mu) - m_l^{(\mu,l-1)}(\mu)] \phi_{h\mu}^{(l)}(i), l \neq 1, \\ m_h^{(\mu,5)}(i) + [x_1(\mu) - m_1^{(\mu-1,5)}(\mu)] \phi_{h\mu}^{(1)}(i), l = 1, \end{cases} \quad (6)$$

where $m_h^{(\mu,l)}(i) = M[X_h(i) / x_\lambda(\nu), \lambda = \overline{1,5}, \nu = \overline{1, \mu-1}; x_j(\mu), j = \overline{1,l}], h = \overline{1,5},$

$i = \overline{k,12}$ - is the linear optimal quantity by the criterion of the minimum of the average square of the error of the prognosis is the estimation of the future values of the investigated sequence under the condition that the values are known $x_\lambda(\nu), \lambda = \overline{1,5}, \nu = \overline{1, \mu-1}; x_j(\mu), j = \overline{1,l}$.

Essential deficiency of the forecast model (6) is the assumption of existence of only linear stochastic relations in the sequence $X_h(i), h = \overline{1,5}, i = \overline{1,12}$, describing the process of change of economic indices of agricultural enterprises. The analysis of statistical data about the work of agricultural enterprises of Nikolaev region showed that the stochastic relations till the fourth order

$M[X_h^o(i) X_m^o(i)] \neq 0, \nu, i = \overline{1,12}, l + s \leq 4, h, m = \overline{1,5}$ are essential for such a

sequence. Non-linear canonical model of the investigated sequence with taking account of non-linear relations takes the form [21]:

$$X_h(i) = M[X_h(i)] + \sum_{\nu=1}^{i-1} \sum_{l=1}^5 \sum_{\lambda=1}^3 W_{\nu l}^{(\lambda)} \beta_{l\lambda}^{(h,1)}(\nu, i) + \sum_{l=1}^{h-1} \sum_{\lambda=1}^3 W_{il}^{(\lambda)} \beta_{l\lambda}^{(h,1)}(i, i) + W_{ih}^{(1)}, i = \overline{1,12}. \quad (7)$$

Random coefficients $W_{\nu l}^{(\lambda)}, \nu = \overline{1,12}, l = \overline{1,5}, \lambda = \overline{1,3}$ and nonrandom coordinate functions $\beta_{l\lambda}^{(h,s)}(\nu, i), \nu, i = \overline{1,12}, l, h = \overline{1,5}, \lambda, s = \overline{1,3}$ are determined with the help of expressions:

$$W_{\nu l}^{(\lambda)} = X_l^\lambda(\nu) - M[X_l^\lambda(\nu)] - \sum_{\mu=1}^{\nu-1} \sum_{m=1}^5 \sum_{j=1}^3 W_{\mu m}^{(j)} \beta_{mj}^{(l,\lambda)}(\mu, \nu) - \quad (8)$$

$$- \sum_{m=1}^{l-1} \sum_{j=1}^3 W_{vm}^{(j)} \beta_{mj}^{(l,\lambda)}(\nu, \nu) - \sum_{j=1}^{\lambda-1} W_{\nu l}^{(j)} \beta_{lj}^{(l,\lambda)}(\nu, \nu), \nu = \overline{1,12};$$

$$D_{l,\lambda}(\nu) = M\left[\left\{W_{\nu l}^{(\lambda)}\right\}^2\right] = M[X_l^{2\lambda}(\nu)] - M^2[X_l^\lambda(\nu)] - \quad (9)$$

$$- \sum_{\mu=1}^{\nu-1} \sum_{m=1}^5 \sum_{j=1}^3 D_{mj}(\mu) \left\{\beta_{mj}^{(l,\lambda)}(\mu, \nu)\right\}^2 - \sum_{m=1}^{l-1} \sum_{j=1}^3 D_{mj}(\nu) \left\{\beta_{mj}^{(l,\lambda)}(\nu, \nu)\right\}^2 -$$

$$- \sum_{j=1}^{\lambda-1} D_{lj}(\nu) \left\{\beta_{lj}^{(l,\lambda)}(\nu, \nu)\right\}^2, \nu = \overline{1,12};$$

$$\begin{aligned}
 \beta_{l\lambda}^{(h,s)}(\nu, i) &= \frac{M \left[W_{\nu l}^{(\lambda)} (X_h^s(i) - M[X_h^s(i)]) \right]}{M \left[\left\{ W_{\nu l}^{(\lambda)} \right\}^2 \right]} = \\
 &= \frac{1}{D_{l\lambda}(\nu)} (M[X_l^{\lambda}(\nu) X_h^s(i)] - M[X_l^{\lambda}(\nu)] M[X_h^s(i)] - \\
 &\quad - \sum_{\mu=1}^{\nu-1} \sum_{m=1}^5 \sum_{j=1}^3 D_{mj}(\mu) \beta_{mj}^{(l,\lambda)}(\mu, \nu) \beta_{mj}^{(h,s)}(\mu, i) - \\
 &\quad - \sum_{m=1}^{l-1} \sum_{j=1}^3 D_{mj}(\nu) \beta_{mj}^{(l,\lambda)}(\nu, \nu) \beta_{mj}^{(h,s)}(\nu, i) - \\
 &\quad - \sum_{j=1}^{\lambda-1} D_{lj}(\nu) \beta_{lj}^{(l,\lambda)}(\nu, \nu) \beta_{lj}^{(h,s)}(\nu, i), \quad \lambda = \overline{1, h}, \quad i = \overline{1, 12}, \quad \nu = \overline{1, i}.
 \end{aligned} \tag{10}$$

Vector algorithm of extrapolation [22-24] for the considered quantity of the components and order of stochastic relations on the basis of canonical expansion (7) takes the form:

$$m_{j,h}^{(\mu,l)}(s, i) = \begin{cases} M[X_h(i)], \quad \mu = 0; \\ m_{j,h}^{(\mu,l-1)}(s, i) + (x_j^l(\mu) - m_{j,j}^{(\mu,l-1)}(l, \mu)) \beta_{j,l}^{(h,s)}(\mu, i), \quad l > 1, j < 5; \\ m_{j,h}^{(\mu,3)}(s, i) + (x_{j+1}(\mu) - m_{j,j+1}^{(\mu,1)}(3, \mu)) \beta_{j+1,l}^{(h,s)}(\mu, i), \quad l = 1, j < 5; \\ m_{5,h}^{(\mu,3)}(s, i) + (x_1(\mu + 1) - m_{5,1}^{(\mu,3)}(3, \mu + 1)) \beta_{1,l}^{(h,s)}(\mu + 1, i), \quad l = 1, j = 5. \end{cases} \tag{11}$$

$m_{j,h}^{(\mu,l)}(1, i) = M[X_h(i) / x_{\lambda}^n(\nu)], \quad \lambda = \overline{1, 5}, n = \overline{1, 3}, \nu = \overline{1, \mu - 1}; x_{\lambda}^n(\mu), \lambda = \overline{1, j}, n = \overline{1, l}$ is optimal by the criterion of minimum of mean-square error of prognosis estimation of future values of economic index with ordinal number h provided that for the prognosis values $x_{\lambda}^n(\nu), \lambda = \overline{1, 5}, n = \overline{1, 3}, \nu = \overline{1, \mu - 1}; x_{\lambda}^n(\mu), \lambda = \overline{1, j}, n = \overline{1, l}$ are used.

Altogether 165 values $x_h^{\lambda}(i), h = \overline{1, 5}, i = \overline{1, 11}, \lambda = \overline{1, 3}$ and 5220 not equal to zero weight coefficients $\beta_{l\lambda}^{(h,s)}(\nu, i), \nu, i = \overline{1, 12}, l, h = \overline{1, 5}, \lambda, s = \overline{1, 3}$ are used in the algorithm of prognosis (11).

The expression for mean-square error of extrapolation with the help of algorithm (11) by known values $x_j^n(\mu), \mu = \overline{1, k}; j = \overline{1, 5}; n = \overline{1, 3}$ is in the form:

$$E_h^{(k,3)}(i) = M[X_h^2(i)] - M^2[X_h(i)] - \sum_{\mu=1}^k \sum_{j=1}^5 \sum_{n=1}^3 D_{jn}(\mu) \left\{ \beta_{jn}^{(h,1)}(\mu, i) \right\}^2, \quad i = \overline{k+1, 12}. \tag{12}$$

This expression is equal to dispersion of a posteriori casual sequence $\{X_h(i) / x_{\lambda}^n(\nu), \lambda = \overline{1, 5}, n = \overline{1, 3}, \nu = \overline{1, \mu - 1}; x_{\lambda}^n(\mu), \lambda = \overline{1, j}, n = \overline{1, l}\}$.

In Fig. 1 the scheme reflecting the peculiarities of functioning of the forecast model (11) is represented.

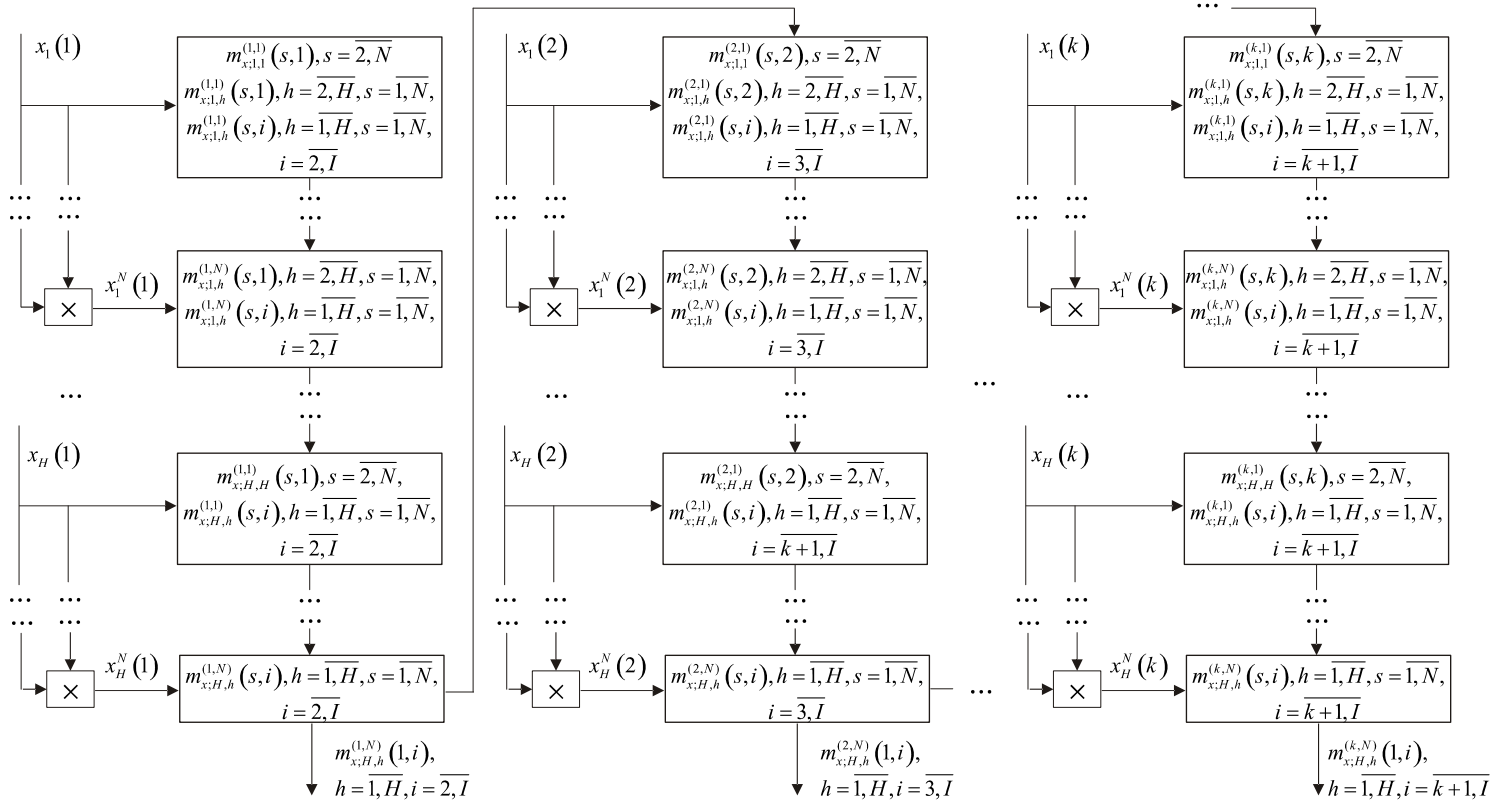


Fig. 1. Scheme of functioning of the forecast model (11) ($N=3$, $H=5$)

Method of prognostication of future values of economic indices on the basis of the forecast model (11) presupposes realization of the following stages:

Stage 1. Gathering of statistical data about the results of enterprises functioning;

Stage 2. Estimation of moment functions $M[X^s(i)]$, $M[X_j^l(\nu)X_h^s(i)]$ on the basis of cumulated realizations of random sequence describing the process of change of economic indices;

Stage 3. Calculating of the parameters of the algorithm of extrapolation (11);

Stage 4. Estimation of future values of economic indices on the basis of the forecast model (11);

Stage 5. Estimation of the quality of the solving of the forecast problem for investigated sequence with the help of the expression (12).

4 Results of Numerical Experiment

Method is approbated on the basis of statistical data of functioning of agricultural enterprises in Nikolaev region during the period 2004-2015 (74 enterprises with gross profit 200-900 thousands grivnas). Moment functions $M[X_h^s(i)]$, $M[X_j^l(\nu)X_h^s(i)]$ were estimated by known formulae of mathematical statistics for sections 2004, 2005, ..., 2014. Data about the work of the enterprises for 2015 were supposed to be unknown and the estimation of moment functions $M[X_h^s(12)]$, $M[X_j^l(\nu)X_h^s(12)]$ for the last section (corresponding to 2015) was carried out on the basis of determinate models with the use of four previous years (2011-2014) in tabular processor Microsoft Excel (instrument "Search for solutions"). For example, in Table 1 the values of autocorrelated function $M[\overset{\circ}{X}_1(\nu)\overset{\circ}{X}_1(i)]$, $\nu = \overline{1,12}$, $i = \overline{1,12}$ for the component $X_1(i)$, $i = \overline{1,12}$ (gross profit) are represented.

For 2015 values $M[\overset{\circ}{X}_h(\nu)\overset{\circ}{X}_h(12)]$, $\nu = \overline{1,11}$ are obtained on the basis of determinate model:

$$M[\overset{\circ}{X}_1(\nu)\overset{\circ}{X}_1(12)] = 0,718M[\overset{\circ}{X}_1(\nu)\overset{\circ}{X}_1(11)] - 0,053M[\overset{\circ}{X}_1(\nu)\overset{\circ}{X}_1(10)] + \quad (13)$$

$$+ 0,2128M[\overset{\circ}{X}_1(\nu)\overset{\circ}{X}_1(9)] - 0,105M[\overset{\circ}{X}_1(\nu)\overset{\circ}{X}_1(8)], \quad \nu = \overline{1,11},$$

Coordinate function $\beta_{11}^{(1,1)}(\nu, i)$ $\nu, i = \overline{1,12}$ (Table 2) corresponds to correlated function $M[\overset{\circ}{X}_1(\nu)\overset{\circ}{X}_1(i)]$, $\nu = \overline{1,12}$, $i = \overline{1,12}$.

Table 1. Autocorrelated function of the component $X_1(i)$, $i=\overline{1,12}$

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2004	1	0,99	0,70	0,42	0,79	0,74	0,49	0,72	0,63	0,46	0,55	0,43
2005	0,99	1	0,72	0,42	0,74	0,74	0,52	0,70	0,64	0,48	0,59	0,46
2006	0,70	0,72	1	0,57	0,67	0,58	0,701	0,69	0,70	0,66	0,78	0,60
2007	0,42	0,42	0,57	1	0,38	0,36	0,45	0,21	0,41	0,36	0,19	0,18
2008	0,79	0,74	0,67	0,38	1	0,81	0,55	0,91	0,80	0,72	0,53	0,41
2009	0,74	0,74	0,58	0,36	0,81	1	0,72	0,73	0,92	0,81	0,51	0,44
2010	0,49	0,52	0,70	0,45	0,55	0,72	1	0,51	0,74	0,73	0,49	0,41
2011	0,72	0,70	0,69	0,21	0,91	0,73	0,51	1	0,77	0,80	0,74	0,55
2012	0,63	0,64	0,70	0,41	0,80	0,92	0,74	0,77	1	0,91	0,60	0,59
2013	0,46	0,48	0,66	0,36	0,72	0,81	0,73	0,80	0,91	1	0,71	0,46
2014	0,55	0,59	0,78	0,19	0,53	0,51	0,49	0,74	0,60	0,71	1	0,71
2015	0,43	0,46	0,60	0,18	0,41	0,44	0,41	0,55	0,59	0,46	0,71	1

Table 2. Coordinate function $\beta_{11}^{(1,1)}(\nu, i)$ $\nu, i = \overline{1,12}$

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2004	1	0,89	0,54	0,55	0,62	0,43	0,45	0,89	0,858	0,90	2,36	2,65
2005	0	1	2,25	-1,46	-2,40	0,27	5,47	-2,71	3,55	1,83	2,85	4,70
2006	0	0	1	5,09	1,17	-1,53	-0,03	-2,77	-0,23	-5,52	2,34	5,07
2007	0	0	0	1	0,17	0,26	0,94	0,18	0,77	1,05	-0,57	-1,17
2008	0	0	0	0	1	0,48	1,27	1,06	1,05	2,01	-2,37	0,69
2009	0	0	0	0	0	1	-1,81	0,74	3,53	0,37	9,31	2,86
2010	0	0	0	0	0	0	1	-0,68	1,44	3,18	-6,74	-3,39
2011	0	0	0	0	0	0	0	1	1,29	2,21	-3,30	0,93
2012	0	0	0	0	0	0	0	0	1	3,88	0,19	-8,44
2013	0	0	0	0	0	0	0	0	0	1	1,99	-4,96
2014	0	0	0	0	0	0	0	0	0	0	1	0,50
2015	0	0	0	0	0	0	0	0	0	0	0	1

In Table 3 weight coefficients $\beta_{13}^{(1,1)}(\nu, i)$ $\nu, i = \overline{1,11}$, $i = \overline{2,12}$ determining the influence of values $x_1^3(i)$, $i = \overline{1,11}$ of gross profit in high-order third degree on future values of this parameter are represented.

Table 3. Values of coordinate function $\beta_{13}^{(1,1)}(\nu, i)$ $\nu, = \overline{1,11}, i = \overline{2,12}$

	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
2004	$9*10^{-6}$	$-1*10^{-7}$	$-3*10^{-7}$	$-5*10^{-7}$	$-3*10^{-7}$	$-2*10^{-7}$	$-8*10^{-6}$	$-3*10^{-6}$	$-1*10^{-6}$	$-3*10^{-7}$	$-4*10^{-7}$
2005	0	$-1*10^{-4}$	$-1*10^{-4}$	$1*10^{-5}$	$-3*10^{-4}$	$3,5*10^{-5}$	$7*10^{-5}$	10^{-4}	$6*10^{-5}$	$-2*10^{-4}$	$8*10^{-5}$
2006	0	0	$-1*10^{-5}$	$2*10^{-6}$	$-3*10^{-6}$	$-9*10^{-7}$	$2*10^{-8}$	$6*10^{-6}$	$-6*10^{-6}$	$-9*10^{-6}$	$-7*10^{-7}$
2007	0	0	0	$2*10^{-8}$	$-9*10^{-8}$	$2*10^{-9}$	$3*10^{-8}$	$7*10^{-8}$	$8*10^{-9}$	$4*10^{-9}$	$3*10^{-7}$
2008	0	0	0	0	$-8*10^{-8}$	$2*10^{-8}$	$2*10^{-8}$	$7*10^{-7}$	$3*10^{-7}$	$4*10^{-7}$	$3*10^{-8}$
2009	0	0	0	0	0	$-8*10^{-7}$	$-6*10^{-7}$	$-7*10^{-7}$	$-4*10^{-8}$	$-3*10^{-7}$	$-9*10^{-8}$
2010	0	0	0	0	0	0	$8*10^{-9}$	$8*10^{-6}$	$5*10^{-7}$	-10^{-8}	$-4*10^{-7}$
2011	0	0	0	0	0	0	0	$2*10^{-5}$	$8*10^{-7}$	$-2*10^{-6}$	$8*10^{-6}$
2012	0	0	0	0	0	0	0	0	$2*10^{-8}$	$4*10^{-7}$	$-3*10^{-6}$
2013	0	0	0	0	0	0	0	0	0	$3*10^{-8}$	$-7*10^{-6}$
2014	0	0	0	0	0	0	0	0	0	0	$9*10^{-6}$

As it can be seen in Table 3 values $\beta_{11}^{(3)}(i)$, $i = \overline{1,11}$ are relatively small but this doesn't mean that given weight coefficients don't influence on the forming of the estimation of future value as $\beta_{11}^{(3)}(i)$, $i = \overline{1,11}$ are multiplied in the process of calculations by values $x_1^3(i)$, $i = \overline{1,11}$ (values of the sixth-seventh order).

For functioning of the forecast model (11) on the basis of statistical data 25 tables of weight coefficients analogous to Tables 2-3 were calculated.

During the application of the method of economic indices prognostication for 2016 optimal order of non-linear relations of the investigated random sequence is unknown. But taking into consideration that $N=4$ is invariable during 11 years there is quite high probability that given parameter will remain on the same level.

Values in Table 4 reflects the change of relative error of prognostication of gross profit of enterprise (component $X_1(i)$, $i = \overline{1,12}$) during 2015 depending on the order of stochastic relations used in model (11).

Table 4. Relative error of prognostication of gross profit

Order of stochastic relations	2	3	4
Relative error	6,9 %	3,3 %	1,5 %

Thus the results of the experiment showed (Table 4) that application of nonlinear relations in the forecast model allows increase considerably the quality of economic indices prognostication.

5 Conclusion

Calculating method of the estimation of future values of economic indices of agricultural enterprises functioning is obtained in the work. The algorithm of extrapolation of vector random sequence based on nonlinear polynomial canonical expansion is assumed as a basis of the method. The optimal algorithm of the extrapolation of the economic indices of agricultural enterprises, which as well as canonical expansion put into its base doesn't impose any essential limitations on the stochastic properties of economic indices. In addition to pre-aggregate indicators (gross output, land resources, manpower, plant and equipment) a range of parameters can be used as the components of investigated random sequence (weather conditions, prices of resources, etc) influencing the effectiveness of the functioning of agricultural enterprises. The results of the numerical experiment showed that the forecast model possesses high accuracy characteristics at the expense of maximal taking into consideration of stochastic qualities of random sequence of economic indices change. Schemes of calculation of the parameters of the forecast model and estimations of future values of economic indices on its basis are introduced in the work. Expression for the mean-square error of extrapolation allows to estimate the quality of the forecast problem solving.

References

1. Drozd, J., Drozd, A.: Models, Methods and Means as Resources for Solving Challenges in Co-Design and Testing of Computer Systems and Their Components. Proceedings of 9th International Conference on Digital Technologies, Zhilina, Slovak Republic, pp. 225-230 (2013)
2. Gil-Aluja, J. Investment in Uncertainty. Kluwer Academic Publishers, Dordrecht, Boston, London (1999)
3. Gil-Lafuente, A.M. Fuzzy Logic in Financial Analysis. Studies in Fuzziness and Soft Computing 175, Springer, Berlin (2005)
4. Imanov, G., Barquero Cabrero, J.D.: Economy & Society 29. ESERP, Madrid, Barcelona, Palma de Mallorca (2015)
5. Kondratenko, Y.P., Encheva, S.B., Sidenko, E.V.: Synthesis of Intelligent Decision Support Systems for Transport Logistic. Proceeding of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS'2011. Vol. 2, Prague, Czech Republic, pp.642-646 (2011)
6. Kondratenko, Y.P., Sidenko, E.V.: Decision-Making Based on Fuzzy Estimation of Quality Level for Cargo Delivery. In book: Recent Developments and New Directions in Soft Computing. Studies in Fuzziness and Soft Computing 317, Zadeh, L.A. et al. (Eds), Springer International Publishing Switzerland, pp. 331-344 (2014)
7. Altman E.I., Marco G., Varetto F.: Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks. Journal of Banking and Finance, Vol. 18, pp. 505-529 (1994)
8. Altman E.I., Narayanan P.: An international survey of business failure classification models. J. Financial Markets, Institutions and Instruments, Vol. 6, №2, pp. 81-130 (1997)
9. Granger C.W.J., Newbold P.: Forecasting economic time series. Academic Press, (1986)
10. Hall S.G.: Applied economic forecasting techniques. Harvester Wheatsheaf, 224 p. (1994)

11. Ryabushkin B.T.: Application of statistical methods in an economic analysis and prognostication. Finances and statistics, Moscow, 175 p. (1987)
12. Trifonov Yu.V., Plehanova A.F., Yurlov F.F.: Choice of effective decisions in an economy in the conditions of vagueness. Publishing house NNGU, Nizhniy Novgorod, 140 p. (1998)
13. Połński M.: Prognozowanie czasu zakończenia inwestycji na podstawie jej bieżącego zaawansowania. J. Metody ilościowe w badaniach ekonomicznych, Tom XIII/3, pp. 169-179 (2012)
14. Prędkie A.: Subsampling approach for statistical inference within stochastic DEA models. J. Metody ilościowe w badaniach ekonomicznych, Vol. XIV, No. 2, pp. 158-168 (2013)
15. Teyl G.: Economy prognosis and making decision. Statistics, Moscow, 488 p. (1971)
16. Szmuksta-Zawadzka M., Zawadzki J.: Modele harmoniczne ze złożoną sezonowością w prognozowaniu szeregów czasowych z lukami systematycznymi. J. Metody ilościowe w badaniach ekonomicznych, Tom XIV/3, pp. 81-90 (2013).
17. Pugachev, V.S.: The Theory of Random Functions and its Application. Fitmatgiz, Moscow (1962)
18. Kudritsky V.D.: Filtering, Extrapolation and Recognition Realizations of Random Functions. FADA Ltd., Kyiv, (2001)
19. Sirenko N.N.: Management of strategy innovative development of agrarian sector of economy of Ukraine. MNAU, Mykolaiv (2010)
20. Atamanyuk I., Kondratenko Y., Sirenko N.: Management of an agricultural enterprise on the basis of its economic state forecasting. J. Metody ilościowe w badaniach ekonomicznych: Quantitative methods in economics, Vol. XV, No. 2, pp. 7-16 (2014)
21. Atamanyuk I., Kondratenko Y.: Calculation Method for a Computer's Diagnostics of Cardiovascular Diseases Based on Canonical Decompositions of Random Sequences. ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer. Proceedings of the 11 International Conference ICTERI-2015, S.Batsakis, et al. (Eds.), CEUR-WS, Vol. 1356, pp. 108-120 (2015)
22. Atamanyuk I.P.: Algorithm of extrapolation of a nonlinear random process on the basis of its canonical decomposition. J. Cybernetics and Systems Analysis, №2, pp. 131-138 (2005)
23. Atamanyuk, I.P., Kondratenko, V.Y., Kozlov, O.V., Kondratenko, Y.P.: The algorithm of optimal polynomial extrapolation of random processes, In: Modeling and Simulation in Engineering, Economics and Management, K.J.Engemann, A.M.Gil-Lafuente, J.L.Merigo (Eds.), International Conference MS 2012, New Rochelle, NY, USA, Proceedings. Lecture Notes in Business Information Processing, Volume 115, Springer, pp. 78-87 (2012)
24. Atamanyuk, I.P. Optimal Polynomial Extrapolation of Realization of a Random Process with a Filtration of Measurement Errors. J. Automation and Information Sciences. Volume 41, Issue 8, Begell House, USA, pp. 38-48 (2009)

Machine Learning Methods and “Real-Time” Economics

Volodymyr Dorozhinsky¹, Arsen Khadikov¹, and Nina Zholtkevych²

School of Math. and Comp. Sci., V.N. Karazin Kharkiv National University¹

School of Economics, V.N. Karazin Kharkiv National University²

`vdorozhinsky@gmail.com`

Abstract. In the paper some machine-learning method for synthesis an on-fly event acceptor is proposed. Such an acceptor has become necessary for the technique of the on-fly event processing. This technique is increasingly being used in the design of information systems of economic destination. In the paper the synthesis problem for such an acceptor is considered in a rigorous mathematical formulation. The method for solution of the problem and the computer experiment to study the method are described. The directions for the further research are proposed.

Keywords: on-fly event processing, acceptor, regular language, prefix-free language

Key Terms: Component, MathematicalModel, MachineIntelligence

1 Introduction

Modern Information Communication Technology (ICT) causes intensive changes in, practically, all areas of human activity. These changes, in particular, impose non-traditional restrictions for decision processes in the field of economy. New approaches for modelling economic processes such as “econophysics” [6] are responses on these challenges in the information society. The key reason of a huge number of the changes is a reduction of the time scale for decision making to manage of economic processes. This reduction can be argued that business management systems gain more and more features inherent in real-time control systems for complex technical objects. Thus, we may talk about real-time economic processes and use the corresponding technique to analyse and design them on the base of ICT. Today Event-Driven Architecture (EDA) is the generally accepted architecture solution to construct an information system that is scalable, adaptable, and capable to operate in real-time [3]. Mathematical fundamentals to analyse component behaviour of such a system have been established in the papers [1, 7, 8]. The practically important class of

message detectors for systems built on EDA has been defined and considered in [2]. The implementation of these ideas and theoretical results in the development process of information systems for business a priori constrained by the complexity of identifying message patterns of system events. In such situations, the usage of the machine learning methods is one possible way to overcome the complexity of the problem in progress of solving it. The advantage of this approach is needlessness to construct a general theory, covering all logically possible occasions. Instead, it creates a mechanism to adapt the system to new situation that is not consistent with the current knowledge of the system. The main goal of the paper is to present the principal suggestions of our approach and the preliminary results.

2 Basic Mathematical Model and Problem Statement

Here we describe the model proposed in [7] for a component of a system based on EDA.

2.1 Basic Notation and Definitions

Below we use the following notation:

$f: X \dashrightarrow Y$	denotes that f is a partial mapping from X into Y ;
$f(x) \uparrow$	denotes that $f(x)$ is not defined for the member x of X ;
$f(x) \downarrow$	denotes that $f(x)$ is defined for the member x of X ;
$f(x) \Downarrow y$	denotes that $f(x) \downarrow$ and $y = f(x)$ for the member y of Y ;
ϵ	denotes the empty (zero-length) sequence;
X^+	denotes the set of all non-empty finite sequences composed of elements of X ;
X^*	denotes the set $\{\epsilon\} \cup X^+$;
X^ω	denotes the set of all infinite sequences composed of elements of X ;
X^∞	denotes the set $X^* \cup X^\omega$;
$ \mathbf{x} $	denotes the length of the finite sequence \mathbf{x} ;
$\mathbf{x}[0]$	denotes the first element of a finite or infinite sequence \mathbf{x} ;
$\mathbf{x}[1:]$	denotes the sequence obtained by removing the first element of the sequence \mathbf{x} .

Definition 1. For a finite set X a subset L of X^* is called a language and, in the context, X is called an alphabet, and its members are called symbols.

We interpret symbols as a prime messages informing that the corresponding elementary event has happened. Some finite sequences of symbols inform about complex events and below these sequences are called events. Other finite sequences of symbols do not carry information about complex events and below we call them words. Sets of complex events must meet the certain conditions. The most important of these conditions is that any stream of elementary events is uniquely subdivided into a series of complex events by directed viewing the stream from left to right. This condition leads to the requirement that L is prefix-free [7]. Now we can require that any set of complex events related with a system would be prefix-free.

2.2 Mathematical Model

Let us briefly remind the principal tenets of the used model. Any component of a system based on EDA is modelled by using the concept of a **CEP-machine**.

Definition 2 (Structure of CEP-machine). Any CEP-machine is a quintuple $M = (X, Y, H, h_0, \alpha)$ with the following constituents:

- the **alphabet of atomic messages** X , which is a finite set;
- the **alphabet of machine responses** Y , which is a finite set;
- the **set of handlers** H , which is a finite set, whose each member is a partial mapping $h: X^+ \dashrightarrow Y$ such that its domain is a prefix-free language;
- the **initial handler** h_0 , which is some fixed element of H ;
- the **response function** α , which is a mapping with domain Y and codomain H .

The general behaviour of any CEP-machine is described in [7]. However in the general case it is possible that a CEP-machine can become hung while trying to recognize an event. But in our study we consider simpler case, which was first considered in [2]. In this case the situation, when the CEP-machine is hanging, is impossible. To specify this special case we are in need in the following definition.

Definition 3 (Regular handler). A handler $h: X^+ \dashrightarrow Y$ is called regular if there exist some finite set Z with the marked element $z_0 \in Z$

and some mapping $\delta: Z \times X \rightarrow Z \cup Y$ such that for any $\mathbf{u} \in X^+$ and $y \in Y$ the condition $h(\mathbf{u}) \Downarrow y$ is fulfilled iff there exist $z_1, \dots, z_{|\mathbf{u}|-1} \in Z$ such that

$$\begin{aligned} z_{i+1} &= \delta(z_i, u[i]) \quad \text{for } 0 \leq i < |\mathbf{u}| - 1 \quad \text{and} \\ y &= \delta(z_{|\mathbf{u}|-1}, u[|\mathbf{u}|]). \end{aligned}$$

In this case we say that the handler h is realized by the triple (Z, z_0, δ) .

Remark 1. It is evident that a handler is regular if there exist a finite state machine realising it.

Definition 4 (Regular CEP-machine). A CEP-machine is called regular if all its handlers are regular.

2.3 Problem Statement

In practice we propose restrict our technique by methods of synthesis for regular CEP-machines. This restriction is caused by technical and theoretical difficulties of more general technique. Further, it is evident that the synthesis problem for a regular CEP-machine is decomposed into series of synthesis problems for regular handlers each of which has only one possible response “accepted”. In this case we use the term “a regular acceptor” instead the term “a regular handler”. Thus, a machine-learning problem for synthesis process of a regular handler can be formulated in the following manner.

Problem. Let $E = \{\mathbf{u}_1, \dots, \mathbf{u}_M\}$ be a finite prefix-free set of finite sequences composed by elements of X and $C = \{\mathbf{v}_1, \dots, \mathbf{v}_N\}$ be a finite set of finite sequences composed by elements of X such that $E \cap C = \emptyset$ then we interpret elements of set E as examples and elements of set C as counterexamples; we need to find a regular acceptor $h: X^+ \dashrightarrow \{\text{accepted}\}$ such that

1. $h(\mathbf{u}_i) \Downarrow \text{accepted}$ for all $0 \leq i < M$;
2. $h(\mathbf{v}_i) \Uparrow$ for all $0 \leq i < N$; and
3. the corresponding set Z has the least number of elements among all possible.

3 Progress Review

Here we present a method to build a regular acceptor and describe a computer experiment that instills confidence in the existence of a mathematical justification for this method.

3.1 Some Theoretical Background

We premise our presentation with a few simple theoretical results. The techniques used to prove these results are quite common in automata theory therefore we omit the corresponding proofs for simplicity of the presentation. First of all, let us return to Remark 1 and discuss more detail the interrelation between regular acceptors and finite state machines. Namely, if we consider for any regular acceptor $h : X^+ \dashrightarrow Y$ that realized by the triple (Z, z_0, δ) the machine $(Q = Z \cup Y \cup \{q_{trap}\}, X, \bar{\delta} : Q \times X \rightarrow Q, q_0 = z_0, Q_{accept} = Y)$ where $q_{trap} \notin Z \cup Y$ and $\bar{\delta}(y, x) = q_{trap}$; $\bar{\delta}(q_{trap}, x) = q_{trap}$ for any $x \in X$ and $y \in Y$ then the regular language accepted by this machine coincides with the set of events accepted by the regular acceptor. After this brief theoretical review, we can return to our problem.

3.2 Proposed Method

The general view of the method to find an acceptor is presented by Alg. 1. This method consists of the series of redirections for acceptor transitions leading into the trap starting with the minimal acceptor for the set of examples. Two functions used by the algorithm **init** and **modify** are specified separately.

To complete the specification of the proposed method we need to describe algorithms for the function **init** (see item 3 of the Alg. 1) and for the function **modify** (see item 7 of the Alg. 1).

Function init. To build the minimal acceptor the following ideas are used:

1. states of the acceptor are defined recursively as special sets of words;
2. we choose the set E as z_0 and add it to Z ;
3. we choose the empty set as **trap**;
4. if for $x \in X$ in E there is not a word with the first symbol x then assign $\delta(z_0, x) = \mathbf{trap}$ else the set $\{u \in X^* \mid xu \in E\}$ add to Z ;
5. repeat recursively this consideration for all member of Z until Z is stabilized;
6. denote the set $\{\epsilon\}$ by **accepted**.

The acceptor obtained in this manner is assigned as a result of the function **init**.

Algorithm 1. Specification of the proposed method

```
1 def learning_method( $E, C$ ):
    Require : the finite prefix-free set of events  $E$ 
              the finite set of words that are not events  $C$ 
    Ensure : the required acceptor
2   do that:
3       initiate the learning process by applying function init to the set  $E$ 
        and denoting the result by acceptor
    # initialize the set of transitions that cannot be redirected
4   frozen_transitions = set()
5   while halting condition is not fulfilled:
6       do that:
7           modify acceptor by redirecting a transition leading into the trap
            and minimize the resulting acceptor wherein the redirected
            transition cannot belong to frozen_transitions
8       do that:
9           check that acceptor is admissible in the sense that it does not
            accept any word from  $C$ 
10      if the checking is successful: continue
11      else:
12          do that:
13              roll-back the modification and add the redirected transition
                into frozen_transitions
```

Function modify. To select a transition for redirection we use the following simple remark: the minimal regular acceptor has at most one state that is an attractor, i.e. any transition that goes out from this state has this state as a target. Moreover, if this acceptor accepts a finite language then the existence of the attractor is guaranteed. Further, to minimize the new acceptor we use standard Hopcroft’s algorithm [4].

3.3 Computer Experiment Schema

Thus, we assume that the described above method gives a solution of our problem. To check reasonability of this assumption we designed the computer experiment for searching counterexamples to the assumption. The Alg. 2 specifies the schema of the experiment.

3.4 Case Study

To implement the mentioned experimental schema we have used language Python with libraries “scipy” and “numpy” [5]. Particularly, all random

Algorithm 2. The schema of the computer experiment

```
1 for _ in range(given_number_of_experiments):
2     do that:
3         | generate randomly a regular acceptor acceptor
4     do that:
5         | generate randomly sets  $E$  and  $C$  using acceptor
6     acceptor' = learning_method( $E, C$ )
7     do that:
8         | compare acceptor' and acceptor
```

choices have been provided by the standard function **random.choice** contained in the library “numpy”. To randomly generate a regular expression the following schema has been used. An expression is represented by a syntactic tree. Each leaf of this tree is marked by a token and each internal node of the tree is marked by a functor. Moreover, the number of children for an internal node equals the arity of the corresponding functor. The recursive structure of the syntactic tree that represents some regular expression indicates the way of this tree random generation. To make decision whether the root of the current subtree is an internal node or a leaf we propose to use the following function $p(n)$ that determines the conditional probability to mark the current node as internal if its depth is equal to n

$$p(n) = \begin{cases} 1 - \frac{1}{2} \left(\frac{n}{\mu} \right)^2 & \text{if } 0 \leq n < \mu \\ \frac{1}{2} \exp \left(1 - \frac{n}{\mu} \right) & \text{if } n \geq \mu \end{cases}.$$

The results of the more than 10,000 experiments have shown that for a randomly generated acceptor with the obtained sets E and C , the presented method of learning was restoring this acceptor using these sets. Thus, we can assume that the proposed method is precise on regular acceptors. The last assumption can be considered as evidence in favour of the validity of the proposed method of machine learning.

4 Conclusion and Future Study

The presented paper cannot be considered as a complete research. The obtained results are preliminary, but they are very important because they

demonstrate a chance to substitute the complete logical analysis of situation by the learning on the examples and counterexamples during software development of the on-fly processing systems. Also these results make the need for further research in the following directions: future experimental study of the proposed method in order to clarify the boundaries of its applicability; finding rigorous formulations of the method convergence conditions; mathematical justification of the method; evaluation of the effectiveness of the method.

Acknowledgement. The authors thank Prof. Gregoriy Zholtkevych for the idea proposed to them: to use machine learning methods for the synthesis of the logical structure of the on-fly event processing mechanism.

References

1. Dokuchaev, M., Novikov, B., Zholtkevych G.: Partial Actions and Automata. *Alg. Discr. Math.* 11(2), 51–63 (2011).
2. Dorozhinsky, V.: Regular Complex Event Processing Machines. *Sistemy obrobky informacii.* 8, 82–86 (2015).
3. Etzion, O., Niblett, P.: *Event Processing in Action*. Manning Publications (2010).
4. Hopcroft, J.: An $n \log n$ algorithm for minimizing states in a finite automaton. In: *Proc. Internat. Sympos., Technion, Haifa. Theory of machines and computations*, pp. 189–196. Academic Press, New York (1971)
5. Scientific Computing Tools for Python. <http://scipy.org/>
6. Stanley, H.E.: Interview on Econophysics. Published in: “IIM Kozhikode Society & Management Review”, Sage publication (USA). 2(2), pp. 73–78 (2013) <http://www.saha.ac.in/cmp/camcs/Stanley-interview.pdf>.
7. Zholtkevych, G., Novikov, B., Dorozhinsky, V.: Pre-Automata and Complex Event Processing. In: V. Ermolayev et al. (eds.) *ICTERI 2014. CCIS*, vol. 469, pp. 100–116. Springer International Publishing, Switzerland (2014).
8. Zholtkevych, G.: Realisation of Synchronous and Asynchronous Black Boxes Using Machines. In: V. Yakovyna et al. (eds.) *ICTERI 2015. CCIS*, vol. 594, pp 124–139. Springer International Publishing, Switzerland (2016).

Government Financial Accountability and Transparency in the Digital World

Lidiia Hladchenko

Kyiv National Economic University named after Vadym Hetman

`l_gladchenko@kneu.edu.ua`

Abstract. This article examines the issues of accountability and transparency of government finance in Ukraine. The research explains the significance of financial accountability for the economic improvement and civil society development. The study defines the instruments that contribute to government accountability in the digital world. The key problems of ensuring financial accountability and transparency in Ukraine are analyzed. The study focuses on information and communication technology capacity in supporting government performance. The conclusions summarize the main findings of the research.

Keywords. Transparency, government accountability, public finance, civil society, information and communication technology.

Key Terms. InformationCommunicationTechnology, Collaboration, Decision-Support, Model

1 Introduction

The current tendencies of Ukrainian civil society development have great influence on economic performance both in private and public sectors. The evolution of civil society and democratic institutions necessitates strengthening of government accountability. As information and communication technology (ICT) has permeated all the areas of people's life, they are able to improve practice of public governance. Government accountability and transparency have created institutional background for sustainable development. State and local budgets play the leading economic role in this process as they express public policy, society's aims, needs and intensions.

Ukraine demonstrates substantial progress in developing of ICT in the government sector. At the same time, the practice of government financial accountability and transparency in Ukraine are still weak. This situation requires research and developing the ways of improving public governance.

The aim of this study is to reveal influence of ICT on government financial accountability and transparency.

2 Theoretical and Methodological Background

The conceptual framework for government transparency and accountability was developed by research in political science. The common approach considers them as relationships promoting democracy. Thus, O'Donnel [5] focuses on different types of accountability in the context of a political system. The special attention is devoted to the horizontal accountability in counties that have recently become political democracies (mainly, in Latin America). They demonstrate reasonably good vertical accountability, and have problems with horizontal accountability.

In recent research, much attention is paid to the government horizontal accountability, i.e. accountability to parliament. Rahman [9] has examined parliament control and government accountability in the South Asia countries, and concluded that although parliaments have been able to perform the key parliamentary tasks, they do not perform as successfully as their counterparts in the Western world in controlling the government and holding it to account.

Ukraine belongs to the group of countries with young democracy, and demonstrates the common tendencies for this group. This explains the significance of experience of democratic developing countries for Ukraine.

Pelizzo and Stapenhurst [7] use the broad approach to the accountability, and stress that the concept of accountability refers not only to the financial control and reporting. Accountability means a relationship between agent and principal where the agent provides information or justification for its actions. Authors identify four types of it: vertical, horizontal, diagonal and social.

Some international organizations permanently examine financial or budget government transparency and accountability. They usually use indices and ratings to assess countries progress in this area, and take into account different dimensions including financial aspects and information technologies for public governance.

Transparency and accountability studies of Ukrainian academics demonstrate the same pattern as foreign scientists' research. They traditionally concern political science. There are few research devoted to the budget aspects of transparency and accountability. For instance, Lytvynchuk [4] has explored public control over local budget execution in Ukraine, and has concluded that public control although has legislation framework, but does not lead to the better governance. Belets [1] has paid attention to the experiment of providing participatory budgets in Ukraine. This project started in the middle of 2015, and it is too early to have the results.

The using of information and communication technology to promote government financial transparency and accountability has not yet obtained proper attention. The cases of ICT providing for government needs in Ukraine are covered in mass media. They deserve research in the context of dynamic development of a civil society as a background for horizontal accountability relationships.

3 Types of Government Accountability

Government accountability is a concept of relationships where government institutions provide information or justification for its actions [7]. The key questions of such approach are "What kind of information?" and "For whom should information be

provided?” The answers determine different types of accountability. Depending on the principal in the relationships of accountability, there can be horizontal and vertical accountability. *Horizontal accountability* exists within the public authorities and supposes the capacity of higher level of public authorities to control the lower level authorities. These are the strong hierarchical processes, usually determined by laws. They encompass various forms: compliance with different rules and regulations, monitoring and auditing, reconciliation of budget revenues and expenditures, and budget reporting. The executive branch is based on some principles of administrative science such as hierarchy, rules, impersonality and other [9]. They naturally lead to the centralized character of relations within the government institutions.

The structure of public power in democratic countries consists of three branches: the executive, the legislative and the judicial. It causes three dimensions of horizontal accountability: government institutions can be responsible to the higher levels of government within the executive branch, to the legislative institutions, and in some cases to the judicial institutions. Parliament may be aided by other institutions: supreme audit institutions, anti-corruption commissions, ombuds offices and human rights institutes.

Vertical accountability is the ability of citizens, civil organizations and groups, and mass media enforce standards of good performance on officials. The key instrument of vertical accountability is elections. The strength of democracy as a social institution constitutes a precondition for high effectiveness of government accountability. There are a number of other instruments supporting vertical accountability: participatory budgets, public hearings at budget process, budget reporting, monitoring, civil boards at government authorities, petitions, and other.

Diagonal accountability is a relatively new concept of government accountability. It supposes that citizens are engaged directly in the performance of government institutions. The main instruments of it concern the idea of citizens' direct participation in public governance. The abovementioned instruments, such as participatory budgets and civil boards at government authorities, contribute to diagonal accountability.

There are also other types of government accountability (political and legal, social, personal and group), which are beyond this research. Horizontal, vertical and diagonal accountability are the components of logical model of government financial accountability (fig. 1).

4 Government Financial Accountability: Challenges for Ukraine

Financial information that concerns government budget, public debt and fiscal policy determines the *financial (or budget) accountability*. There are three pillars of budget accountability: transparency, public participation, and formal oversight institutions that create together a budget accountability ecosystem [6] (fig. 2).

Budget transparency refers to the extent and ease with which citizens can access information about and provide feedback on government revenues, expenditures, fiscal deficits and public debt. Transparency also supposes the understandable for citizens and clear (1) procedures of decision-making at budget process, and (2) budget information.

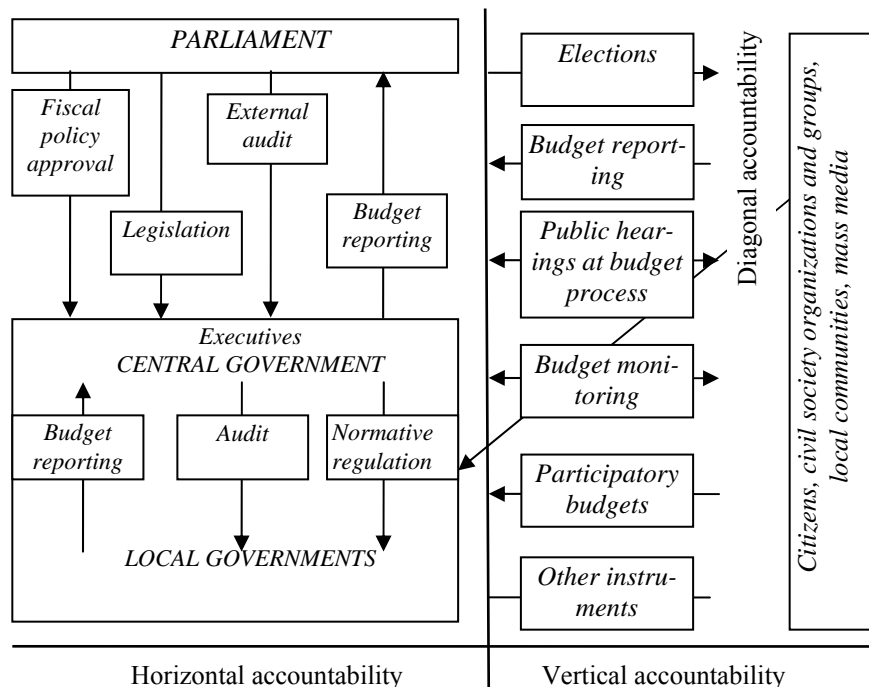


Fig. 1. Logical model of government financial accountability

Further research requires attention to the main features of government sector in Ukraine. It has several components: (1) budget system and (2) social security funds (Figure 3). The distinguished feature of government sector is a wide range of budget organizations. They provide goods and services to the community and individual households: elementary, professional and high education, healthcare services, scientific researches, cultural development, art, sport etc. All budget organizations are under full government control and are funded by central and local budgets. Social Security Funds are included in budget, except of Pension Fund of Ukraine.

The structure of budget system demonstrates a high level of centralization. Thus, according to the reporting data of the Ministry of Finance and the State Treasury, 76.25 % - 81.5 % of all budget revenues (excluding intergovernmental transfers) were concentrated in the State budget during last four years. At the same time, more than 90 % of the local budgets are subsidized. The share of official transfers in the structure of local budgets revenues has increased from 31.2 % in 2002 to 57.1 % in 2015. This creates basis for the low level of the local governments' financial accountability to the local communities and high level of accountability to the higher levels of government.

For a long time Ukraine had a centralized system of governance, where citizens had no real opportunity to influence the decision-making process. This practice has a strong negative influence on national economy: unwillingness to pay taxes and tax avoidance, high level of shadow economy, the very low level of trust to all public authorities. Thus, according to the Ministry of Economic Development and Trade of

Ukraine, shadow economy in 2015 equals 47 % of official GDP of Ukraine. The Ilko Kucheriv Democratic Initiatives Foundation [11] explored the political results of 2015, according to which the level of citizens' trust to the central public authorities – parliament and central government – equals 50 % and demonstrates the declining tendency for the last two years. That is why public participation should become the crucial point in reforming public governance and budget sector in Ukraine.

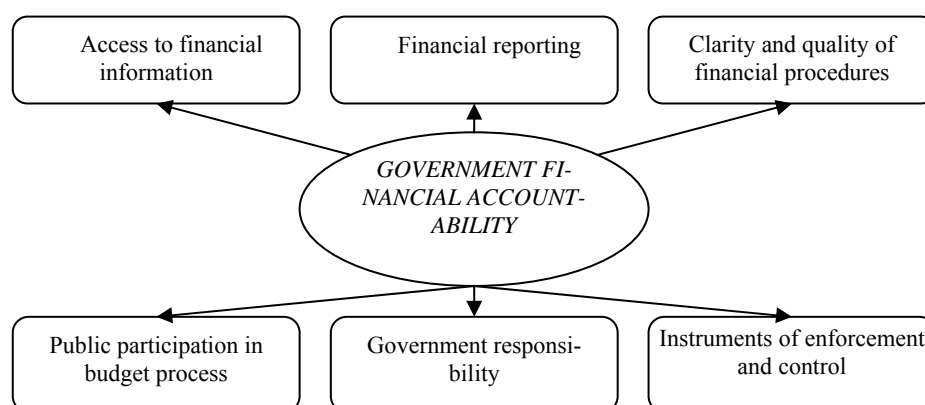


Fig. 2. Key features of Government financial accountability

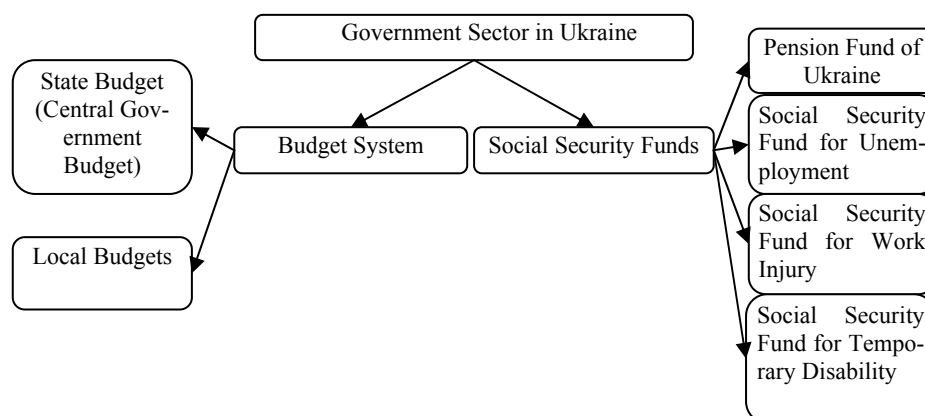


Fig. 3. Structure of Government Sector in Ukraine

More detailed information about government financial accountability and transparency can be analyzed on the basis of Open Budget Index (OBI) methodology. The Open Budget Survey [6] demonstrates that Ukraine obtained 46 of 100 points in 2015. It means insufficient level of budget openness. OBI is an integrated indicator, which takes into account the three components of a budget accountability system [6]:

(1) Availability of budget information to the public. Ukrainian government demonstrates sufficient level of this indicator, failing only two positions [6]. Available

documents included Pre-Budget Statement, Executive's Budget Proposal, Enacted Budget, In-Year Reports, Year-End Reports, and Audit Report; while Ukraine did not have Citizens Budget and Mid-Year Review. This creates the pretty good basis for financial transparency.

(2) Strength of formal oversight institutions shows high points for Ukraine: Oversight by Legislature equals 79, and Oversight by Supreme Audit Institution equals 83 (with a score from 0 to 100). This indicator confirms high level of horizontal accountability in Ukraine, which is natural for the country with centralized system of public governance.

Comprehensive budget information and formal oversight institutions are those tools that afford citizens to monitor and supervise decision-making process, budget planning and implementation.

(3) Opportunities for the public to participate in the budget process. The Open Budget Survey estimated *Public Participations* at the very low level - 23. Public participation characterizes diagonal accountability, which proves to be a real challenge for Ukraine. Information and communication technology can play the key role in combating this problem.

5 Information and Communication Technologies for Improving Government Financial Transparency and Accountability

ICT offers solutions for governments to improve the situation with engaging citizens to public governance processes. ICT also supports public administration, public financial management, and provision of government services. ICT has potential beneficial effects on governance: (1) ICT is able to increase the quality of governance to meet citizens' demands, reducing the cost of government operations, the access to and delivery of government services; (2) the use of ICT enhances improving transparency and accountability; (3) ICT allows to engage citizens and develop democracy; (4) ICT improves government ability to collect taxes and even is able to increase the level of voluntary tax payments [3]; (5) ICT provides ready access to information without visiting government entities; (6) ICT provides feedback to government implementing institutions in real time; (7) ICT ensures easily reporting in real time; (8) ICT curtails corruption in service delivery.

This wide range of positive effects is confirmed by tendencies in public administration. The distinctive feature of the current period is "the generalization of IT systems from only affecting back-office processes to conditioning in important ways the whole terms of relations between government agencies and civil society" [2]. The United Nations E-Government Survey claims that "the transformative changes entail not only the design and implementation of innovative practices, but more fundamentally a transformation of government's role, functions, institutional frameworks and processes" [13].

The index of e-government development demonstrates the same tendency in Ukraine as the Open Budget Indicator. This index estimates the readiness of national governments to use the Internet and mobile technologies for government functions. E-Government Development Index (EGDI) rates the e-government performance of United Nations Member States on the basis of integrated indicator which takes into

account [13]: (1) Online Service Index (OSI); (2) Telecommunication Infrastructure Index (TII); and (3) Human Capital Index (HCI):

$$EGDI = 1/3 (OSI_{\text{normalized}} + TII_{\text{normalized}} + HCI_{\text{normalized}}) \quad (1)$$

Ukraine was in the group of countries with high EGDI in 2014 (the last issue of the Survey). It ranked 87th place of 193 countries with EGDI equaled 0.5032 (in 2010 Ukraine was on the 54th place of 184 countries). The Survey showed that in 2014 there was a tremendous gap between Online Service Index 0.2677 and Telecommunication Infrastructure Index 0.3802, from the one hand, and Human Capital Index 0.8616 - from the other hand. It means that online services and communications were substantially behind the society's capacity of participating in public governance.

Civil society development in Ukraine is expressed through the vast volunteer movement. Volunteers are engaged in the developing of e-Democracy and e-Governance, anticorruption legislation, ensuring government transparency and openness. Among the most prominent results of volunteers work is the system of public procurements ProZorro. The volunteers' idea of reforming the system of public procurements appeared at the Open University of Maydan in the beginning of 2014. The pilot project started to work in February 2015. Ukrainian parliament adopted the law on public procurements reform in September 2015 and it was put in force since April 1st, 2016. ProZorro was developed by IT-Volunteers. Ukrainian government obtained it free of charge. The total market of public procurements equals to about 250 billion UAH per year. Trade volumes on the new e-platform have reached more than 15 billion UAH since ProZorro was launched as the pilot project. ProZorro managed to save 1.4 billion UAH (it equals 9.2 %) and 30 % of time for tenders participants for the first year of its existence in the test regime [8].

At the same time, IT-Volunteers actively participate in developing e-Governance and e-Democracy in Ukraine. Among the recent projects, that have successfully started, are e-Petitions, the "Digital Cabinet", e-Delivery of government services, online budget monitoring. The Association of Ukrainian cities is gathering the best practices of public initiatives on the local level.

IT-Volunteers participation in improving of public management witnesses about active development of diagonal accountability processes in Ukraine. The integration of participatory elements in the process of decision-making is a crucial for improving efficient public governance.

6 Conclusions

Government financial accountability and transparency have the crucial influence on the performance of political system and the quality of public administration. Ukraine demonstrates sufficient development of horizontal accountability. By contrast, the vertical accountability is weak and needs substantial improvement. The recent activation of civil society in Ukraine gave examples of diagonal accountability. The prominent role in ICT embedding in the Ukrainian government sector belongs to volunteers. They initiated, developed and raised money to fund several projects. The gap between horizontal and vertical accountability can be overcome by using information and communication technology. ICT offers a wide range of instruments: e-

Petitions, the “Digital Cabinet”, participatory budgeting, online monitoring, online voting, online budget reporting, and delivery of government services. The e-Government allows for broad changes in collaboration between a government and its stakeholders. The integration of participatory elements in the process of decision-making is a crucial for improving efficient public governance.

References

1. Belets, Zh.: A. Participative Budgeting as Innovative Democratic Instruments of Civic Participation in Administration of Territorial Communities. Theory and Practice of State Governance and Local Self-Governance 1, http://el-zbirn-du.at.ua/2016_1/19.pdf (2016) (in Ukrainian)
2. Dunleavy, P., Margetts, H., Basto, W. S., Tinkler, J.: New Public Management Is Dead – Long Live Digital-Era Governance. *Journal of Public Administration Research and Theory* 16 (3), 468-494 (2006)
3. Hoekstra, W., Pereira, F. G., Queijo, J.: Ten Quick Steps to Unlocking Tax-Revenue Collection in Rapidly Growing Markets. <http://www.mckinsey.com/industries/public-sector/our-insights/ten-quick-steps-to-unlocking-tax-revenue-collection-in-rapidly-growing-markets> (2013)
4. Lytvynchuk, I.: Public Control Over Local Budgets Execution. *Visnyk ZhDYU* 4 (70), 155-158 (2014) (in Ukrainian)
5. O'Donnell, G. A.: Horizontal Accountability in New Democracies. *Journal of Democracy* 9 (3), 112-126 (1998)
6. Open Budget Survey, <http://www.internationalbudget.org/wp-content/uploads/OBS2015-Report-English.pdf> (2015)
7. Pelizzo, R., Staphenurst, F.: *Government Accountability and Legislative Oversight*. Routledge (2013)
8. Public Procurements: How the System ProZorro 2.0 Will Work, <http://delo.ua/ukraine/publichnye-zakupki-kak-budet-rabotat-sistema-prozorro-20-314681/> (2016) (in Ukrainian)
9. Rahman, T.: *Parliamentary Control and Government Accountability in South Asia: A Comparative Analysis of Bangladesh, India and Sri Lanka*. Mitra, S. K. (ed). South Asia Institute, University of Heidelberg, Germany, Routledge (2008)
10. Slynko, T. M.: Transparency of Public Authorities as a Guarantee of Democratic Governance. In: *PRÁVNÁ VEDA A PRAX V TREŤOM TISÍCROČÍ. Zborník príspevkov z medzinárodnej vedeckej konferencie*, 28-31 (2015) (in Ukrainian)
11. The Ilko Kucheriv Democratic Initiatives Foundation, <http://www.dif.org.ua/en/polls/2015a/2015-i-politichni-pidsumki---dumka-naselennja-.htm> (2015) (in Ukrainian)
12. United Nations E-Government 2014. E-Government for the Future We Want, https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2014-Survey/E-Gov_Complete_Survey-2014.pdf (2014)
13. United Nations E-Government Survey and E-Government Indicators, 2014, https://www.itu.int/en/ITU-D/Regional-Presence/AsiaPacific/Documents/Events/2014/October-MIS/Session3_UNPOG.pdf (2014)

Using an Evolutionary Algorithm to Improve Investment Strategies for Industries in an Economic System

Vitaliy Kobets¹ and Maksym Poltoratskiy¹

¹Kherson State University, 27, 40 rokiv Zhovtnya st., Kherson, 73000 Ukraine

vkobets@kse.org.ua, max1993poltorackii@gmail.com

Abstract. Evolutionary investment strategies have been found to adapt distribution of investors' fund to technological changes of economic industries. In this paper we use evolutionary algorithm (EA) to improve investment strategies in economic system comparing shares of most profitable investors with other ones. Proposed algorithm selects an optimal combination of investment share in different industries for improving of investment performance. In our experiment we use statistical data from official sources to examine effectiveness of proposed investment strategy. We find the investment strategies using evolutionary algorithm to teach investors how to increase their profit as much as possible and obtain more profit than investor with best start possibilities.

Keywords. Evolutionary algorithm, industry, technology, input-output model, profitability

Key Terms. EvolutionaryAlgorithm, InputOutputModel, InvestmentStarategy

1 Introduction

Increasing the diversity of the population (investors' decisions) is achieved by means of evolutionary algorithm through discovery of investment share for all investors, except for the most successful one according to criterion of profit that encourages innovations in investment strategies. Reducing the diversity of the population by selecting the most successful investors according to the criterion of profit and fixation of their fund shares stimulates to a quality of investment strategies.

In the evolutionary metaphor *phenotype* is investor, *genotype* is set of investor's decisions (investor strategy means how much money is invested in an asset) (Fig. 1). Defining the best investment strategies using evolutionary algorithms takes place in the space of genotypes. Some phenotypic characteristics (share investment allocation between industries) are beneficial as they generate more profit. These features are rewarding for the progeny (investors' strategies of succeeding periods).

Solution of the problem for investors takes pace on the level of genotype. Population (set of investors) is the basic unit of evolution, i.e. population develops

rather than investors. Selection operators affect the set of investors. Variation operators act on the level of individual investor.

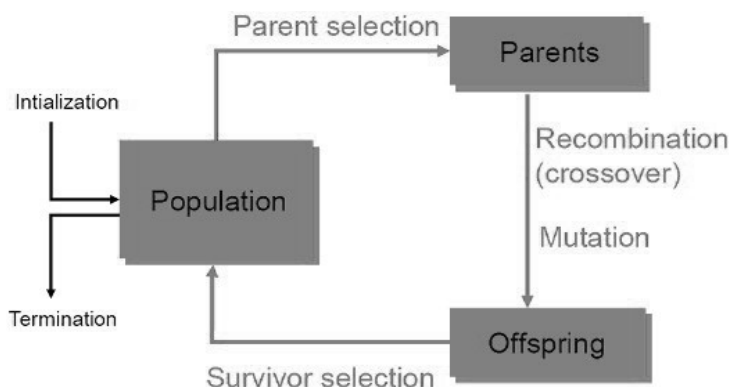


Fig. 1. Evolutionary selection of investment strategies

Evolutionary selection: (i) gives the best investors a better chance to increase their own profit (start the next year with growth of wealth) and not to become bankrupt; (ii) moves the set of investors to better off technological adaptability.

Evolutionary strategies are usually applied to numerical optimization as a fast and good optimizer for real numbers with peculiarity of self-adaptability parameters through mutations [1].

The general problem is the effect of dynamics of technological coefficients on evolutionary strategies of investors in the economy. Basic input-output model for economic system has been developed by us, where the evolution of investment strategies depends on the variety of strategies, potential of industries and investors' fund. Numerical analysis shows how investors adapt their strategies to the changes of economic environment in order to increase their own profits. The question is how investors should change their strategy when there is uncertainty (there is a lack of data or volatility) as regards to a specific type of the dynamics of economic environment.

2 Economic Evolution in a Dynamic Environment

Evolutionary explanations and agent simulation are standard practice of research in economic sciences. Heterogeneity and cooperation of investors and their strategies are better treated by evolutionary agent model. An evolutionary model uses a set of agents for selection and changes. Evolutionary dynamics concerns behavioral interactions in evolutionary economics, which consists of several agents with heterogeneous economic strategies. We investigate how irregular and unpredictable dynamic economic environment changes the behavior of investors who make decisions in terms of such environments. Evolution in a changing environment does not occur when reaching equilibrium (steady state). Often it is more important not

only how *well* agents adapt i enough time, but how *quickly* they adapt to new challenges.

Environment variables include macroeconomic conditions, technological opportunities, economic policies and institutions (rules, procedures) and natural resources. Aspects of changes in environment's technology are: (i) how quickly they occur; (ii) how often (1 time per period).

The emergence of a dynamic environment usually leads to the fact that the system is no longer subject to analytical solution. Numerical simulations of multi-agent systems form an alternative to analytical approach, because it is much more flexible in the study of the behavior of the economic system. They allow to study the impact of set of investors and effects of dynamic environment on the formation of evolutionary strategies and their effectiveness. Different assumptions can be made about the selection factors and innovative mechanisms (random mutations of technological factors, deterministic trends for best investors, recombination of strategies) and bounded rationality of agents (habits, imitation) [2, 3]. Systems with *endogenous* variables (in contrast of exogenous ones) generate system with complex feedbacks.

Evolutionary algorithms are not sufficiently explored for the development, analysis and application of methods of solving problems in the economy [4]. The evolutionary approach allows us to make analysis and adaptation of an existing population (industries) to new environmental conditions, the algorithm implements the principle of adaptation and learning (fig. 2).

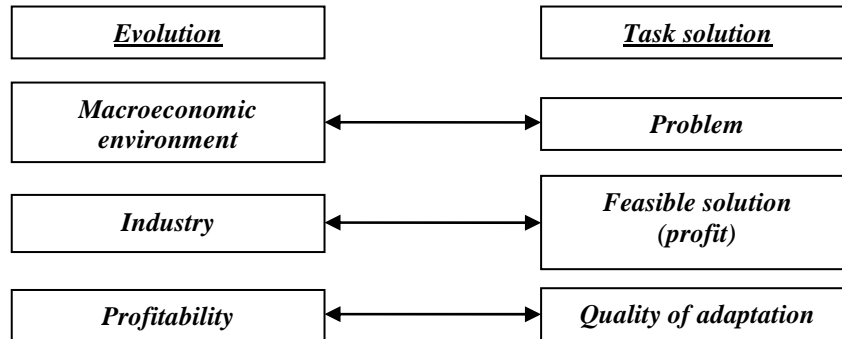


Fig. 2. Correspondence of economic and evolutionary computation

Evolutionary algorithm (method) includes evolution strategies, genetic algorithms and evolutionary programming. In contrast of exact methods, evolutionary methods are approximations and allow for predetermined reasonable time to obtain high-quality solutions, which are very close to optimal and acceptable from a practical point of view [5]. Existing research do not consider investor's decision making using input-output model and evolutionary algorithms [6,7,8]. Thus we developed our own general model to compare how profit maximization strategies of investors adapt to technological changes.

The evolutionary agent model with initial conditions obtained through statistic data has been developed [9] describing a set of n economic industries (final demand and matrix of technological coefficients) and m investors, which are characterized by open exchange of information. Technological matrix coefficients are determined empirically for different industries. We identify the key variables for monitoring, such as (i) industry profitability, (ii) profit of investors, (iii) the share of investment fund distribution among industries for most efficient investors. Changing technology is reflected on the adaptive strategies of investors. Therefore, investors should diversify their portfolio from initial to the most profitable in an environment where technological change and the dynamics of resources may make a variation of production factors.

In the dynamic input-output model the investors form their individual evolutionary strategy, determining the direction of increasing their profits under technological changes. Their goal is to maximize their own profits. Investors prefer evolutionary strategies that generate more profit. Their rational capacity (possibility of predicting changes) and information for investment (priority selection industries) is limited. Investors have access to information about profits of other investors, profitability of all industries and investment shares, distributed by each investor between industries. Microeconomic information about private businesses is usually confidential, while macroeconomic information about industries is public.

Behavioral interaction determines how investors use accessible information to improve their evolutionary strategy through simulation of the most successful investors' behavior. The dynamics of the environment are outside the control of investors. Investors imitate the most effective investment portfolio, but how much it will increase their profit depends not only on the accuracy of information disclosed but on unpredictable dynamics of technological interconnection of industries within economic system. The dynamics of the environment is described by types of business cycle (table 1):

(i) **Kitchin cycle** is a short-term business cycle (40 months), when firms react to improvement of commercial situation by increasing their output through full employment using a fixed capital or decreasing of demand and increasing of stocks of products. Entrepreneurs need time to respond to this changing demand and materialize their decisions in the product. This cycle corresponds to final demand Y change.

(ii) **Juglar cycle** (7-11 years) means fluctuations of investment in fixed capital without changing the level of employment at fixed capital. Business cycles are driven by investments. It aligns to evolutionary strategies of investors in our input-output model.

(iii) **Kuznets swing** (15-25 years) relates with demographic processes, migration flows, reflects the demographic or infrastructure investment cycles. It corresponds to minor changes of technical coefficients in our input-output model.

(iv) **Kondratiev wave** (45-60 years) implies technological change, expressed by intervals of high growth of sectors and intervals of relatively slow growth. These waves conform to significant changes in the technological coefficients of our input-output model.

Table 1. Types of business cycles

Cycle type	Period (years)
Kitchin cycle (inventory, e.g. pork cycle)	3–5
Juglar cycle (fixed investment)	7–11
Kuznets swing (infrastructural investment)	15–25
Kondratiev wave (technological basis)	45–60

The paper **goal** is to develop new general model to compare how profit maximization strategies of investors adapt to changes in technological interrelations between industries under imperfect information by means of evolutionary algorithms.

3 General Model Characteristics

Consider a set of investors, each of them having a goal to achieve a higher level of personal welfare. Each investor may in turn invest his money in a finite number of industries. The procedure for allocation of funds by investor defines his investment strategy. Once the funds are invested, they can not be moved between sectors within one time period (year).

Agents understand the existence of a causal link between the investment strategy in the current period and the profit in the next period, but they can not use a calculation to find an investment strategy that maximizes their profit level. Instead, agents use the evolutionary search method, borrowed from nature: they expand their investment strategies by simulating the most successful strategies for investors as amended. Because investor prefers higher profits to lower ones he imitates two most successful investment strategies in this industry which bring higher profit than he has. Imitation is imperfect as describing the behavior of the investor with bounded rationality. Changes implemented during simulation ensure diversity in the pool of investor strategies and maintain the evolutionary search in action. The investor chooses the strategy of other investors, based on the characteristics (phenotype), indicating its effective ongoing performance and imitate his investment strategy with variations.

In this model growth of investment can be caused by increasing the efficiency of economic industries by reducing of technological coefficients (which corresponds to decrease in average cost of product in industry) due to technological progress.

The **task** of investor is to determine what share of his fund each investor should invest in the industries of economic system in a current period to get the maximum profit on the assumption of uncertainty as regards the next period.

Since investors do not invest their funds $I_{kj}^{(t)}$ in the industries (if any) with non-positive profitability, the sum of all shares of investment for each investor will be less or equal to one. Not all investment funds invested in the industry can give return in the next period due to imperfect information as regards next year's technological coefficients, therefore planned share of investments $d_{kj}^{(t+1)}$ may differ from the actual one.

Perfect information suggests that technologies are static or investor knows for sure in which direction and what exactly quantitative changes occur in technological dependencies between industries. In general it is unknown. Thus under imperfect information and bounded rationality (inability to process all available information to make rational decisions) each investor must adapt to changes, choosing as a standard the strategy of the most successful investor and, following it, seek to increase his own profit on his funds. This approach is an evolutionary metaphor that is used in evolutionary strategies for economic systems.

Investor adapts his current strategy depending on initial conditions (history). Profit function of each investor demonstrates which of the two strategies will survive and extend: either a leader strategy bringing the highest profit in the current period without changing investment shares, or follower strategy imitating the most successful strategies of other investors and changing his own investment shares.

Rational agent that imitates the strategy of the most successful investor seeks to reach same or greater profit under imperfect information. It leads to imitation by rational agent for the funds share of most successful investor and second best one in the current period. Criterion of strategy efficiency is the largest profit of the investor and the best (close to optimal) investment shares of most successful investors.

Strategies which used to be successful can lead to drop of investor's profit in the next periods. It requires adaptation of investors' strategies to new technological coefficients. To find the optimal investment level under unpredictable dynamics of industries, repeated numerical simulation with variable technology coefficients is applied and investors' profit is measured at the end of each simulation. After determination of highest level of profit we can prepare propositions for investment policy.

Numerical simulation is based on discrete, time-synchronized model, where profits and strategies of each agent are simultaneously updated at fixed time intervals. Each period t corresponds to one financial period (e.g., year). Each step of a simulation is divided into 2 separate renewal operations:

- 1) *economy update* – each agent invests his funds according to his evolutionary strategy taking into account the profitability of industries; profit is calculated for each investor;
- 2) *strategies renewal* – all agents compare their profit with other investors, and then agents simultaneously decide whether to imitate the strategies of other investors to change theirs.

Rules of evolutionary algorithms determine how a specific property (information) about effectiveness of strategies is distributed. These rules determine how participants exchange information in appropriate environments. The rules define the conditions of birth, death or survival of properties (information about directions and share of investment). The depth of influence h_{\max} of history on current investment decisions should also be defined.

To describe the evolutionary algorithm for investors' strategies we apply variables and parameters presented in Table 1.

Table 2. Different approaches of interpretation the category “e-commerce”

Denomination	Variables and parameters
$d_{kj}^{(t+1)}$	capital share of investor k in industry j in period t
n	number of industries (constant)
m	number of investors (constant)
$I_{kj}^{(t)}$	capital of investor k in industry j in period t
$I_k^{(t)}$	total fund of investor k in period t
I_t	total fund of all investors in period t
$A_t = \{a_{ij}^{(t)}\}$	technological coefficients matrix in period t
X_t	gross output of all industries in period t
$X_i^{(t)}$	gross output of industry i in period t
Y	final demand matrix for products of all industries (constant)
Y_i	final demand for product of industry i (constant)
$r_i^{(t)}$	profitability of industry i in period t
$\Pi_{kj}^{(t)}$	profit of investor k from industry j in period t
$S_{kj}^{(t)}$	undisposed fund of investor k that were planned for investment in industry j in period $t-1$ and are transferred as balance in period t
$S_{kj}^{(t)}$	total fund earned by investor k from industry j in period t

4 Algorithm of Investors' Strategies Adaptation to Changes in Technological Interdependence

The algorithm of adaptation of investors' strategies to changes of technological interdependence includes 8 steps (fig. 3):

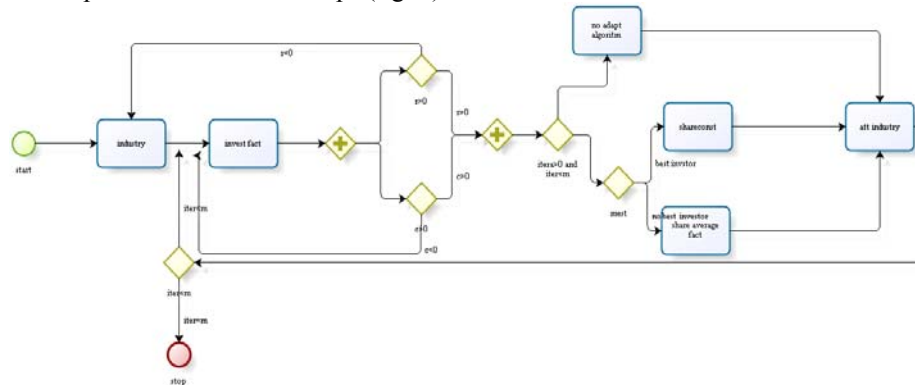


Fig. 3. Algorithm of adaptation of investors' strategies to technological changes

1. Initialization of industries

Economics system includes n industries. To analyze the investment attractiveness of industries in the initial period $t = 0$ we introduce technological coefficient matrix $A_0 = \{a_{ij}^{(0)}\}$, $i, j = 1, \dots, n$, describing the relationship between all industries and the matrix of final demand Y for products of all industries.

2. Calculation of gross output

By means of input-output model the gross output of each industry is calculated as $X_0 = (E - A_0)^{-1} \cdot Y$; and the gross output of the whole economy during the year

$X_0 = \sum_{i=1}^n X_i^{(0)}$ is determined. Further through matrix A_0 we calculate profitability for

each industry i : $r_i^{(0)} = 1 - \sum_{j=1}^n a_{ij}^{(0)}$.

3. Initialization of investors

There are m investors, their number does not change during all simulations. In the initial period, there is egalitarian distribution of all funds between investors, each of them has the same initial money fund $I_k^{(0)} = \frac{X_0}{m}$, $k = 1, \dots, m$ to form his own

investment portfolio. The total fund of all investors $I_0 = \sum_{k=1}^m I_k^{(0)}$ equals gross output

X_0 of the country. Each investor applies one strategy per period, the number of active strategies is the number of investors.

4. Initial investment distribution across industries

Each investor can invest all his fund in one or more industries. The vector of their investment strategies is non-negative. If several investors allocate the same sum, the right to priority investments is determined by a chance.

The richest investor begins investment and invests in the most profitable industry. First investing in most profitable industry the investor may invest no more than 50% of his fund (to prevent his dominance in the most profitable industries). Then second richest investor invests no more than 50% of his funds in this most profitable industry. Once all potential investors fill the most profitable industry *in full*, there comes the turn of second profitability industry to be invested by the investor who is next in size of his funds. Investors do not invest in non-positive profitability industries. The investment process will continue until funds of investors or potential for investment in industries is completed (in period $t = 0$ the second option is not fulfilled by

assumptions of the model). Wherein $I_k^{(0)} = \sum_{j=1}^n I_{kj}^{(0)}$, where $I_{kj}^{(0)}$ is a fund of investor k in industry j in period $t = 0$.

5. Technological changes

In general, changes of technological coefficients are determined by replacement of one vector of technological coefficients by another one. Each vector is formed independently and randomly from a uniform distribution.

Changes in technological coefficients are characteristics of industries that depend on: - unreliable resources, such as biotic resources vulnerable to climate change (forest or fish); - reserves of reliable resources (iron or coal) that are known by the end of the decade or century; - technological changes or changes in human capital [2].

Technological innovations are guided by research and developments, the effects of which on technological coefficients are treated as environment dynamic, to which investors have to adapt. Some industries gradually change with low frequency (mining and manufacturing), others – gradually with high frequency (telecommunications and communications), some others - accidentally with low frequency (hotels and restaurants), the rest - by accident with high frequency (IT industry). Simulation mechanism has one free parameter - technological coefficients which are beyond the control of investors.

Low frequency changes are simulated by transition from one vector of technological coefficient to another every 50 years (Kondratiev long waves) which is characteristic for the industries when changes of production methods are not a driving force of innovation (fisheries). To model high frequency changes transitions begin every 3 years, in response to rapid Juglar business cycle. These rapid changes are characteristics of industries that invest powerfully in research and development (communications and telecommunications). Both Kondratiev wave and Juglar cycle are related with fund dynamics and technological restructuring, as opposed to simple changes in employment represented by a shorter Kitchin cycle [10].

For next period $t=1$ each technological coefficient of matrix $A_0 = \{a_{ij}^{(0)}\}$ will randomly change from interval $p \in [\alpha; \beta]$, where $-1 < \alpha; \beta < 1$. As a result, we obtain matrix $A_1 = \{a_{ij}^{(1)}\}$, where $a_{ij}^{(1)} = (1 + \frac{p}{100}) \cdot a_{ij}^{(0)}$. Under unchangeable finish demand Y we obtain new gross output $X_1 = (E - A_1)^{-1} \cdot Y$. In general case $X_1 \neq X_0$. For investors it means that not all of the funds invested by them in industries in period $t=0$ will bring them profits in period $t=1$.

6. Calculation of investors' profits

For period $t=1$ starting from the first industry investor who invested the largest capital in the industry is determined. The following condition is used to calculate the profit for largest investor (to be specific – the first) in the industry:

$$\begin{aligned} & \text{If } I_{1j}^{(0)} \leq \min(X_j^{(1)}; 0) \text{ (if investment of 1st investor does not exceed the gross} \\ & \text{output of the industry } j) \\ & \text{then } \Pi_{1j}^{(1)} = r_j^{(1)} \cdot I_{1j}^{(0)} \\ & \text{else } \Pi_{1j}^{(1)} = r_j^{(1)} \cdot \min(X_j^{(1)}; 0). \end{aligned} \tag{1}$$

For the next largest investors in the industry ($k \neq 1$):

$$\begin{aligned} &\text{If } I_{kj}^{(0)} \leq \min(X_j^{(1)} - \sum_{h \neq k} I_{hj}^{(0)}; 0) \text{ (if investment of } k\text{-th investor does not} \\ &\text{exceed the gross output decreased by previously invested fund of larger} \\ &\text{investors)} \\ &\text{then } \Pi_{kj}^{(1)} = r_j^{(1)} \cdot I_{kj}^{(0)} \\ &\text{else } \Pi_{kj}^{(1)} = r_j^{(1)} \cdot \min(X_j^{(1)} - \sum_{h \neq k} I_{hj}^{(0)}; 0). \end{aligned} \quad (2)$$

Similarly we calculate profits of the other investors in the rest of industries. The objective function of investor in period $t+1$ of investor $k=1, \dots, m$ for profitability in period t :

$$\Pi_k^{(t+1)} = r_1^{(t+1)} \cdot I_{k1}^{(t)} + r_2^{(t+1)} \cdot I_{k2}^{(t)} + \dots + r_n^{(t+1)} \cdot I_{kn}^{(t)}, \quad k=1, \dots, m. \quad (3)$$

where $I_k^{(t)} = I_{k1}^{(t)} + \dots + I_{kn}^{(t)}$ is exogenous (fixed) investment funds of investor $k=1, \dots, m$ in period t that can bring returns in period $t+1$. Each period investor makes decision on distribution of this fund between various industries of economic system. So $I_{kj}^{(t)}$, $k=1, \dots, m$, $j=1, \dots, n$ is the fund from which the investor makes decision to invest in industry j in period t that can bring returns in the period $t+1$.

7. Determining the most effective investor

Imitating investor strategy changes, while imitated investor strategy (investor with highest profit) does not. The choice of agent strategy to be imitated depends on relative profit. Imitating investor always chooses the investor with the largest current profit. Only if the investor has no peers in profit value, he does not revise his own strategy. Imitation is the only mechanism affecting the change of agents' strategy. There should be *initial state* of income distribution between investors before evolutionary selection strategies. Real imitation is never without errors. Errors are called mutations in evolutionary theory. They are fundamental to the evolutionary process, because they create and maintain diversity in which selection can work [2].

As a result of calculating profit of all investors the investor who received the greatest profit is determined, then the shares of its investment to all industries is

$$\text{calculated: } d_{kj}^{(1)} = \frac{i_{kj}^{(0)}}{\sum_{j=1}^n i_{kj}^{(0)}}, \quad j=1, \dots, n, \text{ where } i_{kj}^{(0)} = I_{kj}^{(0)} \text{ (when fund of investors is less}$$

than industry's potential), or $i_{kj}^{(0)} = \min(X_j^{(1)} - \sum_{h \neq k} I_{hj}^{(0)}; 0)$ (when industry's potential is less than investor's fund, $\sum_{h \neq k} I_{hj}^{(0)}$ - investment of all previous investors, except k). In general shares of investment will equal to:

$$d_{kj}^{(t+1)} = \frac{I_{kj}^{(t)}}{I_k^{(t)}}, \text{ де } \sum_{j=1}^n d_{kj}^{(t+1)} \leq 1, k = 1, \dots, m, j = 1, \dots, n, t = 0, \dots, T. \quad (4)$$

The total investment for all investors in period t is $I^{(t)} = I_1^{(t)} + \dots + I_m^{(t)} = \sum_{k=1}^m I_k^{(t)}$.

In next investment period the most effective investor will not change his strategy, which has brought him the greatest profit. He leaves unchanged his share of investment from all his wealth in the same industries and will distribute the remaining funds (if any) according to stage 4. Wealth of investor is the sum of his investment, profit from his investments and retained fund (if any).

8. Determining evolutionary strategy

The remaining investors with lower profit than the most efficient investor, try to achieve and eventually surpass the best result. To do this they apply the following evolutionary strategy:

- 1) they calculate shares of invested funds for all investors in all industries;
- 2) second largest efficient investor for each industry in which he has invested, defines two most successful investors in this industry by the profit criterion. Then he calculates the average share among these ones:

$$d_{kj}^{t+1} = \frac{d'_{Bj} + d'_{(B-1)j}}{2}, \quad (5)$$

where investors B and $B-1$ received the greatest profit in previous period after investing their funds in industry j . Evolution strategy is arithmetic mean, which also absorbs errors in the decisions of investors.

According to this evolutionary algorithm investor k for the period $t = 2$ invests in industry j sum $I_{kj}^{(1)} = d_{kj}^{(1)} \cdot S_{kj}^{(1)}$, where $S_{kj}^{(1)} = \sum_{j=1}^n (I_{kj}^{(0)} + \Pi_{kj}^{(1)} + s_{kj}^{(0)})$ is total sum of funds available to the investor k , where $I_{kj}^{(0)}$ - investments in industry j , returning to the investor in period $t = 1$, $\Pi_{kj}^{(1)}$ is a profit earned by investor k in period $t = 1$ and $s_{kj}^{(0)}$ - retained investment funds of investor k , for next period $t = 1$.

If only one investor invests in specific industry, then as the second share is taken 0, i.e. evolutionary share of investments will be half the size of previous share. Similarly, the evolutionary algorithm is applied to other industries. This algorithm is repeated for all other investors.

The Evolutionary Algorithm is described below (Table 3).

Table 3. The evolutionary algorithm for improving investment strategy

Algorithm 1 Evolutionary algorithm for investors' strategies	
1	Initialization of industries n
2	Evaluation of the initial gross output X_j and profitability of industries r_i , $i = 1, \dots, n$
3	Initialization of investors m with initial funds $f_j = \frac{X}{m}$, $j = 1, \dots, m$
4	Initial allocation of funds between industries
5	while $I_j > 0$ and $r_i > 0$ do
6	$\min(I_{kj} - \sum_{h \neq k} I_{hj}; X_j \cdot coef)$, $coef = 0.5$ for $t = 1$; $coef = 1$ for $t > 1$
7	Endwhile
8	Profit of investors
9	Change of technological matrix A_t
10	Actual distribution of investors
11	while $I_j > 0$ and $r_i > 0$ and $X_i > 0$ do
12	If $I_{kj} \leq \min(X_j - \sum_{h \neq k} I_{hj}; 0)$
13	then $\Pi_{kj} = r_j \cdot I_{kj}$
14	else $\Pi_{kj} = r_j \cdot \min(X_j - \sum_{h \neq k} I_{hj}; 0)$
15	end while
16	Selection of most profitable investor
17	if $\Pi_B = \max(\Pi_1, \dots, \Pi_m)$
18	then $d_{kj} = const$
19	else $d_{kj} = \frac{d_B + d_{B-1}}{2}$

5 Experimental Results of Desktop Application

Desktop application *Investment Strategy* has been created specifically for simulating of experimental data for macroeconomic input-output model of this paper, using a graphical interface of C# system libraries System.Drawing and System.Windows.Forms. Note that desktop application *Investment Strategy* shows results of two competitive strategies, such as leader, who do not changes his investment shares, and follower, who imitates strategies of two most efficient investors. Results are presented in tabs *Calculation*, *Profit Dynamics*, *Profitability Dynamics*, *Gini Index*. Fig. 4 shows the application window for 100 periods (corresponds to 100 years).

Investor with largest fund has privilege to invest in most profitability industries. If technological coefficient of matrix grow at average (it means that profitability de-

creases), than all investors except first almost have no chance by the means of evolutionary algorithm to be more effective than largest investor (fig. 4). During the experiment, the evolutionary dynamics of investment strategies is described by the following graphs (x-axis – time, ordinate is dynamics index):

1. Industries' profitability variation (after changes of technological coefficients).
2. Profit dynamics of investors to compare two types of strategies: the leader and imitators.
3. Gini coefficient (profit distribution between investors).
4. Dynamics of correlation coefficient between investment and profit of investors.

On fig. 5 for almost all experiment for any number of investors and investment periods all investors except first have no chances to be more effective than greatest one. In this case leader strategy is more effective than evolutionary algorithm for technological changes when average profitabilities of industries rise ($\alpha; \beta > 0$).

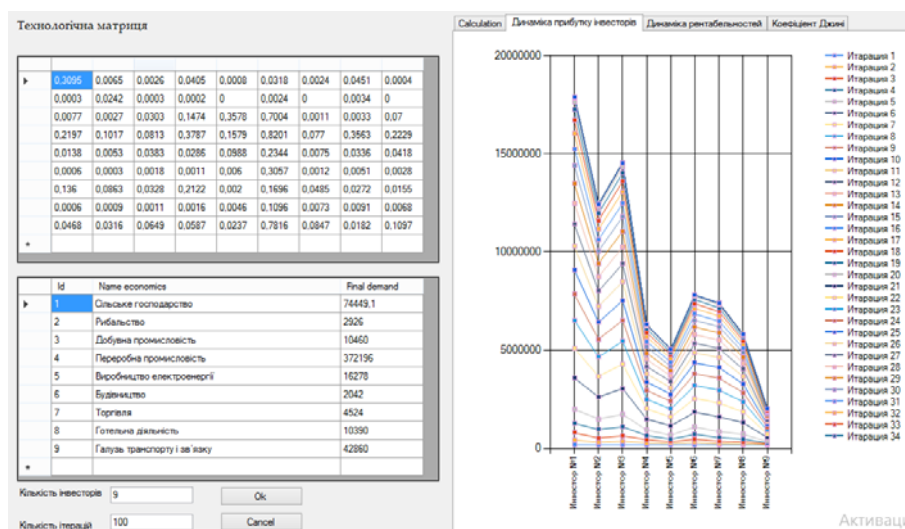


Fig. 4. Desktop application *Investment Strategy*

The more investors the more chances to achieve best results by another investor even technological coefficients increase (fig. 6).

Second investor became leader instead of first investor (fig. 6), that confirms efficiency of evolutionary algorithm for larger number of investors using profitability industry indexes (fig. 7).

On fig. 8 for prevail cases (both small and large number of investors) evolutionary strategies of investors are more effective than unchangeable shares of leader when average profitabilities of industries fall ($\alpha; \beta < 0$).

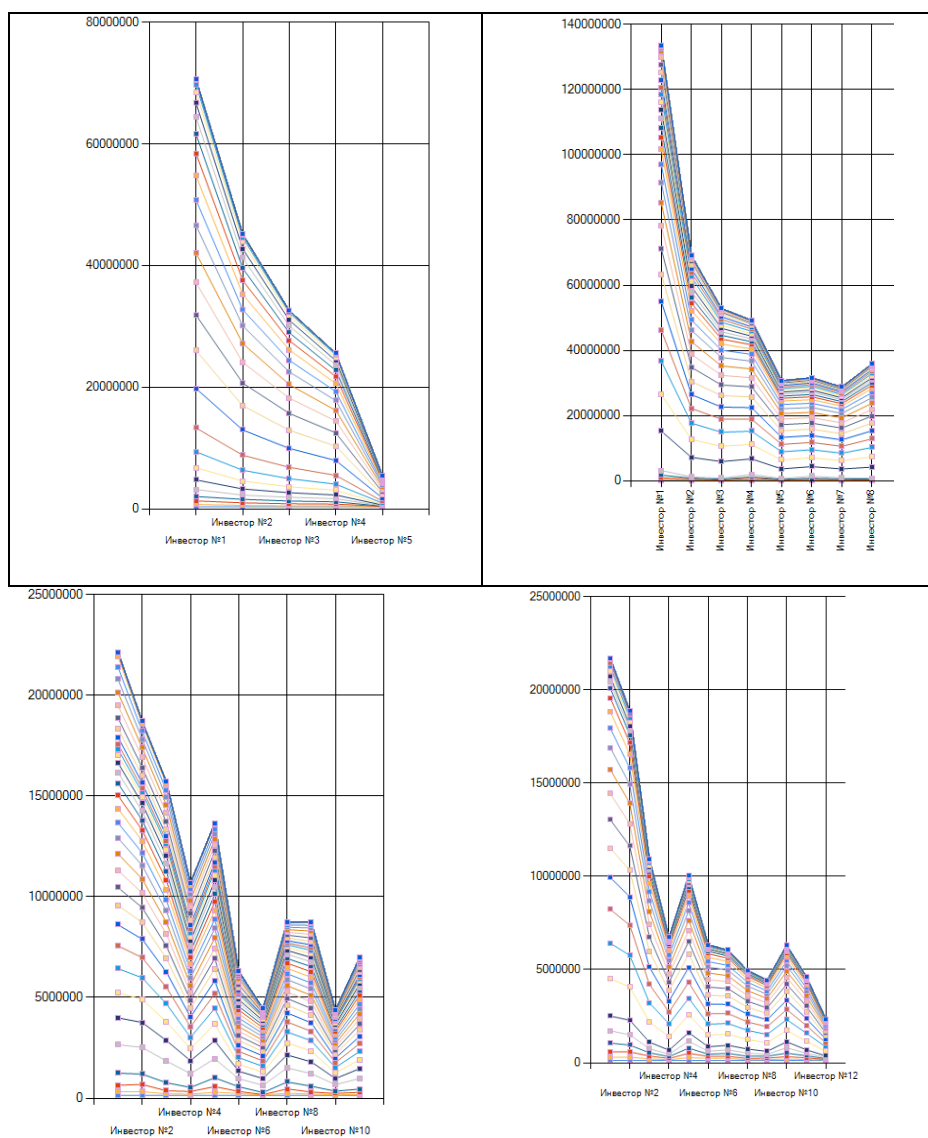


Fig. 5. Effectiveness of leader strategy for 5, 8, 10 and 12 investors when technological coefficients rise

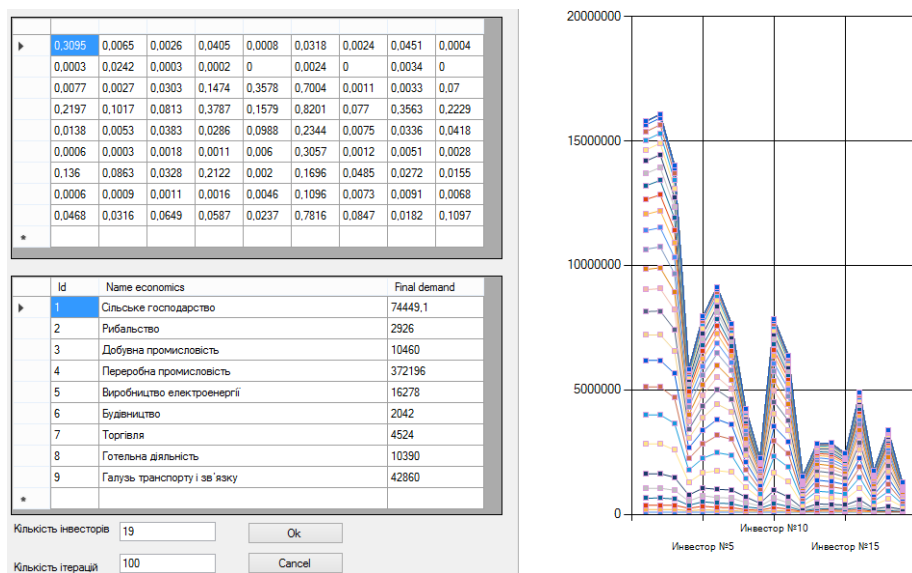


Fig. 6. Effectiveness of evolutionary algorithm for 11 investors

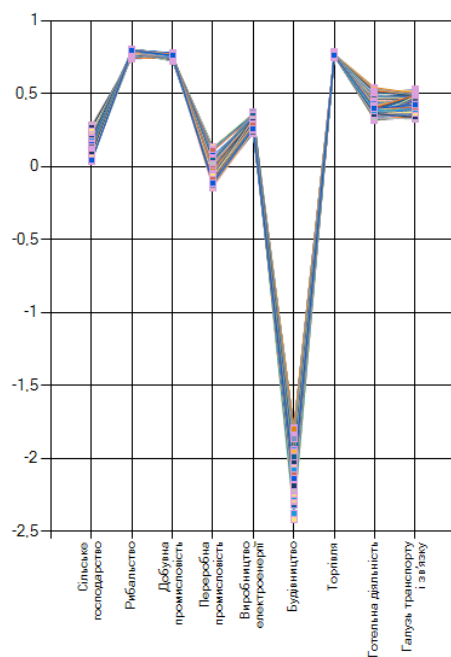


Fig. 7. Dynamics of profitability industry indexes

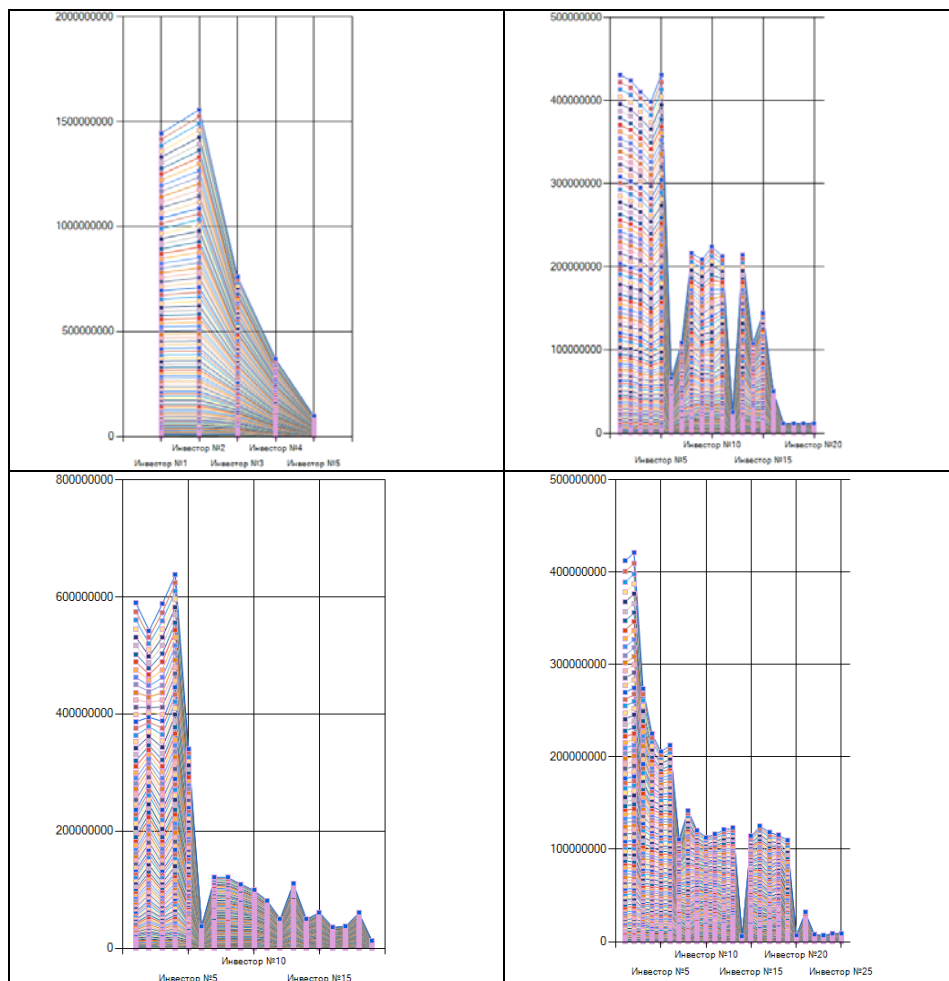


Fig. 8. Effectiveness of evolutionary strategy for 5, 15, 20 and 25 investors when technological coefficients fall

6 Conclusions

Evolutionary investment strategies have been found to adapt distribution of investors' fund to technological changes of economic industries. In this research we use evolutionary algorithm (EA) to improve investment strategies in economic system comparing shares of most profitable investors with other ones. Proposed algorithm selects an optimal combination of investment share in different industries for improving of investment performance. In our experiment we use statistical data from official sources to examine effectiveness of proposed investment strategy. We find the investment

strategies using evolutionary algorithm to teach investors how to increase their profit as much as possible and obtain more profit than investor with best start possibilities.

Desktop C[#] application *Investment Strategy* using a graphical interface with tabs has been created specifically for the numerical investigation of effectiveness of investment strategies using evolutionary algorithm.

As a result of our numerical experiment we have found that under increasing of technological coefficients effectiveness of evolutionary strategy is less than strategy of leader with unchangeable shares. In contrast under decreasing of technological coefficients effectiveness of evolutionary strategy is more than strategy of leader with unchangeable shares. The more investors the more effectiveness of evolutionary algorithm to improve investment strategies for industries of economic system

References

1. Eiben, E.: Introduction to Evolutionary Computing. Computing book in natural computing series (2003). DOI: 10.1007/978-3-662-05094-1
2. Nannen, V., Bergh, J., Eiben, A.E.: Impact of environmental dynamics on economic evolution: A stylized agentbased policy analysis. Technol. Forecast. Soc. Change 80, 329--350 (2013), DOI:10.1016/j.techfore.2012.08.013
3. Kobets, V., Yatsenko, V., Poltoratskiy. Dynamic model of double electronic Vickrey auction. In: Ermolayev, V. et al. (eds.) Proc. 8-th Int. Conf. ICTERI 2012, Kherson May 22, 2012, CEUR-WS.org/Vol-848, ISSN 1613-0073, pp.259--266, online CEUR-WS.org/Vol-848/ICTERI-2012-CEUR-WS-Volume.pdf
4. Gulayeva, N.M.: Evolutionary Algorithms. Bulletin of Taras Shevchenko National University of Kyiv 2, 141--150 (2013).
5. Semenov, M.A., Terkel, D.A.: Analysis of Convergence of an Evolutionary Algorithm with Self-Adaptation using a Stochastic Lyapunov function. Evolutionary Computation 4(11), 363--379 (2003.)
6. Allen, F., Karjalainen, R. Using genetic algorithms to find technical trading rules. Journal of Financial Economics 51, 245--271 (1999).
7. Chen, G., Firth, M., Rui, O.M. The dynamic relation between stock returns, trading volume, and volatility. Financial Review, 36, 153--174 (2001).
8. Maringer, D., Ramtohul, T. Regime-switching recurrent reinforcement learning for investment decision making. Computational Management Science, 9, 89--107 (2012).
9. Kobets, V.: Direct and Indirect Impact Analysis of Ukrainian Industries on Gross Output and Labor Market in Leontief Model. In: Ermolayev, V. et al. (eds.) Proc. 8-th Int. Conf. ICTERI 2012, Kherson May 22, 2012, CEUR-WS.org/Vol-848, ISSN 1613-0073, pp.259--266, online CEUR-WS.org/Vol-848/ICTERI-2012-CEUR-WS-Volume.pdf
10. Zhang, J., Maringer D.: Using a Genetic Algorithm to Improve Recurrent Reinforcement Learning for Equity Trading. Computational Economics 4(47), 551-567 (2016).

Nonlinear Dynamic Model of a Microeconomic System with Different Reciprocity and Expectations Types of Firms: Stability and Bifurcations

Vitaliy Kobets¹ and Alexander Weissblut¹

¹Kherson State University, 27, 40 rokiv Zhovtnya st., Kherson, 73000 Ukraine

vkobets@kse.org.ua, veits@ksu.ks.ua

Abstract. This paper analyzes the dynamic interaction between selfish and reciprocity firms in the market of homogeneous product. The decisions of both types of firms in respect of their output strategies are investigated under naive, adaptive and generalized expectations. The standard postulate for competitive firms' model has been extended by the assumption that there is a share of reciprocity firms which, unlike selfish firms, maximize both private and social benefits as consumer surplus. It has been proved that the unique Nash equilibrium is stable for all affordable values of parameters in the model with adaptive expectations, and is unstable for the model with naive expectations at sufficiently large number of firms in the market. A special desktop application has been created for animation of model trajectory and demonstration of stable quantity trajectories and bifurcation diagrams of firms' output. Naive expectations of two-thirds of firms result in a state of dynamic chaos in the market leading to degeneration of the existing competition model between the two types of firms. The crucial factor which ensures the stable equilibrium in the market and the ability to predict firms' output is the adaptive approach which takes into account the adaptive expectations of firms planning their product quantity.

Keywords: microeconomic system, reciprocity, naïve expectations, adaptive expectations, consumer surplus, stability, bifurcation.

Key Terms: DynamicSystem, DesktopApplication, NashEquilibrium, Expectations

1 Introduction

In recent years the researchers are renouncing the assumption of perfect rationality as unconditional basis of economic agents' behavior. The neoclassical 'rational man' does not exist in reality; economic agents act according to established rules, without being fully informed and maximizing their own utility [1].

Karl Polanyi identified the alternative economic organization where social norms are not generated by economic self-interest of the individual. This network of reciprocal relations is based on mutual economic cooperation of efforts and resources between the members of non-economic network, dominated by cultural norms rather than market laws. Under reciprocity relations the exchange donor and recipient can be transposed. So this is a symmetrical relationship of gifts exchange between members of horizontal social networks [2].

This relationship is not regulated by formal institutions but based on informal commitments giving moral right to mutual help and reciprocal exchange on sustainable basis in the long run period. But this is a relationship with minimal risk for participants and the penalty of loss of social capital (reputation and trust) and social isolation. Society supports the stability of the exchange to ensure their survival during crises and wars. Reciprocity is not altruism which does not create reciprocal obligations in quantitative, qualitative or time respects, just vague commitment (e.g., you give, if you can).

Actually, reciprocity relations, commodity exchange and hierarchical subordination exist at the same time. But it is reciprocity that underlies most decentralized corrections of diverse shortcomings and failures of markets and firms. These relations are long-run factors of economic efficiency; they set most social obligations of firms towards individuals without government intervention. No society can exist without reciprocal relationship [3].

Reciprocity or social responsibility implies that the firms not only pursue their selfish goal of increasing profits, but are also ready to sacrifice some of their own profits for the benefit of consumers without direct compensation for it by the state [1]. Such targets can be stipulated by the firms' desire to get stable profits in the long run rather than maximal short-run profits. Such forward-thinking firms-reciprocators are considered in the model of this paper. Their objective function is a weighted average of the profits and consumer surplus of their market segment.

The real economic processes make a clear demonstration that neoclassical "rational man" is not their subject. In real economy "optimal imperfect decisions" are taken by simple and non-expensive calculations, well adapted to frequent repetitions, to evolution; it is more efficient for perfectly rational firm to perform multiple experiments with quantity to estimate the demand function rather than search for nonrecurrent, instantaneous achieving of equilibrium. New paradigm of nonlinear economics is a mix of qualitative theory of nonlinear dynamical system, optimal control theory, game theory, and theory of stochastic processes [4, 5].

The evolutionary approach and analysis of the dynamics allow to explain why one type of firm ousts another from the market, why sometimes the economic system is stable, but in other cases is unstable [6, 7]. If the system has multiple equilibria, the dynamics and evolution is the selection mechanism of best equilibrium according to certain criteria [8]. The evolutionary process is analogous to social learning. An example of its application is the pricing mechanisms for auctions that occur in agents social networks, e-commerce and trade through the Internet [9, 10].

The study of the evolution of the markets with the strategic interaction usually uses the following assumptions: (a) two firms or two types of similar firms operate in the market; (b) firms produce homogeneous goods in quantities of $x_1(t)$ and $x_2(t)$; (c)

no firm knows the rivals' quantities; (d) the firms seek to predict the output of the competitors using the adaptive scheme.

Planning of quantity for the following period firms resolve optimization problem: $Max \Pi_i(x_i; x_j^e(t+1))$, where Π_i is the objective function of firm i , $x_j^e(t+1)$ - expected quantity of a competitor j ($i, j = 1, 2$).

Examples of bounded rationality of firms are: ignoring the impact of competitors' actions on their own output (local monopoly approximation LMA), naive expectation (assumption of unchangeable behavior of competitors for a long time and using $x_j(t)$ instead $x_j^e(t+1)$). [8] Of course both decision making approaches (adaptive and naive, bounded rational) coexist in the market with a certain probability.

Analysis of nonlinear oligopoly with heterogeneous players reveals that a higher degree of product differentiation may destabilize the Cournot-Nash equilibrium. Authors showed that a cascade of flip bifurcation may lead to periodic cycles and chaotic motions [11]. Stability conditions of Nash equilibrium and complex dynamics are also studied for heterogeneous duopoly with isoelastic demand function. For such heterogeneous players a cascade of flip bifurcation leads to periodic cycles and chaos and the Neimark-Sacker bifurcation generates attractive invariant closed curve [12].

Such scheme serves as the basis for mathematical model of this paper, which distinguishes from the other models in that: (a) firms use more than one way of decision-making, and combine different ones; (b) except their own selfish interests, firms take into account social ones.

The **paper goal** is to consider the impact of naive, adaptive and generalized expectations of egoist and reciprocator firms on stability of equilibrium and the conditions of transition to dynamic chaos in the numerical experiment using a specially designed desktop application.

The paper is organized as follows: part 2 describes two-dimensional market model with naïve and adaptive expectations; part 3 is devoted to dynamics model in general case; part 4 demonstrates C#-application model for numerical investigation; part 5 concludes.

2 Two-Dimensional Market Model

We consider the market of homogeneous product, which consists of n firms, including k identical firms-reciprocators, each of them producing x units of product and $n - k$ identical selfish firms, each of them producing y units of product. Thus, the industry quantity of the two types of firms is $Q = k \cdot x + (n - k) \cdot y$. Product price P in the market is given by the inverse market demand function $P = P(Q) = \frac{b}{Q}$ ($b > 0$).

The objective function of a firm-egoist is profit $\pi_y = (P - v) \cdot y$, where v is the firm's costs per unit in the market. Firm-reciprocator maximizes both its own profit $\pi_x = (P - v) \cdot x$ and consumer surplus CS of its own market segment:

$CS = \frac{\gamma}{k} \left(\int_{\varepsilon}^Q P(q) dq - PQ \right)$, where γ is the parameter defining the segment of the market, which the reciprocator firm believes its own and optimizes it ($0 < \gamma \leq k$), ε is the minimal technologically possible quantity of product. Then

$$CS = \frac{\gamma}{k} \left(b \cdot \ln \left(\frac{Q}{\varepsilon} \right) - \frac{b}{Q} \cdot Q \right) = \frac{b\gamma}{k} \left(\ln \left(\frac{Q}{\varepsilon} \right) - 1 \right) = \frac{b\gamma}{k} \ln \left(\frac{Q}{\varepsilon} \right),$$

where $\hat{\varepsilon} = \varepsilon e$. The specific choice of ε does not affect the dynamics of the model because the objective function, as any potential, is set up to an arbitrary constant accuracy, so further we will write ε instead of $\hat{\varepsilon}$. Then the objective function of firm-reciprocator is:

$$\Pi_X = \alpha(P - v)x + (1 - \alpha)CS = \alpha(P - v)x + (1 - \alpha) \frac{b\gamma}{k} \ln \left(\frac{Q}{\varepsilon} \right),$$

where α is share of private interest PI (reciprocator's profit), $1 - \alpha$ is share of social interest (responsibility) SR (consumer surplus from its own market segment) in the objective function.

2.1 Dynamics Model Equations with Naive Expectations

Consider the dynamics of this two-dimensional model in discrete time $t = 0, 1, \dots$; where x_t, y_t are the outputs at time t of reciprocator and egoist firm, respectively. On the basis of these values at time t each firm finds the optimal value for its own production quantity in the next moment $t+1$, maximizing its objective function. It distinguishes this model, in which the firm responds to changes in output of both their and other types of firms from traditional competition models, where one type of firm responds to changes in other types only. So each selfish firm is looking for such value of y_{t+1} at which it maximizes its own profits, suggesting that SR firms and the other $n - k - 1$ PI firms leave their quantities unchanged:

$$\pi_Y = \left(\frac{b}{y_{t+1} + k \cdot x_t + (n - k - 1) \cdot y_t} - v \right) \cdot y_{t+1}. \quad (1)$$

Obviously, the maximum point for y_{t+1} is found from the condition $\frac{\partial \pi}{\partial y_{t+1}} = 0$, whence:

$$v(y_{t+1} + kx_t + (n - k - 1)y_t)^2 = b(kx_t + (n - k - 1)y_t). \quad (2)$$

From equation (2) we obtain the response function of the PI firm:

$$y_{t+1} = \sqrt{\frac{b}{v}(kx_t + (n - k - 1)y_t - kx_t - (n - k - 1)y_t)}. \quad (3)$$

Similarly, firm-reciprocator finds such value of x_{t+1} at which the maximum value of its objective function is:

$$\Pi_X = \alpha \cdot \left(\frac{b}{x_{t+1} + (k - 1)x_t + (n - k) \cdot y_t} \cdot x_{t+1} - vx_{t+1} \right) + (1 - \alpha) \cdot \frac{b\gamma}{k} \ln \left(\frac{x_{t+1} + (k - 1)x_t + (n - k)y_t}{\varepsilon} \right) \quad (4)$$

Here the maximum point for x_{t+1} is found from the condition $\frac{\partial \Pi_x}{\partial x_{t+1}} = 0$:

$$\frac{\partial \Pi_x}{\partial x_{t+1}} = \alpha \left(\frac{b(x_{t+1} + (k-1)x_t + (n-k)y_t) - bx_{t+1}}{(x_{t+1} + (k-1)x_t + (n-k)y_t)^2} x_{t+1} - v \right) + (1-\alpha) \cdot \frac{b\gamma}{k} \cdot \frac{1}{x_{t+1} + (k-1)x_t + (n-k)y_t} = 0$$

Further, without loss of generality, we assume here $\gamma = 1$, otherwise we redefine the

share of profit as $\hat{\alpha} = \frac{\alpha}{\alpha + (1-\alpha)\gamma}$. Then

$$v(x_{t+1} + (k-1)x_t + (n-k)y_t)^2 = b((k-1)x_t + (n-k)y_t) + \frac{1-\alpha}{\alpha} \frac{b}{k} (x_{t+1} + (k-1)x_t + (n-k)y_t) \quad (5)$$

Assuming $z = x_{t+1} + (k-1)x_t + (n-k)y_t$, we present (5) as:

$$z^2 = \frac{b}{v} ((k-1)x_t + (n-k)y_t) + \frac{1-\alpha}{\alpha} \frac{b}{vk} z.$$

Hence, in view of (3), we obtain a system of dynamics equations of this model:

$$\begin{cases} x_{t+1} = \sqrt{\frac{b}{v} ((k-1)x_t + (n-k)y_t) + \left(\frac{1-\alpha}{2} \frac{b}{\alpha vk} \right)^2} - (k-1)x_t - (n-k)y_t + \frac{1-\alpha}{2} \frac{b}{\alpha vk}, \\ y_{t+1} = \sqrt{\frac{b}{v} (kx_t + (n-k-1)y_t - kx_t - (n-k-1)y_t)}. \end{cases} \quad (6)$$

2.2 Equilibrium Conditions for the Model with Naive Expectations

In the Nash equilibrium point $x_{t+1}=x_t=x$, $y_{t+1}=y_t=y$ for all $t = 0, 1, \dots$. Therefore, at this point, by (2) and (5) we obtain:

$$(kx + (n-k)y)^2 = \frac{b}{v} (kx + (n-k-1)y) = \frac{b}{v} ((k-1)x + (n-k)y) + \frac{1-\alpha}{\alpha} \frac{b}{vk} (kx + (n-k)y) \quad (7)$$

From the last equation we obtain $x - y = \frac{1-\alpha}{\alpha} (x + \frac{n-k}{k} y)$, whence it follows that

$\frac{2\alpha-1}{\alpha} x = (1 + \frac{1-\alpha}{\alpha} \frac{n-k}{k}) \cdot y$, i.e. the response functions of both types of firms are respectively:

$$x = \frac{\alpha k + (1-\alpha) \cdot (n-k)}{(2\alpha-1) \cdot k} \cdot y, \quad y = \frac{(2\alpha-1) \cdot k}{\alpha k + (1-\alpha) \cdot (n-k)} \cdot x \quad (8)$$

To calculate the coordinates of a fixed point, we substitute the expression of y through x in the first equation (7). Thus the following proposition is proved.

Proposition 1. There is unique Nash equilibrium point in a dynamic system (6):

$$\begin{cases} x^* = \frac{b}{vn} \left(1 - \frac{2\alpha-1}{\alpha n} \right) \left(1 + \frac{1-\alpha}{\alpha} \frac{n-k}{k} \right), \\ y^* = \frac{b}{vn} \cdot \frac{2\alpha-1}{\alpha} \cdot \left(1 - \frac{2\alpha-1}{\alpha n} \right). \end{cases} \quad (9)$$

However, is this point stable?

Proposition 2. For any set $b, v > 0$ and α ($0 \leq \alpha \leq 1$) Nash equilibrium (9) is unstable for sufficiently large number of firms n if $\left| \frac{k}{n} \right| > \varepsilon$ and $\left| \frac{k}{n} - \frac{3}{4} \right| > \varepsilon$ for any $\varepsilon > 0$.

The destabilizing role of number of players n is well known for the evolution of firms' strategies in oligopoly games [8]. However, in this case, according to calculations, point (9) is unstable even at $n \geq 5$.

Proof. We show that in dynamic system (6) at Nash equilibrium point (9) modulus of Jacobian J is greater than 1: $|\det J| > 1$. This implies that at least one eigenvalue of the Jacobian is greater than 1 in absolute value, which means instability of the fixed point (9). Here, the Jacobian of the system (6):

$$J = \begin{pmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{pmatrix} = \begin{pmatrix} \frac{\partial x_{t+1}}{\partial x_t} & \frac{\partial x_{t+1}}{\partial y_t} \\ \frac{\partial y_{t+1}}{\partial x_t} & \frac{\partial y_{t+1}}{\partial y_t} \end{pmatrix}.$$

$$\begin{aligned} J_{xx} &= \frac{\frac{b}{v}(k-1)}{2\sqrt{\frac{b}{v}((k-1)x_t + (n-k)y_t + d^2)}} - (k-1), & J_{xy} &= \frac{\frac{b}{v}(n-k)}{2\sqrt{\frac{b}{v}((k-1)x_t + (n-k)y_t + d^2)}} - (n-k), \\ J_{yx} &= \frac{\frac{b}{v}k}{2\sqrt{\frac{b}{v}(kx_t + (n-k)y_t)}} - k, & J_{yy} &= \frac{\frac{b}{v}(n-k-1)}{2\sqrt{\frac{b}{v}((k-1)x_t + (n-k-1)y_t)}} - (n-k-1), \end{aligned}$$

where $d = \frac{1-\alpha}{2} \frac{b}{\alpha vk}$, then $\det J = J_{xx} \cdot J_{yy} - J_{xy} \cdot J_{yx} =$

$$= (1-n) \cdot \left(\frac{\frac{b}{v}}{2\sqrt{\frac{b}{v}((k-1)x_t + (n-k)y_t + d^2)}} - 1 \right) \cdot \left(\frac{\frac{b}{v}}{2\sqrt{\frac{b}{v}((k-1)x_t + (n-k-1)y_t)}} - 1 \right).$$

But for point (9) in the denominator $\frac{b}{v}((k-1)x^* + (n-k-1)y^*) =$

$$= \left(\frac{b}{v}\right)^2 \left[\frac{1-\alpha}{\alpha} \left(1 - \frac{k}{n}\right) + \frac{2\alpha-1}{\alpha} \left(1 - \frac{k}{n}\right) \right] + o\left(\frac{1}{n}\right) = \left(\frac{b}{v}\right)^2 \cdot \left(1 - \frac{k}{n}\right) + o\left(\frac{1}{n}\right),$$

where $o\left(\frac{1}{n}\right) \rightarrow 0$ for $n \rightarrow \infty$. Similarly, we obtain for the second denominator:

$$\frac{b}{v} ((k-1)x^* + (n-k)y^*) + d^2 = \left(\frac{b}{v}\right)^2 \cdot \left(1 - \frac{k}{n}\right) + o\left(\frac{1}{n}\right).$$

But by the data $\left| \frac{k}{n} - \frac{3}{4} \right| > \varepsilon$ at a certain $\varepsilon > 0$, which guarantees that the factors

$$\frac{\frac{b}{v}}{2\sqrt{\frac{b}{v}((k-1)x_t + (n-k)y_t + d^2)}} - 1 \quad \text{and} \quad \frac{\frac{b}{v}}{2\sqrt{\frac{b}{v}((k-1)x_t + (n-k-1)y_t}} - 1$$

do not equal zero for all possible $n, k, b, v > 0$ and α ($0 \leq \alpha \leq 1$), Q.E.D.

2.3 Dynamic Model Equations with Adaptive Expectations

Since all selfish firms are assumed as identical, it is natural to suggest that they have the same planning at moment t , so their production quantities y_{t+1} at moment $t+1$ will be equal too. Given these expectations, each selfish firm is looking for such value y_{t+1} at which it obtains the highest profit, suggesting that production quantity of *SR* firms will remain unchanged:

$$\pi_Y = \left(\frac{b}{kx_t + (n-k)y_{t+1}} - v \right) \cdot y_{t+1}. \quad (10)$$

Obviously, the maximum point for y_{t+1} is found from the condition $\frac{\partial \pi_Y}{\partial y_{t+1}} = 0$, which gives us:

$$(kx_t + (n-k)y_{t+1})^2 = \frac{b}{v} kx_t. \quad (11)$$

Then $kx_t + (n-k)y_{t+1} = \sqrt{\frac{b}{v} kx_t}$, from here response function of *PI* firms is:

$$(n-k)y_{t+1} = \sqrt{\frac{b}{v} kx_t} - kx_t. \quad (12)$$

Similarly, firm-reciprocator naturally expects that the quantity of production of all these firms at moment $t+1$ would be the same. Based on this expectation, each firm-reciprocator finds the value of x_{t+1} at which the objective function is maximal, assuming that the output of *PI* firms does not change:

$$\Pi_X = \alpha \left(\frac{b}{kx_{t+1} + (n-k)y_t} x_{t+1} - vx_{t+1} \right) + (1-\alpha) \frac{b\gamma}{k} \ln \left(\frac{kx_{t+1} + (n-k)y_t}{\varepsilon} \right). \quad (13)$$

Here we can find the maximum point for x_{t+1} from the condition $\frac{\partial \Pi_x}{\partial x_{t+1}} = 0$, hereof:

$$(kx_{t+1} + (n-k)y_t)^2 = \frac{b}{v}(n-k)y_t + \frac{b\gamma}{v} \frac{1-\alpha}{\alpha} \cdot (kx_{t+1} + (n-k)y_t). \quad (14)$$

Let $z = kx_{t+1} + (n-k)y_t$, represent (14) as:

$$\left(z - \frac{1}{2} \frac{b\gamma}{v} \frac{1-\alpha}{\alpha}\right)^2 = \frac{b}{v}(n-k)y_t + \left(\frac{1}{2} \frac{b\gamma}{v} \frac{1-\alpha}{\alpha}\right)^2.$$

Hence $z - \frac{1}{2} \frac{b\gamma}{v} \frac{1-\alpha}{\alpha} = \sqrt{\frac{b}{v}(n-k)y_t + \left(\frac{1}{2} \frac{b\gamma}{v} \frac{1-\alpha}{\alpha}\right)^2}$. Thus, in view of (12), we obtain a system of dynamics equations of the model, taking into account the forecast:

$$\begin{cases} kx_{t+1} = \sqrt{\frac{b}{v}(n-k)y_t + \left(\frac{1}{2} \frac{1-\alpha}{\alpha} \frac{b\gamma}{v}\right)^2} - (n-k)y_t + \frac{1}{2} \frac{1-\alpha}{\alpha} \frac{b\gamma}{v}, \\ (n-k)y_{t+1} = \sqrt{\frac{b}{v}kx_t - kx_t}. \end{cases} \quad (15)$$

2.4 Equilibrium Conditions for the Model with Adaptive Expectations

In the Nash equilibrium point $x_{t+1} = x_t = x$, $y_{t+1} = y_t = y$ for all $t = 0, 1, \dots$. Therefore, at this point in view of (11) and (14) we get:

$$(kx + (n-k)y)^2 = \frac{b}{v}kx = \frac{b}{v}(n-k)y + \frac{b\gamma}{v} \frac{1-\alpha}{\alpha} \cdot (kx + (n-k)y). \quad (16)$$

From the second equation we get $x - \frac{1-\alpha}{\alpha} \gamma x = \frac{n-k}{k} \cdot \left(1 + \frac{1-\alpha}{\alpha} \gamma\right) y$, whence response functions for selfish and reciprocator firms are, respectively:

$$y = \frac{k}{n-k} \frac{\alpha - (1-\alpha)\gamma}{\alpha + (1-\alpha)\gamma} \cdot x \quad x = \frac{n-k}{k} \frac{\alpha + (1-\alpha)\gamma}{\alpha - (1-\alpha)\gamma} \cdot y \quad (17)$$

To calculate the coordinates of the fixed point, we substitute this expression y in terms of x at first equation (16):

$$\left(kx + (n-k) \frac{k}{n-k} \frac{\alpha - (1-\alpha)\gamma}{\alpha + (1-\alpha)\gamma} \cdot x\right)^2 = \frac{b}{v}kx \quad (kx)^2 \cdot \left(\frac{\alpha - (1-\alpha)\gamma}{\alpha + (1-\alpha)\gamma} + 1\right) = \frac{b}{v}kx$$

Hence, we obtain:

Proposition 3. There is unique Nash equilibrium point in the dynamic system (15) with adaptive expectations:

$$\begin{cases} x^* = \frac{b}{vk} \left(\frac{\alpha + (1-\alpha)\gamma}{2\alpha}\right)^2, \\ y^* = \frac{b}{v(n-k)} \frac{\alpha^2 - ((1-\alpha)\gamma)^2}{(2\alpha)^2}. \end{cases} \quad (18a)$$

As before, without loss of generality, let $\gamma = 1$, otherwise we can override the share of profit as $\hat{\alpha} = \frac{\alpha}{\alpha + (1-\alpha)\gamma}$. At $\gamma = 1$ system (18) takes the form:

$$\begin{cases} x^* = \frac{b}{vk} \left(\frac{1}{2\alpha} \right)^2, \\ y^* = \frac{b}{v(n-k)} \frac{2\alpha-1}{(2\alpha)^2} = \frac{b}{v(n-k)} \frac{1}{2\alpha} \left(1 - \frac{1}{2\alpha} \right). \end{cases} \quad (18b)$$

Proposition 4. The equilibrium point (18) is stable for all possible values of the parameters.

Proof. To prove the stability of dynamic system (15) in Nash equilibrium point (18) it is necessary and sufficient to demonstrate that for Jacobian J of this system in (18) the following conditions named after Shur were satisfied:

$$\begin{cases} 1 + tr J + \det J > 0, \\ 1 - tr J + \det J > 0, \\ 1 - \det J > 0. \end{cases}$$

Here, the Jacobian of system (15)

$$J = \begin{pmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{pmatrix} = \begin{pmatrix} \frac{\partial x_{t+1}}{\partial x_t} & \frac{\partial x_{t+1}}{\partial y_t} \\ \frac{\partial y_{t+1}}{\partial x_t} & \frac{\partial y_{t+1}}{\partial y_t} \end{pmatrix},$$

obviously, $J_{xx} = J_{yy} = 0$ where $tr J = J_{xx} + J_{yy} = 0$. Thus, to test Shur conditions it is

sufficient to establish that $\det J < 1$. But at point (18) $y^* = \frac{b}{v(n-k)} \frac{\alpha^2 - ((1-\alpha)\gamma)^2}{(2\alpha)^2}$

and therefore

$$kJ_{xy} = \frac{\frac{b}{v}(n-k)}{2\sqrt{\frac{b}{v}(n-k)y^* + d^2}} - (n-k) = \frac{n-k}{2\sqrt{\frac{\alpha^2}{4\alpha^2}}} - (n-k) = 0$$

Consequently, $\det J = J_{xx} \cdot J_{yy} - J_{xy} \cdot J_{yx} = 0$, Q.E.D.

The price of product P in the market is given by the inverse market demand function $P = P(Q) = \frac{b}{Q}$ ($b > 0$), and the price is not less than a cent, i.e. $P \geq 0.01$.

Therefore, the product quantity of each firm-reciprocator is $x \leq \frac{100b}{k}$. Similarly, the

product quantity of each selfish firm is $y \leq \frac{100b}{n-k}$.

Corollary. The trajectories of the dynamical system (15) converge to a Nash equilibrium (18) for any initial values $x_0 \leq \frac{100b}{k}$, $y_0 \leq \frac{100b}{n-k}$.

3 Dynamic Model Equations in a General Case

Suppose that in planning under the given market model adaptive expectations are used with probability p , naïve ones - with probability $q = 1 - p$. Then the profit function for a typical (representative) firm-egoist has the form:

$$\pi_Y = \left(\frac{b}{y_{t+1} + kx_t + p(n-k-1)y_{t+1} + q(n-k-1)y_t} - v \right) \cdot y_{t+1}, \quad (19)$$

and the objective function for the representative firm-reciprocator

$$\begin{aligned} \Pi_X = & \alpha \cdot \left(\frac{b}{x_{t+1} + p(k-1)x_{t+1} + q(k-1)x_t + (n-k)y_t} x_{t+1} - vx_{t+1} \right) +, \\ & + (1-\alpha) \frac{b\gamma}{k} \ln \left(\frac{x_{t+1} + p(k-1)x_{t+1} + q(k-1)x_t + (n-k)y_t}{\varepsilon} \right). \end{aligned} \quad (20)$$

Obviously, for $p = 0$ ($q = 1$) objective functions π_Y and Π_X are consistent with the results of naïve model (1) and (4), for $p = 1$ ($q = 0$), they are consistent with the results of the adaptive model (10) and (13) respectively. Let us assume

$$z_{\pi_X} = y_{t+1} + kx_t + (n-k-1)(py_{t+1} + qy_t) \quad z_{\Pi_X} = x_{t+1} + (n-k)y_t + (k-1)(px_{t+1} + qx_t)$$

$$\text{In this notation } \pi_Y = \left(\frac{b}{z_{\pi_X}} - v \right) \cdot y_{t+1}; \quad \Pi_X = \alpha \cdot \left(\frac{b}{z_{\Pi_X}} x_{t+1} - vx_{t+1} \right) + (1-\alpha) \frac{b\gamma}{k} \ln z_{\Pi_X}.$$

Then the point y_{t+1} of maximum profit function π_X is found from the condition

$$\frac{\partial \pi_X}{\partial y_{t+1}} = 0, \text{ here}$$

$$z_{\pi_X}^2 = \frac{b}{v} (kx_t + q(n-k-1)y_t) \quad (21)$$

whence

$$y_{t+1} \cdot (1 + p(n-k-1)) = \sqrt{\frac{b}{v} (kx_t + q(n-k-1)y_t - k \cdot x_{t+1} - (n-k-1) \cdot q \cdot y_t)} \quad (22)$$

The maximum point x_{t+1} for the objective function Π_X is found from the first order condition $\frac{\partial \Pi_X}{\partial x_{t+1}} = 0$. Forth without loss of generality we assume here $\gamma = 1$,

otherwise as above we redefine the share of profit as $\tilde{\alpha} = \frac{\alpha}{\alpha + (1-\alpha)\gamma}$. Then

$$z_{\Pi_X}^2 = \frac{b}{v} \cdot ((n-k) \cdot y_t + (k-1) \cdot qx_t) + \frac{1-\alpha}{\alpha} \frac{b(1+p(k-1))}{vk} z_{\Pi_X} \quad (23)$$

Thus, in view of (22), we obtain the dynamics model of equations system of in the general case:

$$\begin{cases} (1+p(k-1))x_{t+1} = \sqrt{\frac{b}{v}w_x + d^2} - w_x + d, \\ (1+p(n-k-1))y_{t+1} = \sqrt{\frac{b}{v}w_y} - w_y, \end{cases} \quad (24)$$

where $\begin{cases} w_x = q(k-1)x_t + (n-k)y_t, \\ w_y = kx_t + q(n-k-1)y_t. \end{cases}$

3.1 Equilibrium Conditions in a General Case

Since Nash equilibrium point is $x_{t+1}=x_t=x$, $y_{t+1}=y_t=y$ for all $t=0,1,\dots$, then at this point in view of (21) and (23) we obtain:

$$\begin{aligned} z_{\pi_y} = z_{\pi_x} = (kx + (n-k)y)^2 &= \frac{b}{v}(kx + q(n-k-1)y) = \frac{b}{v}((k-1)qx + (n-k)y) + \\ &+ \frac{1-\alpha}{\alpha} \frac{b(1+p(k-1))}{vk} \cdot (kx + (n-k) \cdot y) \end{aligned} \quad (25)$$

From the second equation we get:

$$y \left[(n-k) + \frac{1-\alpha}{\alpha} \frac{1+p(k-1)}{k} (n-k) - q(n-k-1) \right] = x \left[k - (k-1)q - \frac{1-\alpha}{\alpha} \frac{1+p(k-1)}{k} k \right]$$

Thus,

$$y \left[p \frac{n-k}{\alpha} + q \left(1 + \frac{1-\alpha}{\alpha} \cdot \frac{n-k}{k} \right) \right] = x \left[(1+p(k-1)) \cdot \frac{2\alpha-1}{\alpha} \right],$$

where response function in this case

$$\frac{x}{y} = G = \frac{p(n-k) + q(\alpha + (1-\alpha) \frac{n-k}{k})}{(2\alpha-1)(1+p(k-1))}. \quad (26)$$

To calculate the coordinates of the fixed point we substitute from (26) expression for $\frac{x}{y}$ in the first equation of (25) $y^2(kG + (n-k))^2 = \frac{b}{v}y(kG + q(n-k-1))$. Hence

Proposition 5. There is unique Nash equilibrium point in a general dynamical system (24):

$$y^* = \frac{\frac{b}{v}(kG + q(n-k-1))}{(kG + (n-k))^2} \quad x^* = Gy^* = \frac{\frac{b}{v}(k + q(1/G)(n-k-1))}{(k + (1/G)(n-k))^2} \quad (27)$$

where the function $G = G(p, q, n, k, \alpha)$ is given in (26).

Proposition 6. For $p=0$ ($q=1$) the equilibrium point (x^*, y^*) coincides with point (9) of a dynamic system with naive expectations. When $p=1$ ($q=0$) the equilibrium point coincides with point (18b) of the dynamic system with adaptive expectations.

4 C[#] - Application Model for Numerical Investigation

C[#] window application *Model* has been created specifically for the numerical investigation of the model of this paper, using a graphical interface of C[#] system libraries System.Drawing and System.Windows.Forms. Note that all the calculations associated with the model, are localized in the method *calc* of the application *Model* that makes it easy to modify the equations of the model and use the *Model* to study the other two-dimensional dynamical systems. Fig. 1 shows the application window.

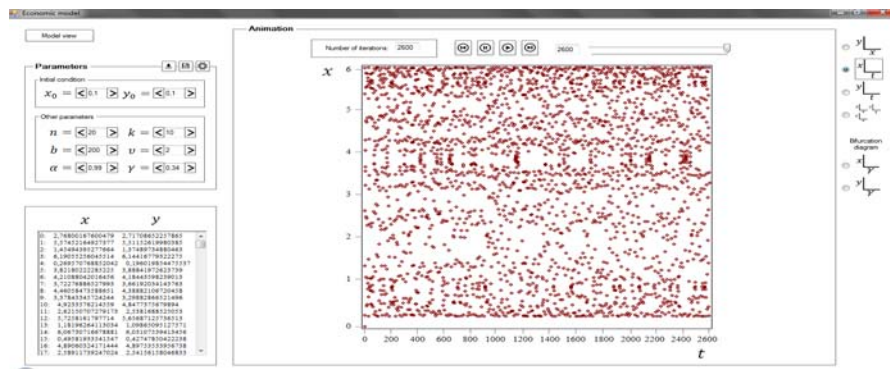


Fig. 1. Application *Model* for two-dimensional model

The right side presents 6 kinds of graphs displayed by the application; their examples are set forth in the paper. Selected switch indicates that here the graph of trajectory $x(t)$ is selected. On the left side counters allow us to specify the parameters of the model and the initial values of the trajectory. After their setting the calculation results of the iterations' coordinates below and their image in the center of the window. This displays an animation of a selected path, the number of iterations been set on the scroll bar above. Pressing the button *Model view* left displays information about the model, its equations and parameter information.

4.1 Numerical Experiment: from Stability to Chaos with Increasing of Naive Expectations

With the increasing probability of naive expectations q , that is with decreasing p , the market becomes unstable, evolving from simple dynamics (15) with a single stable equilibrium point to the unpredictable behavior of system (6). From the proof of Proposition 2 it follows that the market volatility is proportional to the number n of firms in the market. Therefore, for fixed q market instability increases with increasing n . Thus, model (24) has two parameters: the number of firms n and the probability of a naive approach q , whose growth leads to instability. The transition from stability to chaos is the same in both cases. Consider this transition for parameter q .

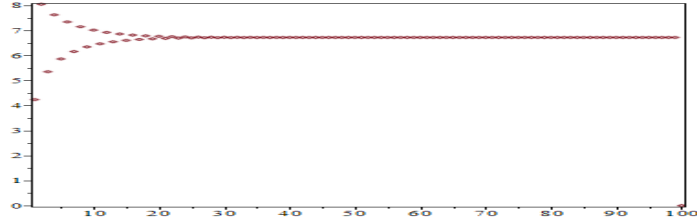


Fig. 2. Quantity trajectory of selfish firm under probability of naive expectations $q = 0.5$

Let $n=20$, $k=6$, $b=200$, $v=2$, $\alpha=0.9$, $q=0.5$. The trajectory of the dynamical system (24) with the following parameters and the initial point $x_0=0.1$, $y_0=0.1$ is shown in the following figures 2 and 3. In Fig. 2 on the x -axis of the system are given iterations of system (24) from $m = 1$ to $m = 100$, on the y -axis – corresponding quantity product of selfish firm y_m .

As we can see from the graph, the path quickly converges to the equilibrium value $y^* \approx 2.488$. The graph for the trajectory of firm-reciprocator x_m on y -axis is similar. The equilibrium value of x^* is about 6.72. Let us consider the graph of the trajectory for the same parameters except q . Now $q = 0.55$ (Fig. 3).

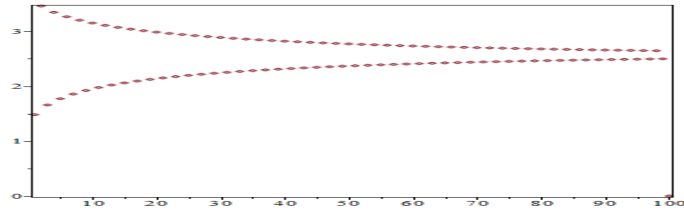


Fig. 3. Quantity trajectory of selfish firm under probability of naive expectations $q = 0.55$

It still has stable Nash equilibrium, but 100 iterations does not suffice for convergence. Further, let $q = 0.6$ (Fig. 4).

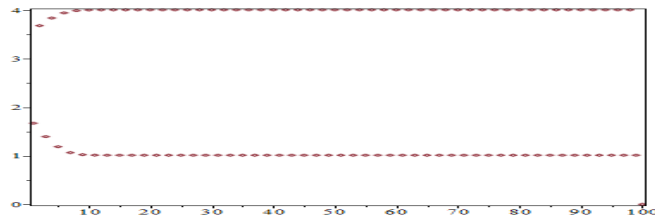


Fig. 4. Bifurcation quantity trajectory of selfish firm under probability of naive expectations $q = 0.6$

As we can see, bifurcation occurred, and instead of equilibrium point there was a steady cycle, where values of y_m are approaching the point of $y^* \approx 4$ for even m and the point of $y^* \approx 1$ for odd m . By doubling the lag between iterations only even or only

odd iterations will be considered, and thus either point $y^* \approx 4$, or $y^* \approx 1$ respectively would be the equilibrium steady state.

Stable cycle has four cycles for $q=0.64$ (fig. 5). There was a new cycle doubling bifurcation. Calculations show that with increasing parameter q doubling bifurcation cycle continues, following Sharkovskii's scale. According to this scale, when $q \approx 0.675$ there is the state of dynamic chaos (fig. 6). Similarly, the graph of product x_m on y-axis by firm-reciprocator looks like trajectory of a selfish firm.

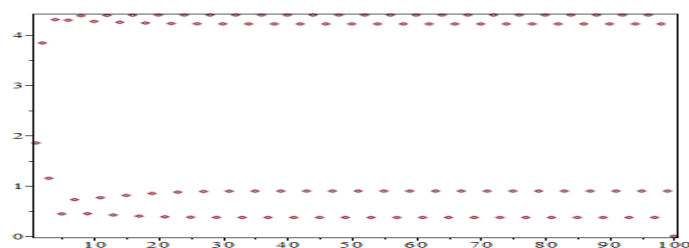


Fig. 5. Doubling bifurcation cycle of quantity by selfish firm under probability of naive expectations $q = 0.64$

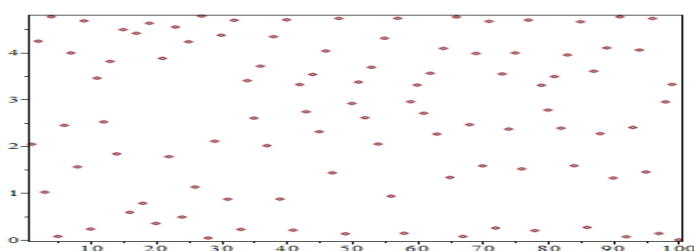


Fig. 6. The state of dynamic chaos of quantity by selfish firms under probability of naive expectations $q \approx 0.675$

Note that the ratio between the quantity of output by selfish firms and reciprocators remains almost unchanged. It is demonstrated in the graph of fig. 9, where each iteration on x-axis shows the value of output by firms-reciprocators x_m , and the vertical axis - the appropriate output of quantity y_m of selfish firms (fig. 7).

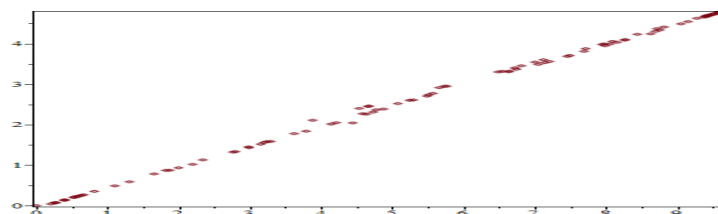


Fig. 7. The ratio between the quantity of product of selfish firms (horizontal axis) and reciprocator ones (vertical axis)

4.2 Bifurcation diagram

In detail the process of loss of stability and transition to chaos of dynamic system (24) can be presented in the following bifurcation diagram (fig. 8).

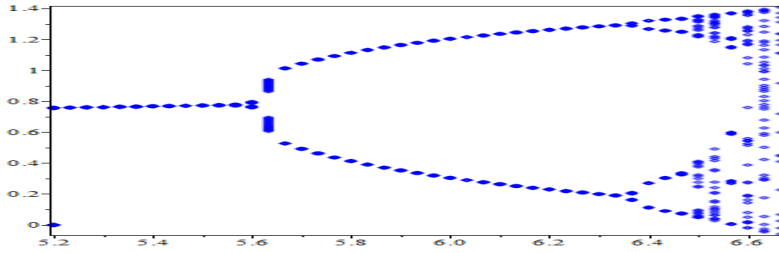


Fig. 8. The bifurcation diagram of dependence quantity product of selfish firm (y) on the probability of naive expectations (q) in a general dynamical system

Here the horizontal axis represents the parameter value of q multiplied by 10. The ordinate values quantity volumes of selfish firm on stable cycle, multiplied by 0.3. This rescaling is done for the sake of clarity. The values of the other parameters are the same as above. The bifurcation diagram, where on vertical axis are placed the values of output of firms-reciprocators x_m looks similar.

As noted in numerical simulations, the bifurcation may be interpreted as separation of equilibrium into several ways, one of which is selected by the market due to evolution of firms' strategies, such as repeated interactions and adaptations. Numerical experiments with n firms as the variable parameter are analogous to those described above.

5 Conclusion

Thus, we have designed the strategic model of cooperation between the two types of firms in the market of homogeneous product, where reciprocator and selfish firms plan their output using the adaptive approach with probability p and naïve (bounded rationality) one with a probability of $q = 1-p$, which distinguishes this model from existing analogues, where each type of firm adheres to one strategy rather than their combination and maximizes only its own profit rather than social welfare.

Desktop C# application *Model* using a graphical interface to animate the model trajectories has been created specifically for the numerical investigation of the model.

It has been proved that in the model with adaptive expectations the unique Nash equilibrium in a dynamic system is stable for all possible values of the parameters. The trajectories of the dynamical system converge to the fixed point for any possible initial values. In the model with naive expectations the unique Nash equilibrium is unstable for sufficiently large values of n for all possible values of other parameters. According to the calculations, this point is unstable even at $n \geq 5$.

As a result of numerical experiment we have found that bifurcations of cycle doubling occur with an increase in naive expectations. This bifurcation can be interpreted as separation of equilibrium state into several ways, one of which is selected by the market in the evolution of firms' strategies. If two-thirds of firms use naive expectation ($q \approx 0.675$), then in accordance with the Sharkovskii scale there appears the state of dynamic chaos in the market, leading to degeneration of the existing competition model between two types of firms.

Thus, the crucial factor, which ensures sustainable equilibrium in the market and the ability to predict the product quantity of firms, is the adaptive approach, i.e. the one taking into account adaptive expectations of the firms when they plan their production.

Similar results are obtained if instead of q we use parameter n - number of firms in the market, where system also moves from stability to chaos if n increases.

References

1. Barsukova, S.Y.: Reciprocal interactions: the nature, function, specificity. *Sociological studies* 8. 20--30 (2004)
2. Polanyi, K.: Economics as institutional process. *Economic sociology* 2. 62--73 (2002)
3. Kolm, S., Ythier, J.M.: 2006. *Handbook of the Economics of Giving, Altruism and Reciprocity*. Elsevier: Worth-Holland (2006)
4. Bischi, G.I., Lamantia, F. A dynamic model of oligopoly with R&D externalities along networks. Part I. *Mathematics and Computers in Simulation*, 84. 51--65 (2012).
5. Puu, T., Panchuk, A. *Nonlinear economic dynamics*. New York: Nova Science Publishers (2011).
6. Puu, T. Chaos in duopoly pricing. *Chaos, Solitons&Fractals* 1. 573--581 (1991).
7. Szidarovszky, F., Okuguchi, K. An Oligopoly model of commercial fishing. *Seoul Journal of Economics* 11. 321--330 (1998)
8. Bischi, G.I., Naimzada, A.K., Sbragia, L.: Oligopoly games with Local Monopolistic Approximation. *Journal of Economic Behavior&Organization* 62. 371--388 (2007)
9. Kobets, V., Poltoratskiy, M: Forming an Evolutionarily Stable Firm Strategy under Cournot Competition Using Social Preferences. In: Ermolayev, V. et al. (eds.) *ICT in Education, Research, and Industrial Applications. Revised Extended Papers of ICTERI 2014, CCIS 469*, pp. 343-361, Springer Verlag, Berlin Heidelberg (2014) DOI: 10.1007/978-3-319-13206-8_17
10. Kobets, V., Yatsenko, V., Poltoratskiy. Dynamic model of double electronic Vickrey auction. In: Ermolayev, V. et al. (eds.) *Proc. 11-th Int. Conf. ICTERI 2015, Kherson, Lviv, May 14-16, 2015, CEUR-WS.org/Vol-1356, ISSN 1613-0073, pp.236--251, online CEUR-WS.org/Vol-1356/ICTERI-2015-CEUR-WS-Volume.pdf*
11. Fanti, L., Gori, L. The dynamics of a differentiated duopoly with quantity competition. *Economic Modelling*, 29. 421--427 (2012).
12. Tramontana, F. Heterogeneous duopoly with isoelastic demand function. *Economic Modelling*, 27. 350--357 (2010).

Adjusting Business Processes by the Means of an Autoregressive Model Using BPMN 2.0

Vitaliy Kobets¹, Valeria Yatsenko²

¹Kherson State University, 27, 40 rokiv Zhovtnya st., Kherson, 73000 Ukraine

vkobets@kse.org.ua

² Taras Shevchenko National University of Kyiv, 90-A, Vasulkivska st., Kiev, 03022 Ukraine

ValeriaYatsenko5@gmail.com

Abstract. Process approach in the management of modern business entity activities opens a wide range of opportunities, ways and methods of activity reorganization, improving the quality of customer interaction, optimization of internal and external business processes. An effective tool in analyzing business processes as a central category of the process approach is modeling that increases their adaptability and mobility. We propose approach to improve and adjust business process to time and profit increasing by the means of autoregressive model using BPMN 2.0 in contrast existing approaches without econometrics analysis of experimental data for BPMN.

Keywords. Business process, regression, BPMN, profit

Key terms. BusinessProcess, EconometricModel, DynamicModel

1 Introduction

When we construct a business process by means of BPMN 2.0 we will set the initial conditions of the operation, such as execution time and the cost of resources (e.g., human resources in cost per hour or piece wage). During operating of a business process such timing and costs of the resources are chosen as would satisfy the restrictions of a customer or designer of the business process. To support decision-making in which direction should change the initial parameters of the business process, it is proposed to use dynamic distributive-lag and autoregressive econometric models. These models allow to take into account the impact of previous indices' values of business process on its succeeding values to assess short-run and long-run impact on the efficiency of a business process as a whole according to specific criteria.

The goal of the paper is to develop a procedure of business processes adaptation according to cost and profitability criteria by means of autoregressive model using BPMN 2.0.

Paper has following structure: section 2 is devoted to related works, section 3 demonstrates designing and adjustment of business processes, section 4 concludes.

2 Related Work

2.1 Dialectical Essence of the Process Approach

The paradigm of the functional approach, which has long been used for organization and management of companies, appeared to be unable flexibly and effectively to address current challenges faced by each business entity. Focus on the functional approach has led to the isolation of the top management from the current economic situation that caused information asymmetries, bureaucratization of companies, leveling of the system of incentives and reward schemes for human capital assets, excessive localization of functions within individual units, distortion of control and analysis system. The next stage of the management concept evolution was marked by the emergence of process management approach that provides for a high degree of mobility and adaptability in highly turbulent current market conditions (table 1).

Table 1. Invariant character of the “business process” definition

Criterion of classification	U. Deming	J. E. Short.	T. Davenport	I. Fedorov	O. Volkov	V. Kepin, V. Yeliferov	ISO-9000:2001	L. Frolova	Y. Zinder	J Martin	B. Nemurivskiy	V. Yevdokiyenko	V. Poronarenko
Any types of activity													
A logic range of interfacing													
Distribution of working operations in the space													
Human engineering													
Getting the results at the output based on resources at the input													
Customization													
Method of solution of the business target													
Technology													

This approach allows a company to easily integrate into the concept of the client-oriented market that meets the actual needs of society and economy (fig. 1).

Finally, the analysis of the previous scientific researches demonstrated that the process approach is the mainstream of the “business process” definition which is considered as the logic range of interfacing in the field of customization of the company activity for the purpose of convergence customers and producers targets.

2.2 Business Process as a Central Category of the Process Approach

The object of the process management concept is a “process” in which the activity of a company is a combination of business processes, management of which allows to significantly increase the transparency and manageability of business and to improve its effectiveness as per specified criteria. The process is described by the parameters

Strengths	Weaknesses
-----------	------------

<ul style="list-style-type: none"> ❖ taking into account the structural features of business processes ❖ customer-oriented strategy ❖ continuous optimization of cross business processes ❖ delegation of authority and responsibility ❖ high qualification required ❖ specialization of administrative personnel in the strategies development ❖ simplification of multilevel hierarchical organizational structures ❖ simplified exchange of information ❖ time saving ❖ reduction of overhead costs ❖ simplified coordination, organization and control systems ❖ vertical integration of resource flows ❖ high flexibility and adaptability of the management system ❖ high degree of transparency in relations ❖ priority of processes with a high degree of added value ❖ risks leveling 	<ul style="list-style-type: none"> ❖ strict social requirements ❖ complex design of management system for the account of detailing and formalization
<p>Opportunities</p> <ul style="list-style-type: none"> ❖ change of organization functioning logic and mechanism of its control ❖ competitive growth ❖ comprehensive effectiveness evaluation of functioning ❖ possibility to create systems to monitor the quality of company activities ❖ preparing for benchmarking ❖ process approach as a reserve for enhancement of efficiency ❖ resources optimization ❖ high staff motivation ❖ the possibility of team-working ❖ possibility of integrated automation 	<p>Threats</p> <ul style="list-style-type: none"> ❖ delegation of authority and responsibility ❖ serious requirements for qualification of contractors

Fig. 1. SWOT- analysis of the process approach in a company management

of time, sequence, transformation of object states during its movement to the final state [2], which helps the company to flexibly respond to the dynamic changes, to switch from one process to another, to comprehensively analyze the entire operational cycle. Business process, as a part of the process approach, has its “input” and “output” in the form of production of resources and products (goods, services or works), respectively (fig. 2).

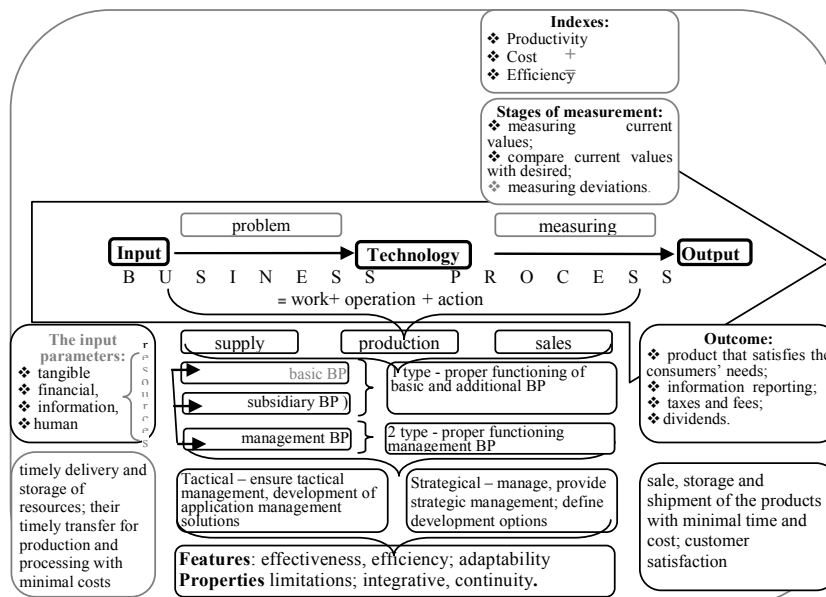


Fig. 2. Components of Business Process, [1, 2, 5]

Technology, mechanism of interaction between functions and control system are additional components of the process approach (fig. 3) [6].

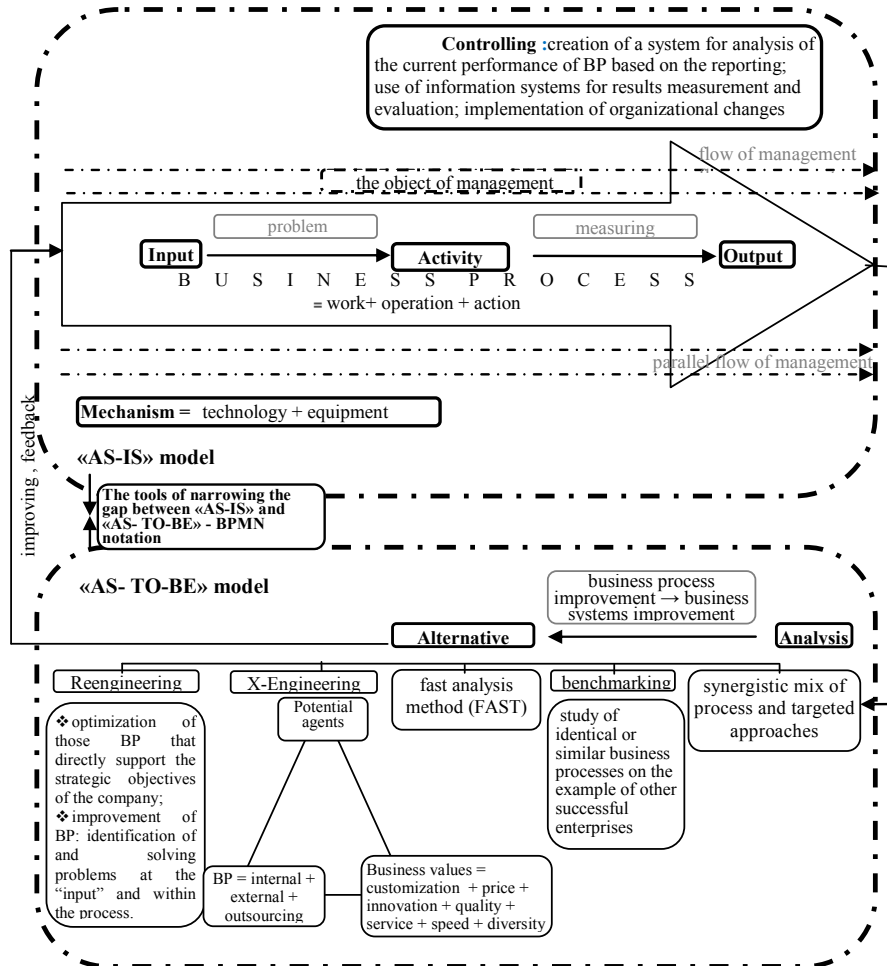


Fig. 3. Model of process management concept [1, 2, 4, 6-10]

2.3 Correspondence of Business Processes with the Strategic Objective of the Business Entity

The mechanism of leveling the differences between "AS-IS" and "AS-TO-BE" models is the "Road map" that sets benchmarks for optimization of business processes. "Input" resources during the process of transformation into goods, services and works at the "output" form "AS-IS" model, the main purpose of which, according to L. Dryuchenko, is aimed at identifying of "bottlenecks" in business process: current problems, differences, inconsistencies, threats that can be overcome subject to radical modernization of existing organizational and administrative decisions and procedures.

Instead, the “AS-TO-BE” model is aimed at improving the existing practice of business process implementation [6]. Improving the effectiveness of business processes and transition of “AS-IS” model to “AS-TO-BE” model are achieved by correlation of current business processes with strategic installations and objectives of the business entity, that leads to a radical rethinking of business philosophy. The theoretical basis for the implementation of this process involves multiple possible scenarios: reengineering, X-engineering, FAST technique, benchmarking and synergistic combination of process and target-oriented approaches. Simulation modeling and economic analysis are technical tools to achieve this objective.

The most popular of these scenarios is the concept of reengineering, which, according to the definition of the founders of the concept, namely M. Hammer and J. Champy, is treated as a fundamental rethinking and radical reconstruction of business processes. Reengineering should not be identified only with the processes of reorganization or automation. It shall be considered in the context of a component of a larger category – Total Quality Management (TQM) since its main function is the fundamental transformation of the essence of processes implementation and performance of operations at the stages of “input” and direct process, but not only increase in efficiency or profitability of the “output”. However, in our opinion, X-engineering theory of J. Champy is a more progressive one. This theory, unlike reengineering, provides for a comprehensive restructuring of not only internal operations and relations between internal staff, but also the transformation of external relations, including not only consumers, but also other stakeholders – competitors, contact groups, state, global economy entities [10], that will make it possible to adequately meet the urgent needs of the economic system as a whole.

2.4 Business Processes in the Coordinates of Evolutionary Economics

In the context of complicating of the forms and types interfacing between counterparties and market indeterminacy economic systems are able to self-organization and reach the temporary equilibrium, which transforms the system from the chaotic state to the equilibrium within the sphere of evolutionary economics, which identifies stochastic development script.

Given the actualization of the key tenets of the evolutionary economics of J. Schumpeter, the analysis of business processes through the prism of the theory of jokers is an interesting one. According to this analysis, some parallels can be drawn between a business process and economic and physical method of “channels” and “jokers” that simplifies the model, highlighting the main parameters and discarding irrelevant ones [11]. “Input” of a business process can be compared with the “source” from jokers theory, similarly, “Output” – the “mouth”. Conditional distinction of parameters and processes in this area is performed for the account of determining the degree of dynamism: the system in “sources” and “mouths” is slow, which can be explained by the relatively stable partnership relations that are often built on long-term cooperative basis; while the “jokers” are characterized by high dynamism and unpredictability as production operations and actions tend to increasing volatility because of strengthening of scientific and technical progress development (fig. 4).

Riverbeds are the methods of simplification the complicates systems or processes.

Riverbeds separate the unite to the various parts and also, determine the possibility of development process forecasting. Diverse fields of the space which are characterized by the high level of the time rate of change and unpredictability are called jokers (specific bifurcations) in which the system is tested for sustainability.

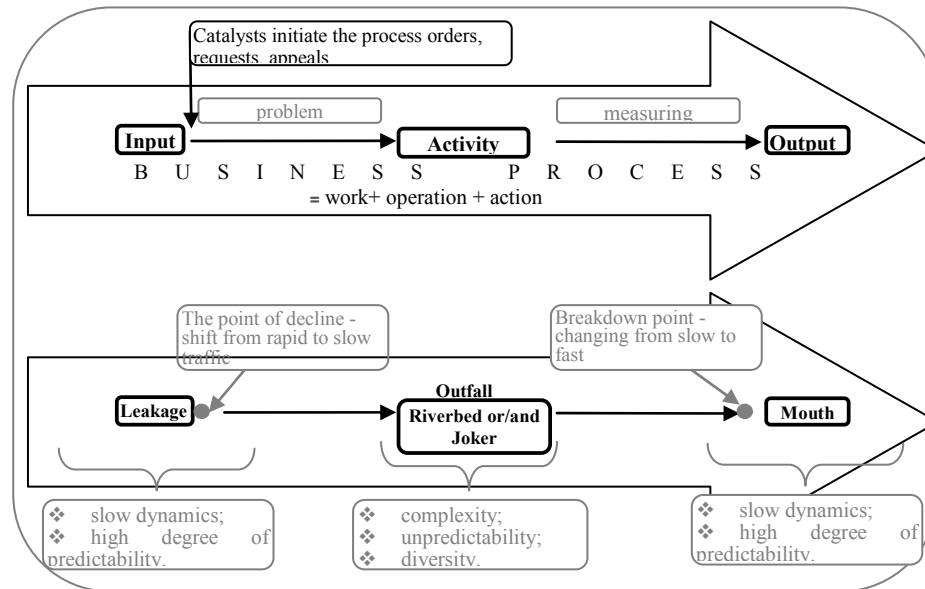


Fig. 4. Overlapping between business process and jokers theory in evolutionary economics, [4, 11]

The mechanism of decision-making in the process approach can be represented by the following algorithm (fig. 5).

Despite the benefits of process management, empirical implementation of this approach is characterized by a number of difficulties, including those identified by J. Riley: fragmentarity of processes that adversely affects the flexibility and adaptability of business; low level of automation of processes resulting in increased operating costs; the basis of enterprise motivation is maximization of its own profits, but not satisfaction of customers' needs; lack of evidence-based research on the advantages and disadvantages of modeling and optimization of business processes; rejection of modeling and optimization through formal introduction of process management [12].

Technical leveling tool for these shortcomings and limitations of the process approach is a mechanism for predicting and adapting business processes through modeling.

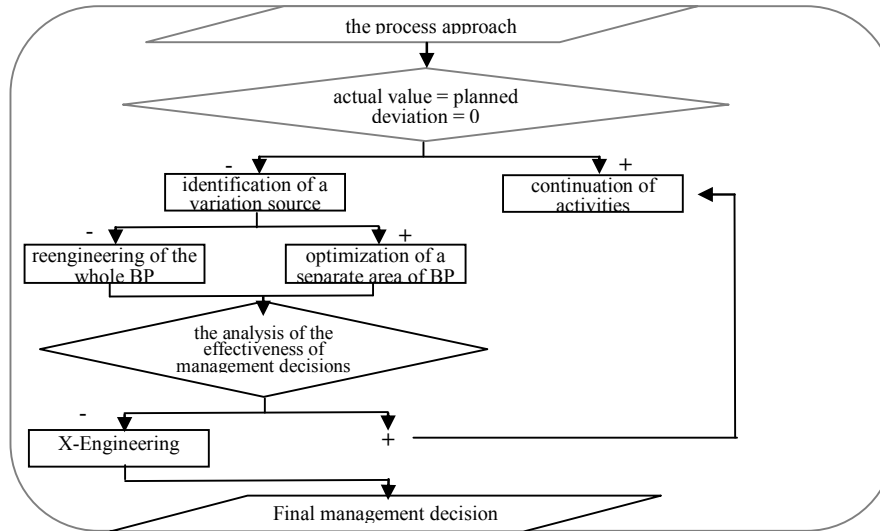


Fig. 5. Logic of the process approach

3 Modeling of Business Processes

We believe that business process modeling is a tool for reflecting the aggregate of logically related operations or processes to improve the quality of business processes through improving their forecasting, optimization, and adaptation (fig. 6).

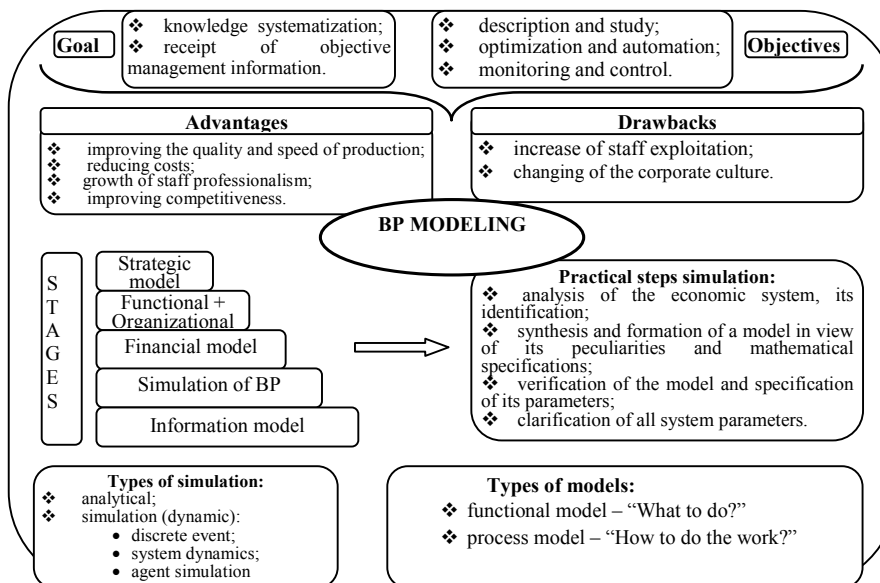


Fig. 6. Essence of the “Business Process Modeling” category, [4, 11, 13 - 16]

We should note that under the optimization of business processes, the interpretation of J. Harrington, a pioneer in the field, is meant. Under the specified category J. Harrington understands improvements aimed at increasing productivity, efficiency and adaptability of business processes [3].

According to D. Kozenkov, the modern management concept reduces the formalization of business process modeling to the process of building of enterprise architecture with a three-level structure (fig. 7) [2].

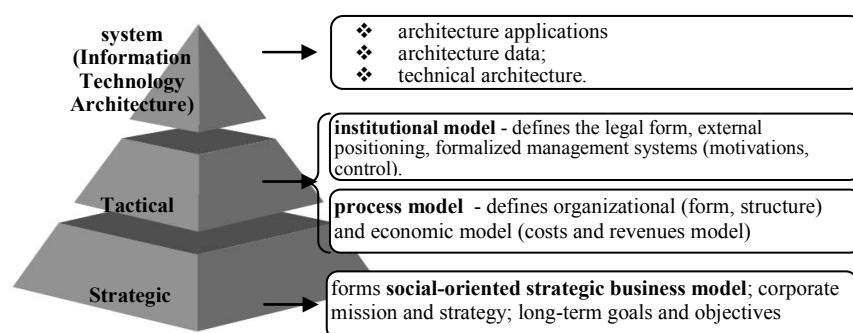


Fig. 7. Levels of enterprise architecture

3.1 Designing of business processes

Consider the business process, which describes the manufacture of soft drink (Fig. 8).

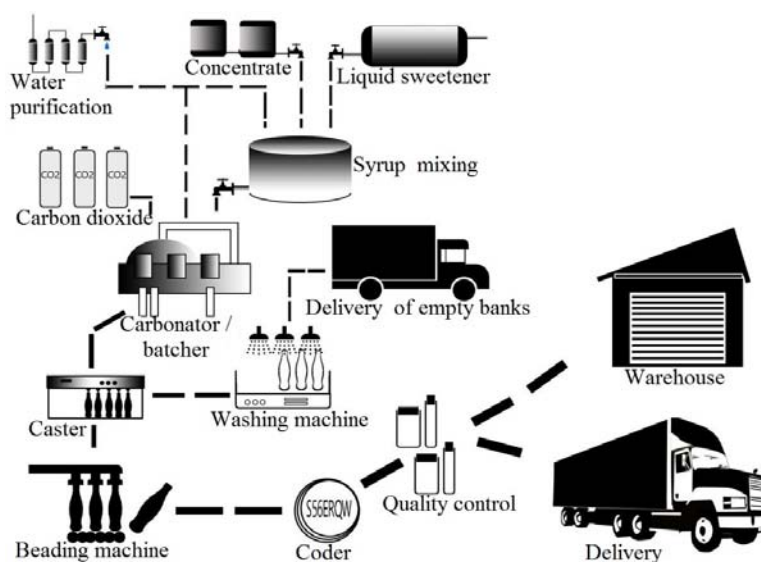


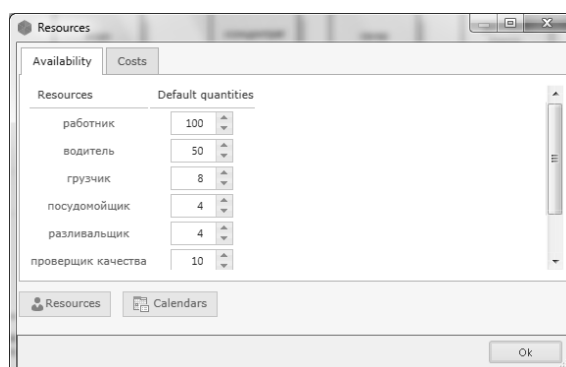
Fig. 8. Preserving cycle of bottling on the manufacturing of soft drink

Positions of workers performing all business process operations and in-payment forms are described in Table 2.

Table 2. Staff categories of business process

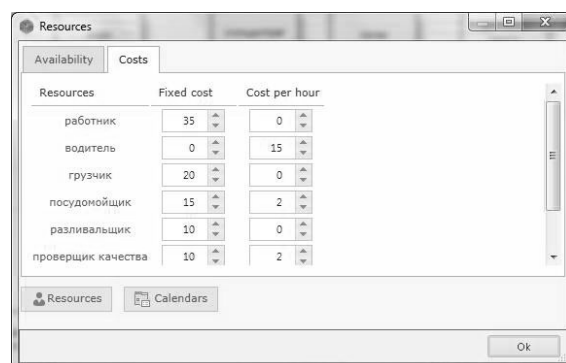
Resource	Total fixed cost	Total unit cost
worker		+
driver	+	
loader		+
dishwasher		+
quality tester		+
storekeeper		+

In the menu Resources in tab Availability a maximum number of employees available to perform business process is determined (fig. 9), and in tab Costs – piece (fixed cost) or hourly (cost per hour) wage (fig.10).



Resources	Default quantities
работник	100
водитель	50
грузчик	8
посудомойщик	4
разливальщик	4
проверщик качества	10

Fig. 9. Determining the employees' number for all positions



Resources	Fixed cost	Cost per hour
работник	35	0
водитель	0	15
грузчик	20	0
посудомойщик	15	2
разливальщик	10	0
проверщик качества	10	2

Fig. 10. Setting in-payment forms for each category of employees

In BPM notation business process will look like (fig. 11)

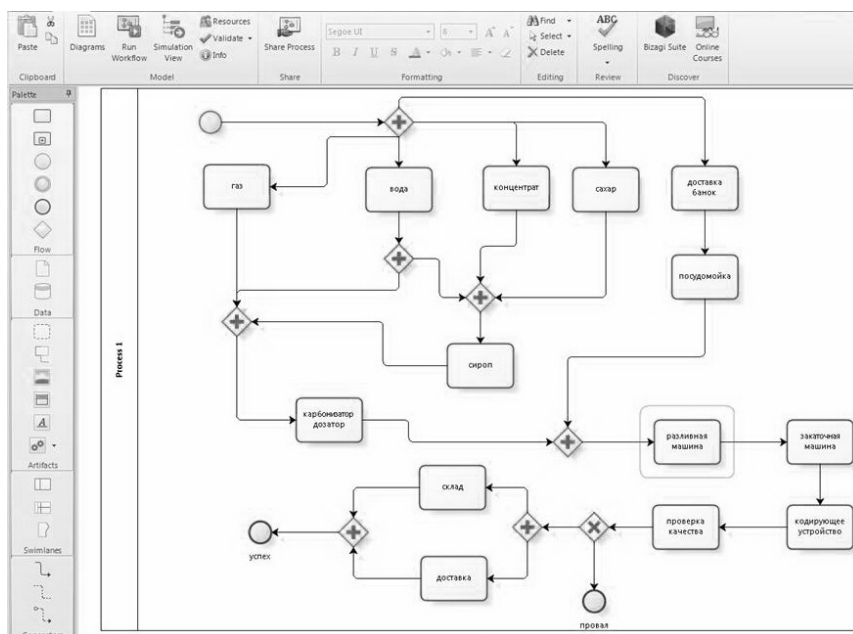


Fig. 11. Business process of soft drinks production in BPM notation

Standard activity time, the number of performers and activity cost are defined for each business process operation (Fig. 12).



Fig. 12. Determining time, executors and activity costs for a business process operation

The number of actually executed orders that are shipped to the end customer; average time of each operation; delay time in operations; a share of time during which the employee performs his work (on a scale from 0 to 100%) (fig. 13) is obtained as a result of the business process (fig.13).

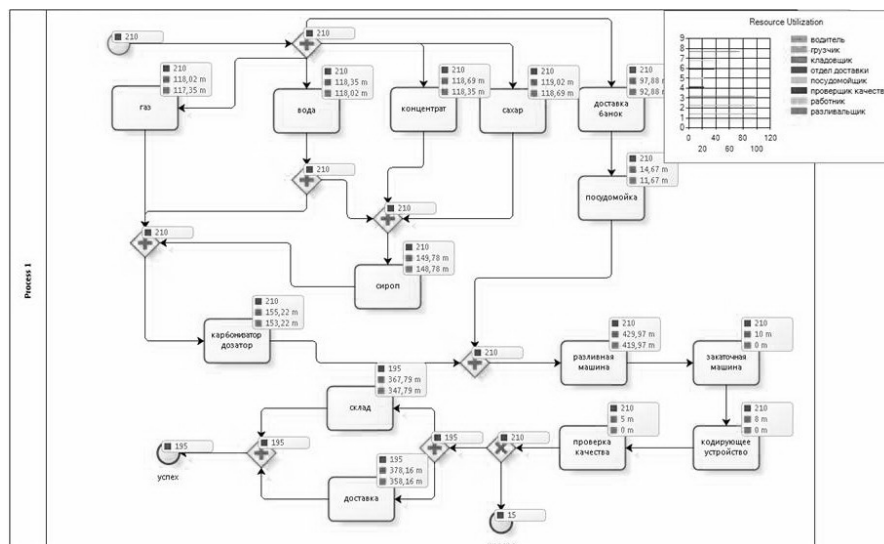


Fig. 13. The results of the business process in BPMN

Export of experimental data concerning staff costs (Fig. 14) provides an econometric estimation of payroll costs on cost per unit (average cost) of soft drink parties.

Resource	Utilization	Total fixed cost	Total unit cost	Total cost
работник	35,32 %	44 100	0	44 100
водитель	100,00 %	0	34 687,5	34 687,5
грузчик	98,02 %	2 720	0	2 720
посудомойщик	22,70 %	3 150	84	3 234
разливальщик	75,68 %	2 100	0	2 100
проверщик качества	37,84 %	2 100	350	2 450
кладовщик	98,02 %	4 080	0	4 080
отдел доставки	22,88 %	1 270	529,17	1 799,17
Total		59 520	35 650,67	95 170,67

Fig. 14. Experimental data for payroll staff

Export of business process activity costs (Fig. 15) from Bizagi Modeler in MS Excel makes it possible to build an econometric model for estimating the impact of total time of work performed on the profit of business process owner.

Name	Type	Instances completed	Instances started	Min. time	Max. time	Avg. time
Process 1	Process	210	210	1h 13m	1d 15h 17m	20h 2
NoneStart	Start event	210				
газ	Task	210	210	40s	3h 57m	1h 56
вода	Task	210	210	1m	3h 57m 19s	1h 56
концентрат	Task	210	210	1m 19s	3h 57m 40s	1h 56
сахар	Task	210	210	1m 40s	3h 58m	1h 56
доставка банок	Task	210	210	5m	3h 39m	1h 31
сироп	Task	210	210	1m	3h 58m	2h 26

Fig. 15. Experimental data concerning timing of business process operations

3.2 Adjustment of Business Processes

We estimate the impact of timing on the efficiency of business processes in the short-run and long-run periods. To assess the effectiveness of business processes consider the following distributive-lagged model where current cost (y_t) depends on the payroll of the current and all prior periods (x_t, \dots, x_{t-n}):

$$y_t = c + b_0 \cdot x_t + b_1 \cdot x_{t-1} + \dots + b_n \cdot x_{t-n} + \dots \quad (1)$$

where short-run (SR) impact in period t is $b_t = b_0 \cdot \delta^t$, $t = 1, 2, \dots, n, \dots$. Long-run impact is $LR = \sum_{t=0}^{\infty} b_t$ or after substitution we get: $LR = \frac{b_0}{1-\delta}$. After substitution of SR-impact in equation (1) we get:

$$y_t = c + b_0 \cdot x_t + \delta \cdot y_{t-1} \quad (2)$$

To determine the regression parameters b_0 and δ conduct a series of experiments for different numbers of orders. The task of the company is to analyze the effectiveness of the wage fund on the average cost of production. For that firm's budget (TC) is determined as sum of payroll and activity costs for each order (q), then we calculate cost per unit of production (AC). After 10 experiments we obtained the following data (table 3):

After construction of regression model using experimental data from table 3 we have:

$$y_t = 14.5 + 0.0007 \cdot x_t + 0.888 \cdot y_{t-1} \quad (R^2 = 0.71) \quad (3)$$

The model is adequate to reality according to determination coefficient, i.e. chang-

ing the current wage fund and activity cost and average cost of prior periods explain 71% change in the average cost of a current period. Parameter $b_0 = 0.0007$ indicates how much the cost per unit will increase with the growth of the wage fund and activity cost of the current period to \$ 1. For previous periods the influence of wage fund and activity cost on average cost shrinks rapidly. According to Student test parameter δ is statistically significant ($t(\delta) = 3.4 > 2.45 = t_{kr}$), confirming the impact of per unit cost of the previous period on the next period. That is after decreasing of total cost in previous periods, average cost will reduce in the current period. Long-run effect $LR = \frac{0.0007}{1-0.888} = 0.0058$ describes the impact of reduction of total cost for by \$1 in all previous periods on current average cost. If b_0 is positive then we can gradually decrease total cost (to reduce cost per unit) until b_0 became negative.

The next task of the company is to analyze the impact of the business process duration on the company profits from soft drink production via the model of partial adjustments. It is assumed that profit depends on the expected duration of business process. This profit y_t is observable, and the time duration of employees (x_t^*) is unobservable value; u_t is residual term of the model:

$$y_t = b_0 + b_1 \cdot x_t^* + u_t. \quad (4)$$

Software BPMN allows to simulate the time and resources required to perform business process, but require additional analysis of statistical data for its adaptation by the criterion of profitability. Procedure of determining an adaptive time performance for operations is carried out by the econometric model of partial adjustment. We introduce the hypothesis of adaptive expectations for the leader of the business process, seeking to define the time of operations within which the profit for entire business process reaches its maximum:

$$x_t^* - x_{t-1}^* = \gamma \cdot (x_t - x_{t-1}^*), \quad 0 < \gamma < 1, \quad (5)$$

where γ shows the speed of adjustment (corrections) of leader's expectations to differences between the actual time of performing the operations and previously expected time. Expression $x_t^* - x_{t-1}^*$ shows how expectations are formed and the expression $x_t - x_{t-1}^*$ explains how expectations are corrected. Rewrite equation (4) as:

$$x_t^* = \gamma \cdot x_t + (1-\gamma) \cdot x_{t-1}^*, \quad (6)$$

where the expected time of performing the operations at time t is a weighted average of its actual value in the current period and its expected value in the previous period with weights γ and $1-\gamma$ respectively.

After the substitution of (6) in (4) we get:

$$y_t = b_0 + \gamma \cdot b_1 \cdot x_t + (1-\gamma) \cdot b_1 \cdot x_{t-1}^* + u_t \quad (7)$$

Then multiply by $1-\gamma$ both sides of equation (1), recorded with a delay of one lag, and find the difference between this equation and equation (7), and then get:

$$y_t = \gamma \cdot b_0 + \gamma \cdot b_1 \cdot x_t + (1-\gamma) \cdot b_1 \cdot y_{t-1} + v_t, \quad (8)$$

where $v_t = u_t - (1-\gamma) \cdot u_{t-1}$

In BPMN environment for simulation data in each experiment we will change the duration of all the company's operations at a fixed time. Then calculate the profit of the company as the difference between its revenue and cost of resources for each time period (t). After 10 experiments we obtain the following data (table 4):

Table 4. Adaptation model where profit depends on time of performance of business process

№	Profit y_t, \$	Time performance x_t, min.	Previous profit y_{t-1}, \$
1	9611,65	346453,67	
2	22030,50	325469,99	9611,65
3	29940,00	1500015,00	22030,50
4	5940,15	66908,00	29940,00
5	7937,89	172778,00	5940,15
6	8264,50	210119,00	7937,89
7	49822,20	10939490,00	8264,50
8	11346,00	519113,00	49822,20
9	8024,65	257556,00	11346,00
10	8847,80	38597,50	8023,65

Using experimental data of Table 4, we obtain the following regression between the current profit and explanatory variables, such as current time performance and profit of the previous period:

$$y_t = 11098.9 + 0.0037 \cdot x_t + 0.0025 \cdot y_{t-1} \quad (R^2 = 0.79). \quad (9)$$

Regression (9) is adequate to reality according to the coefficient of determination, i.e. changing the current performance time and profits in prior periods explain by 79% change in profits in the current period. Taking into account the values of regression parameters, we obtain: $\gamma = 0.9975$, $b_1 = 0.0037$, $b_0 = 11126.6$. Parameter $\gamma = 0.9975$ shows the speed of adjustment of expectations to the profit on the basis of its actual value. The velocity adjustment shows that the business process expected time performance x_t^* almost instantly adapts to the actual time performance x_t . Time performance is expected, but not observable value, whose impact on profits, in view of (9), can be estimated from expression $y_t = b_0 + b_1 \cdot x_t^*$. It means that with increased operation time by 1 minute the margin profit increases to \$0.0037. Parameter b_1 is statistically significant according to Student's test ($t(b_1) = 4.73 > 2.45 = t_{kr}$), confirming the significant effect of time performance on the profit of the company. Thus, for examined project the company must adjust (increase) standards of time operations,

enabling it to increase the value of the resulting profits until parameter b_1 became negative.

4 Conclusions

Thus, the process approach in the management of modern business entity activities opens a wide range of opportunities, ways and methods of activity reorganization, improving the quality of customer interaction, optimization of internal and external business processes. An effective tool in analyzing business processes as a central category of the process approach is modeling that increases their adaptability and mobility.

Analysis of business process modeling has an analogy with evolutionary economics in the theory of jokers and neurosystem theory. Among the future prospects of business process modeling is the use of ontologies, which, according to T. Gruber's opinion, in business modeling sphere is reduced to a formal specification of conceptualization [16] that involves the use of OWL (Web Ontology Language).

Another promising research vector is the use of neurosystems involving the use of neural networks that can be effectively used in the study and analysis of dynamic processes.

It is demonstrated how we can improve and adjust business process to time and profit increasing by the means of autoregressive model using BPMN 2.0 in contrast existing approaches without econometrics analysis of experimental data for BPMN.

References

1. Bezgin, K. S.: Quality management of business processes in the enterprise. Manuscript (2009)
2. Kozenkov, D.: Business Processes Planning as the Basis of the Creation Enterprise Architecture. The Visnyk of the SSU 3. 126--136 (2011)
3. Harrington, H. J., Esseling, E. K.C., van Nimwegen. H.: Business Process Improvement Workbook: Documentation, Analysis, Design, and Management of Business Process Improvement. London: McGraw-Hill, 1996
4. Fedorov, I.: Modeling business processes in BPMN 2.0 notation. M., MESI (2013)
5. Davenport, T.: Process Innovation: Reengineering Work through Information Technology., Boston, MA.: Harvard Business School Press (1993)
6. Druchenko, L.: Implementation of Business Processes Modeling and Optimization in the Local Government. Public Administration: Theory and Practice 1(3) (2010)
7. Ponomarenko, V., Minuhin, S., Znahur, S.: Theory and Practice of Business Process Modeling: Monograph. Simon Kuznets Kharkiv National University of Economics (2013)
8. Manganelli, R., Klein, M.: The Reengineering Handbook: A Step-By-Step Guide to Business Transformation. New York: Amacom (1994)
9. Hammer, M., Champy, J.: Reengineering the Corporation: A Manifesto for Business Revolution. London : Nicholas Brealey Publishing (1993)
10. Taranyuk, L., Zaporozhchenko, O.: Organizational Principles of Reengineering Business Processes of Enterprises. Mechanism of Economic Regulation 4, 122--131 (2011)

11. Kuzmin, O., Sidorov, Y., Kozik, V.: Achievements and Problems of Evolutionary Economics: Monograph. Publishing House of Lviv Polytechnic National University (2011)
12. Reilly, J. P., Creaner, M. J.: NGOSS Distilled, The Lean Corporation (2005)
13. Kalyanov, G.: Business Processes Modeling, Analysis, Reorganization and Automation. Finance and Statistics (2007)
14. Melnikov, A.: Bases of Business Organization. Center of educational literature (2013)
15. Korzachenko, O.: The concept of business processes modeling and optimization of telecommunications companies. Scientific and theoretical journal Khmelnytsky University of Economics "Science and Economics" 4 (32), 247--253 (2013)
16. Gruber, T.: A Translation Approach to Portable Ontologies. Knowledge Acquisition, 5 (2), 199--220 (1993)

Intelligent Cyber Defense System

Myroslav Komar¹, Anatoliy Sachenko^{1,2}, Sergei Bezobrazov³, Vladimir Golovko³

¹ Ternopil National Economic University,
3 Peremoga Square, Ternopil, 46020, Ukraine

{mko, as}@tneu.edu.ua

² Silesian University of Technology,
26 Roosevelta st, Zabrze, 41-800, Poland

sachenkoa@yahoo.com

³ Brest State Technical University,
256 Moskovskaja st, Brest, 224017, Republic of Belarus

gva@bstu.by, bescase@gmail.com

Abstract. In this paper a novel method for detection of network attacks and malicious code is described. The method is based on main principles of Artificial Immune Systems where immune detectors have an Artificial Neural Network's structure. The main goal of proposed approach is to detect unknown, previous unseen cyber attacks (malicious code, intrusion detection, etc.). The mechanism of evolution of the neural network immune detectors allows increasing the detection rate. The proposed Intelligent Cyber Defense System can increase the reliability of intrusion detection in computer systems and, as a result, it may reduce financial losses of companies from cyber attacks.

Keywords. Artificial Neural Networks, Artificial Immune Systems, Malicious Code Detection, Intrusion Detection, Intelligent System, Cyber Attacks, Cyber Defense Financial Losses

Key Terms. Software System, Research, Cyber Defense, Intelligent System

1 Introduction

The up-to-date computer system cannot be imagined without safety equipment. The work in the Internet is accompanying by high risks to be attacked by network intrusions and malicious code. As a result, cybercrime continues to do more financial damage to companies: a company's costs for preventing cybercrimes are estimated approximately \$ 15 million per year. Thus, the costs of each company may vary from \$ 1.9 million to \$ 65 million per year. In absolute terms, the damage from cyber at-

tacks increased by 82% in the last six years [1]. In 2011, the direct annual global losses from cybercrime were estimated \$ 114 billion. Taking into account the financial losses of companies from cyber attacks and the costs of downtime and recovery, cybercriminal activity is worth of \$ 388 billion per year to the world economy [2].

According to the opinion of Forbes experts [3], one of the high-profile crimes in the field of information security was the Anonymous attack on the MasterCard, Visa and Paypal payment systems in late 2010. Damage from this attack was worth of \$ 5.5 million. The other high-profile cybercrime was an attack on Citibank in June, 2011. Hackers had stolen \$ 2.7 million from the accounts of 3,400 customers of the bank. Breaking into Sony PlayStation Network took place in April, 2011. The total damage to the companies was estimated \$ 171 million. As a result of hacking, there was a leakage of confidential information of 138,000 Internet users.

According to this, the costs for preventing cyber attacks are increasing. In average, the consequences of the attack can be eliminated during 46 days. Companies, participating in the research, in average, spend more than \$ 1.9 million during this period. Thus, there is a growth of costs to 22% in comparison with 2014, when the amount of the costs has averaged to \$ 1.5 million during 45 days [1]. As it states above an urgent problem is to have effective methods defending against cyber-attacks.

2 State-of-the-Art

Depending on used techniques, experts define four basic types of network attacks: denial of service attacks, user-to-root attacks, remote-to-local attacks and probe attacks, and several subtypes of these attacks [4].

Nowadays many methods for solving the problem of network attacks detection where developed. The essential part of these methods is based on artificial intelligence such as: artificial neural networks, methods of fuzzy logic, artificial immune systems.

Intrusion Detection Systems (IDS) on the ANN can be divided into four categories [5]. The first category (earlier studies) of IDS is based on Multi-Layer Feed Forward Neural Network (MLFF) [6, 7], such as the Multi-Layer Perceptron (MLP) [8, 9] and Back Propagation (BP). The second category of IDS is based on Cerebellar model Articulation Controller (CMAC) [10] neural networks and Elman neural networks [11, 12]. The third category of IDS is based on unsupervised learning the neural networks for classifying and visualizing the input data to recognize the normal behavior from abnormal one. Most systems in this category use the Kohonen Self-Organizing Map (SOM) [13]. The fourth category of IDS is based on hybrid neural networks [14, 15].

Intrusion detection models on fuzzy logic are using the fuzzy rules or fuzzy classifiers [16]. Dickerson et al [17] proposed a Fuzzy Intrusion Recognition Engine (FIRE) for the detecting the malicious activity in the network. Data portions are classified using static metrics and enabling to generate fuzzy rules for classification of the input network data. The main disadvantage of this approach is that the rules are cre-

ated manually, but not automatically. Moreover the process of rules generating is laborious, and it imposes serious constraints on system development.

In the field of Artificial Immune Systems several basic algorithms were proposed: Negative Selection algorithm [18] Clonal Selection algorithm [19, 20], Idiotypic Network [21, 22] and Dendritic Cell algorithm [23]. A. Perelson and S. Forrest in 1994 proposed the Negative Selection algorithm for solving the anomaly detection problems [18]. It's based on the process of lymphocytes maturation in the thymus – biological organ that plays the basic role in the human immunity [24, 25].

Despite its successful application, the negative selection algorithm has several serious weaknesses [22, 26]. The first, it needs to create a randomly-generated initial detector population. If the dimensionality of the future space increases, then a number of detectors is growing exponentially. The second, the definition of “normal” is not updated as the time progress. And the third, the negative selection algorithm can cause excessive numbers of false alerts.

F. Burnet in 1959 proposed the Clonal Selection algorithm based on the Clone Selection theory [19, 20]. This theory explains the basic response of the adaptive immune system to an antigenic stimulus during proliferations of B-cells.

In our opinion, main problems of most AIS applications for data mining and anomaly detection tasks are the complex structure of immune detectors and no representatively matching methods. For example, in [18, 27] the binary structure of detectors is employed. Such structure of detectors requires the use of a contiguous bit matching method (r -contiguous bit [18], r -chunks [28]) that reduces the space and time complexity. Several works [26, 29] outline the unacceptable computational complexity of such methods by reason of the exponential relationship between the size of the data set(to be used) and the number of detectors that it is possible to generate. Also Gonzalez et al. [30] showed that matching rules between two binary strings cannot represent a good generalization of a self-space, and detectors demonstrate the insufficient good coverage of a nonself-space [31].

In comparison with mentioned methods above we propose the immune detectors based on neural network. An artificial neural network is an adaptive system that changes its structure based on external or internal information, and it flows through the network during the learning phase, and it's characterized by learning capability, generalizing ability and self-organization. Implementation of ANN permits to avoid the listed weaknesses above and increases the self-adaptation and self-evolution abilities of detectors in the tasks of data mining and anomaly detection.

In this paper we investigate the ability of immune detectors with neural network architecture to adapt to the changeable software environment and self-evolution in order to detect the unknown (before invisible) threat. The adaptability and self-evolution of detectors consists in modification of its structure for increasing the detection rate of unknown cyber attacks.

3 Generalized Architecture of Intelligent Cyber Defense System

The Artificial Immune System for Cyber-defense is the set of “intelligent” immune detectors and rules that describe their behavior. The structure of immune detectors, the algorithm of their training and evolution are described in [32, 33]. The system consists of modules that perform the control of immune detectors. The immune detectors are going through the different stages during the lifetime. There is creation, training, selection, detection etc. stages. Each stage can be represented as a module of the defense system. Thus, the developed system for computer attacks detection consists of several interacting modules. Figure 1 shows the architecture of the proposed system.

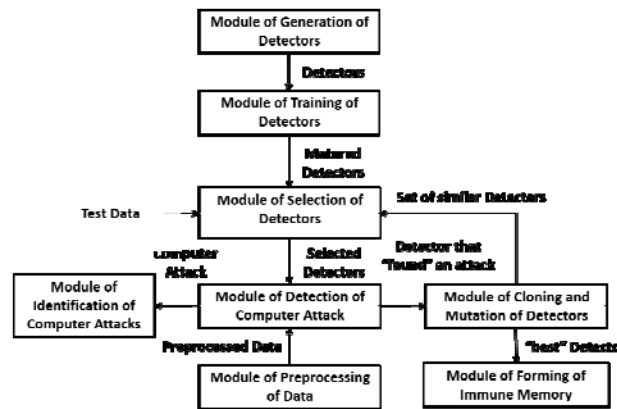


Fig. 1. Generalized architecture of intelligent cyber defense system

The module of generation of detectors produces the set of so-called pre-detectors that go through the several stages before they acquire the ability of correct classifications of objects. Every immune detector has the limited lifetime during which it “lives” in the system. At the end of the lifetime the detector is replaced by another new detector. This mechanism provide the system the continuously inflow of new immune detectors with different structure and different ability that can be more powerful than its precursors.

After creation, during the train stage, the immune detectors acquire the ability of correct classification of different objects and processes in computer environment and to detect the cyber attacks.

After training all immune detectors going through the selection stage where detectors pass the checking for correctness to minimize the erroneous work, while legitimate objects (files, processes, connections etc.) is considered as the threat. For this purpose, the preliminary created test sample – which consists only of legitimate objects – is given to detectors. If a i -th detector classifies one of test objects, as an attack, then it is destroyed and replaced by a new detector. If a i -th detector does not

generate the erroneous work during the test selection, then it is considered as a correct one and admitted to the analysis of computer environment. As a result the set of immune detectors for the analysis of environment is created, and it can be filled up due to the detectors of immune memory and generating new detectors after the end of their lifetime. The module of selection allows decrease the false alarm rate and increase the defense level.

All selected detectors can defense a computer system against cyber attacks. The set of active immune detectors forms a multi-agent system, where each immune detector is an intelligent agent with its own list of tasks. It selects the target of scanning, makes clones and evolves. At the functioning stage (or detection stage) all the information – which is getting by the computer – is primarily analyzed by immune detectors. If none of detectors found an anomaly, then data are processed by the operating system and the proper software. In addition, some period of life is given to each detector which analyses the environment during given period. If upon termination of given time the detector didn't find an anomaly, it is destroyed, and a new detector is created but on its place. If the any object is classified by immune detectors as an attack, then such detectors react on this attack. For example, they can block the proper connection, and, as a result, it is not processed by the operating system and software. The user receives the message simultaneously about the attack attempt on the computer system.

If detector found a threat then processes of cloning and mutation are activated. The goal of cloning and mutation module is to produces copies of the immune detector that found an attack. Such “clones” that are similar to the “parent” are very useful for example for defense against the family of malware, where each example have the similar malicious code. Such clones are capable to react on the found malware and check all the objects in the computer environment in a short period.

When clones are creating, some changes in their structure are taking place. As a result, the clones are not exact copies of the parent but with small differences. This process is called mutation. It allows immune detectors acquire the new ability, adapt to new attacks and increase the detection rate. In our case, when detectors base on neural networks, each clone is training on information that is abstracted from the detected attack. It allows adapting to new attack with the purpose of increasing of quality detection.

During the detection and eliminating of attacks, it is expedient to save their parameters and samples with the purpose of further detailed analysis. The point is that immune detectors are trained on the limited set of data, which can not include all possible cyber attacks. Therefore at the samples of attacks, that was classified as an unknown, are saved and added to the trained sample. It enables to increase the authenticity of attacks detection and classification as well as provide the flexibility of system, so this process updates the information. Newly generated detectors will be trained already, on new data.

At the end, the best detector is chosen and transformed to “memory” detector. Memory detectors have unlimited lifetime and provide a quick reaction on repeated cyber attacks. Thus the set of memory detectors forms the “immune memory” and keeps the information about all met cyber attacks and provides the high level of reaction on repeated attempts of attacks.

Finally, the module of identification of threats is used for classification of detected threat. The knowledge about the class of the detected threat allows taking correct response.

4 Results and Discussions

4.1 Ability of Artificial Cyber Defense System Detecting the Network Attacks

In order to eliminate disadvantages of existing works and improve the reliability of detecting the network attacks authors developed the intrusion detection systems using the proposed methods above (Fig. 2).

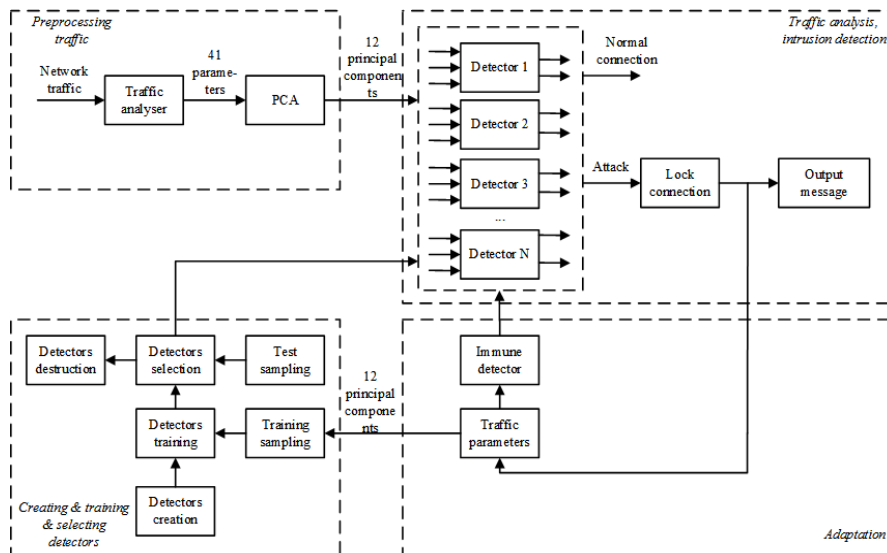


Fig. 2. Structure of Intrusion Detection System

Preprocessing traffic module is designed to represent traffic parameters in a convenient form for analysis. It consists of two modules – module capturing network traffic on the computer network and module of the principal components calculation from parameters of the captured traffic.

To capture the network traffic there used the specialized software called a sniffer – a software network analyzer of the traffic designed to capture and make the subsequent analysis of network traffic.

The network traffic – captured by sniffer – is analyzed. As a result there are extracted the 41 parameters of the network connection which characterize this connection and include time of the connection work, the protocol type, service type, a number of transferred bytes, etc.

Note the different parameters of the network connection have a different type of information, for example, the parameter "time of connection work" is set in seconds,

"protocol type" is set in a symbolic form, and "number of bytes from source to receiver" is set in bytes. Hence, in order to analyze such heterogeneous data, they have to be reduced to a general view.

To reduce the dimension of the analyzed data the unit PCA is used (see Fig. 2), performing the principal components selection and leading to improved quality of network attacks detection as well as increasing the analysis speed of network packets.

The PCA unit receives data from the sniffer – 41 parameters for network connections and, after all transformations and calculations, forms the 12 principal components. It was experimentally proved the 12 major components of network traffic parameters contain more than 99% information (Table 1).

A module of creating and training and selecting detectors (see Fig.2) is designed to create neural network immune detectors $1...N$, which are considered as the basic elements of Intrusion Detection(each individual immune detector represents an artificial neural network).

Table 1. Distribution of information according to the components

Number of components	1	2	3	4	5	6
The amount of information, %	52,40	71,67	88,37	91,49	94,21	95,90
Number of components	7	8	9	10	11	12
The amount of information, %	96,96	97,71	98,27	98,73	99,00	99,18
Number of components	13	14	15	16	17	18
The amount of information, %	99,33	99,47	99,59	99,67	99,75	99,81
Number of components	19	20	21	22	23	24
The amount of information, %	99,87	99,90	99,93	99,94	99,95	99,96
Number of components	25	26	27	28	29	30
The amount of information, %	99,97	99,98	99,98	99,99	99,99	99,99
Number of components	31	32	33	34	35	36
The amount of information, %	99,99	99,99	99,99	99,99	99,99	99,99
Number of components	37	38	39	40	41	
The amount of information, %	99,99	100	100	100	100	

Unit of detectors creation is assigned to generate neural networks that are considered as the basis of the detectors. The input of this unit is supplied data such as: the number of neurons in the first layer and the number of neurons in the hidden layer. The unit then generates a neural network with the given parameters and initializes the weights coefficients between neuron elements according to a random distribution. Such neural network is going in the input of training detectors unit, whose task is to train how to classify images of normal network connections and network attacks. For this purpose there (in input of training detectors) are appeared data from the training set which are selecting in the random way. Therefore, such data are unique for each neural network that provides the great structural diversity of neural immune detectors.

To train the one neural network data are used, consisting of 64 connections from one of four classes of network attacks (that is, 80 percent of all training data for the neural network), and 16 connections belonging to the class of legitimate network con-

nections (representing 20 percent of the training sample). This relation of classes in the training set was obtained by experiment and showed the best results.

As a result, each neural network is trained on parameters of the 80 network connection. While a neural network is training, there is used the controlled competitive training in accordance with the rule of "winner takes all". I.e. data from the training set – formed exactly for this neural network – are supplied to its input sequentially. Then, in the output of the neural network the weights coefficients are adjusted depending on the coherence of submitted data. Trained neural networks must pass the verification process to prove their correctness at the classification of the various images for the network traffic. A verification function is running by the unit of selection detector (see Fig. 2).

The trained neural network is checked on a specially prepared test sample consisting of the parameters of legitimate network connections. The neural network (NN) is analyzing and classifying data of the test sample. If NN detects a network attack then such NN is considered as incorrect one (because the test sample contains the legitimate traffic only) and it's destroyed by detectors destruction unit (see Fig. 2). If NN does not detect network attacks in the test sample then it "passes" the selection phase and transforms into immune neural network detector and implements into module of the traffic analysis and intrusion detection.

Verification mechanism of NN functioning – passed a training stage – eliminates the beta errors i.e. those errors when the legitimate connection is classified as a network attack. Neural network immune detectors that have successfully passed the selection stage of training are considered as the basis of the module for the traffic analysis and intrusion detection. As it comes from above an individual NN immune detector is adjusting to detect network attacks of a particular type the set of such detectors provides intrusion detection belonging to any of the classes.

To analyze the network traffic the 12 parameters of network connection are going from the module of preprocessing network traffic to the input of each NN functioning immune detector in parallel. Detectors analyze these parameters and make a decision. If all the detectors identified analyzed connection as legitimate one (not attack), it is permitted for processing and implementation. If one of the detectors at least classified the current connection as a network attack, it is blocked and the message comes about network attack.

Each neural network immune detector has the so-called "lifetime", during which it can analyze network traffic. This limited period of detectors existence is needed to get rid of the "weak" detectors. Because there is no guarantee the detector is able to detect network attacks after the training and selection of neural network detectors. More precisely, it may be a situation where the detector will be classify an unknown image as an attack in a case when the image is exactly the same as the training sample.

This situation may arise due to the fact that the training sample is randomly generated per each detector. Then the situation may occur when the neural network – which is a base for the detector on the training data – cannot identify patterns in the parameters of compounds belonging to different classes.

A mechanism that limits the operation time gets rid of the "useless" detectors, so such detectors are destroyed if they did not recognize the attack within the specified

time. A new immune detector – which was just trained and selected – substitutes the destroyed one.

The modern system of the Intrusion Detection has to protect from known network attacks as well as from unknown ones which are not previously encountered. In other words the system must have the ability for self-adapting to changed "signatures" of network attacks and methods of their organization and implementation. Such kind of functions are provided by the adaptation module (see Fig. 2) based on a study of the detected network attack and ability of neural network immune detectors to additional training.

When the network attack is recognized by one of the detectors, there is a blocking of the network connection, and a message is generated to the user. Note, apart from the above actions, the system memorizes the characteristics of the network connection, classified as the attack by immune detector. Further, the parameters of this attack are compared with the parameters of the attacks that are in the data warehouse for detectors training. If such an attack or sufficiently similar characteristics already exists, nothing happens.

However, if an attack with such characteristics does not exist in the database or the characteristics of the detected attacks are quite different from those already known, then there are made the following operations:

1. A new detector – called the immune memory detector – is created on the basis of the neural network detector that detected a network attack.
2. Additional training of the new detector is running on the parameters of the new detected attack.
3. The new up-training detector is implementing into the module of the traffic analysis and intrusion detection.
4. Parameters of the new network attack are entered into the database that stores the data for training the neural network immune detectors.

The above described algorithm enables to analyze the detected network attack. If this attack is considered as the new one then the system is adapting to it by creating the immune memory detectors and entering the characteristics (signatures) of the new network attack into the training set for next new detectors. The results of conducted experiments exploring the summarizing properties of neural network immune detectors showed that trained detectors can detect and classify not only attacks – on which they are trained – but the new attacks as well. The authenticity of detection and classification of new attacks can approach 100 percent sometimes (Tables 2).

For example a detector 1 quite enough detects an attack <dos_back> where it was trained. In particular there were detected the 100 percent attacks of such kind at the low level of beta error equal 0.2 percent. Moreover this detector can detect new attacks as well, in particular 99,1%, of <dos_neptune> attacks and 100% of r2l_spy attacks and 88,9% of <u2r_loadmodule> attacks correspondingly. There were conducted experimental researches of adaptation neural network immune detectors to the new attacks using both cloning and mutation operations. For this purpose we selected a paternal detector which was studied on the example of <DoS_land> attack. This detector found out the two new attacks: <R2L_imap> and <Probe_portsweep>. As-

suming that such attacks are absent in the database, let's generate the two new detectors of A1 and A2, and add the parameters of found attacks in the trained sample for these detectors and thus train the proper detector <clonals> (Table 3).

Table 2. Detection of network attacks by detectors 1-3

Type of attack	Detector 1 (trained on DoS_back), %	Detector 2 (trained on Probe_Nmap), %	Detector 3 (trained on R2L_ftpname), %
DoS-attacks			
Back	100,0	99,1	0,3
Land	0,0	9,5	23,8
Neptune	99,1	99,9	0,0
Pod	0,0	12,9	1,9
Smurf	0,0	0,1	0,0
Teardrop	0,0	11,0	0,0
Probe-attacks			
Ipsweep	0,1	5,7	1,0
Nmap	80,5	100	0,0
PortswEEP	2,1	30,8	0,1
Satan	13,3	96,1	2,1
R2L-attacks			
Ftp_write	0,0	0,0	100
Guess_passwd	0,0	0,0	5,7
Imap	0,0	0,0	0,0
Multihop	0,0	0,0	57,2
Phf	0,0	0,0	0,0
Spy	100,0	100	0,0
WareZclient	1,1	0,6	65,0
WareZmaster	0,0	0,0	90,0
U2R-attacks			
Buffer_overflow	0,0	0,0	83,4
Loadmodule	88,9	100	0,0
Perl	33,4	0,0	0,0
Rootkit	0,0	0,0	20,0

Table 3. Adaptation of the neural network immune detectors to new attacks

Type attacks	Detector 2 Sp(TNR)=99,0% Se (TPR), %	Detector of A1 Sp(TNR)=99,1% Se (TPR), %	Detector of A2 Sp(TNR)=98,9% Se (TPR), %
DoS-attacks			
Land	100,0	100,0	100,0
Pod	2,3	0,0	31,8
Probe-attacks			
Ipsweep	7,22	0,2	33,9
PortswEEP	15,9	2,6	55,3
Satan	11,0	31,3	11,0
R2L-attacks			
Imap	83,3	91,7	83,3
Multihop	0,0	0,0	14,3
U2R-attacks			
Perl	0,0	66,7	0,0
Rootkit	0,0	20,0	0,0

As it can be seen from a table 4 the detector A1 began better find attacks of separate classes, in particularly <Probe_satan> – in 2,8 times, and <R2L_imap> on 8,4%, and it began also to find both the <U2R_perl> attacks and the <R2L_rootkit> attacks. On the other hand the detector A2 showed better results detecting the following attacks: Probe_ipsweep, DoS_pod, Probe_portsweep, R2L_multihop.

4.2 Adaptation Ability of Cyber Defense System Detecting the Malware

The goal of this experiment is to show the adaptation ability of the proposed artificial Cyber Defense system on the example of malware detection. Let us discuss briefly the experimental conditions. Initially, we generate several immune detectors. In the Table 4 there are five detectors D1... D5 that went through training and selection phase.

Table 4. Immune Detectors

Detectors	Learning set
D1	eventvwr.exe, dllhost.exe, eventvwr.exe, fixmapi.exe, Trojan-Downloader.Win32. Bagle.f
D2	finger.exe, eventvwr.exe, loadfix.com, proxycfg.exe, Email-Worm.Win32. Brontok.q
D3	control.exe, proxycfg.exe, systray.exe, regwiz.exe, Trojan-Proxy.Win32.Lager.d
D4	forcedos.exe, rspndr.exe, share.exe, lpq.exe, Net-Worm.Win32.Bozori.k
D5	regedt32.exe, redir.exe, loadfix.com, control.exe, Trojan-Downloader.Win32. Small.dde

The learning set for each detector is unique and consists of randomly chosen legitimate files and malicious code. After training and selection neural network immune detectors check the set of malware. A Table 5 shows the detection ability of each immune detector. P_T and P_F are interdependent values that characterize the membership of under-test file into legitimate or malicious class. P_T is the probability that checked object is legitimate. P_F is the inverse value and shows the belonging of checked object to the malware class. The equations 1 and 2 describe the calculation of P_T and P_F :

$$P_T = \frac{\overline{Y_1}}{L},$$

$$P_F = 1 - P_T = \frac{\overline{Y_2}}{L},$$
(1)

$$\begin{aligned}\bar{Y}_1 &= \sum_{k=1}^L Y_1^k, \\ \bar{Y}_2 &= L - \bar{Y}_1 = \sum_{k=1}^L Y_2^k.\end{aligned}\tag{2}$$

where P_T – the probability of legitimate file; P_F – the probability of malware; Y_1 and Y_2 – the number of legitimate and malicious fragments of under-test file correspondingly; L – the total amount of fragments from under-test file, Y_{ik} – i -th output of immune detector for k -th input pattern.

Table 5. The Detection Ability of Detectors

Malware	D1, P_T/P_F	D2, P_T/P_F	D3, P_T/P_F	D4, P_T/P_F	D5, P_T/P_F
Worm.Brontok.q	0,78/0,22	0,83/0,17	0,83/0,17	0,85/0,15	0,78/0,22
Worm.NetSky.q	0,74/0,26	0,95/0,05	0,97/0,03	1,00/0,00	0,90/0,10
Worm.Rays	0,96/0,04	0,86/0,14	0,85/0,15	0,79/0,21	0,82/0,18
Worm.Bozori.a	0,78/0,22	0,93/0,07	0,95/0,05	0,99/0,01	0,88/0,12
Worm.Bozori.k	0,77/0,23	0,92/0,08	0,93/0,07	0,96/0,04	0,88/0,12
Packed.Tibs	0,77/0,23	0,96/0,04	0,97/0,03	0,99/0,01	0,92/0,08
Trojan.Dialer.eb	0,89/0,11	0,81/0,19	0,80/0,20	0,83/0,17	0,79/0,21
Trojan.Bagle.f	0,83/0,17	0,87/0,13	0,89/0,11	0,91/0,09	0,89/0,11
Trojan.INS.bl	0,86/0,14	0,86/0,14	0,84/0,16	0,86/0,14	0,81/0,19
Trojan.INS.gi	0,86/0,14	0,79/0,21	0,75/0,25	0,73/0,27	0,75/0,25
Trojan.Ladder.a	0,89/0,11	0,92/0,08	0,94/0,06	0,96/0,04	0,93/0,07
Trojan.Small.da	0,80/0,20	0,94/0,06	0,95/0,05	0,99/0,01	0,90/0,10
Trojan.Small.dde	0,77/0,23	0,95/0,05	0,96/0,04	1,00/0,00	0,90/0,10
Trojan.Small.dg	0,86/0,14	0,97/0,03	0,99/0,01	1,00/0,00	0,97/0,03
Trojan.Daemon.a	0,89/0,11	0,89/0,11	0,89/0,11	0,93/0,07	0,88/0,12
Trojan.Lager.d	0,83/0,17	0,88/0,12	0,75/0,25	0,93/0,07	0,79/0,21
Trojan.Mitglied.o	0,90/0,10	0,87/0,13	0,87/0,13	0,91/0,09	0,84/0,16
Trojan.Small.a	0,89/0,11	0,96/0,04	0,98/0,02	1,00/0,00	0,97/0,03
Virus.Bee	0,97/0,03	0,79/0,21	0,77/0,23	0,77/0,23	0,80/0,20
Virus.Neshta.a	0,90/0,10	0,74/0,26	0,72/0,28	0,72/0,28	0,72/0,28
Virus.VB.d	0,93/0,07	0,69/0,31	0,65/0,35	0,65/0,35	0,69/0,31

The fragments of files appear when we divide the checked file into the chunks with the size equals to number of input neurons of neural network immune detectors. For example, if the file size equal 16 Kbyte and we use 128 input neurons in neural network immune detector the number of fragments equal

$$L = 16 \cdot 1024 / 128 = 128\tag{3}$$

We are using the threshold that defines the belonging of checking objects to class of malware. If $P_T > 0.8$ ($P_F < 0.2$ correspondingly) then the checked object is legitimate. Conversely if $P_T < 0.8$ ($P_F > 0.2$ correspondingly) then the checked file is malicious.

A Table 6 demonstrates the ability of detectors to recognize the new unknown malware. In Tables 5 and 6 the gray cells indicate the detection of malware. As can be seen each immune detector is capable to detect not only the known malware (known malware is the malware that is from training set, for example, Trojan-Proxy.Lager.d for D3 neural networks immune detector) but also the unknown malware (for example, there are Trojan.INS.gi, Virus. Bee and Virus.VB.d. for D3 detector).

Table 6. The Detection Ability of Clones

Malware	C1, P_T/P_F	C2, P_T/P_F	C3, P_T/P_F	C4, P_T/P_F
Worm.Brontok.q	0,86/0,14	0,88/0,12	0,87/0,13	0,85/0,16
Worm.NetSky.q	0,95/0,05	0,93/0,07	0,94/0,06	0,90/0,10
Worm.Rays	0,87/0,13	0,86/0,14	0,87/0,13	0,85/0,15
Worm.Bozori.a	0,97/0,03	0,95/0,05	0,97/0,03	0,94/0,06
Worm.Bozori.k	0,94/0,06	0,91/0,09	0,92/0,08	0,90/0,10
Packed.Tibs	0,96/0,04	0,98/0,02	0,95/0,05	0,91/0,09
Trojan.Dialer.eb	0,89/0,11	0,90/0,10	0,91/0,09	0,81/0,19
Trojan.Bagle.f	0,79/0,21	0,78/0,22	0,79/0,21	0,77/0,23
Trojan.INS.bl	0,77/0,23	0,79/0,21	0,78/0,22	0,76/0,24
Trojan.INS.gi	0,64/0,36	0,66/0,34	0,65/0,35	0,68/0,42
Trojan.Ladder.a	0,75/0,25	0,76/0,24	0,77/0,23	0,77/0,23
Trojan.Small.da	0,83/0,17	0,81/0,19	0,79/0,21	0,80/0,20
Trojan.Small.dde	0,84/0,16	0,79/0,21	0,85/0,15	0,80/0,20
Trojan.Small.dg	0,77/0,23	0,77/0,23	0,78/0,22	0,76/0,24
Trojan.Daemon.a	0,73/0,27	0,79/0,21	0,78/0,22	0,74/0,26
Trojan.Lager.d	0,78/0,22	0,78/0,22	0,78/0,22	0,77/0,23
Trojan.Mitglied.o	0,78/0,22	0,77/0,23	0,77/0,23	0,78/0,22
Trojan.Small.a	0,81/0,19	0,82/0,18	0,82/0,18	0,78/0,22
Virus.Bee	0,91/0,09	0,89/0,11	0,90/0,10	0,87/0,13
Virus.Neshta.a	0,85/0,15	0,85/0,15	0,86/0,14	0,81/0,19
Virus.VB.d	0,78/0,22	0,81/0,19	0,88/0,12	0,79/0,21

As an example of adaptation ability of detectors let's chose the detector D3 that detects several malware: Trojan-Proxy.Lager.d, Trojan.INS.gi, Virus.Bee, Virus.Neshta.a and Virus.VB.d. After detection of the first malware Trojan.Win32.INS.gi detector D3 undergoes on the cloning process and as a result several detectors-clones C_i are generated. Every clone goes through the relearning process where data from the detected malware Trojan.Win32.INS.gi are included into the learning sample.

The Table 6 above demonstrates the detection ability of four clones C_i . As it can be seen each clone C_i detects Trojan.Win32INS.gi with higher rate than detector D3. In

addition the detectors-clones demonstrate the ability to detect such the new unknown malware (Trojan.Bagle.f, Trojan.INS.bl, Trojan.Ladder.a, Trojan.Small.da, Trojan.Small.dde, Trojan.Small.dg, Trojan.Daemon.a, Trojan.Mitglied.o, Trojan.Small.a) that are stayed undetectable in the case of detector D3 scanning. The process of re-training of the immune detectors where data from detected malware are used allows immune detectors to improve the detection ability, increase detection rate and permits the whole system to adapt to changeable environment by means of evolution. However while detectors-clones C_i acquire the ability to detect Trojans with high rate they practically lost the ability to detect malware from other classes. It is the back side of the processes of cloning and mutation – the detectors-clones are tuned in to the specific malware or series of specific malware.

Thus the detected new malware is added to the training data set in order to train new neural network immune detectors and it enables obtaining the new detectors with the different structure, tuned into new malware. These two mechanisms, namely cloning and training dataset modification, allow increasing the detection quality and adapting detectors to the new unknown malware. Obviously, the use of the proposed Intelligent Cyber Defense System in a number of applications mentioned in the Introduction above, will improve the reliability of detecting intrusions into computer systems, and as a result – can lead to a significant reduction in financial losses of companies from cyber attacks.

5 Conclusion and Future Work

In this paper we proposed the system for cyber defense which can detect not only already known network attacks but previously unknown, new cyber threats. This system is characterized as the intelligent, adaptive, evolutionary and self-organizing one. We examined the ability of immune detectors with neural network structure to adapt to changeable cyber-attacks trend and as result to evolve with the scope increasing the rate of unknown cyber attacks. We have run experiments proofing that detectors can adapt to the new threat.

The ability of the immune detectors to evolve – by exploring of the new malicious material - allows the intelligent security system to adapt to the new threat and provide an effective defense against the known and unknown cyber attacks. The adapted detectors acquire the ability to detect some new attacks with a higher quality. The new detected attack is adding to the training sample that increases the difference in immune detectors and enables detecting unknown attacks.

Thus, it is experimentally confirmed, that proposed neural network immune detectors are able to discover the unknown types of attacks and adapt to them.

The proposed Intelligent Cyber Defense System can increase the reliability of intrusion detection in computer systems and therefore it may reduce financial losses of companies from cyber attacks.

In the nearest future we expect to implement a part of the cyber defense system units using programmable logic arrays. Such solution, unlike the software protection, will allow eliminate the influence of the software intrusions on the cyber defense

system. To make decisions about counter invasion methods it is expected to use the Mamdani's fuzzy inference rules.

Acknowledgements

This work is running under a grant by the Ministry of Education and Sciences, Ukraine, 2016–2017. This work is supported by the Belarusian State Research Program “Informatics and Space”, 2011–2015.

References

1. Cybercrime in the world, <http://www.tadviser.ru/index.php> (In Russian)
2. Symantec it estimated the annual losses from cybercrime at \$ 114 billion, <http://www.companion.ua/articles/content?id=162264> (In Russian)
3. Forbes experts have chosen the most high-profile cyber attacks recently, <http://www.securitylab.ru/news/444700.php> (In Russian)
4. Tavallaee, M., Bagheri E., Lu, W., et al.: A detailed analysis of the KDD CUP 99 data set, In: Proceedings of the IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA'2009), 1–8. DOI: 10.1109/CISDA.2009.5356528 (2009)
5. Laheeb, M. I.: Anomaly network intrusion detection system based on distributed time-delay neural network (DTDNN). *Journal of Engineering Science and Technology*, 5 (4), 457-471 (2010)
6. Grediaga, A., Ibarra, F., García, F., Ledesma, B., Brotons, F.: Application of neural networks in network control and information security. *LNCS*, (3973), 208-213 (2006)
7. Zhang, C., Jiang, J., Kamel, M.: Comparison of BPL and RBF network in intrusion detection system. *LNCS (LNAI)*, (2639), 460-470 (2004)
8. Lorenzo-Fonseca, I., Maciá-Pérez, F., Mora-Gimeno, F., Lau-Fernández, R., Gil-Martínez-Abarca, J., Marcos-Jorquera, D.: Intrusion detection method using neural networks based on the reduction of characteristics. *LNCS*, (5517), 1296-1303 (2009)
9. Kang, B.D., Lee, J.W., Kim, J.H., Kwon, O.H., Seong, C.Y., Park, S.M., Kim, S.K.: A mutated intrusion detection system using principal component analysis and time delay neural network. *LNCS*, (3973), 246-254 (2006)
10. Cannady, J.: Applying CMAC-based online learning to intrusion detection, In: Proceedings of the International Joint Conference on Neural Networks, (IJCNN'2000), IEEE-INNS-ENNS, (5), 405-410 (2000)
11. Debar, H., Becker, M., Siboni D.: A neural network component for an intrusion detection system, In: Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, 240-250 (1992)
12. Cheng, E., Jin, H., Han, Z., Sun, J.: Network-based anomaly detection using an Elman network. *Networking and Mobile Computing, Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, (3619), 471-480 (2006)
13. Höglund, A.J., Hästönen, K., Sorvari, A.S.: A computer host-based user anomaly detection system using the self-organizing map, In: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00), (5), 411-416 (2000)
14. Jirapummin, C., Wattanapongsakorn, N., Kanthamanon, P.: Hybrid neural networks for intrusion detection system, In: Proceedings of the International Technical Conference on Circuits/Systems, Computers and Communications, Thailand, 928-931 (2002)

15. Horeis, T.: Intrusion detection with neural networks – Combination of self-organizing maps and radial basis function networks for human expert integration, Tech. report, University of Passau, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.106.191&rep=rep1&type=pdf>
16. Chimphee, W., Abdullah, A.H., Sap, M.N.M.: Anomaly-based intrusion detection using fuzzy rough clustering, In: Proceedings of the International Conference in Hybrid Information Technology (ICHIT'06), (1), 329-334 (2006)
17. Dickerson, J.E., Juslin, J., Koukousoula, J., Dickerson, J.A.: Fuzzy intrusion detection, In: Proceedings of the 20th International Conference of the North American Fuzzy Information Society (NAFIPS'01) and Joint the 9th IFSA World Congress, 3, Vancouver, Canada, (3), 1506-1510 (2001)
18. Hofmeyr, S., Forrest, S.: Immunity by design, In: Proceeding of the Genetic and Evolutionary Computation Conference (GECCO'1999), 1289-1296 (1999)
19. Burnet, F.: The Clonal Selection Theory of Acquired Immunity. Cambridge University Press (1959)
20. Burnet, F.: Clonal selection and after. Theoretical Immunology, Marcel Dekker Inc, 63-85 (1978)
21. Jerne, N.: Towards a network theory of the immune system. Ann. Immunology (Inst. Pasteur), (125C), 373-389 (1974)
22. Greensmith, J., Whitbrook, A., Aickelin, U.: Artificial Immune Systems. Handbook of Metaheuristics, 2nd edition, Springer, chapter 14, 421-448 (2010)
23. Greensmith, J., Aickelin, U.: The deterministic dendritic cell algorithm. In: Proceedings of the 7th International Conference on Artificial Immune Systems (ICARIS'2008), Springer, 291-302 (2008)
24. Coico, R., Sunshine, G., Benjamini, E. Immunology: A Short Course. Wiley-Liss (2003)
25. Murphy, K., Travers, P., Walport, M. Janeway's Immunobiology. Garland Science, 7th edition (2008)
26. Stibor, T., Mohr, P., Timmis, J., Eckert, C.: Is negative selection appropriate for anomaly detection? In: Proc. GECCO'2005, Springer, pp. 321--328 (2005)
27. Harmer, P., Williams, P., Gunsch, G., Lamont, G.: An artificial immune system architecture for computer security applications. IEEE Transactions on Evolutionary Computation, 6(3), 252--280 (2002)
28. Balthrop, J., Esponda, F., Forrest, S., Glickman, M.: Coverage and generalization in an artificial immune system, In: Proc. GECCO'2002, pp. 3-10 (2002)
29. Stibor, T., Bayarou, K., Eckert, C.: An investigation of r-chunk detector generation on higher alphabets, In: Proc. GECCO'2004, pp. 299-307 (2004)
30. Gonzales, F., Dasgupta, D., Gomez, J.: The effect of binary matching rules in negative selection, In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2003), 198-209 (2003)
31. Kim, J., Bentley, P., Aickelin, U., Greensmith, J., Tedesco, G., Twycross, J.: Immune system approaches to intrusion detection – A review. Natural Computing, Springer, 6 (4), 413-466 (2007)
32. Komar, M., Golovko, V., Sachenko, A., Bezobrazov S.: Intelligent system for detection of networking intrusion. In: Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS-2011), Prague, Czech Republic, (1) 74-377 (2011)
33. Komar, M., Golovko, V., Sachenko, A., Bezobrazov, S.: Development of neural network immune detectors for computer attacks recognition and classification, In: Proceedings of the 7th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS-2013), Berlin, Germany, (2) 665-668 (2013)

Fractal Analysis of Currency Market: Hurst Index as an Indicator of Abnormal Events

Olena Liashenko¹, Tetyana Kravets¹

¹ Taras Shevchenko National University of Kyiv

lyashenko@univ.kiev.ua, tankravets@univ.kiev.ua

Abstract. The article is devoted to analysis of currency quotes behavior on the currency market by defining dynamic changes over time. The main tool of fractal analysis is the Hurst under the hypothesis of fractal market. For 17 major currency pairs on closing prices and the prices maximum-minimum the Hurst index is calculated by formula for the adjusted R/S analysis. Values at the currency markets of different countries in different economic conditions are compared during 2008-2015. For currency pairs Hurst index tends to maintain its average value in stable economic situation, while it is an indicator of events affecting directly or indirectly on the state's economy and the rate of its national currency. Application of sliding window method allows to simulate the dynamics of Hurst index for the currency pairs USD/JPY, GBP/JPY, EUR/USD, GBP/USD and establish certain patterns of conduct series of quotes due to appropriate reaction to economic, political and natural disturbances.

Keywords. Fractal analysis, exchange rates, crisis, Hurst index

Key Terms. Model, Research, Management

1 Introduction

The transition of Ukraine to the market economy, creation of modern infrastructure, evolution of ownership and increasing independence of participants of market relations inevitably leads to the need for constant monitoring of trends and identification of features of financial and monetary system functioning. Foreign exchange market as one of the main elements of the system in the last decade characterized by increasing globalization and transformation processes.

Due to the fact that international economic relations generate the corresponding cash requirements and obligations of the parties, a prerequisite for their settlement is national currency using, as the only universal global means of payment does not exist yet. It leads to the need to exchange one currency for another in the form of purchase of foreign currency by payer or recipient of funds in international operations. The international payment transactions related to payment of receivables and liabilities of

businesses and individuals around the world are serviced by foreign exchange market, defining its objective necessity.

The features of contemporary currency markets are internationalization, globalization, standardization and automation of communication facilities in the implementation of foreign exchange transactions. Thus there is the instability of exchange rates. Predicting future behavior of exchange rates is important because it allows to reduce currency risks and ensure the efficiency of various solutions in international financial management.

The aim is to study the behavior of the currency pairs using Hurst index monitoring as one of the tools of fractal analysis, under the hypothesis of fractal market.

2 Analysis of Recent Research

During the last three decades an efficient market theory was the most famous theory of financial markets. The statement of this theory is that changes in asset prices reflect the important new information release fully and immediately. In addition, through a flow of information that can be provided between the current and the next trading period, changes in asset prices are independent. In other words, the unpredictable release of information drives asset prices in a random order, and price fluctuations comply the normal distribution [1-6].

Efficient market hypothesis (EMH), like all other economic concepts is based on linear paradigm, whereby each economic action (event) causes linearly proportional reaction that produced some cause and effect relationships. However, economic theory, based on the principles of balance, couldn't explain many complex financial phenomena. Revolution was needed and put into nonlinearity analysis.

Based on the nonlinear paradigm the fractal market hypothesis (FMH) emerged and was developed, whereby a certain action (or event) causes a nonlinear response that is exponential, unexpected, extremely strong and no one expected reaction. In contrast to the efficient market hypothesis the fractal market hypothesis states that the information is evaluated depending on the investment horizon of the investor. As different investment horizons evaluate information differently, dissemination of information is uneven also. At any certain time moment the price may not reflect all existing information, it can display only the part that is important for this investment horizon. FMH admits that chaotic mode occurs when investors lose confidence in the long-term fundamental information [1-5].

One of the results of exchange and stock markets research made by Mandelbrot, Peters [3, 1-2] is that the distribution of price changes are fractal Pareto distribution. This distribution has the property of statistical self-similarity in time (the presence of long memory). In addition, it was shown that the financial markets are nonlinear dynamic systems, which opened up opportunities for the study of financial markets by means of the theory of dynamical systems and deterministic chaos [5].

There are several alternative approaches to assessing the fractal structure of the time series: R/S-analysis; method based on the determination of cell dimensions; standard fluctuation analysis; detrend fluctuation analysis (DFA); multifractal DFA. The description and practical application of these methods can be found in [3, 5-16]. For example, using the R/S-analysis in [6] the Hurst index was estimated and it was

proved that hypothesis of FMH could be "reasonable" generalization of the efficient market hypothesis. In [12], the authors conducted an empirical study of scaling and multifractal properties of currency pair USD/DEM.

Multifractal spectra singularities for different currency pairs were studied in [13-16]. The use of different approaches to assessing the performance of fractal time series supports the hypothesis of FMH and allows to generalize it to multifractal market hypothesis (MFMH). In [8] it is demonstrated that a change of fractal properties of returns in exchange rates are an indicator of a currency crisis.

3 Research method

R/S-analysis method of study of fractal time series was proposed by Mandelbrot [3] and it is based on research conducted by the British explorer Hurst. It is based on the analysis of accumulated magnitude deviation of observations series and standard deviation. Hurst offered new statistics - Hurst index, which is widely used in the analysis of time series due to its stability. Its calculation requires minimum assumptions about studied system and time series can be classified by the type and the depth of memory on its basis. It can distinguish a random series of non-random one, even if the random series has non-Gaussian distribution [18].

Calculation of Hurst index can be carried out as follows: $R/S = (aN)^H$ or $H = \frac{\log(R/S)}{\log(aN)}$, where H - Hurst index; S - standard deviation of observations number; R - variation of accumulated deviation; N - number of observation periods; a - positive constant [9, 17].

The scale of the accumulated deviation R is the most important element of Hurst index formula: $R = \max_{1 \leq u \leq N} (Z_u) - \min_{1 \leq u \leq N} (Z_u)$, where Z_u - the accumulated deviations number of values x from the mean \bar{x} , i.e. $Z_u = \sum_{i=1}^u (x_i - \bar{x})$. Formula for Hurst index shows that the increasing scale, reducing standard deviation and reducing number of observations influence on its growth.

For further calculations we use $a = \frac{\pi}{2} \approx 1,5708$ as the choice of another constant for calculating. Hurst index inflates its value significantly. It will lead to erroneous conclusions about persistence of random series [9].

While a small number of observations, actual calculations of normalized variation R/S for random series give much too low results compared to theoretical ones

$R/S = \sqrt{\frac{N\pi}{2}}$. This contradiction leads to lower values of Hurst index when the number of observations $N < 250$. To avoid this contradiction, it is necessary to transform actually calculated values of normalized variation using the formula [9]:

$$R / S_T = R / S \times 0,998752 + 1,051037. \quad (1)$$

However, due to feature of logarithmic calculations of Hurst index, the adjusted value of normalized variation will contain a minor error also. On the basis of the correlation between the number of observations and the ratio of the standard and actual Hurst index it is necessary to adjust the formula for Hurst index calculation so that its value was close at most to a standard one for random series ($H = 0,5$) for all N . The final formula is [9]:

$$H_T = \frac{\log(R / S_T)}{\log(\pi \times N / 2)} \times (-0,0011 \times \ln N + 1,0136) \quad (1)$$

So Hurst index above 0.5 confirms the presence of long-term memory of the market: current depends on the past and the future depends on the present.

The economic literature is usually gives recommendation to calculate the accumulated variation on closing prices. However, for practical market trade minimum and maximum prices set in the middle of interval are also important. To calculate the Hurst index on the maximum-minimum prices we use the formula [9]:

$$H_M = \frac{\log(R / S_T)}{\log(\pi \times (N - 1) / 2)} \times (-0,0011 \times \ln(N - 1) + 1,0136) \quad (3)$$

Testing hypotheses about market on the basis of Hurst index can be done in case of data mixing. If the result of calculations on randomly mixed data is Hurst index close to 0.5, and it is different from the actual calculations, it may indicate that some data is not Brownian motion.

4 Results

The studies calculated the Hurst index by adjusted formulas for 17 major currency pairs on annual data within 2008 - 2014. Figure 1 presents the dynamics of Hurst index change, which is calculated on closing prices for different currency pairs.

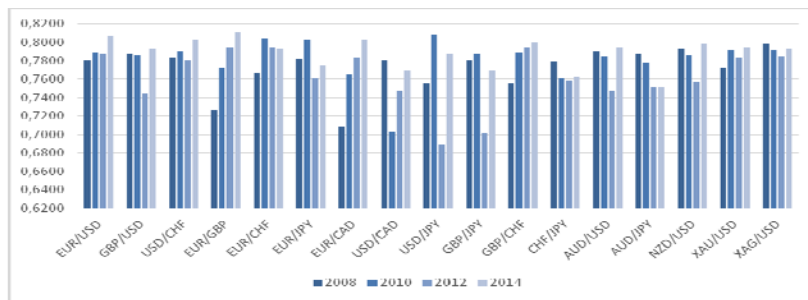


Fig. 1. Dynamics of Hurst index, calculating on closing prices

Note that the values of the Hurst index are higher than 0.68, i.e. series are persistent. Herewith there is the tendency to change the depth of long-term memory for different periods. Most pairs characterized by increase in the degree of persistence an average over the period 2008-2014, which means the stabilization of the economic situation in the world, overcoming of the global economic crisis.

However, for some currency pairs, especially USD/JPY, USD/CAD, GBP/JPY, Hurst index dynamics is characterized by significant fluctuations throughout the study period.

Currency pairs EUR/USD, GBP/USD, USD/CHF, USD/JPY are highly liquid financial instruments that are characterized by significant volatility and therefore have great potential for profit. The most popular is EUR/USD. The EUR is highly dependent on interest rates, economic conditions in the euro area, policies of central banks of the US and EU, political stability in the world [19].

Currency pair GBP/USD is one of the most moving and aggressive currency pairs. The dynamics of the currency pair largely follows the trend of the currency pair EUR/USD.

The feature of the currency pair USD/CHF is more dependence of its change of information on the economic situation in the United States than in Switzerland. The growth trend of US dollar against the Swiss franc, the tendency of weakening of other currencies against the US dollar is observed. This is especially true for currency pairs USD/CAD, AUD/USD [18].

Japanese yen - one of the world reserve currency, a tool for international settlements of countries with the lowest short-term interest rates. There is a significant correlation between currency pairs USD/JPY, EUR/JPY and CHF/JPY. Since the bulk of Japan's funds invested in European assets, changing the course of the currency pair EUR/JPY to a large extent depends on the level of interest rates of Eurozone and Japan.

Figure 2 presents the dynamics of Hurst index change, which is calculated by the maximum-minimum prices for various currency pairs.

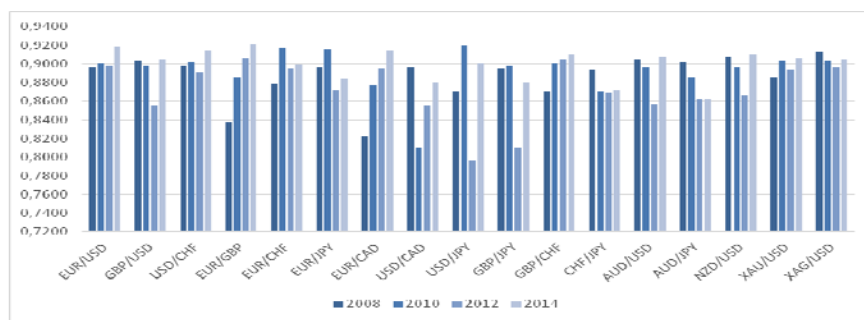


Fig. 2. Hurst index dynamics, calculated by the maximum-minimum prices

Comparison of the results (Fig. 1, 2) suggests a similarity diagrams and increasing values of Hurst index, designed for maximum-minimum prices compared with the Hurst index for closing prices. Calculation of average values of years gives identical pattern shift 0.1.

For further study consider a pair of the most pronounced drop Hurst index: USD/JPY and GBP/JPY. Using the method of sliding windows simulate of dynamic change of Hurst index for these pairs in the time period 2010-2015 (Fig. 3). On the horizontal axis the right end of the time window is marked. There is a consistency of behavior expected of these pairs, due to the presence of the yen and the interconnect-edness of economies of the US and Britain as the developed countries, members of the Group of Seven.

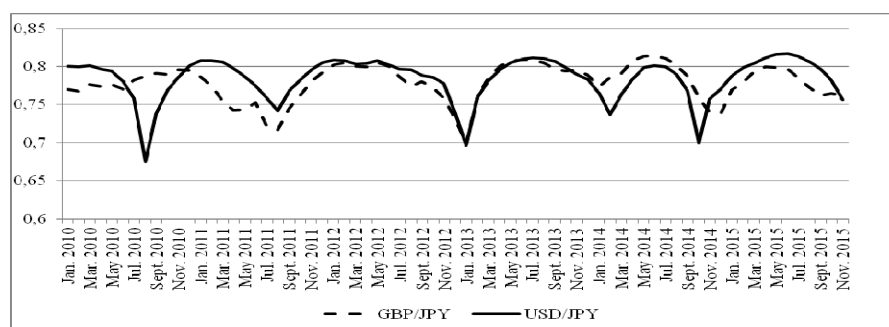


Fig. 3. The dynamics of Hurst index for closing prices of USD/JPY and GBP/JPY

The graphs in Fig. 3 are typical manifestations of cyclical and periodic drop of Hurst index in relatively insignificant period of time. Consider in more detail the dynamics of Hurst index for pair USD/JPY, because it is one of the most influential in the foreign exchange market.

The first decline of Hurst index and thus persistence weakening can be seen in the summer of 2010. It is a consequence of the global currency crisis 2008. The event which caused the decrease in the degree of predictability of the market currencies, was active company fight for the preservation of the national economies. It was in 2010, the government of Japan has allocated 1 trillion yen for forming a reserve to combat the crisis and restore regions, and in summer 2010 "new growth strategy" was adopted and a significant reform of the economy of rising sun was conducted. Large inflows of funds, coupled with the instability of the economic situation led to a drop in Hurst index, reducing the persistence [20].

Thus, the events were a shock for the foreign exchange market and caused a drop in Hurst index. During these time periods it was difficult to predict further developments for the currency pair USD/JPY, as its behavior has not been dictated by the internal events in the market, but external, such as monetary policy in Japan and the US.

Figure 4 presents the dynamics of Hurst index change for relatively stable and very influential pairs in the foreign exchange: EUR/USD and GBP/USD.

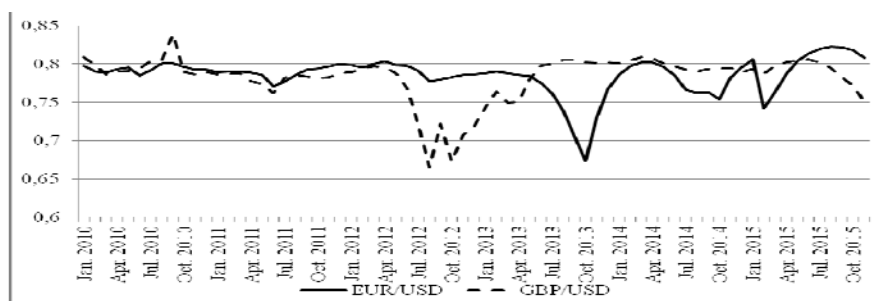


Fig. 4. The dynamics of Hurst index, calculated at closing prices for EUR/USD and GBP/USD

For the currency pair EUR/USD we could see only one, but significant "failure" of Hurst index, which happened in the fall of 2013. The cause of this could be the announcement by governments of the US and EU about the start of negotiations for the establishment of so-called Transatlantic trade and investment partnership (TTIP). However, the very mention of this agreement has caused some instability in the economy, many meetings and public criticism from the media.

5 Conclusions

Hurst index as a tool of fractal analysis allows to determine the degree of persistency of financial series, the presence of the long-term memory at foreign exchange market. In stable economic situation the Hurst index for currency pairs tends to maintain their average. However, this index is very responsive to events that directly or indirectly affect the state's economy and the rate of its national currency. The biggest jump of index can be seen when the country holds planned intervention to improve their economic situation, thus reducing its currency.

Another group of events affecting the persistence of the currency market are important for the country's situations, such as changing the government, announcement of a new policy or natural disasters. All this violates the usual course of events in the currency markets, and changing of the Hurst index of exchange quotations is the indicator of this. The foreign exchange market disturbance passes over the index approaches to its average value inherent in each currency pair.

References

1. Peters, E.E.: Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility, John Wiley&Sons, Inc. (1996)
2. Peters, E.E.: Fractal Market Analysis: Applying Chaos Theory to Investment and Economics, John Wiley&Sons, Inc. (1994)
3. Mandelbrot, B.B., Hudson, R.L.: The (Mis)behavior of Markets: A Fractal Views of Financial Turbulence, Basic Books, (2005)
4. Soros, G.: The New Paradigm for Financial Markets: The Credit Crisis of 2008 and What It Means, NY: Public Affairs (2008)

5. Crownover, R.M.: Introduction to Fractals and Chaos, Jones and Bartlett (1995)
6. Corazza, M., Malliaris, G.: Multifractality in Foreign Currency Markets. *Multinational Finance Journal*, 6, 387–401 (2009)
7. Benbachir, S. & El Alaoui, M.: A Multifractal Detrended Fluctuation Analysis of the Moroccan Stock Exchange. *International Research Journal of Finance and Economics*, 78, 6–17 (2011)
8. Mansurov, A.K.: Forecasting Currency Crises by Fractal Analysis Techniques. *Studies on Russian Economic Development*, 19(1), 96–103 (2008)
9. Naiman, E.: The Hurst Index Calculation to Identify Persistence of the Financial Markets and Macroeconomic Indicators. *Ukrainian Journal Ekonomist*, 10, 18–28 (2009)
10. Nava, N., Di Matteo, T., Aste, T.: Anomalous Volatility Scaling in High Frequency Financial Data. *Physica A*, 447, 434–445 (2015)
11. Morales, R., Di Matteo, T., Gramatica, R., Aste, T.: Dynamical Generalized Hurst Exponent as a Tool to Monitor Unstable Periods in Financial Time Series. *Physica A*, 391(11), 3180–3189 (2011)
12. Xua, Z., Gencay, R.: Scaling, Self-similarity and Multifractality in FX Markets. *Physica A*, 323, 578–590 (2003)
13. Drozd, S., Kwapien, J., Oswiecimka, P., Rak, R.: Foreign Exchange Market: Return Distributions, Multifractality, Anomalous Multifractality and Epps Effect. *New Journal of Physics*, 12, online <http://iopscience.iop.org/article/10.1088/1367-2630/12/10/105003/pdf> (2010)
14. Oh, G., Eom, C., Havkin, S., Jung, W.-S., Wang, F., Stanley, H.R., Kim, S.: A Multifractal Analysis of Asian Foreign Exchange Markets. *Eur. Phys. J. B.* online <http://polymer.bu.edu/hes/articles/oehjwsk12.pdf> (2012)
15. Norouzzadeh, P. & Rahmani, B.: Multifractal Detrended Fluctuation Description of Iranian Rial – US Dollar Exchange rate. *Physica A*, 367, 328–336 (2006)
16. Gunay, S.: Source Of The Multifractality In Exchange Markets: Multifractal Detrended Fluctuations Analysis. *Journal of Business & Economics Research*, 12(4), 371–384 (2014)
17. Schroeder, M.: *Fractals, Chaos, Power Laws: Minute from an Infinite Paradise*, New York: W.H.Freeman and Co. (1991)
18. Cannon, M.J., Percival, D.B., Caccia, D.C., Raymond, G.M., Bassingthwaigte, J.B.: Evaluating Scaled Windowed Variance Methods for Estimating the Hurst Coefficient of Time Series. *Physica A*, 241(3-4), 606–626 (1997)
19. Feldstein, M.: The Failure of the Euro. *The Foreign Affairs.* online <https://www.foreignaffairs.com/articles/europe/2011-12-13/failure-euro> (2012)
20. Harari, Daniel. Japan's Economy: from the 'Lost Decade' to Abenomics. House of Commons Library, Standard Note SN06629. London: Oct 24, 2013

Using Hybrid Algorithms Based on GMDH-Type Neural Networks for Solving Economic Problems

Olha Moroz

International Research and Training Centre of Information Technologies and Systems
of NAS and MES of Ukraine, Kyiv, Ukraine

olga_moroz@irtc.org.ua

Abstract. This survey deals with up-to-date results in the field of hybrid algorithms development of GMDH-type Neural Networks (GMDH-NN) and other methods of Artificial Intelligence (AI) which are successfully used for solving complex economic problems. Such hybrid algorithms are now only in its early stage of active research. General characteristics and main weaknesses of GMDH-NN are firstly presented. The paper further gives brief information on some AI paradigms such as Swarm Intelligence and Evolutionary Computation. Then known hybridization cases of GMDH-NN and certain methods of these paradigms (Genetic Algorithms, Differential Evolution, Particle Swarm Optimization, and Genetic Programming) are considered eliminating a number of the networks shortcomings. In the future it is worth to study the most promising ways of GMDH-NN hybridization with other methods of AI in order to increase their efficiency and extend applications.

Keywords. GMDH-type Neural Networks, Hybridization, Particle Swarm Optimization, Genetic Algorithms, Differential Evolution

Key Terms. MachineIntelligence, Model, Computation, BioInspiredApproach

1 Introduction

Methods of AI have been used successfully in real-world applications, for example, predicting, data mining, classification, pattern recognition, knowledge discovery, system identification, clustering, control etc. However, there are many problems of such complexity that no method of AI can solve them alone.

The current trend is in developing composition of AI methods to obtain synergic effect, or hybridization in other words, since no one method is superior to the others in all situations. In doing so, methods respectively aim to enhance strengths of the components of the hybrid system and eliminate weaknesses of individual components. Due to the possibilities of hybrid systems solve many real problems that are characterized by inaccuracy, unpredictability, uncertainty, high dimension and variable environment, they attract much attention. Thus, they can be used to solve problems in a non-standard and promising way.

The Group Method of Data Handling (GMDH) is a powerful method based on the principles of self-organization. Basic Multilayered Iterative Algorithm MIA GMDH is an original Polynomial Neural Network of perceptron type. It should be noted that research in the field of hybridization GMDH-NN with other methods of AI are now in its early stage. But the first results of these studies indicate already a strong potential, high efficiency, diversity of applications of hybrid models. For example, GMDH not suitable for modeling of multi-output systems, while hybrid one with Multi Objective Genetic Algorithm is able to solve this problem.

This survey is aimed to analyzing and structuring typical known approaches to hybridization of GMDH-NN and AI methods as powerful tool for solving complex problems of economic modeling as well as determining most expedient ways of developing research in this field.

2 GMDH as a Specific Type of Neural Network

The GMDH was developed by Ivakhnenko [1] as a multivariate analysis method for complex systems modeling and identification. The main idea of GMDH is to build an analytical function in a feedforward network based on a quadratic node transfer function whose coefficients are obtained using regression technique. Its structure is very similar to that of multilayer feedforward neural networks but the numbers of layers as well as nodes are objectively defined by an external criterion in accordance with the incompleteness theorem. To compare and select the best model, external criteria are used based on dividing the sample into two sets (training and testing), that eliminates the overfitting problem [2].

GMDH is ideal for complex, unstructured systems where the investigator is only interested in obtaining a high-order input-output relationship. Hence this can be treated as a good data mining tool where data is transformed into knowledge for decision making. The most pronounced feature of GMDH is that it can choose the really significant input variables among dozens of these, thus actually reducing the dimension of a solved problem.

There are two main problems in GMDH-NN modeling such as training parameters and selection of an optimal topology, which have a great impact on their performance and encourage searching more efficient methods for their solution. The mathematical basis for the vast majority of training algorithms for NNs is to utilize gradient information to adjust the connection weights between nodes in the network. Its use implies drawbacks such as slow convergence (training) rates, neglected multiple extremum points, infinitesimally small step sizes (e.g. learning rates), costly computation, no guarantee that algorithm converges to an optimum point, entrapped in local minimum points, necessarily imply a least-mean-squared-error criterion, non-differentiability of many error function [3].

Selection of the optimal topology often has the following problems:

- the search space of possible topologies is infinitely large, complex, multimodal, and not necessarily differentiable;
- there is little reason to expect that Neural Network can find a uniformly best algorithm for selecting the weights in a feedforward artificial neural network. At pre-

sent, neural network design relies heavily on human experts who have sufficient knowledge about the different aspects of the network and the problem domain;

- as the complexity of the problem domain increases, manual design becomes more difficult and unmanageable.

Anastasakis and Mort [4] have carried out a comprehensive study of the shortcomings of GMDH, the most problematic can be stated such as:

- ill-suited to solve complex problems with many inputs, almost equally important;
- uses local methods to find optimal solutions;
- tends to create complex polynomials for relatively simple systems;
- tends to create highly complex networks (models) when it comes essentially nonlinear systems, limitations due to its overall structure;
- does not objectively evaluate coefficients by least squares;
- has a fixed structure and determined search character for a better model;
- does not effective in addressing the multi-task.

3 Methods of Evolutionary Computation and Swarm Intelligence Paradigms

Evolutionary Computation [5], [6] is a paradigm in the Artificial Intelligence domain that aims at benefiting from collective phenomena in adaptive populations of problem solvers utilizing the iterative progress comprising growth, development, reproduction, selection, and survival as seen in a population. EC are the most well-known, classical and established algorithms among nature inspired ones based on the biological evolution in nature that is being responsible for the design of all living beings on Earth, and for the strategies they use to interact with each other. EC employ this powerful design philosophy to find solutions of hard problems. EC are nondeterministic or cost based optimization algorithms.

Usually grouped under the term EC, the domains of Genetic Algorithm (GA), Genetic Programming (GP), Differential Evolution (DE), Evolutionary Strategy (ES), Learning Classifier Systems (LCS), Estimation of Distribution Algorithms (EDA) and most recent Paddy Field Algorithm (PFA) are founded. They all share a common conceptual base of simulating the evolution of individual structures and they differ in the way the problem is represented, processes of selection and the usage/implementation of reproduction operators. The processes depend on the perceived performance of the individual structures as defined by the problem.

Swarm Intelligence [7] is a collective behavior of decentralized, self-organized systems, natural or artificial. The term swarm is used for an aggregation of animals such as fish schools, bird flocks and insect colonies such as ant, termites and bee colonies performing collective behavior. The individual agents of a swarm behave without supervision and each of these agents has a stochastic behavior due to her perception in the neighborhood. Local rules, without any relation to the global pattern, and interactions between self-organized agents lead to the emergence of collective intelligence called swarm intelligence.

Swarm Intelligence can be described by considering five fundamental principles. Proximity Principle: the population should be able to carry out simple space and time

computations. Quality Principle: the population should be able to respond to quality factors in the environment. Diverse Response Principle: the population should not commit its activity along excessively narrow channels. Stability Principle: the population should not change its mode of behavior every time the environment changes. Adaptability Principle: the population should be able to change its behavior mode when it is worth the computational price. The most popular Swarm Intelligence algorithms are Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony Optimization (ABC), Cuckoo Search (CS), Fiery Algorithm (FA), Intelligent Water Drops (IWD), Gravitational Search Algorithm (GSA) and Charged System search (CSS).

4 Hybrids of GMDH-NN with Some Computation Intelligence Methods

Hybrid with DE. In [8], [9], the hybrid of DE and GMDH systems have created and clearly showed that this structure is superior to the conventional GMDH approach. The architecture of model is not predefined but can be self-organized automatically during the design process. The hybrid Differential Evolution and Singular Value Decomposition is used for simultaneous parametric and structural design of GMDH networks used for modeling and prediction of various complex models. The DE-GMDH approach has been applied to the problem of developing predictive model for tool-wear in turning operations; the exchange rate problem; the Box-Jenkins gas furnace data; with experimental results clearly demonstrating that the proposed DE-GMDH-type network outperforms the existing models both in terms of better approximations capabilities as well as generalization abilities.

Hybrid with GP. In [10], [11] GMDH-based approach to Genetic Programming, which integrates a GP-based adaptive search of tree structures, and a local parameter tuning mechanism employing statistical search is presented. In traditional GP, recombination can cause frequent disruption of building blocks or mutation can cause abrupt changes in the semantics. To overcome these difficulties, traditional GP with a local hill climbing search using a parameter tuning procedure is supplemented. More precisely, the structural search of traditional Genetic Programming with a multiple regression analysis method and establish adaptive program called "STROGANOFF" (i.e. Structured Representation On GAs for Nonlinear Function Fitting) is integrated. The fitness evaluation is based on a "Minimum Description Length (MDL)" criterion, which effectively controls the tree growth in Genetic Programming. Its effectiveness by solving several system identification (numerical) problems and compare the performance of STROGANOFF with traditional GP and another standard technique (i.e. "radial basis functions") is demonstrated. The effectiveness of this numerical approach to Genetic Programming is demonstrated by successful application to computational finances.

Hybrid with GA. In [12], [13], [14] a hybrid of GA and GMDH systems have created which is superior to the conventional GMDH method. These papers present specific encoding schemes to genetically design GMDH-NNs based on using hybrid GAs and Singular Value Decomposition to design the coefficients as well as the connec-

tivity configuration of GMDH-NNs used for modeling and prediction of various complex models in both single and multi-objective Pareto based optimization processes. Such generalization of network's topology provides near optimal networks in terms of hidden layers and/or number of neurons and their connectivity configuration, so that a polynomial expression for dependent variable of the process can be achieved consequently.

Hybrid with PSO. Such algorithms have been proposed in [15], [16]. The principal approach to modeling of PSO-GMDH and traditional GMDH is more or less same. In PSO the whole population (of constant size) of swarm particles progresses iteratively until the optimum solution is found. Iterative process of Particle Swarm Optimization can be compared with layered approach of GMDH where swarm particles search for better position in each iteration as GMDH nodes look for better solution in each layer. The proposed hybrid PSO-GMDH uses a heuristic search process which makes it more attractive for efficiently searching large and complex search spaces. It is likely that solution found by traditional GMDH is trapped into local minimum whereas PSO's domain of search space is infinitely large and it has its internal mechanism to position is used for simultaneous parametric and structural design of GMDH networks used for modeling and prediction of various complex models.

Group of Adaptive Model Evolution (GAME). P. Kordik [17] created the GAME as a hybrid GMDH-based self-organizing modeling system which uses neurons (units) with several possible types of transfer functions (linear, polynomial, sigmoid, harmonic perceptron network, etc.). The GAME is an original data mining method. It can generate models for classification, prediction, identification or regression purposes. It works with both continuous and discrete variables. The topology of GAME models adapts to the nature of a data set supplied. The GAME is highly resistant to irrelevant and redundant features, suitable for short and noisy data samples. The GAME engine further develops the MIA GMDH algorithm. A GAME model has more degrees of freedom (units with more inputs, interlayer connections, transfer functions etc.) than MIA GMDH models. GAME engine also use genetic search to optimize the topology of models and also the configuration and shapes of transfer functions within their units.

Drawbacks of the up-to-date hybrids. The hybrid structures developed by researchers abroad Ukraine are typically based on classical MIA GMDH algorithm. Those researchers generally do not take into account the fact that during last decades in Ukraine there was developed and widely used another types of GMDH algorithms, namely combinatorial COMBI [18] and relaxational-iterative RIA [19] ones. Moreover, recently it was shown that hybrids of these three types of GMDH algorithms generate new, more effective GMDH structures named generalized iterative algorithm GIA [20]. Accordingly, these new developments can be used to create more powerful hybrid structures with various types of Computation Intelligence paradigms.

5 Some Examples of Successful Use of Hybrid Structures for Solving Complex Economic Problems

Example 1. The cement industry is one of the most important and profitable industries in every developed country and great content of financial resources are investing in this sector yearly. In [21] GMDH-type neural network and genetic algorithm is developed for stock price prediction of cement sector in Iran. Genetic algorithm is arranged in a new approach to design the whole architecture of the GMDH-type-NNs. It provides the optimal number of neurons in each hidden layer and their connectivity configuration to find the optimal set of appropriate coefficients of quadratic expressions to model stock prices. The results are very encouraging and congruent with the experimental results.

Example 2. The prediction of stock price is an important task in investment and financial decision-making since stock prices/indices are inherently noisy and non-stationary. In [22] a GMDH type-neural network based on Genetic algorithm is used to predict stock price index of petrochemical industry in Iran. The results obtained by using GMDH type neural network are in excellent agreement with the experimental results and has high performance in stock price prediction.

Example 3. Forecasting of short-term water demand is useful in planning and management of water and wastewater facilities such as pump scheduling, control of reservoirs volume, pressure management and water conservation program. This helps network managers to decrease vulnerability of the system and consumers and to increase network reliability. In [23] two types of neural networks: Multi Layered Feed-Forward with backpropagation learning algorithm and GMDH with genetic learning algorithm were investigated to model the water demand forecast in Tehran city. The comparison reveals that the GMDH neural network with GA produces results that are close to the actual data.

Example 4. Wheat is the chief carbohydrate source which is taken to be human's food too. Considering the irregular population augmentation in the world the provision of food stuffs is a predicament of intense significance. Prediction of wheat yield using different quantity microelements is an important issue as a key parameter in increase of wheat production. In [24] wheat yield in Iran is modeled and predicted using GMDH-type neural networks based on some experimental data. The aim of such modeling is to show how wheat yield change with the variation of levels of fertilizers (Zinc & Iron). In this way, GA are deployed in a new approach to design the whole architecture of the GMDH-type neural networks, i.e., the number of neurons in each hidden layer and their connectivity configuration, in combination with using regression method to find optimal set of appropriate coefficients of quadratic expressions for modeling and prediction of wheat yield.

Example 5. Accurate forecasting of electricity load is one of the most important issues in the electricity industry, economic and reliable power systems operation such as unit commitment, reducing spinning reserve, maintenance scheduling, etc. It is essential part of an efficient power system planning and operation. Short-term load forecasting (STLF) is a difficult task because the accuracy of forecasting is influenced by many unpredicted factors such as economic, temperature, etc. whose relationships are commonly complex, implicit and nonlinear. The modified locally weighted group

method of data handling (M-LWGMDH) based on genetic algorithm is proposed in [25] to solve the STLF problem. The GA is used to design the whole architecture of M-LWGMDH network.

6 Conclusion

Hybridization of GMDH-NNs and methods of Artificial Intelligence is a new field of investigation that is actively developed by researchers around the world. As a main result of this paper, it may be pointed out that the most promising ways of research activities in the field of modeling structures hybridization are the followings:

- creating new hybrid structures based on other types of GMDH algorithms and methods of Artificial Intelligence which have not yet been used for hybridization;
- exploring which kinds of hybrids are more successful when solving a specific class of applied problems.

References

1. Bezgin, K. S.: Quality management of business processes in the enterprise. Manuscript (2009)
2. Kozenkov, D.: Business Processes Planning as the Basis of the Creation Enterprise Architecture. The Visnyk of the SSU 3. 126--136 (2011)
3. Harrington, H. J., Esseling, E. K.C., van Nimwegen. H.: Business Process Improvement Workbook: Documentation, Analysis, Design, and Management of Business Process Improvement. London: McGraw-Hill, 1996
4. Fedorov, I.: Modeling business processes in BPMN 2.0 notation. M., MESI (2013)
5. Davenport, T.: Process Innovation: Reengineering Work through Information Technology., Boston, MA.: Harvard Business School Press (1993)
6. Druchenko, L.: Implementation of Business Processes Modeling and Optimization in the Local Government. Public Administration: Theory and Practice 1(3) (2010)
7. Ponomarenko, V., Minuhin, S., Znahur, S.: Theory and Practice of Business Process Modeling: Monograph. Simon Kuznets Kharkiv National University of Economics (2013)
8. Manganelli, R., Klein, M.: The Reengineering Handbook: A Step-By-Step Guide to Business Transformation. New York: Amacom (1994)
9. Hammer, M., Champy, J.: Reengineering the Corporation: A Manifesto for Business Revolution. London : Nicholas Brealey Publishing (1993)
10. Taranyuk, L., Zaporozhchenko, O.: Organizational Principles of Reengineering Business Processes of Enterprises. Mechanism of Economic Regulation 4, 122--131 (2011)
11. Kuzmin, O., Sidorov, Y., Kozik, V.: Achievements and Problems of Evolutionary Economics: Monograph. Publishing House of Lviv Polytechnic National University (2011)
12. Reilly, J. P., Creaner, M. J.: NGOSS Distilled, The Lean Corporation (2005)
13. Kalyanov, G.: Business Processes Modeling, Analysis, Reorganization and Automation. Finance and Statistics (2007)
14. Melnikov, A.: Bases of Business Organization. Center of educational literature (2013)

15. Korzachenko, O.: The concept of business processes modeling and optimization of telecommunications companies. Scientific and theoretical journal Khmelnytsky University of Economics "Science and Economics" 4 (32), 247--253 (2013)
16. Gruber, T.: A Translation Approach to Portable Ontologies. Knowledge Acquisition, 5 (2), 199--220 (1993)

Modeling Input Financial Flows of Insurance Companies as a Component of Financial Strategy

Tatsiana Verezubova¹, Tetiana Paientko²

¹ Belorussian State Economic University

² Kyiv National Economic University Named After Vadym Hetman

verezubova@mail.ru, tpayentko@mail.ru

Abstract. The financial strategy of an insurance company has a significant impact on the success of that company's activity. A well-developed financial strategy is able to decrease financial risks and guarantee the discharge of insurance companies' liabilities in the long-run. One of the main components of financial strategy is financial flow modeling. Mathematical modeling allows a more accurate determination of the financial support of the tasks envisaged in the corporate strategies of insurers. Insurance companies located in the post-Soviet region have less experience with financial flow management. These companies are more vulnerable because they do business in an economically and sociologically very undeveloped environment. The aim of this article is to explain the role of financial flow modeling in order to develop financial strategies for insurance companies in the post-Soviet region. The methodology of this research is correlation and regression analysis. The result of the research is a regression equation which is useful to predict input financial flows of an insurance company. The proposed method of calculation will be used to support forecasts of premiums and revenues from other types of insurance.

Key words. Financial flow, insurance company, financial strategy, forecast

Key terms. Model, regression analysis, correlation

1 Introduction

The financial strength of insurance companies depends on efficient financial strategies and management of financial flows. Optimal movement of financial flows is able to create favorable conditions for the development and competitiveness of an insurance company. Imbalances in the movement of financial flows increase the risks for insurance companies and may undermine their ability to discharge liabilities.

Developing a financial strategy is closely linked with financial flow modeling. In accordance with modern economic science, financial strategy development can include mathematical modeling. This is achieved mainly through the formalization of the main features of the activities of the entity, identifying the relationship of its parameters in a mathematical form. Thus it is possible to determine their dynamics

with mathematical formulas that describe different economic processes, as well as measuring their reaction to external and internal factors. Information technology and the increased use of innovative financial instruments of analysis and forecasting are transforming economic relations. This increases the usefulness of simulation to optimize current and future challenges in the macro and microeconomic levels of management. The importance of mathematical modeling is also gaining in the financial sector, which is currently characterized by constant changes in market conditions, subject to frequent fluctuations in exchange rates, as well as uncertainty in the growth of cash turnover. With the increasing influence of many external and internal factors on the stability of insurers' risk, this complicates making clear predictions of final results of their activity. Insurance companies which are located in the post-Soviet region have less experience in financial flow management. These companies are more vulnerable because they act in an economically and sociologically undeveloped environment. The aim of this article is to explain the role of financial flow modeling for developing financial strategies for insurance companies in the post-Soviet region.

2 Theoretical and Methodological Background

One important path in the development of a reliable financial strategy of any organization is having methods of economic-mathematical modeling of financial flows. They allow a more accurate determination of the needed financial support for the tasks envisaged in the corporate strategies of insurers. It should be noted that the development of economic process modeling has a long history. The concept of constructing mathematical models from different forms of administrative processes appeared in foreign literature in the middle of the nineteenth century. Modern publications are connected to different fields of mathematical modeling. Some of them are connected to optimization tasks of different economic processes (Sethi and Thomson; 2000). The most important for our research are the mathematical models which allow predicting different financial indicators. Statistical methods for forecasting were very well explained by Abraham and Ledolter (1983). The need and possibilities of a new approach for forecasting is shown in the work of Wieland and Wolters (2013). The main characteristics of capital in the twenty-first century are explained in a publication by Piketty (2014). New views on capital also require new approaches to financial flow modeling. This was confirmed by Nedopil (2009) and Cornelius (2003). The role of macro financial modeling was explained by Bernanke, Gertler, and Gilchrist (1999). Insurance companies have their own special features for financial flow modeling as was explained by Kuester and Wieland (2010). In this paper we used correlation and regression analysis to predict input financial flows of insurance companies.

3 Efficiency Estimation Procedure

We consider it appropriate forecasting the financial capacity of insurers by beginning with a study of trends in the dynamics of related income and identification of the impact of the dominant factors on the process. This allows us to develop a more realistic approach to the forecasting of insurance premiums in the future compare to approaches which were developed by Belarussian analysts. For example, the forecasts of Belarussian Ministry of Finance (Ministry of Finance, 2015) and some investment companies (Yupiter, 2015) are more optimistic. They believe that insurance premiums will continue to increase. The growth of insurance premiums will be higher for non-life insurance companies than for life-insurance.

In order to achieve the intended purpose, a methodology of determining the value of insurance premiums in the future was developed. It includes the following stages:

- study of the organizational and assortment structure of a particular insurer and identify its priority types of insurance;
- select the basic kinds of insurance services (one to three) and assess the dynamics of income premiums in the past period, as well as depending on the primary exogenous (external) factors influencing changes in their volume;
- calculate predictable amounts of premiums on the proposed perspective on selected major types of insurance by using correlation and regression analysis;
- determine the projection of revenue premiums of other types of insurance through the calculation of the arithmetic mean values between their average annual growth rate for the previous period;
- study of a complex economic and mathematical model that describes the total amount of the forecast of revenues for the future scheduled amount of insurance payments on all types of insurance.

In our view, for forecasting the whole magnitude of premiums in the medium term it is advisable to use the proposed methodology for generating a moderate financial strategy for the insurance companies for a five-year development period.

Based on an examination of the terms of priority insurance services, exogenous factors affecting the amount of premiums income were identified. Among them are: employed population (x_1), nominal wages accrued (x_2), and the number of legal entities (x_3).

The model is based on the information on insurance premiums received on a quarterly basis by Belarussian insurance companies between 2003-2013. All data has been adjusted for inflation and organized as panel data. Panel analysis was used because a large number of observations increases the number of degrees of freedom. This reduces the dependence between the variables and the degree of errors.

Correlation and regression analysis to identify linkages were used. It confirmed the dependence of insurance contributions on selected factors. Verification of the existence of satisfactory relationship variables was done through a pair-wise evaluation of correlation. As a result, a correlation matrix was received (Table 1). Established ratios of correlation confirm that the dependent variable y_1 is intertwined with all variables x (over 50%). The closest link identified with variable x_2 is 99.9%, because the premiums directly depend on salary size. Wages are the dominant source of income for

Belarussian citizens. Also, 60% of insurance services are compulsory in Belorussia and premiums depend on wage amount. The study found that independent variables x_1 , x_2 , x_3 do not have a significant impact on each other, i.e. There is no multicollinearity. Thus, all three factors can be used to build an econometric multiple linear regression model.

Table 1. Matrix coefficients of correlation

<i>Indicator</i>	<i>Insurance premiums</i>	<i>Employed population, total, thou. people</i>	<i>Nominal average monthly wage, rub.</i>	<i>The number of registered legal entities operating, unit</i>
	y_1	x_1	x_2	x_3
y_1	1			
x_1	-0,68079	1		
x_2	0,999468	-0,65846	1	
x_3	0,568015	0,18915	0,588188	1

As a result of regression analysis a formula of insurance premiums for this insurance type and selected factors was defined (1):

$$y_1 = 2929724 - 645,769x_1 + 0,208x_2 + 0,634x_3, \quad (1)$$

where y_1 -insurance contributions, million rubles;

x_1 -the number of employed population in the Republic of Belarus, total, thou. pers.;

x_2 -nominal average monthly wages, rub;

x_3 -the number of existing legal persons, u.

2329724 – random variable describing the deviation factor X from the regression line.

The results of the statistical significance evaluation of the parameters of the equation as a whole on heteroscedasticity of critical values and t-F-statistics (t-test and Fisher), and for autocorrelation (Durbin-Watson test) demonstrate that the construction of a stochastic model has a positive quality. Its parameters confirm the plausibility of the impact of selected factors on the change in the volume of premiums income. Therefore, this feature can be applied to calculate their predictions for the future.

Scatter charts for each of the factors were constructed and set the trend line, as well as defined the equation squares and odds (Table 2).

Table 2. Number of employed population in the Republic of Belarus, thous. people (completed by authors)

Indicator	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
t	1	2	3	4	5	6	7	8	9	10	11
x_1	4 347	4 326	4 414	4 470	4 518	4 611	4 644	4 666	4 655	4 577	4 518

Based on the data in Table 2, the trend line accounted for more appropriate functions: linear, exponential, logarithmic, polynomial, and exponential. For the x_1 factor, coefficients of determination of the relevant function were determined (Table 3). Coefficient of determination R^2 differentiates effective communication layer indicator x_1 and the independent variable t for the analysis of selected polynomial functions in which were the highest coefficient of determination (0.8818).

Table 3. Function and coefficients of determination of the dependence of the employed population by time period

Type of function	Equation	Coefficient of determination
Exponential	$x_1 = 4353,6e^{0,0063t}$	$R^2=0,5953$
Linear	$x_1 = 28,166t+4353,3$	$R^2=0,5915$
Logarithmic	$x_1 = 137,96 \ln(t)$	$R^2=0,7162$
Polynomial	$x_1 = -7,0661t^2 + 112,96t + 4169,5$	$R^2=0,8818$
Power trend	$x_1 = 4304,7t^{0,0308}$	$R^2=0,7217$

Functional dependency on the adequacy of the model was checked using the criterion of Goldfeld-Kvandt. It showed that the residues are homoscedastic. Homoskedasticity of residues means that for each value of the factor x_j residues have the same dispersion (the confidence level is 95%). This allows us to recognize a resulting regression equation adequately reflecting the relationship of variables, vindicating his use population projections for future periods.

$$x_1 = -7,0661t^2 + 112,96t + 4169,5. \quad (2)$$

Other selected factors were identified and verified the adequacy of equations in the same way. For x_2 (nominal, accrued monthly wage), despite the higher coefficient of determination on exponential model ($R^2 = 0.975$), after checking for heteroscedasticity, the best model proved to be a power function with $R^2 = 0.959$.

$$x_2 = 445,96t^{3,876}. \quad (3)$$

Using the criteria of Fisher and Student ratio statistics, it was proven that the third factor x_3 (number of existing legal persons) of the logarithmic function model is best when $R^2 = 0.9685$.

$$x_3 = 9959,4 \ln(t) + 156505. \quad (4)$$

On the basis of the calculations, interdependent revenue premiums and the influence of primary factors, as well as selected functions to establish themselves as factors trends was defined. They are used in justifying the value of the forecast of premiums in the developed financial strategy for Belgosstrakh (Table 4). Reducing the number of employees can be explained as a result of the improvement of technologies that reduce the need for workers specialties. Promoting business development will contribute to growth in the number of enterprises, mostly small. This explains the growth of nominal wages. Forecasts reality is achieved mainly due to a planned increase of priority indicators of the national economy development. As already noted, premiums are a priority, but not the only source of financial base of the

development strategy of any insurance company. In recent years the investment activity of insurance companies increased. However, the amount of investment income of Belarussian insurance companies are relatively low compared to developed countries.

Table 4. Forecast revenue of premiums

Forecast	2014	2015	2016	2017	2018	Rate of growth for last 5 years? %
Employed population, thous. people (x ₁)	4 508	4 444	4 370	4 274	4 168	0,92
Nominal wage, rub. (x ₂)	6 790 139	9 259 897	12 340 829	16 124 000	20 706 358	3,05
The number of legal entities, un. (x ₃)	174 350	175 885	177 214	178 388	179 437	1,03
Insurance premiums, mln rub. (y ₁)	1 491 889	2 028 481	2 698 438	3 520 778	4 517 726	3,03

As already noted, the projected volume of insurance premiums in direct insurance and investment income are crucial sources of financial resources for any insurance company. With regard to other income from reinsurance, regression to the perpetrators of insurance claims, property rental, sale of fixed assets, positive exchange rate differences and other input financial flows, they are less likely to affect the overall size of financial support the strategic objectives of insurers. This is largely due to their economic nature, a kind of occurrence of sources, as well as the unpredictability of their occurrence in the activities of insurance companies. The study of the sources of these revenues in the Republic of Belarus insurance sector confirms the value of their oscillation in time and the complexity of identifying persistent factors influence their dynamics. However, these circumstances do not give reason to abandon their account as other income is able to some extent to expand the financial strategy of insurance companies. Therefore, it becomes a more reasonable use of the simplified method of predicting the future. It is based on the determination of the average annual dynamics of related income in the previous period and their relation to the total volume of insurance premiums (5):

$$K_{pri} = \frac{a_i + b_i + c_i + d_i}{z_i} \quad (5)$$

where K_{pri} - the ratio of other income to total insurance premiums in the i-th year;
a_i - the amount of reinsurance premiums on risk-taking in the i-th year;
b_i - the amount of insurance indemnities received from reinsurers in the i-th year;
c_i - the amount of reinsurance commission on ceded reinsurance in the i-th year;
d_i - the amount of other revenues to the i-th year;

z_i - the overall outlook of the input cash flow from premiums in the i -th year.

Calculation results for the "Belgosstrakh" are represented in Table 5. Belgosstrakh was used as an example because this is a major Belarussian insurance company, holding more than 50% of the insurance market. Also this company offers more than 70 different insurance services, which allows it to diversify its financial flows.

Table 5. Other input financial flows Belgosstrakh for the years 2009-2013, mln

Indicators	2009	2010	2011	2012	2013
1. Reinsurance premiums on risk-taking (a_i)	583	1 065	2 841	7 200	4 878
2. Insurance reimbursement received from reinsurers (b_i)	816	356	468	1 425	3 814
3. Reinsurance commission on ceded reinsurance (c_i)	1 041	1 024	1 516	5 534	12 423
4. Other income (d_i)	12 652	12 743	126 721	56 189	99 215
5. Total other income (Sum of rows 1+2+3+4)	15 092	15 188	131 546	70 348	120 330
6. Insurance premiums for direct insurance (z_i)	602 874	698 301	1 166 542	2 142 599	3 219 996
7. Coefficient K_{pri} , %	2,50	2,17	11,28	3,28	3,74

As can be seen from Table 5, other input financial flows to the total amount of insurance premiums Belgosstrakh for the last period under review range from 2.17% to 11.28% (2011 increased due to exchange rate differences). Their average share in the total volume of premiums is 4.6% over 5.

By combining the targets of all sources of income financial flows and using the mathematical model that determines the total amount of input financial flows, the Belgosstrakh financial strategy for the 2014-2018 was built. (Table 6).

Table 6. General model of financial support of Belgosstrakh strategy in perspective for the period 2014-2018, mln rubles

Projected source of formation of the input of financial flows	2014	2015	2016	2017	2018	The growth rate in 5 years time (Reference)
1. The total amount of insurance premiums	4473089	6477844	9352 033	13487923	19472004	4,35
2. Investment returns	354 357	512 247	740 486	1 070 420	1 547 361	4,37
3. Other input financial flows	205 762	297 981	430 194	620 444	895 712	4,35
4. Total input financial flows	5033209	7288071	10522712	15178787	21915077	4,35

The successful solution of the tasks planned depends on the exact definition of the input of financial flows in the first year of the formation of financial strategy. For subsequent years, the projections of revenues are only approximate, and require constant refinement based on the actual achievement of the projected parameters for the previous period, as well as adjustments based on the occurrence of a new situation.

4 Conclusions

The dynamics of relevant income are primarily influenced by an increase of nearly 3.05 times the amount of gross wages and salaries. It is to a lesser extent affected by changes in the number of operating entities and employees. The proposed method of calculation is used to support forecasts of premiums and revenues from other types of insurance. Modern computer technology and software allow greater speed of simulations and accuracy. This can ensure greater continuity and effectiveness for an organization in the long run. A correct solution of the issues involves the simultaneous support of optimal value and costs. For forecasting financial flow inputs in the future, insurance companies are encouraged to use a moderate financial strategy. However, each individual financial strategy of a particular insurer should be based on the identification of the features of its activities. It should consider alternatives for its planned strategic objectives. Also it is necessary to adapt them to a choice of verification methods of the projected input and output dynamics of financial flows.

The main problem of the Belarussian insurance company's financial strategy development is the weak diversification of input financial flows. Domestic analysts mainly pay attention to the ratio of incoming flows from the financial capacity of compulsory and voluntary insurance, life insurance or other risks. However, with low investment activity, Belarussian insurance companies will be deprived of flexibility. If the identified trends continue, insurance companies quickly reach their limits of growth and face the challenge of long-term scarcity.

References

1. Suresh P.Sethi, Gerald L.Thomson. Optimal Control Theory. Applications to management Science and Economics, Second Edition, Boston/Dordrecht/London, Kluwer Academic Publishers, (2000).
2. Ricardo Azevedo Araujo and Tito Belchior S. Moreira. A dynamic model of production and tra_c of drugs. Econ. Lett., 82(3):371 { 376, (2004).
3. Abraham B. and Ledolter J. Statistical methods for forecasting. New York: Wiley, (1983).
4. Piketty, T. Capital in the TwentyFirst Century. Cambridge, Mass. (2014).
5. Nedopil Christoph. International Capital Flows and the Corporate Governance Environment. Business& Economics. (2009).
6. Cornelius, Peter K. Corporate Governance and Capital Flows in a Global Economy. Oxford University Press, USA, (2003).
7. Kuester, K., Wieland, V. Insurance policies for monetary policy in the euro area. Journal of the European Economic Association 8(4), 872–912, (2010).

8. Ministry of Finance of Belarus. – [Electronic source]. – Available at:
<http://www.minfin.gov.by/ru/supervision/stat/>. Accessed 20.03.2016.

Simulation of the Impact of Social Media on Promoting Education Services

Iryna Ushakova

Simon Kuznets Kharkiv National University of Economics,
Lenin Avenue, 9, 61166 Kharkov, Ukraine

`iryna.ushakova@m.hneu.net`

Abstract. The role of education services in the global competitiveness of Ukraine was shown. A specific of educational services, educational services market structure and characteristics of its subjects was reviewed. Analysis of survey results on the effect of different channels of knowledge enrollees on the university was held. The advantage of impact the social of the channel from the inner circle of enrollees, namely parents, friends, classmates was shown. Bass model parameters for the purpose of educational services, simulation of the dynamics of the spread of information among universities AnyLogic is made.

Keywords. Educational services, higher education institution (HEI), social media, Bass model, Agent based modeling, AnyLogic

Key Terms. Academia, MathematicalModel, Simulation, AgentBasedSystem

1 Introduction

Ukraine is deeply integrated into the world community. As the global financial crisis deepens, so it affects Ukraine. It challenges all spheres of socio-political and economic life, and the education sector can't be an exception. The level of educational services is one of the important indicators of country's competitiveness.

Education is a process of learning, teaching. In a broad sense, education is the product of the formation of intelligence, character and physical abilities of the individual. The society is purposefully transmitting accumulated knowledge, values and skills through education by certain institutions: schools, colleges, universities and other institutions. Higher education is the highest level of professional education, following the general secondary and vocational education. It includes a set of systematized knowledge and practical skills that can solve the theoretical and practical problems of the professional profile, using and creatively developing the modern achievements of science, technology and culture.

Global competitiveness ranking "The Global Competitiveness Report 2015–2016" was presented at the International Economic Forum in Davos. Global Competitiveness Index is a global study, which accompanied by country ranking in terms of eco-

conomic competitiveness. This index is calculated according to the methodology of the World Economic Forum "World Economic Forum". According to this rating by one of the strengths of Ukraine is education and innovations [16].

The place of Ukraine in the field of educational services from 140 countries:

- primary school enrollment – 45th place;
- secondary education enrollment – 33rd place;
- higher education enrollment – 14th place;
- quality mastering math and science – 38th place;
- Internet access in schools – 44th place.

Innovations are the result of educational services. Ukraine takes the following positions in this area:

- the quality of scientific research institutions – 43rd place;
- availability of scientists and engineers – 29th place;
- the number of registered patents (intellectual property rights) – 50th place.

So education is one of the few spheres in which our country takes a worthy place in the world ranking. Therefore, positive transformation in this area is one of the key aspects to overcome the economic crisis and forming a positive presentation of the country and its European integration [6].

The Market for education services is socially important. This market was formed in the country on the post-Soviet territory. Therefore, the process of market relations formation is quite complex and sometimes have spontaneous and uncontrollable character.

Under market conditions, the essence of relationships between the citizen and the state is changed. Education becomes the sphere of educational services. The citizen as the taxpayer is entitled customer and requires improved quality of educational services, taking into account the needs of enrollees and their parents.

The aim of this paper is to build and test a dynamic model of information dissemination on HEIs as a tool to attract students. For achievement of aim it is necessary:

- to consider and to describe the structure of the education market;
- to analyze the impact of various information channels to choose from enrollees in Ukraine;
- to explore the dynamics of the spread of information among university enrollees in social networks.

2 Features of the Education Market

In recent years, the close integration is between Ukraine and the EU. It is naturally reflected in the educational relations. Today 15% of the students leave Ukraine to study abroad. Every year the competition increases in the education area. So now it is needed to use all available tools to promote higher education institutions and their services to attract new students.

Educational service is a specific product, characterized by its features. Educational services are services that develop and improve the consumer. In general there is not plenty of such services, for instance: medical, cosmetic etc. Besides, educational services are characterized by complex influence on various aspects of a person's behavior.

In the education market, there are producers, consumers and intermediaries. The specificity of the educational market has four types of subjects:

- consumers – graduates, students who directly receive educational services;
- customers – parents who pay for educational services and influence the choice of educational institution;
- customers – employers who are interested in upgrading and further training of personnel, and who pay for staff training;
- customer and consumer – a state that guarantees the quality of education and is interested in the formation of skilled labor.

The main subjects of this market are: the consumer – enrollee or student and producer – school. Each subject has its own specific interests: student – tends to get a certain level of knowledge that will allow them to take a worthy place in the labor market; school – offers such educational services that are competitive in the market. The state should create such legislative conditions in education in general, to ensure the competitive growth of highly skilled professionals, and to improve the quality of human capital.

The education market is also inherent in the competition. But educational services have the peculiarity.

The first, the customer is the main subject of choice. It can be parents of students, but not the students themselves. First of all, this aspect affects the target audience advertising of HEI.

Second, to promote educational services the HEI cannot offer all traditional marketing strategies. This service has specificity. Offered intellectual product to consumers does not guarantee that received education will bring the expected results. But the promotion of educational services in market conditions is necessary, because a successful marketing strategy helps to attract more students, increase competitiveness and give more opportunities to attract the best students.

3 Analysis of the Impact of Social Channels on Decision Making to Choosing a Higher Education Institution

So to attract specific educational services consumer we must find what influences the enrollee in choosing higher education institution, what sources of information are most authoritative. For this purpose two surveys in social networks was administered. This format was chosen because, the first, social networks offered convenient tools for carrying out such surveys. Second, complete anonymity is ensured. This is the necessary psychological factor that allows giving unbiased answers to questions. Respondents who know what his answer will be anonymously rather answer the question

truthfully, even if he inwardly doesn't like his answer. Third, students are a very active audience in social media, and they are interested in obtaining objective information on the channels that are most popular.

The marketing strategy of promoting educational services has many channels of communication that influences the decision on the choice of higher education institute. To determine the preferences of enrollees concerning these channels the survey was carried out on social networks. Questions in the questionnaire were set so as to identify from which most enrollees receive information about the university.

The survey was addressed directly to the enrollees. So could find out how to enhance the effectiveness of certain areas in the promotional campaign. But also to understand most effective directional distribution way of advertising among students and their parents is needed to maximize efficiency.

This question was asked in the questionnaire: "How did you know about our university?" The following answers were given:

- from classmates, parents or friends;
- from ad campaign of the University (open day, University presentation at school , printed advertising materials);
- from friends in social networks;
- advertising on social networks;
- official website of the University.

Two surveys were carried out: in the urban social network (Kharkiv) and in the social network at the Web-page group of University department related to IT.

The first survey was carried out to get information for the overall situation of the city in general. The survey was carried out at the end of May, when the future enrollees generally have not yet made up their minds. Outcomes of the responses were grouped as follows (Fig. 1):

- Group 1 – "Parents, Classmates, or Friends" – answering options: from classmates, parents or friends; from friends in social networks;
- Group 2 – "Advertising of Universities" – answering options: ad campaign of University (open day, presentation of the University at school, print advertising materials); advertising on social networks;
- Group 3 – "Universities web sites" – answering options: the website of the university.

About 2 thousand of respondents took part in the survey. Such a number of respondents were made possible by the high popularity of social groups mainly among younger audiences: high school students, college students, graduates and working people who recently graduated from high school. The answers were distributed as follows. More than two-thirds of respondents received information and feedback from friends, classmates and parents. Other answers almost evenly distributed between the other two groups of sources.

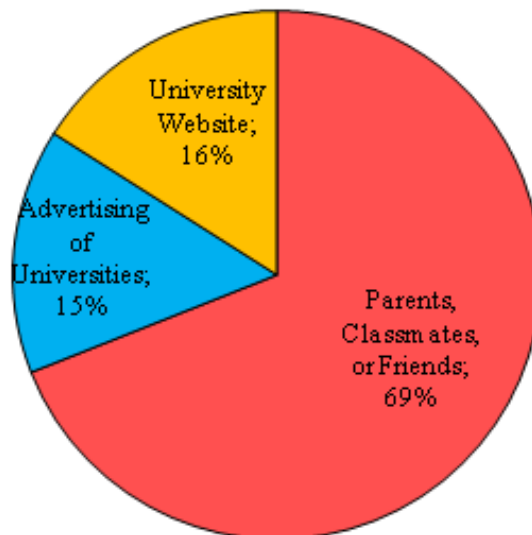


Fig. 1. Distribution of channels in the survey in the social network of the city

It is not an accident that social groups organized the way where friends in a network are in the first position among the list of group members, then – friends of friends (acquaintances), and so on. Considering the significant number of respondents, we must recognize that social channel influences final decision in a greatest way. The vast majority of successful marketing strategies in promoting various products and services have been actively using these channels for a long time. Promotion got popularity as through the traditional social channels and also Internet social networks lately. [15, 19].

The second survey was published in the social network of the University department group which provides education in IT (Fig. 2). As we see in this survey the overwhelming majority of respondents gave preference to the social channel from the inner circle (48%). A feature of this audience is its focus on the IT sector; therefore, a higher percentage was obtained for the advertising through the site university (20%). Also, among this audience, a large number of enrollees were previously motivated to go this university through the professionally-oriented institutions work. Therefore, this channel has obtained more than twice higher percentage of answers than traditional advertising of universities. As we see in this survey the overwhelming majority of respondents gave their preferences to the social channel from the inner circle (48%). A feature of this audience is in its focus on the IT sector; therefore, a higher percentage was obtained for the advertising through the site university (20%). Also, among this audience, a large number of enrollees were previously motivated to go to this university through the professionally-oriented institutions work. Therefore, this

channel has obtained more than twice higher percentage of answers than traditional advertising of universities.

The difference between those two surveys may indicate following. When enrollee chooses the professions related to IT sphere, he relies largely on his own experience, and on his idea of what he wants to achieve in the future.

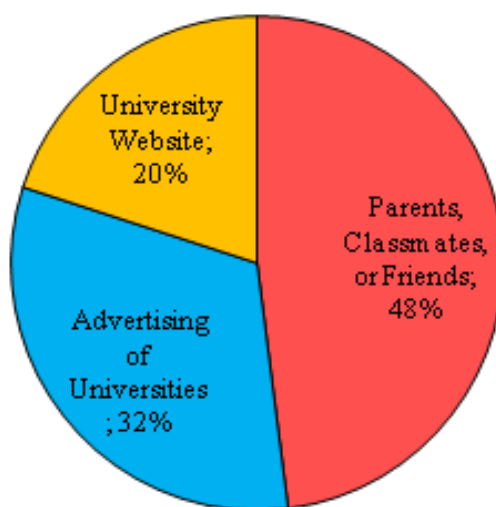


Fig. 2. Distribution of channels in the survey in the social network of the University Department group

Social channel "Advice of family and friends" is important for the enrollee in both surveys. This channel belongs to the personal channels (direct) communication. Promoting educational institution and its services may be performed through university teachers and students meetings with school graduates and their parents, as well as open door days, summer schools, competitions, work in small academies and more.

4 Simulation of Dissemination of Information on HEI among Students

4.1 Agent-Based Modeling as a Method to Study Consumer Behavior

An important task of promoting educational services is to analyze the impact of dissemination of information on HEIs up to enrollees. Simulation modeling methods are widely used to carry out this analysis. Simulation is a method for Applied Systems Analysis, which is a powerful research tool of complex systems and processes, including those associated with making decisions under uncertainty. In comparison with other methods simulation modeling allows us to review a large number of alter-

natives in order to improve the quality of management decisions and to make more accurate predictions of their effects.

Agent-based modeling is a method of simulation modeling, which examines the behavior of decentralized agents and how such behavior determines the behavior of the system as a whole. When developing agent model, parameters of agents are entered, defined their behavior, put them in a certain environment, established possible links, and then ran the simulation. The parameters of agents may be people, such enrollees. The individual behavior of each agent creates a global behavior of the simulated system. The main feature of the agents is the possibility of taking independent decisions, due to the active role of agents in the model.

In practical terms, agents must have certain characteristics [13]:

- the agent is identified. It is the essence with the set of characteristics and rules that governs its behavior and decision-making capabilities;
- the agent is autonomous. The agent has a border and can be easily identified, that is part the agent, and that is not part of the agent;
- the agent is in an environment where it interacts with other agents. Agents have protocols for communication with other agents, and the ability to respond to the environment;
- the agent must be purposefully based on the logic of their behavior. This allows the agent to compare the results of their behavior towards their goals;
- the agent is autonomous and independent. The agent can independently function in their environment and in their relations with other agents, at least in a limited range of situations;
- the agent is flexible, It has learning ability and adapts its behavior based on experience. It requires some form of memory;
- the agent may have rules that change their own rules of conduct.

On this basis agent can be formally presented as it is in Fig. 3:

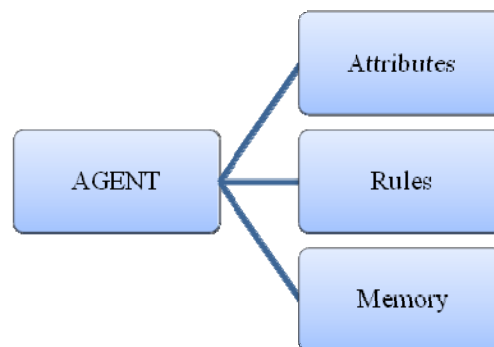


Fig. 3. Formal presentation agent

In a dynamic, competitive and complex market environment the choice of customers often depends on the individual, congenital activity of consumer, networking, and

external influences. This is best described by using agent-based modeling. Agent-based modeling will help to highlight and simulate processes in social networks and various contacts between people. As a result, objective information distribution forecasts will be obtained.

4.2 Bass Model to Determine the Dynamics of Dissemination of Information on HEI among University Enrollees

Bass model describes the process of dissemination of the product (service) when the product (service) is unknown for buyer at first. Then the product is advertised to start selling. [2–5]. As a result, a certain share of people will buy the product under the influence of advertising. This category of people is called "innovators". Others can buy the product only under the influence of communication with other people who already bought this product. This category is called "imitators". The process of purchasing of a new product under the influence and conviction of people who already bought it something is like an epidemic. The probability of acceptance of the idea of those who have not accepted is a linear function of those who previously accepted idea. [2].

To make the laws a reality it is necessary to have absolute interest in product of the entire tested audience. How quickly innovation is spread depends on how quickly and from which sources audience will get information about it. We should also keep in mind the trust factor in advertising and general psychological characteristics of customers and their groups. Some will be more likely to try new product under the influence of advertising, others will make any decisions only when they saw the experience of friends or colleagues.

The formal description of this model is.

$$\frac{f(t)}{1-F(t)} = p + q \cdot A(t) \quad (1)$$

where $f(t)$ - the rate of change of set basic fractions, $F(t)$ is a basic fraction; p , and q - constant, reflecting the influence through advertising and through society; $A(t)$ - a cumulative function of the number of followers (adopters).

The parameter p does not depends on the number of followers $A(t)$, It is therefore called "coefficient of innovation". The parameter q is influence factor on the number of followers $A(t)$. Therefore, we can clearly define its role as a "coefficient of imitation", factor increasing the number of new followers influenced by of the old followers.

Later, instead of the constant q was introduced constant m , which displays the potential market size. In this case, the equation will work much more accurately. Thus, the Bass model has become as follows.

$$\frac{f(t)}{1-F(t)} = p + \frac{q}{m} \cdot A(t) \quad (2)$$

While working on his idea Bass oftentimes changed and improved mathematical reflection of its concept. But despite the fact that the formula was repeatedly redesigned by him, and then was greatly expanded by his followers, the main idea remained in its original form. It is only more confirms author's deep and accurate understanding of marketing processes.

Currently, the most popular is the following formula:

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 - \frac{q}{p} \cdot e^{-(p+q)t}} \quad (3)$$

Equation (3) allows getting graphic representation of the function. The function that displays the cumulative number of followers is more informative. It is get the following:

$$A(t) = m \cdot F(t) \quad (4)$$

We take a higher education institution as a service provider and the enrollee as a customer. Dissemination of information is carried out at each of the university admission campaigns and can be attributed to the processes described by Bass.

All of the features of the process are:

the audience of potential customers (enrollees) interested in using the learning service of higher education. Not interested, it is rather an exception to the rule and not a trend;

in the final phase, all enrollees who make their choice based on available information;

the choice of the institution is the defining event for life, so just advertising or open days at universities is not sufficient a of information for making decision by enrollees. They are also highly affected by the reviews of friends, advises of elders, and even user discussions on the forums and social networks.

So, we conclude that the Bass classical model is acceptable for analytical description of this process, and can be used for simulation of the behavior of enrollees on choosing HEIs.

4.3 Defining the Parameters of Base Model Based on Statistical Data

Bass model is widely used to predict the spread of new products and technologies. [7, 14] In this paper we propose to use this model for specific products that are educational services. Parameters of external and internal influence significantly affect on the results of the Bass model. In order to use the model we must first define its basic parameters: m , p , and q . Full enough analysis of methods is given in [2–5]. So, to determine the parameters of external (innovation) and internal (imitation) impact of p and q , there is four most convenient methods: Based on the conventional method of least squares (the ordinary least squares – OLS); based on nonlinear least squares method (the nonlinear least squares – NLS); based on

maximum likelihood (the maximum likelihood estimation – MLE); based on the algebraic method (the algebraic estimation – AE). The ordinary method of least squares is the most simple and effective if there are statistics. The main advantage of OLS–assessment procedure is the ease of implementation. The necessary changes needed to conduct Bass equation are shown below. Bass formula will present in the following form:

$$P(t) = p + \frac{q}{m} N(t-1) \quad (5)$$

where N – the cumulative function of followers, P – a probability of acceptance of ideas in the time interval $(t-1; t)$.

If the number of not followers at time $(t-1)$ is $(m - N(t-1))$, then the expected number of followers in the interval $(t-1; t)$ can be marked by $X(t)$ and formula is rewritten as follows:

$$X(t) = pm + (q - p) N(t-1) - \frac{q}{m} N^2(t-1) \quad (6)$$

If in function (6) replace some constants related to the equation, we can get the next function:

$$X(t) = a_1 + a_2 N(t-1) + a_3 N^2(t-1) \quad (7)$$

The resulting function is well suited for use the least squares method to calculate the coefficient functions. When a_1, a_2 and a_3 are obtained parameters, p and q can be easily obtained as follows:

$$p = \frac{a_1}{m} \quad (8)$$

$$q = -m a_3 \quad (9)$$

To determine p and q we will use data from a survey which was held in a popular group of social network. Seasonality, which adversely affects the accuracy of the results, was removed from empirical data. Then time series was created.

Table 1. The survey results

t	$N(t)$	$X(t)$
0	610	–
1	1352	742
2	1622	270
3	1704	82
4	1774	70

t	$N(t)$	$X(t)$
5	1805	31
6	1842	37
7	1848	6
8	1861	13
9	1865	4
10	1857	2
11	1868	1
12	1869	1
13	1869	0
14	1869	0

Calculation of model parameters is performed by mathematical package MathCAD:

$$\text{count} := 15$$

$$G(n, a) := a_0 + a_1 n + a_2 n^2$$

$$\Phi(a) = \sum_{t=0}^{\text{count}-1} (X_t - G(N_t, a))^2$$

$$a_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad a_1 = \text{Minimize}(\Phi, a_0) \quad (10)$$

$$a_1 = \begin{pmatrix} 211.654 \\ 0.836 \\ -5.165 \times 10^{-4} \end{pmatrix}, \quad R = X - G(N, a_1)$$

$$p = 0.113$$

$$p_i = \frac{a_{1i}}{m},$$

$$q_i = -m \cdot a_{1i},$$

c

The obtained \mathbf{p} and \mathbf{q} themselves confirm the results of the survey. We see a great influence of internal factors (parameter \mathbf{q}), and little impact of external factors (parameter \mathbf{p}). This completely coincides with the results. Consequently chosen method gave a fairly accurate result of the model parameters. Now they can be used to customize the model and its further exploitation.

4.4 Simulation of Information Dissemination in the Environment AnyLogic

AnyLogic provides a wide range of modeling tools, and has many specialized modules devoted to specific methods and approaches. Especially for the needs of agent-based modeling AnyLogic there is the appropriate ways to create specialized models, as well, as broad information support. Full documentation is provided [10, 17, 18].

The first step in creating the agent model is to determine the parameters of agents who will take part in the simulation and the environment where they are located. The class name is Enrollee. Agents can inhabit in various types of spaces (environments): two-dimensional continuous, two-dimensional discrete and geospatial (GIS).

GIS space allows placing agents in the geospatial environment by using this card. [12]. Continuous space gives us ability to change the location of the agent and receive information about its current location, move the agent with a given speed from one place to another, perform actions on his arrival at the destination, display an animation (static or moving) of agent, make connections under the template of agent's locations, etc. Part of the functionality of continuous space does not even require agents to belong to explicitly given for environment. If the environment is not specified, so then, by default, that space will be continuous. But if agents belong to the environment, the type of space must be defined clearly.

Two-dimensional discrete space is a rectangular array of cells, fully or partially occupied by agents. One cell may contain no more than one agent. Support for this type of space in AnyLogic includes the possibility of distribution agents on the cells, their movement to neighboring or any other cell, determination that agents are neighbors (according to the chosen model of neighborhood), determination of free cell etc.

To simulate the interaction of subjects the educational market will use continuous space. The communications networks between agents in this area may be the following types:

random (*Agent.NETWORK_{RANDOM}*) – Agents randomly connected with a given average number of connections to the agent;

based on the distance (*Agent.NETWORK_{ALLINRANGE}*) – any two agents are connected if the distance between them is less than a specified maximum;

ring (*Agent.NETWORK_{RING-LATTICE}*) – connection of agents forms a ring, where the agent is connected with a given number of closest agents;

small world (*Agent.NETWORK_{SMALLWORLD}*) – can be viewed as a ring where some links were "re-made" for agents located far from each other;

scale free (*Agent.NETWORK_{SCALE-FREE}*) – some agents are "hubs" with lots of connections and some of them have only a few connections.

The analysis of the nature of relations in social networks was carried out [1, 11, 12]. It showed that in the case of educational services it is better to use network-based distance. The distance, in this case, is a mechanism for modeling the realest and very complex nature of connections between enrollees themselves and their "friends" on the network.

The next step is to determine behavior of agents (algorithm of actions). In Any-Logic, this is possible with the help of creating Java code or by creating an algorithm by using visual elements of Statechart panel. This panel allows you to create an algorithm of actions of the agent by using the agent transitions between states.

The agent will have three states:

NotInformed – the enrollee has not received information about HEI;

Informed – the enrollee received information about HEI;

Decided – the enrollee who submitted the original documents to the HEI.

To implement the Bass diffusion model, we must add the appropriate transitions between states. Three variables are added:

AdEffectiveness – coefficient of external impact, $p = 0,113$;

ContactRate – the average number of contacts of the agent with the environment ($ContactRate = 15$);

AdoptionFaction – coefficient of internal impact, $q = 0,966$.

Now, we create transitions between states, according to Bass diffusion model. The transition to a state of "Informed" will be done in two ways: through external and internal impact.

Transition **Ad** simulates the dissemination of information through outside influence, i.e. advertising, media, and so on. Transition is parameter p .

Transition **Interaction** simulates disseminating of information through internal impact, i.e. acquaintances, friends, parents, and others. It is realized through the third transition in a state of "Informed", and it is named "IntMsg" (Interaction Message), and it works according to the expression $ContactRate \cdot AdoptionFaction$.

As a result of the simulation model with a population of 700 agents was derived curves that are characteristic of the Bass diffuse model.

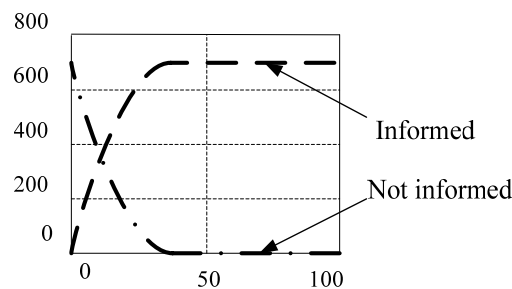


Fig. 4. Agents dynamics chart: two state

Now, we create a transition to the third state of agent by entering additional variable **DecisionReady** and appropriate transition.

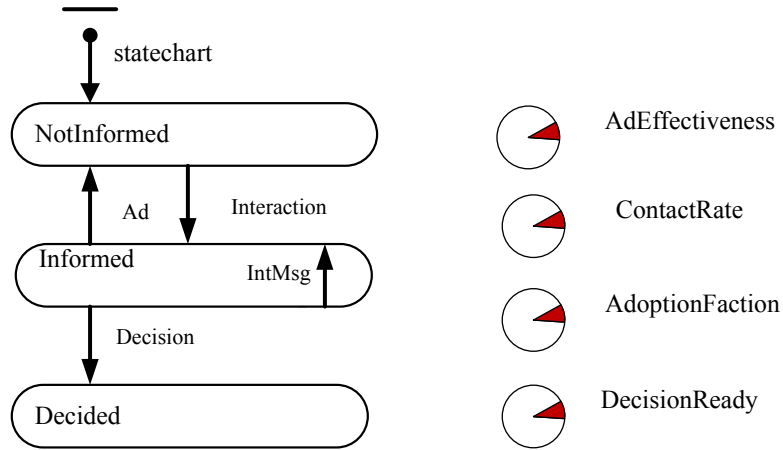


Fig. 5. The final model in AnyLogic

When we are performing a model with a population of 700 we can see three phases of behavior of agents, three changes of state: first, the lack of information, next, re-view of the options of available actions, and after that, making a decision by the agent according to available information (Fig. 6).

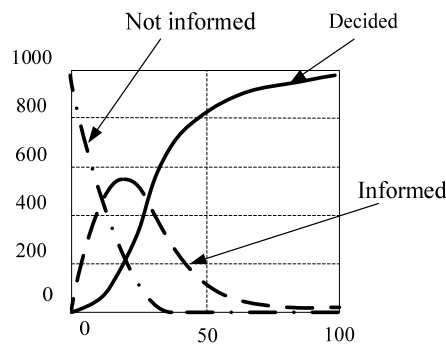


Fig. 6. Agents dynamics chart: three state

The chart shows the time when "Informed" curve is at maximum – this is a period of time with maximum uncertainty in making a decision among entrants when they already obtained all necessary information, but the final decision on choosing a particular institution of higher education is not adopted yet.

The "Decided" curve in the chart can be interpreted as making decision in general and not in favor of particular HEIs. In general advertisement has an impact on enrollees who are going to make a decision ("Informed" curve) or do not have enough information for making the adequate decision ("Not informed" curve).

Parameters of the model are based on a survey that was carried out in social networks. But traditional channels have its own specific of dissemination of information.

So the simulation results show just the dissemination of information in the Internet environment. Social networks and traditional channels have some common. Some time is needed for the dissemination of information. It is much less in social networks. So if you need to quickly convey information to a considerable number of enrollees are to use social networks.

5 Conclusion

Education in Ukraine is one of the competitive trends, as evidenced by leading world ratings. Therefore, promotion of educational services is a priority task in the country. But education is a specific service. In the education market, the leading place is not only for the enrollee but also for the intermediary that affects the choice of the enrollee. This is the parents who may be paying in the future for the educational services and therefore, they have an impact on the enrollee. Also significant influence is made by friends, classmates, acquaintances, and recent graduates who succeed in their careers. Thus, the greatest impact on the enrollee is made by the channels of their inner circle. So, to better promote educational services, Universities should strive to create a positive image of the institution among the inner circle of enrollees.

In the recent time, the channel that makes the most impact on customers, are social networks. They are just very popular among young people. Therefore, HEIs should use these channels to form a positive image of the institution among participants in social networking, use it to provide information of the promotional activities and events. Because a significant impact on the enrollees makes the inner circle, it is necessary to create group of the university in the social networks. Students and graduates of university have to involve in these groups. So, these groups will cover two areas of promotion: friends and acquaintances, advertising HEI.

The paper has presented an approach to define the strategy of promoting HEI among future enrollees based on Bass diffusion model of innovation. This model is widely used to predict the dissemination of new products and technologies. The analysis showed that the model of Bass is appropriate to describe the process of information dissemination on education services and can be used to simulate the behavior of students choosing HEI. Model of Bass is implemented through agent-based simulation modeling in the environment AnyLogic. The results indicate that the period of time for the most effective long-term strategy is a time when the future enrollee selects items for UPE. Starting career guidance is necessary first of all among those graduates who are still undecided in their abilities and interests due to the lack of adequate and complete information about the specialty.

References

1. Abraham, A., Hassanien, A., Snasel, V.: Computational Social Network Analysis: Trends, Tools and Research. Springer-Verlag London Limited, London (2010)
2. Bass, F.M.: A Dynamic Model of Market Share and Sales Behavior. In: Toward Scientific Marketing, pp. 263–276. American Marketing Association, Chicago (1964)

3. Bass, F.M.: A New Product Growth for Model Consumer Durables. *J. Management Science*. 15, 215–227 (1969)
4. Bass, F.M.: Heterogeneity in Purchase Intention, Diffusion and Exogenous Influence. In: *Research Traditions in Marketing*, pp. 224–226. Springer, Dordrecht (1994)
5. Bass, F.M., Jain, D., Krishnan, T.: Modeling the Marketing-Mix Influence in New-Product Diffusion. In: *New-Product Diffusion Models*, pp. 99–122. Springer, New-York (2000)
6. Bobuh, I.: Smarter Europe: how to Create in Ukraine a Market of Scientific and Educational Services. *J. Forbes Ukraine*. 11 (2015) (in Russian). <http://forbes.net.ua/opinions/1406116-umnee-evropy-kak-sozdat-v-ukraine-rynok-nauchno-obrazovatelnyh-uslug>
7. Churkin, V.I.: Sales Forecasts of Innovative Products within the Macroeconomic Factors (Using the Example of Small Wind Turbines). *J. St. Petersburg State Polytechnical University Journal. Economics*. 1–1', 104–112 (2013) (in Russian)
8. Freeman, L.: Centrality in Social Networks Conceptual Clarification. *J. Social Networks*. 1(3), 215–239 (1979)
9. Girvan, M., Newman, M.: Community Structure in Social and Biological. *J. Proceedings of the National Academy of Sciences*. 99(12), 7821–7826 (2002)
10. Grigoryev, I.: AnyLogic 7 in Three Days: A Quick Course in Simulation Modeling. CreateSpace Independent Publishing Platform, New York (2015)
11. Leskovec, J., Lang, K., Dasgupta, A., Mahoney, M.: Statistical Properties of Community Structure in Large Social and Information Networks. In: *Proceedings of the 17th International Conference on World Wide Web in Beijing (China)*, pp. 695–704. WWW2008, Beijing (2008)
12. Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P., Tomkins, A.: Geographic Routing in Social Networks. *J. Proceedings of the National Academy of Sciences*. 102(33), 11623–11628 (2005)
13. Macal, C. M., North, M. J.: Tutorial on Agent-Based Modeling and Simulation. *Journal of Simulation*. 4, 151–162 (2010)
14. Mahajan, V., Mason, C., Srinivasan, V.: An Evaluation of Estimation Procedures for New Product Diffusion Models. In: *Innovation Diffusion Models of New Product Acceptance*, pp. 203–232. Ballinger Cambridge, Massachusetts (1986)
15. Repich, T.: Role of Social Networks in the Promotion of Educational Services. *J. Scientific Works of NUFT*. 21, 178–185 (2015) (in Ukraine)
16. Schwab, K.: The Global Competitiveness Report 2015–2016. Insight Report. The World Economic Forum (2016)
17. Uhsakova, I.: Information Modeling the Impact of Social Networking on Customer Loyalty. In: *Modern Methods and Model Data in Information Systems*, pp. 333–398. S. Kuznets KhNUE, Kharkiv (2013) (in Ukraine)
18. Uhsakova, I., Drohov, O.: Computer Modeling of the Influence of Social Internet-Networks on the Formation of Consumer Preferences. *J. Medijski Dijalozi*. 15, 61–73 (2013) (in Russian)
19. Uhsakova, I., Drohov, O.: Information Impact of Social Networking on Customer Loyalty. *J. Medijski Dijalozi*. 13–14, 669–683 (2012) (in Russian)

A Company-Specific Business Process “Accounting of Cash on the Bank Accounts” in the Global Electronic Payment System

Viktoria Yatsenko¹

¹ Kherson National Technical University, 24, Beruslavske st., Kherson, 73008 Ukraine

viktorijajacenko@gmail.com

Abstract. In the document, simulation of the company-specific business process (BP) “Accounting of cash on the Bank accounts” in the global electronic payment system is presented based on the process approach built upon the evolutionary reengineering principles and applying the high-tech software. BP “Accounting of cash on the Bank accounts” (Activity Model As-Is) at the modern Ukrainian companies is analyzed. Simulation of BP “Accounting of cash on the Bank accounts” in the global electronic payment system (Activity Model To-Be) is shown.

Keywords. Business process, accounting, accounting system, information system (IS), e-bank.

Key Terms. Primary document, Workflow, Accounting registers, Operational documents, Universal staging mechanism, Documentary authority, Financial peg, Typical business operation (TBO)

1 Introduction

A current trend in the accounting information systems is transition from the functional (hierarchical) approach to the process one, the fundamental principle of which is representation of a company accounting system as a system of the interacting BPs. Transition to the process approach is accompanied by reorganization of the BPs and depends on a clear understanding of essence of the “business process” concept and its role in accounting. For the functional and modular approach to the construction and operation of an company information system (IS), some localization of the accounting BPs (or their individual modules for the end-to-end processes) within certain areas of accounting is typical, consequently, the structure of the accounting BPs overlaps the functional accounting system of a company, so the basic principle to define the accounting BPs will be the principle of the functional hierarchical ordering.

Under an accounting business process (BP) we understand a guided workflow (actions), resulting in the information (accounting and reporting data) of certain value to

the users. The construction of an accounting BP network is based on the hierarchical functional structure of a company's accounting system (accounting of cash, expenses, settlements with the staff, etc.).

Evolution of relations between a company and a bank shows a strong tendency of adaptation to the realities of functioning in the constantly changing environment. Electronic banking as a stage of relations "client-bank" amidst E-economy is primarily treated as modernization of the communication channels and means, and secondly, as a response to the increased customers' requirements to the banking services quality and range. As Hammer rightly notes, such categories as innovation and speed, service and quality are the key concepts of the new decade[1].

The goal of the paper is analysis of the current state of cash accounting on the Bank accounts for a typical Ukrainian company (Activity Model As-Is) and formulate recommendations regarding the strategy for implementation of a process approach and evolutionary economics to the information system by the example of simulation of the company-specific BP "Accounting of cash on the Bank accounts" (Activity Model To-Be) in terms of the global electronic payment system.

In order to achieve this goal, it is necessary to perform the following tasks:

- build a BP model "Funds transactions accounting" on the basis of evolutionary postulates of economy that defines the place of BP in the architecture of the accounting system of the enterprise and provides a conceptual understanding of the essence of the process and its surrounding at the top level of the hierarchy, while empirical implementation represents the bottom hierarchy level;
- step by step presentation of business analytics as a general description of the business process, its modules, regulations, providing results in tabular form, identifying and fixing all possible versions of the regulations fulfillment on the basis of the fundamental accounting principles in the form of algorithms;
- achieving balance of evolutionary changes and revolutionary quality transformations of the BP "Funds transactions accounting" in order to adapt it to the current and the potential challenges of the environment to preserve the viability of the enterprise accounting system in the long run.

Paper has following structure: section 2 is devoted to related works, section 3 demonstrates a description of BP "Accounting of cash on the Bank accounts" (Activity Model As-Is), while section 4 considers the description of BP "Accounting of cash on the Bank accounts" in the global electronic payment system (Activity Model To-Be), section 5 concludes.

The following terms within BP "Accounting of cash on the Bank accounts" shall be applied in the above sense:

Workflow – a flow route (scheme) of the documents accounting funds between the services of the company and the bank.

Operational documents – the primary documents supporting and confirming the actual operations on the funds flow on the Bank accounts. BP "Accounting of cash on the Bank accounts" is accompanied by the following documents: a Payment order and Bank account statement. Each primary document, in light of the evolutionary theory

of paradigm [3], can be considered as a living organism having its own life cycle, which in turn comprises the steps of passing certain services or business units of a company. The main steps (stages) of a primary document's life cycle include: preparation → accounting treatment → archive.

Universal staging mechanism - a mechanism designed to formalize and control the BP stages.

For every stage of the BP "Accounting of cash on the Bank accounts" the following parameters are prescribed:

- the stage owner: can work with a staging object at the respective stage (BUH_DS – accountant of cash operations of Bank accounts, FO – finance department, GL_BUH – head of accounting department);
- access to the actions related to introducing changes to the document (adding, deleting, copying) or amending some of its details;
- access to the settlement mode at the stage;
- setting of the automatic checks defining the document flow route.

Documentary authority – an organizational or financial administrative document governing the actual operations with the funds on the Bank accounts (an agreement, supplier's invoice, sales invoice to ship the goods, etc.).

Financial peg – an operation linking a documentary authority and an operational document (for example, invoice → Payment order) or two documentary authorities determining the amount of executing of one document by another (for instance, an agreement → invoice).

Accounting registers (Journals and Books) - for the chronological, systematic accumulation, grouping and summarizing of information about the business transactions recorded in the relevant primary documents (for instance, Book of payment orders, Bank statements register, Business operations journal).

Typical business operation (TBO) – a uniform instrument of the accounting entries for all subsystems of the accounting system.

2 Related Works

The definition of "Accounting" has been attracting the leading scientists around the world for a long time. Accounting is being considered as a management function, system, process, craft, profession, and in light of the current economic institutional theory – as a social and economic institution. However, professor V. Zhuk says "Even if we speak of accounting in terms of its primitive, technical component, it is probably hard to find a more complex system of actions providing a set of operations from assessment of the economic life elements to the end-to-end documentary evidence of the business operations with formation of the corresponding accountable figures" [2].

In addition, the evolutionary transformation in plane of the economic and social

organization related to the globalization processes, scale electronic communication development, rapid development of the computer equipment and communication technologies, increasing number of the E-business types, determine the vector of the accounting deep transformation and accounting data formation in accordance with the post industrial society needs. Professors Y. Kuzminskyi and S. Svirko emphasize the urgency of updating the methods and approaches to the research studies and practical advice in the field of accounting [3].

The idea of the transfer of the cyber process approach to organizations management appeared in the late eightieth of the 20th century. Hammer and Champy understands the business process as "a set of actions, which create the valuable outcome for the client on the basis of one or more types of source data"[4]. Harrington interprets this term as "logical, coherent, interconnected set of activities that consumes the resources of the supplier, creates the value and gives the result to the consumer"[5]. The definition of BP is closely associated with reengineering, developed by Hammer and Champy, who interprets it as a rejection of the established procedures, a fresh look at the work of creating a product or service and understanding the customer's value" [6].

The process approach as a basic foundation of modern management approaches should be represented in the balance circuit of the management system, but scientists in the areas of accounting haven't awarded it with the unambiguous assessment and practice implementation. Most scientists believe that "technical" process approach leads to inhibition of accounting modeling. Shyhun considers the process approach to be too narrow for accounting model creation, which is based on classifications of accounting objects, and economic activity is subject to modeling (processes, operations) [7].

However, Osmiatchenko proposed a methodological approach to the design of information systems of accounting, which, unlike the traditional approach based on functionally closed organizational schemes, stipulates the rejection of the autonomous logic in favor of strengthening the adaptation mechanism, focus on advanced structural engineering methods and business process reengineering [8].

Evolutionary economics has not acquired a significant spread in the academic community of Ukraine, but there are attempts in accounting to combine accounting procedures with the facility lifecycle [9-12].

Let us analyze the current state of cash accounting on the Bank accounts for a typical Ukrainian company (Activity Model As-Is) and formulate recommendations regarding the strategy for implementation of a process approach to the information system by the example of simulation of the company-specific BP "Accounting of cash on the Bank accounts" (Activity Model To-Be) in terms of the global electronic payment system.

3 A Description of BP “Accounting of Cash on the Bank Accounts” (Activity Model As-Is)

Cash at Ukrainian companies is usually accounted in IS of the company and an individual E-bank module without an established relationship or not using the electronic payment system at all.

3.1 Specifying Problems (“Bottleneck”)

The problems of operation of the existing model of BP “Accounting of cash on the Bank accounts”:

- double entry of the data contained in the primary documents, duplication of information, high risk of errors due to possible amendments in different systems;
- absence of operational information on balance and cash flow on the Bank accounts, a possibility of payments to contractors in real time, as well as planning and cash flow management at the company;
- slow response of the staff with regard to the innovations;
- limited resources for the purchase and operation of the appropriate IT platform supporting the workflow management, for instance, ERP- or CRM-systems.

3.2 Objectives and Metrics of Simulation and Automation of BP “Accounting of Cash on the Bank Accounts” (Activity Model To-Be)

The main objectives of BP “Accounting of cash on the Bank accounts” are:

- organization of operational accounting of funds on the Bank accounts in real-time in the “now for now” mode;
- a possibility to obtain timely information on the balance and cash flow on the Bank accounts.

Among the key metrics of BP “Accounting of cash on the Bank accounts” the following can be mentioned:

- no duplicated input of the primary (accounting and operational) data into the system;
- optimization of time to input and treat the primary documents;
- availability of actual data on the balance on the Bank accounts in the “now for now” mode;
- a possibility of planning and analysis of the cash flow on the Bank accounts.

4 Description of BP “Accounting of Cash on the Bank Accounts” in the Global Electronic Payment System (Activity Model To-Be)

4.1 General Description of BP “Accounting of Cash on the Bank Accounts” in Accounting System of a Company

BP “Accounting of cash on the Bank accounts” in the coordinates of a company’s accounting system and from the position of evolutionary economics is a “branch” of the accounting area “Accounting of transactions in cash” along with BP “Cash accounting”, “Loans accounting”, “Deposit accounting” and others. To ensure operation of the electronic payment system it is necessary to add BP “Electronic payment system. Relationship with the E-bank system”. Relationship with the E-bank system”, consisting of several modules, to IS of the company (fig. 1).

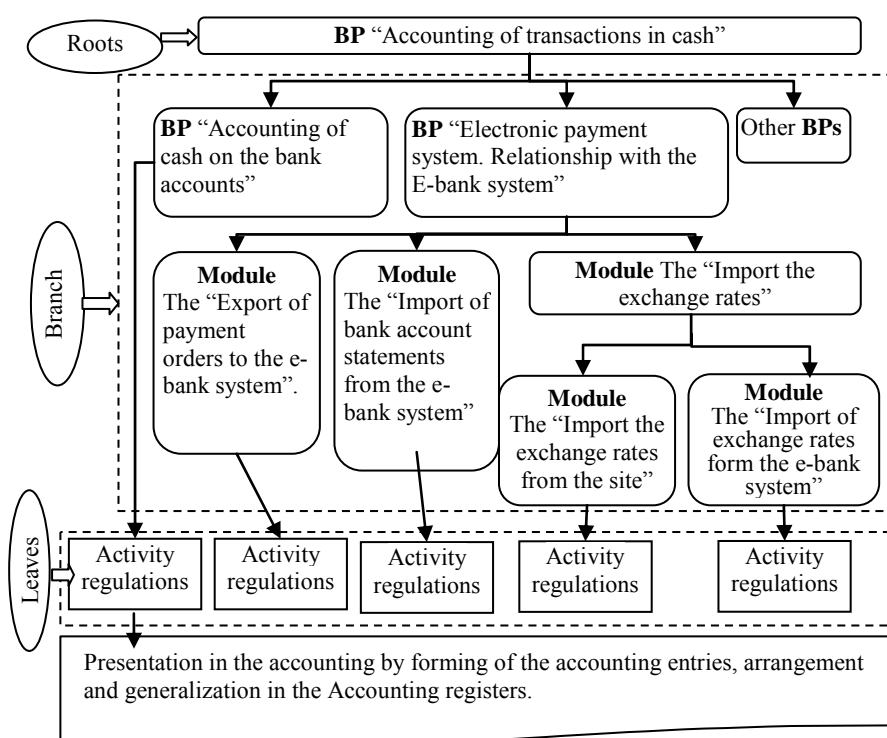


Fig. 1. Simulation of BP “Accounting of transactions in cash” in a company’s accounting system on the evolutionary economics basis

The order for execution of a certain actions sequence defines the procedure for each BP or module. Description of the account parameters and procedure of actions

for BP define the mechanisms of formation and treatment of correspondence of the accounts in the accounting system, the control algorithms, the procedures for compilation of figures for the cumulative statements.

4.2 General Description of BP “Accounting of Cash on the Bank Accounts” in Accounting System of a Company

4.2.1. Procedure and Model of the “Export of Payment Orders to the E-Bank System” Module

According to the procedure of the “Export of Payment orders to the E-bank system” module, the accountant responsible for this area of accounting shall perform the following actions:

1. Prepare the Payment orders.

The Payment orders are prepared manually or generated automatically according to the documentary authority. Registration in the Book of payment orders is automatic in the context of each current account with the bank.

2. Export the Payment orders to an external file.

The duly executed Payment orders are exported to an external fixed-format file.

3. Import of the file to the E-bank system.

The existing file is sent to the E-bank system.

4. Transfer the signed Payment orders to the bank via the E-bank system.

The Payment orders are signed electronically by the responsible persons and sent to the bank via the E-bank system.

So, the Payment orders entered are transferred from the IS enterprise to the E-bank system (fig. 2), these operations shall not be displayed in the financial accounting, in other words, no correspondence of the accounts is formed.

4.2.2. Procedure and Model of the “Import of Payment Orders from the E-Bank System” Module

Every day the accountant responsible for this area of accounting shall perform the below specified actions in accordance with the procedure of the “Import of Bank account statements from the E-bank system” module:

1. Unload the Bank account statements out of the E-bank system.

Using the export procedure, the user unloads a file of the Bank account statements in the given format from the E-bank system through the exchange buffer for the previous day.

2. Download the Bank account statements to the company’s IS.

Import to IS of the company of the file received using the exchange procedure to the interim Bank statements register.

3. Amendment of data (if compliance is not determined automatically).

4. Set the data arrangement mode.

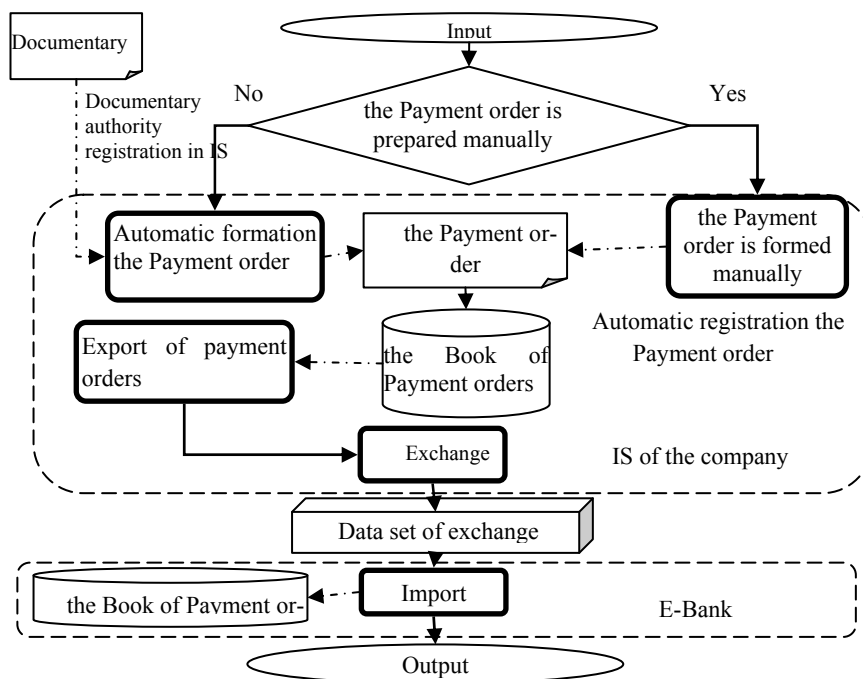


Fig. 2. Model of the “Export of Payment orders to the E-bank system” module

The Bank account statements imported from the E-bank system to IS of the company are processed determining pegging to the relevant Payment orders and/ or Documentary authorities (fig. 3), these data shall not be displayed in the financial accounting, so that no correspondence of the accounts is formed.

Consequently, the operation of BP “Electronic payment system. Relationship with the E-bank system” offers the following advantages:

1. When you download a file to IS of the company from the E-bank system using the exchange procedure, the following actions are performed automatically:

- identification of the organization for its details from the Bank account statements (search in the “Register of organizations” under the Bank codes, account or other details, when the codes match - the organization is identified, the previous statement is filled in from the Register);
- search for the Payment orders matching the Bank account statement in the IS Journal of the company. When the match is found - the previous Bank account statement is marked;
- determination of a typical business operation code for the debit and credit Bank statement (required for possible manual line arrangement in the Bank account statement).

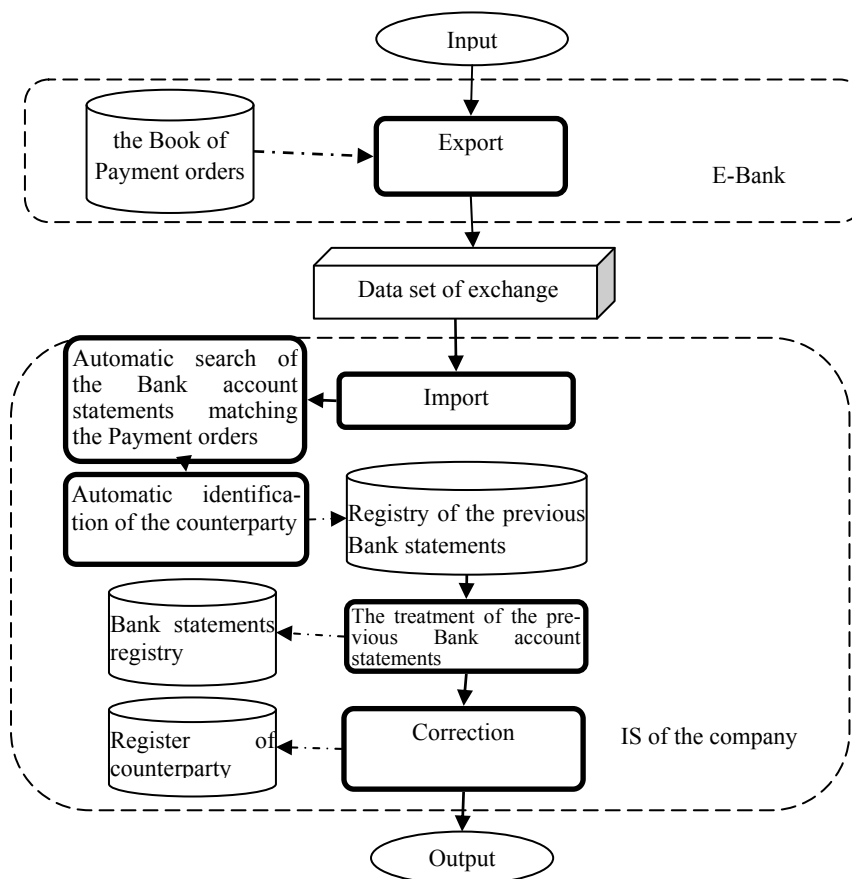


Fig. 3 Model of the “Import of Bank statements from the E-bank system” module

2. The results of treatment of the previous Bank account statements are:

- automatic determination of a typical business operation code for further formation of correspondence of the accounts during arrangement;
- automatic specification of number and date of the documentary authority matching the Bank account statement.

3. The data arrangement mode automatically performs the following actions:

- the Bank account statements from the previous database are transferred to the Bank statements register;
- the “Paid” note is made and date of entry in the Payment orders matching the Bank account statement is specified;
- the financial peg of the Bank account statement to the relevant documentary authority is determined, if the number of the documentary authority is specified during the treatment of the previous Bank account statements.

4.2.3. Procedure and Model of the “Import of Payment Orders from the E-Bank System” Module

The exchange rate published on the website at 14-00 p.m. becomes effective on the following day. The module is configured to daily automatically import the exchange rates if the following requirements are met:

- the user is connected to the internet;
- the time for the daily request for the actual exchange rates is set.

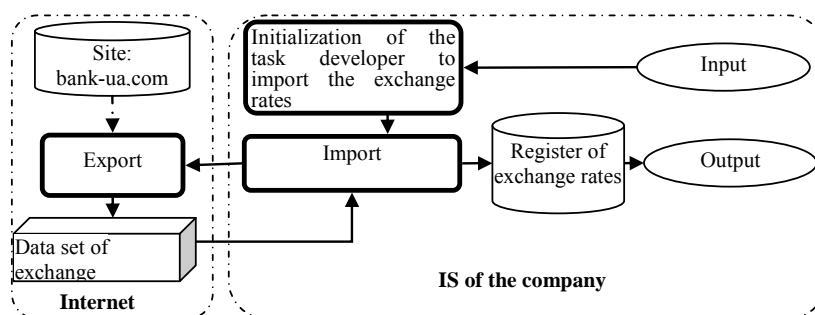


Fig. 4. Model “Import of exchange rates form the site <http://bank-ua.com>.” module

If necessary, all parameters can be changed, and the automatic mode can be completely cancelled.

4.2.4. Procedure and Model of the “Import of Exchange Rates from the E-bank System” Module

This module procedure (fig. 5) includes the daily performance by the employee responsible for updating the currency exchange rates of the following actions:

- export of the exchange rates set by the bank on the current date, to an external fixed-format file from the E-bank system;
- import the file obtained through the exchange procedure to the “Register of exchange rates” of the company’s IS.

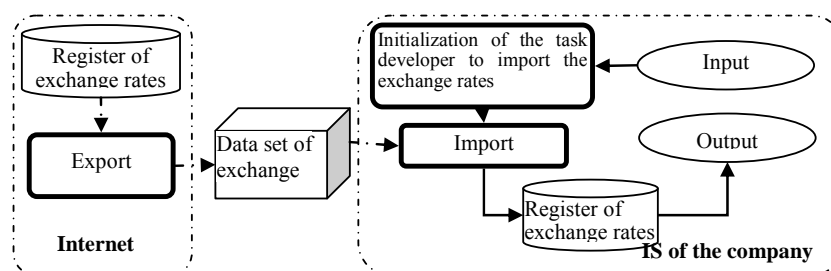


Fig. 5. Model of the “Import of exchange rates form the E-bank system” module

4.3 General Description of BP “Accounting of Cash on the Bank Accounts” in Accounting System of a Company

4.3.1. Procedure and Model of the “Import of Exchange Rates from the E-Bank System” Module

For each current account with the bank, there is a separate sub-account is set up in the work Chart of accounts. For the multicurrency current accounts, one sub-account is set up relating to each current account.

The recommended sub-account format:

311– current account in the national currency (general Chart of accounts)

311XX –bank (ordinal number)

311XXYYY – currency (for instance, code UAH (980)

311XXYYYZZ – current account in the Bank (ordinal number)

312– current account in the currency (general Chart of accounts)

312XX – Bank (ordinal number)

312XXYYY – currency (for instance, code USD (840)

312XXYYYZZ – current account in the Bank (ordinal number)

For instance:

3110198022, 26000000000022 “ Open Joint-Stock Company the name of the bank” (UAH);

3120284023, 26000000000023 “ Open Joint-Stock Company the name of the bank” (USD);

3120397824, 26000000000024 “ Open Joint-Stock Company the name of the bank” (EUR).

4.3.2. Procedure and Model of the “Import of Exchange Rates from the E-Bank System” Module

The owner of BP is an accountant of operations on the Bank accounts that performs the sequence of actions under the Procedure (Table 1).

Table 1. Procedure of BP “Accounting of cash on the Bank accounts”

Action	Essence
1. Preparation of a Payment order. Checking details.	The approved documentary authorities are received by the accountant department to prepare the Payment orders. The accounting department checks the documentary authorities and makes the Payment orders in accordance with the payment period.
2. Making file for export to the E-bank system	Export to the E-bank system according to BP “Electronic payment system. Relationship with the E-bank system”, module “Export of Payment orders to the E-bank system”.

Action	Essence
3. Approval of the Payment orders	Variant I: if payments are made via the E-bank system, the Payment order shall be approved by the person responsible (electronic signature, visa, etc.). Variant II: if payments are not made via the E-bank system, the Payment order shall be printed out, the printed document shall be signed and delivered to the bank, and after the payments are effected one shall receive a printed Bank account statement.
4. Loading of a Bank account statement from the e-Bank system	The Bank account statement for the previous day is exported from the e-Bank system to an external fixed-format file. This file through the exchange procedure is exported to the company's IS into the interim Bank statements registry in accordance with the procedure of BP "Electronic payment system. Relation with the E-bank system", module "Import of Bank statements from the E-bank system".
5. Treatment of the interim Bank statements register	The Bank account statement is checked line by line (counterparty, documentary authority, business operation code, etc.)
6. Financial treatment of a Bank account statement	Financial peg to the following documents is determined: - the expense part of the Bank account statement (if a Payment order in the company's IS, financial peg to the documentary authority is determined automatically); - the revenue part of the Bank account statement to the documentary authority manually.
7. Accounting treatment of a Bank account statement	Automatic arrangement mode (to form the accounting entry). In the electronic document "Payment order" the "Paid" note is made to automatically form the relevant correspondence of the accounts based on the business operation code determined by the peg. A corresponding account to the "Account with the bank" is determined by the documentary authority. The mode provides a variant of a group formation of correspondence of the accounts based on the documents chosen. After treatment of the Bank account statement and formation of correspondence of the accounts for a certain date, the balance-based account may be closed for amendment before the date specified, that is, any entering/amendment/deleting of correspondence to the account before the specified date inclusive is not permitted.

The algorithm of executing the sequence of actions of BP "Accounting of cash on the Bank accounts" is graphically shown in Fig. 6.

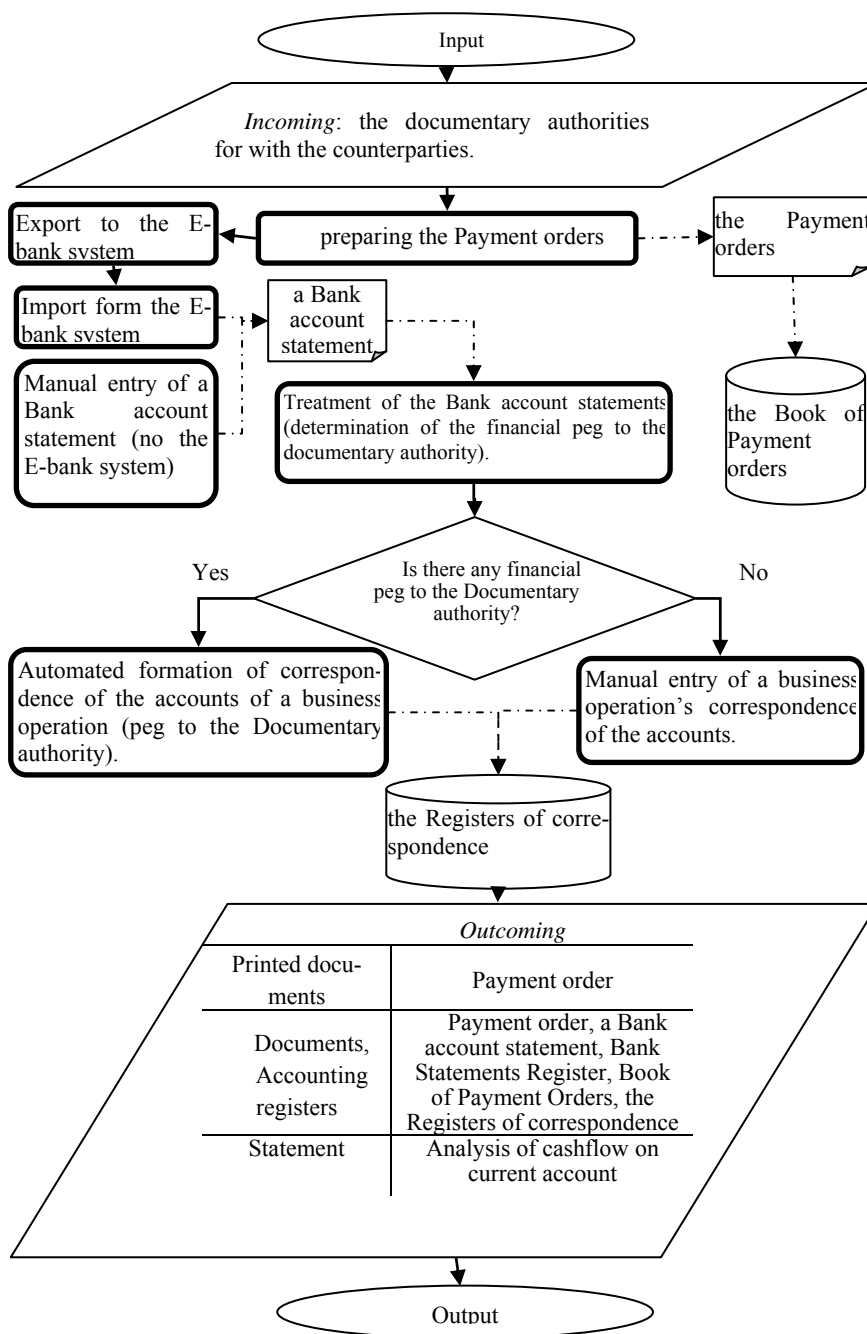


Fig. 6. Model of the BP “Accounting of cash on the Bank accounts”

4.3.3. Procedure and Model of the “Import of Exchange Rates from the E-Bank System” Module

The company’s IS makes a provision for keeping the workflow rules according to the approved procedures. A primary document serves as:

- a basis for taking actions;
- a legal confirmation of the transaction;
- a basis for representation in the accounting (entry);
- an accounting information storage medium.

The primary document details storage is divided into the mandatory details (established by the external standards) and optional (introduced by the company at its own discretion). The proposed module of BP “Accounting of cash on the Bank accounts” enters the required details into the mechanism of a primary document preparation, but does not exclude the possibility to revise and add them at the configuration stage (table 2).

Table 2. Documents of BP. Document details

Detail	Explication
the Payment order	
Number	Automatic formation the numeration. There is a possibility of manually correction.
Date	Date of Payment order
Bank date	Date of Bank statement
Organization	
Agreement	
Accounts	31*
Bank	According to the account
Current account	According to the account
Sum	
Including VAT	There is a possibility of manually correction.
Receiver	
Bank of receiver	
The current account of the receiver	
Payment purpose	
Comment	
the Bank account statements	
Number	Ordinal, automatic formation.
Bank date	Date of Bank statement
Organization	
Agreement	
The type of business operation	To determine a corresponding account
Accounts	31*
Bank	According to the account
Current account	According to the account
Currency	
Sum in the currency	
Sum in the national currency	
Payment purpose	

A primary document details determine the assignment operation – specification of numbers of the accounts that are debited and credited, the directions of search and systematization in the section specified.

Normalization (data tabulation) lets to have the Journals (Registers) as the relational tables with the determined associative links between the elements, the lines (tuples) of which comply with the documents entered, and the details of the documents are put into columns (domains).

Each tuple in the table is usually unique due to a single domain.

For example, in the Book of Payment orders (table 3), the “Document number” domain defines the tuple uniqueness. In the operation, if necessary, you can add domains, thus expanding the data analyticity.

Table 3. The Book of Payment orders

Domain 1	Document status	...								Domain n	
Domain 2	Was paid	Document date	Document number	Type of the document	Account of the organization	Sum	Currency	Counterparty	Type of the transaction	Account of the counterparty	
Conducted	Paid	01.03.2016	0000001	the Payment order	Current account of the organization	5 000,00	UAH	Ltd «A»	Supplier payments	Current account of the counterparty	Tuple 1
Conducted	Paid	01.03.2016	0000002	the Payment order	Current account of the organization	10 000,00	UAH	Ltd «B»	The buyer paid	Current account of the counterparty	Tuple 2

Each tuple of the Bank statements registry (table 4) makes the “Bank account” domain unique, since a Bank account statement is made based on all transactions in one day on one current account.

The financial treatment of a Bank account statement forms the accounting entries in accordance with the settings of a typical business operation by the procedure of the corresponding accounts determination based on the Documentary authorities. The result is entered into the Business operations journal, which actually combines the information in the Bank account statement cross-cast and the related correspondence of the accounts (fig. 5).

Table 4. The Bank statements registry

Domain 1	Domain 2	⋮					Domain n
Date	Bank account	Currency	the Beginning Balance	Inpayment	Payout	Balance at END	Account of the counterparty
01.03.2016	Current account of the organization	UAH.	30 000	5 000		35 000	Current account of the counterparty
01.03.2016	Current account of the organization	USD	45 000		10 000	35 000	Current account of the counterparty

Tuple 1

Tuple 2

Table 5. The Business operations journal

The document status	Date	Document number	Type of the document	Counterparty	Sum	Currency	...
Conducted	01.03.2016	0000001	Payment order	Ltd «A»	5 000,00	UAH	
Conducted	01.03.2016	0000002	Payment order	Ltd «B»	10 000,00	UAH	
...							
Debit	Credit	Sum	Substance of Business Operation				...
«Settlements with suppliers »	«Bank account»	5 000,00	The debt to suppliers has been put out				
«Bank account»	«Settlements with buyers»	10 000,00	The buyer paid has been received the current account				

5 Conclusions

Recently, one of the most popular areas of the scientific research is the process approach to the organization and operation of the domestic enterprises, and the advanced foreign and domestic experience of its practical implementation instead of the common functional approach is analyzed. At the same time, we should not forget that apart from the functional approach, the systemic one, which considers a set of the interrelated elements aimed at achieving objectives in the rapidly changing environment, and the situational approach, the essence of which is that the suitability and effectiveness of the system is determined by the situation, as well as the economic theory concepts, which have been forming and developing according to the market conditions, are of a significant influence on the contemporary process approach. Thus, transition to the process approach does not mean a complete give up on other approaches because the process approach is compatible with a variety of approaches,

concepts and theories, so, the question is the change of an imperative, however, the transition shall be based on the reengineering principles and the high-tech software application.

The author worked out the BP model "Funds accounting in the account of the Bank" in the company in the global electronic payment system that summarizes the IC typology involving users with the appropriate level of expertise and competence in IC. The implementation of the BP model reveals the mechanism of documentation and reflection of economic operations according to the principle of double-entry using step by step developed algorithms and providing the results in tabular form. It is proposed to take the results of this research as a basis for simulation of an accounting BPs network for construction of a company's accounting system.

References

1. Bezgin, K. S.: Quality management of business processes in the enterprise. Manuscript (2009)
2. Kozenkov, D.: Business Processes Planning as the Basis of the Creation Enterprise Architecture. The Visnyk of the SSU 3. 126--136 (2011)
3. Harrington, H. J., Esseling, E. K.C., van Nimwegen. H.: Business Process Improvement Workbook: Documentation, Analysis, Design, and Management of Business Process Improvement. London: McGraw-Hill, 1996
4. Fedorov, I.: Modeling business processes in BPMN 2.0 notation. M., MESI (2013)
5. Davenport, T.: Process Innovation: Reengineering Work through Information Technology., Boston, MA.: Harvard Business School Press (1993)
6. Druchenko, L.: Implementation of Business Processes Modeling and Optimization in the Local Government. Public Administration: Theory and Practice 1(3) (2010)
7. Ponomarenko, V., Minuhin, S., Znahur, S.: Theory and Practice of Business Process Modeling: Monograph. Simon Kuznets Kharkiv National University of Economics (2013)
8. Manganeli, R., Klein, M.: The Reengineering Handbook: A Step-By-Step Guide to Business Transformation. New York: Amacom (1994)
9. Hammer, M., Champy, J.: Reengineering the Corporation: A Manifesto for Business Revolution. London : Nicholas Brealey Publishing (1993)
10. Taranyuk, L., Zaporozhchenko, O.: Organizational Principles of Reengineering Business Processes of Enterprises. Mechanism of Economic Regulation 4, 122--131 (2011)
11. Kuzmin, O., Sidorov, Y., Kozik, V.: Achievements and Problems of Evolutionary Economics: Monograph. Publishing House of Lviv Polytechnic National University (2011)
12. Reilly, J. P., Creaner, M. J.: NGOSS Distilled, The Lean Corporation (2005)
13. Kalyanov, G.: Business Processes Modeling, Analysis, Reorganization and Automation. Finance and Statistics (2007)
14. Melnikov, A.: Bases of Business Organization. Center of educational literature (2013)
15. Korzachenko, O.: The concept of business processes modeling and optimization of telecommunications companies. Scientific and theoretical journal Khmelnytsky University of Economics "Science and Economics" 4 (32), 247--253 (2013)
16. Gruber, T.: A Translation Approach to Portable Ontologies. Knowledge Acquisition, 5 (2), 199--220 (1993)

Life-Long Learning: Individual Abilities versus Environment and Means

Oleksandr Burov

Institute of Information Technologies and Learning Tools,
9 Maksyma Berlyns'koho St, 04060 Kyiv, Ukraine

ayb@iitlt.gov.ua

Abstract. This paper describes new and emerging technologies in education, learning environments and methods that have to satisfy life-long learning of person, from school age to retirement, on the basis of the psycho-physiological model of the cognitive abilities formation. It covers such topics as: evaluation of a human (accounting schoolchildren, youth and adults features) abilities and individual propensities, individual trajectory of learning, adaptive learning strategy and design, recommendation on curriculum design, day-to-day support for individual's learning, assessment of a human learning environment and performance, recommendation regards vocational retraining and/or further carrier etc.). The specific goal is to facilitate a broader understanding of the promise and pitfalls of these technologies and working (learning/teaching) environments in global education/development settings, with special regard to the human as subject in the system and to the collaboration of humans and technical, didactic and organizational subsystems.

Keywords. Human abilities, development, learning environment, tools, safety

Key Terms. Capability, Technology, PSI-ULO, ICTEnvironment, PSI Core Time ontology

1 Introduction

Global changes in education aimed at ensuring tasks to account the transition from the post-industrial to a digital era as well as to the knowledge society when a human individual cognitive and creative ability become crucial for the mankind development [11]. Education needs to be more and more individual-oriented, securing individual psychophysiological abilities and development including network (as a main learning/nurturing environment) security and safety [14, 17].

Going beyond learning time and space constraints has been a crucial motivation. Social, businesses, education/learning actions now take place globally and in many time zones. Human knowledge and intelligence is now being systematically transferred to computers, allowing them (e.g. in the form of robots or automated systems) to operate autonomously even far away from the original time or space of its human

originators. Moreover, that notation of space has expanded: new spaces have opened up like the cyber space of Internet, cloud education etc.

Giftedness is associated with creativity, cognitive abilities, intelligence and other psychological qualities. To date, challenges to education are associated with modern possibilities, innovative technologies and higher attention to a student ability to obtain knowledge and skills up to date. Using the concept of context-based cognition it is assumed that learner's individual construction of knowledge should take place within a certain context, which is similar to the context in which this knowledge should be applied in the future [3].

At the same time it is known that human cognitive opportunities can vary from day-to-day depending on his/her functional state and fitness-for-work [4, 5]. As a result, such changes can impact a human ability to adapt in a particular learning environment, to perceive and to conceive new information, to use it, to get knowledge and skills and to prevent its inefficiency. Besides, it is known a negative impact of a human's work with computer over a long period on his/her health and mental efficiency nowadays. It is especially important for children and young people whose organism is more sensitive and unstable comparing to adult one. Health disturbances have impact on a child mental efficiency in learning process as a consequence of this. Some of them have fluctuation nature in their development. As a result, efficiency of learning and talent nurturing depend not only on innate child properties, but on information contour of functional system of learning as a specific type of activity.

Difficulties to assess and to predict students' abilities and achievements are result of lack of recognized definition of giftedness and measurement tools [10], accounting actual learning environment [13] and dynamic nature of giftedness and talent development [12].

It was stated that the most fruitful approach to understand mechanisms of activity is a theory of functional systems proposed by P.Anokhin [1] and his disciples (K.Sudakov, A.Navakatikyan) who proposed the concept of the activity functional system that connects in one model physiological systems state, conditions of the work environment and the goal of activity. According to that a human activity is accompanied by creation and maintenance of functional systems that are activated dominant brain structures and correspond activity of one or another organism systems and are quite enough stable for particular type of the human work. But this is not enough to understand to what degree it can explain reasons of insufficient accuracy of a human performance prediction.

The **purpose** of the article is to describe psycho-physiological basis (concept and model) of a human learning process, as well as human view on network-born threats for a human (users of all ages) health and performance.

2 Method

The method used is based on the model proposed by the author. This model is a development of the basic idea of P.Anokhin regards formation and functioning of the "functional system of activity" [1].

Any human activity of mental type could be analyzed as an operator work, because professional operator (pilot, power unit dispatcher, driver etc.), manager and learner

have to do with the objects indirectly through the information model of process activities. Work (activity) of such a type has specifics consisted, first of all, in a workload, because of not so power-consuming processes, as informational ones that are, by their nature, not so discrete, but continuous [5]. The reason is that the conceptual model of activity (as result of a human psychological adaptation to the work) is expanded in the time independently on external process and an operator activity consists in discrete comparison of the information obtained from outside with the model existing constantly. I.e., conceptual model can be considered as an information stratum of professional work, and physiological chain „afferent inputs – activity acceptor – physiological control – effectors - act” is an energetic stratum. The goal of professional training is forming of the conceptual model of activity of particular type, carrying out of the particular task. It means creation of „information contour” that exists and is maintained in activated state in carrying out process for purposeful activity and embraces afferent inputs, decision making block, activity acceptor and act program, as well as the object of activity (is represented as information model in case of operator-watcher that creates the information contour together with the imagine- conceptual model).

Energetic and information strata can be represented as two contours which partly coincide at the level of morphological structures and functions, but partly differ because of including into the information contour an activity object that does not participated in the energetic contour of the organism regulation, but is an inalienable part of the information contour. Human activity is a mediator between internal and external environment of organism, projection of structural-function specific of professional homeostasis on the operator work. Output parameters of the activity program (activity effectors) stand in the information contour as parameters of capability. In such a context, operator’s activity is an activity program realized as physical and/or mental acts in external environment.

2.1 Research Question / Hypothesis

The specific of learning work consists in the informative processes which are more involved than energetic ones, and conceptual model of learning work is a result of psychophysiological adaptation to this activity. This is relatively close to an operator activity and, from substantial viewpoint, consists in discreet comparison of information received from the educating system with the conceptual learning model used, in carrying out a particular task. In other words, in forming and using an „information contour” that exists in active state in the process of purposeful activity (Fig.1). This contour includes afferent inputs, decision making block, act acceptor and act program, as well as the object of acting (information model for operator). This hypothesized model of psychophysiological maintenance of mental activity was developed as an advancement of P.Anokhin’s theory of functional systems.

Stability level of circulation in the information loop can be provided by its dynamically fluctuation according to changes in levels of activation of certain physiological systems subject to constant resultant of power. In formalized form, this means that if the index of indicator level of information circulation i (eg, speed) in the information loop is constant (ie, periodic with frequency $\omega_i \rightarrow 0$), mental capacity R is

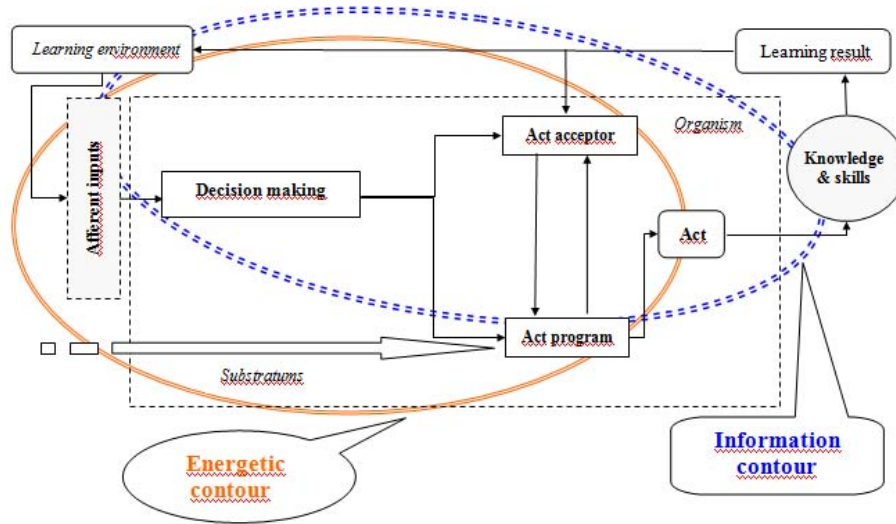


Fig. 1. Theoretical scheme of the functional system of learning activity, where regulation (Anokhin, 1973, [1]) was divided into two contours – information and energetic ones

at one level, and integrated parameters of physiological systems φ_j change of characteristic for their frequency ω_j , then

$$F = Af(\omega) = \sum B_j \varphi_j(\omega_j) \approx \text{const},$$

where A , B_j - average values of estimated parameters in the range of observation.

Description is given from viewpoint of both theory of functional systems and new “vortical model” of ability development as a dynamic transformation [7]. The latter means that natural inclinations transformation into child ability can involve physiological, psychological and external sources as “developing environment” that create particular abilities as a specific combination of abilities existing only in interaction with actual environment. And this interaction has dynamic nature depending on power and rate of three sources circulation that involve information, knowledge and skills as a catalyst saturating the giftedness structure created.

Afferent inputs correspond to well-known sensory systems and could be divided into:

- visual,
- auditory,
- tactile,
- gustatory,
- scents,
- temperature,

- vestibular,
- kinesthetic.

In such a context analyzers' characteristics are as follows: energetic, informational, spatial and temporal.

2.2 Changes of Cognitive Abilities in Day-to-Day Intervals

It was confirmed that human ability to cognitive performance is not stable over time span that impacts on human mental performance efficiency [6]. This corresponds to idea of the central role of cognitive processes in prediction of training and operation. Oscillations of psychophysiological indices of a human state and capacity have exogenous and endogenous nature.

The model can help to answer the following questions: what to measure, assess and predict? Main issues to be accounting in this context are as follows:

- Measurement should be a tool, not a goal.
- Parameters of the information contour should be measured for each stage of a student work:
 - Professional adaptation-dis-adaptation.
 - Day-to-day fitness-for-learning.
 - Current learning performance degradation.

This method and its application regards day-to-day fitness-for-learning are discussed by result of study of 20 intellectually gifted school boys' cognitive abilities on a month day-to-day basis. Subjects participated in two types of observation: (1) preliminary assessment of specific abilities for particular areas of mental activity (math, science, technical etc.), (2) day-to-day monitoring of performance the series of cognitive tasks [7].

The effectiveness of research psycho-physiological techniques considerably rises in case of usage not of a set of tests, but of a psycho-diagnostic system. To solve this task the computer system of psycho-physiological researches of a human psychomotor and cognitive activity in conditions of research laboratories and rehabilitation centers was developed. Experiments included psychological tests performance by subjects at the computer display and simultaneous measurement of physiological parameters. They were used tests as follows: short memory, perceptual (searching of missed numeral), cognitive (logic-combinatorial) [2]. In all tests we registered time of each task performance in milliseconds, correct (expected) and really entered answers. Besides, we used a subjective state assessment of the examinees by means of the reduced variant of the test "General_state - Activity - Mood" (GsAM) at the beginning and at the ending of the test session. (the indices of mood mood, serviceability FfD, attention atten, anxiety anxiety) prior to the beginning (index "0") estimated and upon finishing the tests performance.

As indices of physiological "cost" of activity and the human state we registered a heart rate HR and blood pressure (systolic BPs, diastolic BPd) by means of the car-

diomonitor "Solveig". The indices HR, BPs and BPd we registered during 5 min prior to the tests beginning (index "0") and 5 min after finishing (relaxation).

The specific of the research technique consisted in to check the variability in time of psychophysiological indices registered when subjects performed cognitive and perceptual tests, under impact of infradian rhythms of a various origin. Each examinee took part in experiments with constant workload in the same phase of day to eliminate the circadian effect.

The data on influence of solar activity on a human health and some physiological systems are known, however results of study of cognitive activity associating with heliophysic parameters are not known in the scientific literature to date. In our preliminary pilot researches the precise connection between effectiveness of operator activity and parameters of a solar wind (SW) was revealed. With the purpose to study this phenomenon we registered indices of proton component of a solar wind - velocity SWsp (km/s) and density SWden (proton/sm³) on the data from Internet site NASA [15], as well as parameters of the geomagnetic field (GMF) - planetary index Ks and index of "equivalent amplitude" A.

Average values of physiological parameters testers for testing days indicate the individual character of their dynamics as in the initial state (immediately after school) and after the test activity. Comparison of changes in the nature of physiological parameters as a reaction to stress (cognitive tests are simple and conform to the logical skills of Grade 1 pupils of secondary school), indicates that even these activities can serve as a functional test of the occurrence of fatigue, which was found in previous studies (Fig.2).

It is significant increase in heart rate in the first of the testers, while, as in other myocardial activity is, on average, unchanged. This subjective assessment of mood under the influence of the test after the end is more stable comparing to the original state (just after the last lesson). Thus, different tendencies of objective performance changes, self-assessment and physiological changes in time, under learning and external factors impact need to be accounting as the comprehensive approach in education process to increase its efficiency and to reduce negative its negative impact on students' health.

2.3 Possible Applications of the Model to Control Students Cognitive Performance

Practical realization of these ideas is applied for the cybernetic mechanism of a human performance control. There is the set of psychophysiological parameters of human **P**, which are related to forming and realization of his capacity. The set of his professionally important qualities, parameters of professional senescence and current capacity is a number of the parameters of operator professional activity **D**. The task of synthesis of the system for psychophysiological prediction of operator capacity is the task of optimum reflection **P** on **D**, that provides maximal quality of functioning SLTS **Q** subject to the system cost **C**, which does not exceed a possible level **Clim**: $[p \in P] \rightarrow [D^* \subset D]$, where **m** – operation of optimum reflection of elements **P** on the set of elements **D**; **D*** - optimum set of parameters of professional activity

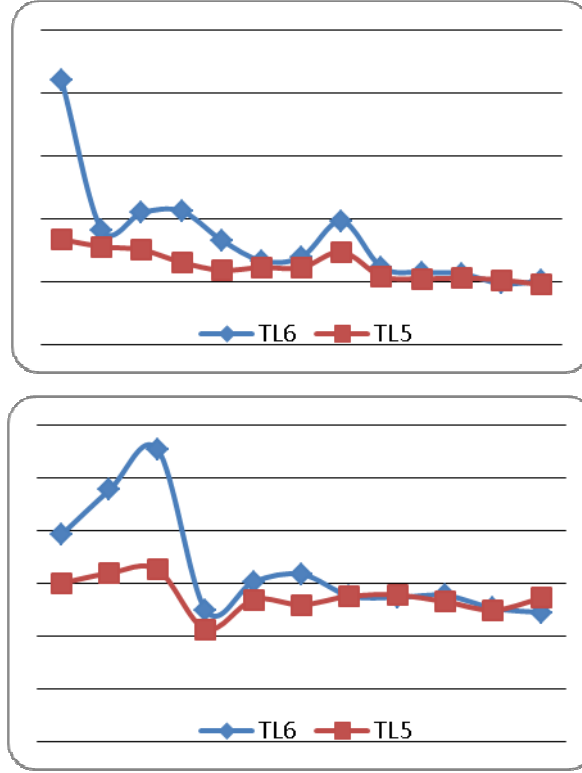


Fig. 1. Daily dynamic run-time tests of two subjects. On the axis: abscissas - days of test performance, ordinate - the average task time performance in tests T6 and T5, ms.

$$D_0 = \arg \max_{D_0 \in D} Q(D) \text{ under } C \leq C_{lim}.$$

Notion “control” is used as a process of adduction of the set object in the state, that answers the task put. Such determination allows considering the capacity of operator as an object of control by organizational-psychological methods, and the system of assessment and prediction considering as a system of the operator capacity control. In the general case of ergonomics approach to the analysis of efficiency of the system Human-Technique-Environment (SHTE) the estimation and purpose of prognosis of functional state and operator capacity U in SHTE is to provide maximal quality of functioning of the system Q , which relies on the realized of operator capacity R , organization of the system O , the state of equipment E , inter-element interface I , dynamics of SHTE changes in time t . Such task is described as:

$$\hat{U}(t) = \arg \max_{\check{R}, \check{O}, \check{E}, \check{I}} Q(t) = \arg \max_{\check{R}, \check{O}, \check{E}, \check{I}} f [R(t), O(t), E(t), I(t), t],$$

where $\check{R}(t) \in R(t)$, $\check{O}(t) \in O(t)$, $\check{E}(t) \in E(t)$, $\check{I}(t) \in I(t)$.

If estimation of capacity works for real $\hat{U} \rightarrow U$, quality of functioning SHTE can achieve the maximal value $Q > Q_{\max}$ thanks to the use of maximal operator capacity.

It was developed the system to optimize students' work with computers in regards to choose optimal form of learning for a particular moment with advanced proposals for a student that gave flexibility of work in particular day accordingly to his/her functional state (to take lectures, to work in interactive mode with computer or teacher, to make applied tasks, to work in library, to work in Internet etc. or to delay an active form of teaching for another time). This could give a tool to intensify the education time by way to re-allocate efforts and to optimize a time use.

The physiological mechanisms of the adaptation of a student's activity functional system to the education environment may vary depending on the conditions, location and time of his/her activity. Hence, proposed approach may contribute to improve human working conditions if it takes into account not only macroergonomic requirements (spatial aspect of human interaction with the environment), but also ergodynamic ones (temporal aspect). Direct or indirect measurement of possible psychophysiological changes when learning makes it possible to review a functional state of the student, which helps to predict his/her individual fitness and reliability for effective learning process.

Task performance measures before current learning activity should be focused on rate and accuracy of information processing. The student's test performance, which precedes a work with the computer system, consists in decision of the same type of cognitive and perceptual tasks. The time and accuracy of each task performance are registered. Task performance times produces during research a time series, which consist in sequence of values of tasks decision time. The further analysis of time series permits to reveal the "waves" that are induced by the regularity of fluctuation of task performance time.

Computer method allows varying flexibly the tempo, volume and complexity of test tasks for different levels of cognitive and perceptual complexity with and without interferences. The research method foresees to investigate and take into account the biorhythms influence on the human behaviour and fitness for duty with periods up to 1 month.

According to schedule a student came to the computerized workplace and performed a cognitive test with the computer sub-system and got as a result a time of his/her effective work with computer for current day. That recommended time could be coordinated with the student's supervisor and a real time could be changed.

Electronic education gives new opportunities such as:

- flexibility of education programs – a student can choose courses, teachers, time of active work, etc.;
- individualization of education process – re-allocate time and education resources in dependence on a student's individual psycho-physiological possibilities to make this process more intensive and to give equal opportunities for both common people and people with disabilities.

This advanced proposals can help a student to make a decision what type of work in particular day is preferable accordingly to his/her functional state:

- to take lectures,
- to work in interactive mode with computer,
- to work in interactive mode with a teacher,
- to make applied tasks,
- to work in library,
- to work in Internet,
- to perform individual tasks etc.
- to delay an active form of teaching for another time.

Although further work is required to gain a more complete understanding of such a tool's use, it is clear that this could give a student an opportunity to intensify a time by way to re-allocate his/her effort and to optimize a time use. The main idea of such approach is "Don't waste a time !". Another aspect of the same problem is an adaptive automation of human-computer interaction in accordance with a student functional state.

The current study provides further evidence that traditionally education managers do not analyze such opportunity to make the education process more effective, because a student must study as it is done. This way is aimed to a "group norm", but not to an individual, especially gifted and talented.

3 Networking Threats

As it was stated [14], not so far ago a human place and role in networks could be described as a terminal element (node) linked to other elements with its specific interface (having human and technical parameters). Currently, when a human life and activity has more virtual nature, information environment (network) becomes independent factor, because process of a human presence as well as results of his/her activity loses their localization in space and in time, as well as could be affected at anytime and anywhere. Even more, those results could «live» inside the network «infinitely», because technical resources holding them are distributed, flexible and supported continuously.

Recommendations for improving a human psychophysiological security have been developed and described [5]. The further development of ergonomics criteria could be developed accounting the new human activity nature and a multiaspect ergonomic analysis [16, 18].

Active usage of networks, especially by children and youth, is accompanying by increase of different kind of threats coming from networks. Particularly acute this problem obtained with development and use of social networks. Most active hidden threats (for children) emanating from the computer network, stacked in the following classification scheme [8]:

- Viruses attacks.
- Cyber-crime (spamming, carding, phishing, botnets etc.).
- Threats from network-surfing (cyber-bulling, "adult" content, illegal content, on-line violence, disclosure of private information, pay services etc.).

It is recommended considering the interaction between schoolchildren and students with computer network as the system "Human-technique-environment" [8]. In this system, computer network serves as a machine that allows us to consider the impact of the network on a human as a threat coming from the machine. Accordingly, the concept of "network effect" can be revealed through the concept of "operator error and reducing the quality of operator activity", "impact of computer games" and "Internet addiction".

Threats coming from networks can be classified into the following types: active and passive, overt and hidden, current and deferred.

Using ergonomic approach and methodology, it is possible to evaluate active hidden dangers as a hierarchical set of indicators:

- one integrated (complex) index - the level of danger as a result of a computer network; index is a dimensionless quantity and is on the upper level assessment system;
- group of three indicators - levels of hazards caused by virus attacks, cyber-crime or internet surfing; indicators are dimensionless quantities and are at average levels of the system assessments;
- a set of individual indicators of the group of one or set of threats; indicators are also dimensionless quantities and are on the lower level grading system.

From educational domain context, target groups of CS could be classified as follows:

- Students as operators
- Educators
- Children/Youth (in general)
- Population (in general, as social environment for children).

The human view regards CS could be a fruitful approach to define tasks, resources and ways of solution the above mentioned challenges [8].

Human View – constraint.

If a system requires a human interface, then the system must be designed to accommodate the human as a passive and as an active element, creating sub-system for safety both for and from a human.

Human View – functions.

Provide a justification for the allocation of tasks and functions between the humans and machines depending on a human current status and capability.

Human View – role.

Describes the roles that have been defined for the human interacting with the system and their possible changes over mission time (f.e., from simple executor to leader and/or commander) accounting his competencies, ability for tasks generation, leadership etc.

Human View - human network.

Team performance impacts, re-allocation, dependencies and communication.

These views should be a basis ergonomically grounded for design and creation new and safety educational tools.

4 Conclusions and Outlook

Digital life and activity gives new opportunities for people and new problems in different domains including education.

Broader understanding of the promise and pitfalls of learning technologies and working (learning/teaching) environments in global education/development settings could be useful with special regard to the human as a subject in the system and to the collaboration of humans and technical, didactic and organizational subsystems.

Global changes in education aimed at ensuring the transition from the post-industrial to a digital era as well as to knowledge society when human individual cognitive and creative abilities become crucial for the further development of mankind.

Psycho-physiological model of learning and cognitive abilities development could be a basis for more effective design of learning organization and process.

They are proposed to discussion: identification of areas in which coordinated research efforts are required to expand an understanding of these network technologies, their effectiveness, the potential risks, and the potential benefits of new ways to educate, learn and collaborate.

References

1. Anokhin, P.K.: Principle questions of the general theory of functional system. Principles of the system organization of function. Science, Moscow, Russian Federation (1973)
2. Reshetyuk A.L., Burov O.Yu, Polyakov O.A. et al.: Automation of experiment for psycho-physiological research in work physiology. Ministry of Health of Ukraine, Kiev, 11 (1993)
3. Bednar, A.K., Duffy, T.M., Perry, J.D.: Theory into practice: How do we link? In: Duffy, T., Jonassen, D. (Eds.): Constructivism and the technology of instruction: A conversation, pp. 17--34, Erlbaum, Hillsdale, NJ, USA (1992)
4. Burov, A.: Evaluation of functional state of operators on parameters of mental serviceability. Human Physiology, 2, 29--36 (1986)
5. Burov, O.: Ergonomics, functional state and human fitness-for-duty. Zastosowania Ergonomii. 1-3(57-59), 203--214 (2005)
6. Burov, O., Tsarik, O.: Educational workload and its psychophysiological impact on student organism. Work, 41(1), 896--899 (2012)
7. Burov, O. Ju. (ed.): Dynamics of development of intellectual abilities of gifted person in teenagers. Tov «Informacijni systemy», Kyiv (2012)
8. Burov, O.Ju.: Educational Networking: Human View to Cyber Defense. Institute of Information Technologies and Learning Tools, 52, 144--156 (2016)
9. Burov, O. Virtual Life and Activity: New Challenges for Human Factors/Ergonomics. In: Proc. Symposium "Beyond Time and Space", STO-MP-HFM-231. STO NATO, pp. 8-1--8-8 (2014)
10. Sternberg, R.J., Davidson, J. E.: Conceptions of Giftedness. 2nd Edition. Clark College, Portland, USA (2005)
11. Future of education: lessons uncertainty (Abstracts of the World Economic Forum in Davos). online: <http://biz.liga.net/upskill/all/stati/3225018-budushchee-obrazovaniya-uroki-neopredelennosti.htm> (2016)

12. Gagné, F.: Transforming Gifts into Talents: The DMGT as a Developmental Theory. In Colangelo, N., Davis, G. A. (Eds.): Handbook of gifted education (3rd ed.), pp. 60-74, Allyn and Bacon, Boston, USA (2003)
13. Horowitz, F. D.: A developmental view of giftedness. *Gifted Child Quarterly*, 31(4), 165--168 (1987)
14. Kleinberg, J.: Analysis of large-scale social and information networks. *Philosophical Transactions of the Royal Society, A*, 371, 1471--2962 (2013)
15. SEC's Anonymous FTP Server (Solar-Geophysical Data). online: <http://sec.noaa.gov/ftpmenu/lists/ace2.html>
16. Pacholsky, L.: A new methodological paradigm of a multiaspect ergonomic analysis. In: Pacholski, L.M., Marcinkowski, J.S., Horst, W.M. (eds.): Dilemmas and issues of modern ergonomics and work safety education and researches, pp., 413--426, Poznan, Poland (2004)
17. Raspopovic, M., Cvetanovic, S., Jankulovic, A.: Challenges of Transitioning to e-learning System with Learning Objects Capabilities. *The International Review of Research in Open and Distributed Learning*, 17(1) (2016)
18. Wilson, J. R., Carayon, P. : Systems ergonomics: Looking into the future – Editorial. Sp. I. on systems ergonomics/human factors. *Applied Ergonomics*, 45(1), 3--4 (2014)

The Use of Cloud Services for Learning Foreign Language (English)

Hennadiy Kravtsov, Olga Gnedkova

The Kherson State University
27, 40 rokiv Zhovtnya St., 73000, Kherson, Ukraine

kgm@ksu.ks.ua, gnedkova84@gmail.com

Abstract. Research goals: determine particularities of using cloud services in English communicative competence forming process and develop the model of distance learning system (DLS) and cloud services interaction that improves the quality of the learning process. **Research objectives:** to identify, describe and develop methodological features of the model of distance learning system and cloud services interaction; consider on the examples the methods of use cloud services in DLS in foreign language training, aimed at English communicative competence forming process of students; experimentally investigate the level of using of cloud services in distance learning and identify ways of improving its use. **Object of research:** methodical system of foreign language students' training of language faculties in high institutions. **Subject of research:** methodological model of the model of distance learning system and cloud services interaction in English communicative competence forming process. **Research methods used:** review and analysis of scientific publications, psychological, educational and instructional materials, modeling of complex systems, questionnaires, conducting pedagogical experiment. **Results of research:** The model of distance learning system and cloud services interaction was developed and the methodological particularities of this interaction are defined. Methods of using cloud services on the example of the distance course "Practical English Course Upper Intermediate" describing the methods of doing the tasks in English communicative competence forming process of language faculties were considered. As a result of experiment it was found there is a low level of use of cloud technologies in distance learning is a consequence of a lack of awareness of teachers in the possibilities of its use.

Keywords. Cloud technology, cloud computing, distance learning system, SaaS, PaaS, electronic educational resources, LMS «Kherson virtual university»

Key Terms. Quality Assurance Methodology, Standardization Process, Knowledge Management Methodology, Knowledge Management Process, Teaching Methodology, Teaching Process

1 Introduction

Due to the rapid development and implementation of the Internet in everyday life and the simultaneous designing of distance learning systems (DLS), there is the opportunity to approximate the traditional learning to distance one. Today, through the designing of Internet-services, which were developed and used a few years ago, there are new capabilities for teachers to organize the distance and mixed (blended) learning to improve the quality of the educational process. Problems of design, implementation and use of cloud technologies in higher education belong to perspective areas of informatization of the education system as a whole [1].

The rapid development of technologies, the necessity to develop models, methods of most appropriate use and implementation of cloud tools and services in educational process are the factors that encourage the further development of this problem [2].

The term «Cloud computing» was first used in 1993 by Eric Schmidt to describe services that support distantly different data and applications hosted on distant servers.

«Cloud computing» is the technology of data processing, in which computer resources and capacities are available to the user as an Internet service. According to the official NIST (National Institute of Standards and Technology) definition, "cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction"[3].

Let's consider the basic directions of development of cloud technologies which are used in education. According to the technology and functionality the services are classified in [4]:

1. SaaS – Software as a Service. This type of cloud services allows using the Internet software the in education. In learning English, it provides access to email, various training sites, blogs, video and audio materials in the Internet. For example, the service Google Apps for Education provides Internet technologies and tools to create information-educational environment of the institution, Microsoft Live@edu is used to equip students with the necessary tools of online interaction without additional costs for software, hardware and support [5].

2. PaaS – Platform as a Service. As the services it is provided by a set of software services and libraries that you can use to develop the own electronic educational resources (EER). It is provided an integrated platform for development, testing and support of web applications that are based on cloud computing. In education, the service is used to develop integrated applications that are used "in the cloud" for managing of educational projects, implementation of joint researches, such as designing the virtual laboratories of share access. Learning Management System (LMS) can be represented in the "cloud" through this type of service.

3. HaaS (Hardware as a Service) – it is provided by the hardware capabilities, such as a certain amount of memory, processor time, bandwidth and so on.

4. IaaS (Infrastructure as a Service) it may be seen as HaaS technology development that provides the certain systems underlying in the construction of other systems, such as virtualization tools, load distribution etc. IaaS may include hardware (servers, storage, client systems and equipment); operating systems and system soft-

ware (virtualization, resource management); communication software between systems (e.g. integration into network, management equipment). The use of the technology in education provides an opportunity to get rid of the need to support complex data infrastructures, client and network infrastructures.

5. CaaS (Communication as a Service) – new service, which is the development of SaaS technology. As a service it is provided by communication services, for example, IP-telephony, mail and chat. For example, as a service to the institution it is used e-mail.

6. DaaS (Desktop as a Service) – users receive the virtualized workplace completely ready for work. This service is the development of SaaS technology, which is widespread in recent years.

Cloud technologies are used in education because its enable significantly reduce the costs, solve the problem of equal access to the information technologies because the powerful resources can be obtained via the Internet. In addition, it is provided the opportunity to meet individual needs of the student and academic trajectory by modeling his activity and selection of appropriate resources based on processing of large amounts of data and mobility training [6].

There are some disadvantages of cloud technologies, which are mainly technical and technological nature. These disadvantages are limiting of functional properties of software compared with local analogues, the lack of native providers of cloud services (Amazon, Google, Salesforce etc. are concentrated in the USA), the lack of national and international standards, and the lack of legal framework applying cloud technologies in education.

Currently there are four models of cloud deployments:

- Private cloud – it is used to provide services within the organization that is both customer and service provider. It is the variant of cloud concept embodiment when the organization creates it's for itself in limited use;
- Public cloud –it is used to deploy of the infrastructure and software to provide access outside of boundaries of institution's infrastructure;
- Hybrid cloud –it is two or more different types of clouds;
- Community cloud –type of infrastructure, designed for use by the specific consumer society of organizations with common objectives. Examples, platform Windows Azure, Web services Amazon, Google App Engine and Force.com.

We study public and community cloud systems.

The main aim in foreign language learning in high school is to develop communicative competence, i.e. the ability to get enough complete information at reading foreign texts, the ability to understand speaker and express own opinion, point of view orally and in writing.

Thus, the feature of the subject "Foreign Language" is not knowledge about the subject, that is knowledge about the language (linguistic competence), but the development of certain skills and abilities of different types of language activities based on knowledge of the way of activity (communicative competence) [7].

Cloud technologies make possible to integrate various active learning methods in the information environment. Let's consider the most popular and affordable for users Internet services:

- Google Docs is an online office to create various documents for share access;
- OneDrive - service of Microsoft, it is similar service of Google Drive;
- Scribd - Internet service of cloud document storage, which allows publishing documents prepared in the most popular formats: Microsoft Office, Open Office, Adobe Acrobat, etc.;
- Slideshare - online storage of presentations;
- Google Scholar - search engine of educational and scientific publications;
- YouTube - a service that lets download and watch videos in the browser;
- Skype - a service that provides audio and video communication;
- Wikipedia - online encyclopedia, based on wiki technology;
- Blogger - Internet-service in the form of online journal (or blog);
- Facebook, VKontakte - a social network that provides the ability to create study groups, communities, etc. [8].

2 Distance Learning System in Language Learning

Information and communication technologies, including distance learning technologies rapidly integrated into the traditional educational process of higher education. The actual problem is the use of distance learning in foreign language training, including the formation of foreign language communicative competence.

The problem of distance learning (DL) in learning languages is studied by number of scientists: V.Yu. Bykov, V.M. Kukhareenko [9], A.A. Andreev, V.I. Gritsenko, K.Yu. Kozhukhov [10], S.P. Kudryavtseva, N.V. Mayer, M.G. Moor, E.S. Polat [11] A.V. Hutorskoy and others. According to K.Yu Kozhukhov, distance learning technologies can realize methodically organized and purposeful leadership of training and cognitive activity of students and it is based on using a wide range of information and communication technologies [10, p.11].

Let's consider the definition «distance learning». There are a set of definitions of «distance learning» We consider only the most accurate definition in our opinion. Distance Learning (DL)– the interaction between teachers and students at a distance that covers all components inherent of the educational process (purpose, content, methods, organizational forms, and teaching aids) specific means of Internet technologies [11, p.15].

DL is complex educational technology, combining the achievements of pedagogy and psychology of didactic opportunities of information and telecommunication technologies that use the computer as a carrier of information and means of communication. DL meets all modern features of society and its purpose is forming the creative person [12, p.20].

However, distance education, e-learning are generally used in combination with other types of teaching, such as face-to-face teaching in a class or mail correspondence. In 2014, more than 6 million students took at least one online course in the world. 60% of four-year private colleges and universities offer online classes as well [13].

There are a set of the advantages and disadvantages of e-learning, especially in language learning. Let's consider them.

Actually, there are a lot of advantages of distance learning. First of all, students can study at their individual pace. They can spend time over something that puzzles or intrigues them before they proceed. That's something they cannot do if the teacher controls the pace of students' learning or the pace of a class group.

Students can arrange their week to suit their work and family commitments. That means they can schedule the studies at times that would be awkward or unsuitable for teachers or their fellows students. Contrariwise they can always delay their studies by a week, month or even more.

It should be mentioned that the positive sides of the commercialized popularity of foreign language distance learning in Japan are, as Toshiya Kawame suggests, that the process of producing teaching materials has been shared not only by teachers concerned, but also openly by non-specialists, and that students are expected to develop their own individual styles of learning, since distance learning expects users to activate and make best use of it [14].

Also students explore the freedom that distance makes possible, they begin to realize that they control the content and emphasis of what they study, to a lesser if not a greater extent. This one is probably the most important advantage because only students decide what they would or would not do.

Let's consider disadvantages of distance learning. In fact, learning at a distance can be very isolated experience. One consequence of that isolation is the absence of social links, whose importance in language learning is surely under-estimated in communicational situations.

When learners are at a distance, teachers cannot do things that they took for granted in the face-to-face situation. For example, it is impossible at a distance for the teacher to keep a casual eye on the learners' performance and progress. Distance is furthermore a severe constraint on the development of a relationship, perceived or otherwise, between the teacher and the learner. And a distance is a disincentive to corporate activity, and the learning together which often happens naturally and informally in a class.

Because learning happens at a distance, it follows that sharing, borrowing, returning, issuing and other such functions involving books, paperwork and assignments, take time. The interruptions to learning which that delay introduces can be frustrating and de-motivating [15].

So, the successful language-teaching is built on a strategy that aims at creating an interest in a language and culture and then uses technology as a means of allowing pupils to communicate with others, use interactive resources and exercises and be creative in the target language. Just like anywhere else, distance learning has created new methods of language teaching; it has promoted new attitudes to the acquisition of foreign languages; and it also has opened up new possibilities of language education.

Thus, distance learning for language is not quite the same as learning of other subjects. Learning a foreign language includes not only the mastering of grammatical knowledge and comprehension skills, but the development of interactive communication ability (competence). With the help of distance learning systems, which has all communicational and virtual tools teachers can possibly create a virtual reality in which students can develop communication skills.

3 The Use of Cloud Services in Learning a Language

We consider the main cloud services that we use to create English communicative competence. So we chose Blogger, Wiki, Google Drive and social network VKontakte.

In our research we are developing the distance course (DC) «Practical English Course Upper Intermediate» based on the textbook with the same title and authors are L.Chernovaty, V. Karaban and others. The course is for third year students of philological faculties and specialty «Translation» of higher educational institutions. This course is created in the distance learning system «Kherson Virtual University» (DLS KVU) [16].

In this DC one of the tasks of forming communicative competence is to create your blog by students and the organization of discussions in blog on the topic «Holidays in Ukraine», where students place the essay, link to video or audio materials in the subject, leave comments in blogs of classmates. Control and evaluation conducts with the use of the rating system of DLSKVU. Thus, there is the interaction between DLS and cloud service «blog».

Consider the use of Wiki-technology on formation of communicative competence in this course. For example, the traditional teaching of the topic entitled «Regional Variations in the USA» are reading the text, consisting of some fragments with titles and students should match titles of fragments with of parts of the sentences of text. In the distance course this exercise is implemented as follows: the text for reading is developed as course's document and all geographical names in it have link to the appropriate page on Wikipedia for a detailed review, study and use in the future. On the next stage, students complete the training task of "Comparison", for strengthening the acquiring knowledge. Thus, there is the interaction EER (document) in distance learning system and cloud service Wikipedia.

It should be noted the growing popularity of using Google Drive service in the learning process. Today Google Drive services are actively used by many teachers and trainers in different disciplines in their professional activities. These services are used as auxiliary tools for forming communicative competence in English. First of all, Google Drive – a free online office, which includes text, spreadsheet, service to create presentations and Internet service cloud storage files. This web-oriented software is a program that works by using a web browser without installation on the user's computer. Docs created by the user are stored on a special server Google, or can be exported to a file. This is one of the key benefits of cloud services, as access to the data entered can be done from any computer connected to the Internet. Let's consider the interaction of distance learning system and SaaS services on the example of Google Docs. Thus, in the process of learning in distance group in DLS tutor has the ability to open access to files for user groups and assign it the appropriate permissions. Users have the ability in Google Docs, depending on their rights, create, edit, share resources, using calendar built into the system to plan tasks etc. For example, in the DC for forming communicative competence of students we use a variety of exercises using text documents (co-create and edit documents on specific topics), spreadsheets (co-creation and editing crossword puzzles, tables, etc.), presentation (co-creating presentations in a particular topic and its further use) service Google Drive. For example, the traditional teaching practice for the topic entitled «Grand Canyon» is lis-

tening to the audio recording (2 times) and filling the blanks in sentences. This exercise aims to develop the ability of listening and understanding foreign speech. In the distance course this exercise is implemented as follows: audio recording is developed as course's document and there is the link on the document in Google Docs made by the teacher (tutor). After listening to the audio recording, the students begin to work with the Google document, fill gaps in sentences, edit and analyze the mistakes. The teacher observes the work of students and makes the comments. This document is saved on Google Drive. Control and evaluation conducts with the use of the rating system of DLS KVU. Thus, there is the interaction between DLS and cloud service Google Docs.

The use of social networks (VKontakte, Facebook) in formation of communicative competence in English has important role. During the learning process in the DC it has been created the group in the social network VKontakte. Organizational information and a large number of links to online resources for information and study are placed in it. Control and evaluation conducts with the use of the rating system of DLS KVU. For example, there is the task for is watching a video on the topic «Holidays in Ukraine» in the group VKontakte, and there is discussion in "Forum" in DLS. Thus, there is the interaction between DLS and social network VKontakte.

4 Model of Interaction of Cloud Services in Education

LMS as part PaaS meets the basic requirements of providing distance learning. At creating and using the distance course of LMS there is provided an opportunity to use internal services such as Whiteboard, Forum, Chat, Virtual Lab, E-mail and more. But there are methodological problems which cannot be solved by standard LMS means. Then LMS as an element of PaaS may use other cloud services to expand the technological means of solving the methodological problems. First of all, it concerns SaaS means, such as file storage Google Docs, OneDrive, media resources YouTube, communication Skype, publications Blogger, social networks Facebook, VKontakte and others.

The model of interaction of distance learning with SaaS cloud services in the English language learning is presented in Fig. 1.

On the basis of the services we specify educational opportunities of cloud technologies confirming the feasibility of their using in learning English:

- simplicity and convenience of teamwork of teacher and students;
- rapid inclusion of developed products in the educational process because of the lack of territorial binding service user to the place;
- organizing interactive sessions and collective training;
- creating of web-oriented laboratories in specific subject areas (mechanisms to add new resources, interactive access to simulation tools, information resources, support for users, etc.);
- access to documents anywhere and anytime;
- organization of various forms of control;
- moving into the cloud learning management systems (e.g. LMS Moodle and others.);

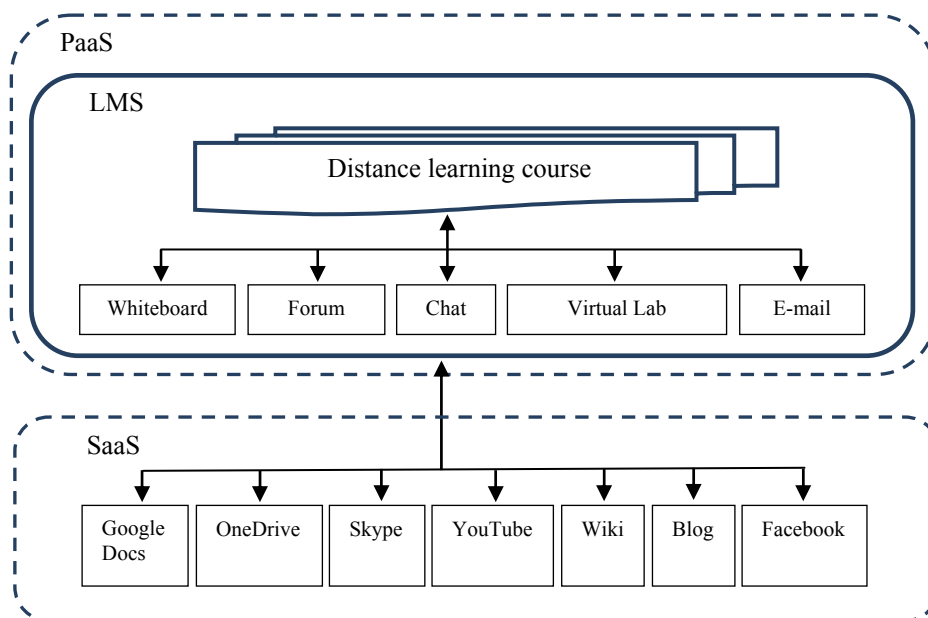


Fig. 1. The model of interaction between LMS and SaaS cloud services

- opportunities for researchers to access, development and dissemination of applied models.

5 Experimental Study

The main objectives of the experimental study are determining the state of use of SaaS cloud services by the teachers and students of Kherson State University (KSU). The method of questioning was used to exposure the attitude of tutors and students to use SaaS cloud services in a blended learning in LMS «Kherson State University». Among the tutors of DLS KVU the questioning was conducted on the study of the level of use of cloud services in the learning process. The purpose of the questioning is to make the analysis of level and quality use of cloud services in education using distance learning technologies.

Questionnaire consists of questions that reflect:

- degree of use of cloud services in a professional activity of the teacher;
- willingness of the teacher to learn new software;
- awareness of teachers about cloud technologies in educational process.

In accordance with the objectives of questioning the poll is designed for teachers of the university, which is located in the Internet using Google Form. Particular attention is paid to the use of cloud technologies in professional work of the teacher.

The list of questions of the poll, answers and results are presented in Table 1 (for

teachers - "T"), and (for students - "S").

Table 1. Questions of polls of teachers and students on the use of SaaS

#	Question	Answers	Quantity	
			T	S
1	Are the cloud technologies in your educational institution?	* Yes * No * Don't know	46 19 15	204 10 6
2	Do you use cloud services in a professional / educational activity?	* use in the classroom * use in distance learning * use in outside classroom (self-study) * use for storing, sharing information and in personal goals * don't use	15 4 10 27 34	45 22 47 68 18
3	What type of services you mostly use in own professional / educational activity?	* SaaS * PaaS * IaaS * Desktop as a Service	7 38 1 0	183 214 17 9
4	Which cloud services you use in the professional / educational activity?	* Internet services to create presentations (Prezi.com, SlideShare.net etc.) * Internet services to create interactive educational applications (LearningApps.org etc.) * Online encyclopedia (Wikipedia) * Common storage system (Google Drive, etc.) * Google services to work together * Internet services for creating flowcharts and diagrams (draw.io, gliffy.com etc.) * Other	25 21 58 9 5 2 1	154 141 220 178 131 115 20
5	Specify Google services that you use in the professional / educational activity?	* Google Disk * Google Docs * Google Calendar * Google Translator * Google Maps * Google Sites * Blogger * Other	12 14 5 52 45 2 19 0	71 60 15 175 205 6 42 33
6	Rate a five-point scale the level of use of cloud services in a professional / educational activity	1 2 3 4 5	7 43 15 12 3	8 30 102 63 17
7	Indicate your willingness of use of new cloud services in a professional activity	* Ready, regularly master new services * Ready, periodically have to master new programs and resources * Ready, but not enough time to master the new cloud services * Not ready	6 29 25 17 3	78 82 41 8 11
8	What is your attitude to use of cloud services in a professional / educational activity?	* Positive * Negative * Neutral	59 6 15	205 4 11

80 from 100 teachers and 220 from 500 students were questioned of the Faculty of Preschool and Primary education and Faculty of Physics, Mathematics and Computer Science. Volume samples provide the margin of error of 5% and 95% level of confidence.

Analysis of the poll shows that cloud services are gradually being introduced in the educational process of KSU, but its use is at a low level. In professional activity the teachers often use cloud services Google, and they also actively use services to create online presentations and online training applications (Fig. 2). At the same time, the willingness of the teacher to learn new cloud services is high, which suggests the feasibility of training teachers in the use of cloud services.

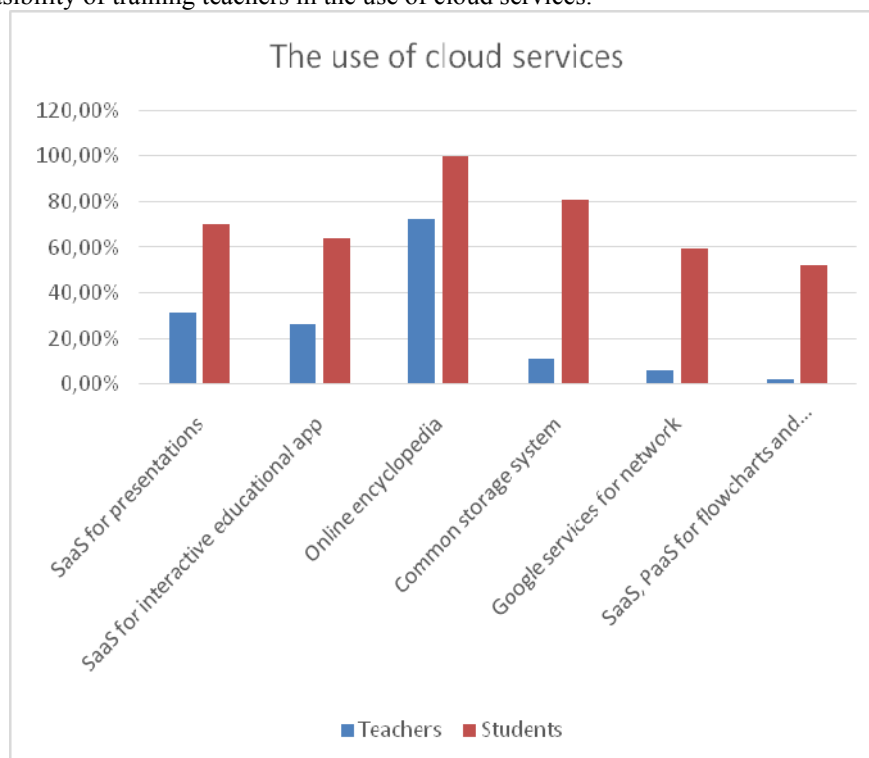


Fig. 2. The use of cloud services in the professional / educational activity

It should be noted that more than half of the students widely used cloud services in independent work in blended learning in LMS, if the teachers use them. The attitude of students to the use of cloud services is positive and the readiness of use them is high enough. Therefore, there is a need of students in increasing the use of cloud services in distance learning.

Thus, the results of experimental study have shown the low levels of use of cloud technologies in distance learning is a consequence of a lack of awareness of teachers in the possibilities of their use. Therefore, the actual work is aimed at popularizing the use of cloud technologies in distance education. It is reasonable to conduct training of university teachers on the use of cloud services, which is part of the quality manage-

ment system of EER [17].

6 Conclusions and Outlook

The peculiarities of cloud services in the process of English communicative competence are identified and examined. The model of the interaction of distance learning system with cloud services, the use of which will improve the quality of the learning process is designed.

1. According to the designed model of interaction LMS with the SaaS resources the technologies of cloud services in distance learning are studied.

2. Methods of using cloud services are considered on the example of DC "Practical English Course Upper Intermediate". They describe the methods of doing tasks aimed on forming communicative English competence of students of language faculties.

3. The results of experimental study have shown the low level of use of cloud technologies in distance learning is a consequence of a lack of awareness of teachers in the possibilities of their use. Therefore, the actual work is aimed at popularizing the use of cloud technologies in distance education.

References

1. Bykov, V.: Models of Organizational Systems of Open Education Atika, Kyiv, Ukraine (2009) (in Ukrainian)
2. Shishkina, M.P., Nosenko, J.G.: Recent Areas of Cloud-based Educational and Scientific Environment Education System. Scientific Journal of NPU of M.P. Dragomanov. Series №2. Computer-oriented education system: Coll. of scientific articles, 16(23), 153--158 (2015)
3. Mell, P.: The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology. Online: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf> (2011)
4. Shishkina, M.P.: Cloud Based Learning Environment of Educational Institutions: Current State and Research Prospects. Information Technologies and Learning Tools, 5(37) online: <http://journal.iitta.gov.ua/index.php/itlt/article/view/903/676> (2013)
5. Kendall, M.: Rank Correlation Methods. Charles Griffen & Company, London, UK (1948)
6. Google Apps for Educational Institutions. online: www.google.com/enterprise/apps/education/products.html
7. Kravtsov, H.M., Chemisova, I.O.: Microsoft Cloud Services in Distance Learning System "Kherson Virtual University". Information Technology in Education, 20, 53--65 (2014)
8. Nikolaeva, S. (ed.): Methods of Foreign Language Learning in Secondary Education. "Lenvit", Kyiv (1999)
9. Gnedkova, O.: Guidelines on the Use of Internet resources in the Distance Learning System «Kherson Virtual University". Information Technologies in Education, 10, 183--187 (2011)
10. Bykov, V.Yu., Kukharenko, V.M., Syrotenko, N.G., Rybalko, A.V., Bohachkov, Y.M.: Technology of Distance Course Development: Navch. Posibnyk. Millenium, Kyiv, Ukraine (2008) (in Ukrainian)
11. Kozuhov, K.Y.: Pedagogical model of application of distance technologies in the process of formation of methodical competence of the future teacher (based on the discipline

- "Theory and methods of teaching foreign languages"): Abstract. Dissertation on completion of PhD in Ped. Sci., Kursk, Russian Federation (2008) (in Russian).
12. Polat, E.S., Buharkina, M.U., et al.: Theory and practice of distance education: Educational material for the students of higher pedagogical educational institutions. Publishing centre "Academia", Moscow (2004) (in Russian).
 13. Bykov, V., Kukhareno, V. (eds.): Distance Learning Process: Textbook. Millennium, Kyiv, Ukraine (2005)
 14. Boyle, R.: Language teaching at a distance: from the first generation model to the third. System 23(3) 283--294.
 15. Breen, M.P. (ed.): Learner Contributions to Language Learning: New directions in research. Longman, Harlow, England (2001)
 16. Carnwell, R.: Distance education and the need for dialogue. Open Learning 14(I), 50--55 (1999)
 17. LMS "Kherson State University". online: <http://dls.ksu.kherson.ua/>
 18. Kravtsov, H.M.: Methods and Technologies for the Quality Monitoring of Electronic Educational Resources. In: Ermolayev, V. et al. (eds.) Proc. 11-th Int. Conf. ICTERI 2015, Lviv, Ukraine, May 14-16, 2015, CEUR-WS.org/Vol-1356, ISSN 1613-0073, p. 311--325. online: http://ceur-ws.org/Vol-1356/paper_109.pdf (2015)

Models and Tools for Information Support of Test Development Process in Learning Management Systems

Olena Kuzminska¹, Mariia Mazorchuk²

¹National University of Life and Environmental Sciences of Ukraine,
16a Heroev Oborony St., Kyiv, Ukraine

`o.kuzminska@nubip.edu.ua`

²National Aerospace University "Khai",
Chkalova str. 17, Kharkiv, Ukraine

`mazorchuk.mary@gmail.com`

Abstract. Many current educational trends are based on the digital technologies support for content modeling and delivering as well as for control services in the teaching and learning processes. E-learning methods allow us significantly to enhance the teaching and learning effectiveness. These methods also show the rapid development of software of a new generation intended for the teaching and studying process support. Nevertheless there still exist problems to deal with in order to make further progress. One of these problems lies in the creating effective systems for student educational progress assessment. This article proposes a new original item bank information system creating method based on the Open Source tools: LMS Moodle and package R. This method was tested in the National University of Life and Environmental Sciences of Ukraine and the National Aerospace University "Khai". Obtained results are presented and practically illustrated.

Keywords. Educational Technologies, Test Items, Testing, Information and Communication Technologies, Experience

Key Terms. Academia, Teaching Process, Development, Information Communication Technologies, Model

1 Introduction

The modern stage of educational system development is characterized by appearance new educational technologies, which are come together with high rate of informatization. Influence analysis of macro, meso and micro trends; designing of educational spaces and models are the subjects of the research [1]. In its annual reports New Media Consortium (NMC) are describing technologies, which will be having significant impact on the educational processes, including higher educational establishments, as in: flipped classroom, learning analytics, blending learning (b-learning), personalizing

learning, bring your own device (BYOD), maker spaces, the Internet of things, adaptive learning technologies, open educational resources and massive open online courses (MOOC) [2, 3]. Concerning e-learning implementation [4] the publications widely represent the experience b-learning application, as well as learning management systems (LMS) for e-learning organization [5, 6]. Simultaneously stay relevant the issues of supporting the objective testing of knowledge quality of students using e-learning systems, which are applied both in MOOC, and in knowledge control systems at universities. Especially burning is the problem of using the qualitative tests in the systems of e-learning, because the users of on-line courses may have different levels of preparation and need individual approach for education, which cannot be ensured by the current e-learning systems. For example, in LMS Moodle, which is used in most Ukrainian universities, provides the functional part of test quality rating [4]. However, the data rates not always are reliable, as they are often determined on the basis of small population of participants and primary test results. The existing methods and models of evaluation of the test quality rating have its own application field; nevertheless they still don't allow solving the tasks connected with performance of effective on-line rating in complex. Virtually, there are no tools that can provide complex support for test designers in the process of testing.

The objective of this research is the development of models and tools of information support of tests forming in systems of distance learning, providing the adequate level of quality depending on the ability level of students.

2 Methods and Quality Analysis Models of Education Tests

For tests quality analysis mainly are used methods of classic theory, based on the calculation of such main parameters as, complexity, ability to differentiate, correlation of tests with total test grade and so[7, 8]. For more detailed analysis the method of threshold group is used, which allows to build the curve and frequency tables of distractor choice for threshold group of test takers, that sufficiently represents the information about the quality of developed test items [9]. In LMS Moodle system, which possesses wide functional, the block of quality tests results interpretation is not established. In case of small populations, the evaluation parameters methods can have serious errors; there is no possibility to compare the results of learners groups.

Today there are a lot of supporters of Item Response Theory, which allows to get tests results in metric scale. The literature [8], [10, 11] gives exhaustive information about methods and models of IRT. The important feature of models of modern tests theory is the limited conditions of its use, which are as follows:

- tests modeling by Rush function (not always possible);
- compatibility of participants response with Gutmann condition (unipath continuum);
- unidimensionality of test (the test must be measured by only one construct);
- items test are independent.

The main disadvantages of this theory are the calculation and results interpretation complexity, and also high demand to study population volume (at least 500 individuals), which is difficult to assure in conditions of functioning of modern educational institutions (tests participants groups may not exceed 15 persons).

User experience of LMS Moodle [12] has shown, that it is possible to guarantee the accumulation of statistical data by tests results in small groups (10-15 persons) during long time period, so it is effectively to use this platform for experiments.

Using LMS Moodle it is possible to save various tests results of students: test parameters (duration of the test, quantity of the used attempts to answer the questions); total points; parameters of test questions (correct and incorrect answers). All results can be presented in convenient format for following analysis.

To analyze the quality of tests it is necessary to evaluate tests parameters, which allow estimating their reliability and validity. Input data for analysis are the matrix of tests results. Matrix of tests results is matrix of $N \times M$ dimension, where N – quantity of tests participants, M – numbers of test items:

$$A = (a_{ij})_{i,j=1}^{N,M}. \quad (1)$$

Evaluation schemes for the majority of tests can be classified into dichotomous and polytomous data [8]. Dichotomous scale is used in this paper.

To evaluate psychometric tests characteristics it has been decided to use Item Response Theory, since it allows to get results in metric scale and to compare groups of participants. For research of psychometric characteristics of tests various IRT models are applied: classical model Rasch, 1PL, 2PL and 3PL. More detailed information about these models may be found in [10, 11].

Let us consider the three-parameter models 3PL, which allows to receive the most fitting results (in fact all other models are the particular case of 3PL models). In this model the conditional probability of correct performing of j -test for examinee with ability level Θ depends on three parameters: difficulty parameter δ_j , discrimination parameter d_j and guessing parameter c_j :

$$P_j(\Theta) = P\{x_{ij} = 1 / \delta_j, d_j, c_j\} = c_j + (1 - c_j) \frac{e^{Dd_j(\Theta - \delta_j)}}{1 + e^{Dd_j(\Theta - \delta_j)}}. \quad (2)$$

Here the constant multiplier $D=1.7$ for better model fitting with the model of normal ogive [10, 11]. On the ground of this dependence are formed characteristic curves for every j task, the position of which in Cartesian plane is determined by the quality of tests. More detailed interpretation of received values and analysis of characteristic curves are also presented in [10, 11].

This way it is possible with the use of 3PL model to calculate probabilistic characteristics of test items, on the ground of which the items, which meet the demands of reliability, may be chosen.

3 Tests Development Model in Learning Management Systems

For supporting the process of decision making for tests formation in the process of distance learning the generalized structural model is suggested, which can be realized

in LMS Moodle (Fig. 1). The main test development steps with account of ability level students are the following:

- qualitative learners ability level analysis by primary testing (initial check);
- selection of relevant category training courses; material learning analysis by intermediate tests (tests 1, 2, ... i, ... N-1);
- qualitative analysis of intermediate tests (using IRT models);
- accumulation of tests base (data of tests are corrected after each entering the test);
- forming the final tests (test N) with appropriate psychometric characteristics by all courses categories;
- the final control and estimation of ability level learners.

The courses may be presented by the separate modules of one discipline or a number of courses, which need to be learnt to acquire specific knowledge.

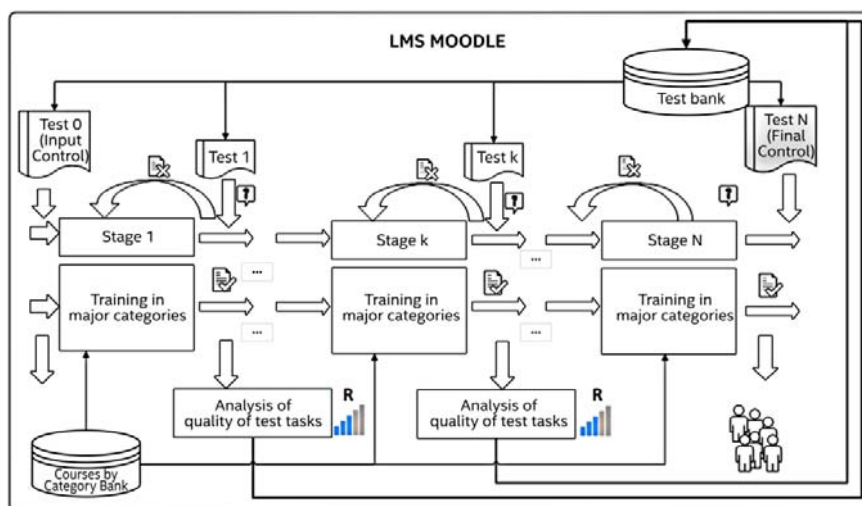


Fig. 1. Structural model of test development in the process of e-learning using LMS Moodle
(Source: Own work)

The collection of empirical results of testing is the preparation step. In fact, the receiving of solid results, reflecting the reliable estimates regarding the preparation level of learners, is possible only after some iteration of system operation with current bank of tests. This is necessary while the information accumulation will be carried out during certain time and psychometric indexes of tests will be recalculated.

Total points, received in login scale (according to the IRT models), allows to estimate the tested by the ability level. The initial difficulty level of test (for the initial check test) is determined by the teacher.

The selection of tests is carried out in random manner, which provides proper evaluation in the process of testing (the students cannot know the answers, cannot pass the information to each other about right answers, all are taking the test in equal conditions). In this research have not yet been considered the issues of automatic selection of

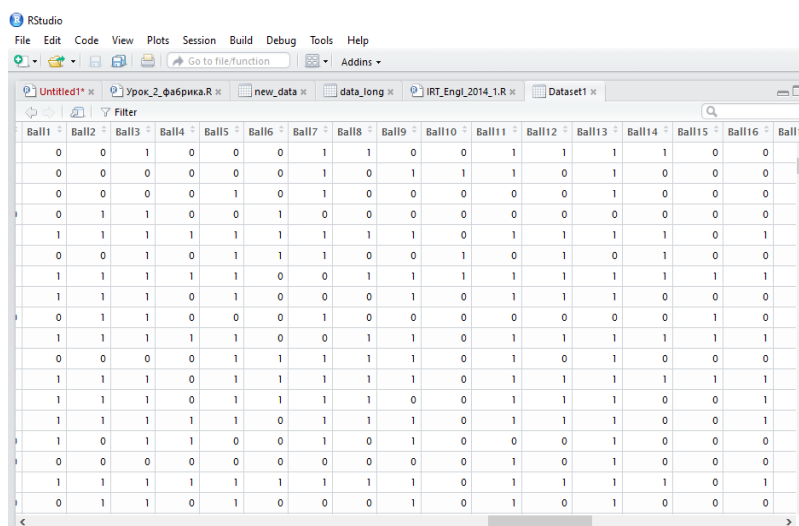
courses in accordance with primary training of learners and the time test characteristics have not been monitored.

It is suggested to carry out all calculations of psychometric test characteristics using software functions with open source code R-Studio. Special package *ltm* exists for the determining of the main characteristics and evaluation of received models quality. Description of the package functions can be found both in R-Studio reference material, and in scientific journal Journal of Statistical Software [13].

4 Quantitative Experiment

Let us consider some calculation results of tests parameters, which were received based on the testing results of intermediate stage of discipline “Information technologies” study cycle in National University of Life and Environmental Science (NULES) of Ukraine (<http://elearn.nubip.edu.ua/enrol/index.php?id=230>) and in National Aerospace University “KhAI” (<http://stm.khai.edu/course/index.php?categoryid=4>). The experimental research was carried out during 2013-2015. The volume of sampling population amounted to 520 students – masters of 1 year of study from the two universities. Students were offered the same sets of testing for the input testing of ICT essentials. The level of initial preparation of students can be considered the same, because the level of IC-competence doesn’t influence entering the masters course.

Testing results are represented in form of the rectangular matrix with zeros and ones, since there were only dichotomous tasks in the tests. The results were received in LMS Moodle and exported to R-Studio (Fig. 2).



	Ball1	Ball2	Ball3	Ball4	Ball5	Ball6	Ball7	Ball8	Ball9	Ball10	Ball11	Ball12	Ball13	Ball14	Ball15	Ball16	Ball17
1	0	0	1	0	0	0	1	1	0	0	1	1	1	1	0	0	
2	0	0	0	0	0	0	1	0	1	1	1	0	1	0	0	0	
3	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	
4	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	
5	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	
6	0	0	1	0	1	1	1	0	0	1	0	1	0	1	0	0	
7	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	
8	1	1	1	0	1	0	0	0	1	0	1	1	1	1	0	0	
9	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	1	
10	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	1	
11	0	0	0	0	1	1	1	1	1	1	0	1	0	1	0	0	
12	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	
13	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	
14	1	1	1	1	1	0	1	1	1	0	1	1	1	1	0	0	
15	1	1	1	1	1	0	1	1	1	0	1	0	0	1	0	0	
16	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	
17	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	
18	0	1	1	0	1	0	0	0	1	0	1	0	1	0	0	0	

Fig. 2. Matrix of testing results (Source: Own work)

There were analyzed 48 test items on the basis of different IRT models. In the process of research were evaluated the parameters of test items by classical theory and using 3PL model. For each tasks were calculated the parameters of difficulty, discrimination

and guessing. Table 1 shows some results of calculation for three-parameter model for the easiest and difficult items. In the last column the probability to give correct answer for each test items ($P(x=1|z=0)$) is given.

Table 1. Psychometric characteristic according to IRT three-parameter models

Test item (number)	Model 3PL			
	Guessing	Difficulty	Discrimination	$P(x=1 z=0)$
7	0,660	1,462	3,236	0,663
11	0,091	-0,921	2,351	0,906
3	0,084	-0,859	2,372	0,894
...
34	0,205	1,168	4,064	0,213
18	0,033	1,551	2,587	0,050
21	0,132	1,760	1,940	0,161
26	0,227	2,872	1,065	0,261
25	0,186	1,907	3,511	0,188

The analysis of the given results has shown that test items 18, 34, 21, 26, 25 and 26 do not satisfy the demands of coherence and reliability of tests by different parameters: they are difficult, have high level of guessing or low discrimination, which, in its turn, causes the low probability to receive the correct answer for given tasks. These tasks were removed from bank of test items and were not included into the final test.

According to the results of calculation also were built characteristic and information curves, which reflect the quality of test items. Fig. 3 and Fig. 4 shows the curves for 3PL model.

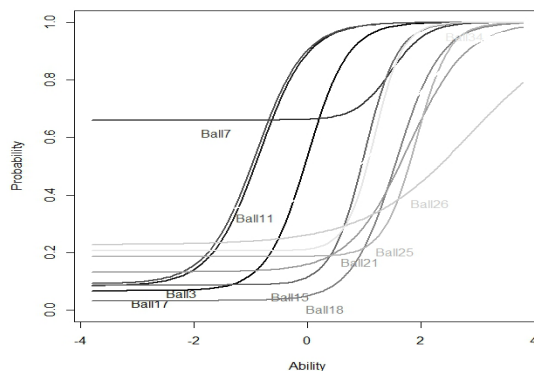


Fig. 3. Item Characteristic Curves for some test items data-set under model 3PL
(Source: Own work)

It can be seen on the graph, that the curves have accumulated to the left of zero, which confirms the received results – the most of test items are difficult. Also, it is apparent that, some items have low parameters of discriminatory power or high parameters of guessing (for example 7, 11).

It can also be seen on the figure, which tasks have the low probability of receiving the correct answer according to model 3PL: 15, 18, 25.

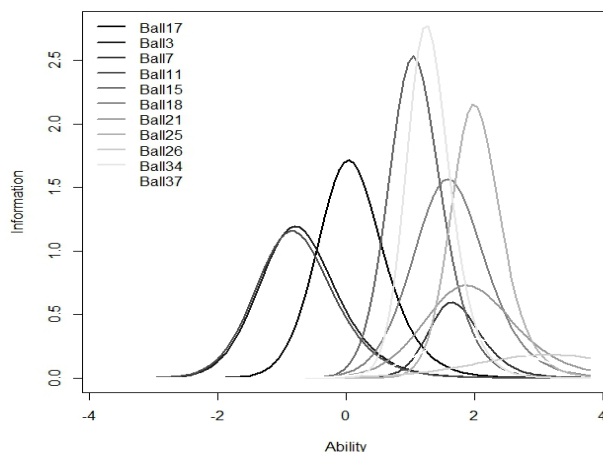


Fig. 4. Item Information Curves for some test items data-set under model 3PL
(Source: Own work)

Thus, the functional R allows us to fully estimate the quality of test tasks and the test in general. In the suggested system of e-learning (Fig.1) the analysis of psychometric characteristics is proposed to be made automatically, i.e. user is offered the list with tasks numbers, which do not meet the demands.

5 Conclusions

This paper describes a method as a technical tool for the quality test items bank forming. Classic test theory and dichotomous logistic item response models are used to estimate the psychometric test items characteristics.

The main assets of this approach use the Moodle system to accumulate and to keep the intermediate results of testing during the process of studying, which allows us to create the calibrated test items bank. That is the reason why using the R package allows us quickly and accurately to calculate the main characteristics of test items and tests overall. Due to the R package, various items analyzing methods become available.

Therefore, due to the logistic IRT model performed in the R package, this method is flexible and easy to administrate, as far as it requires no special knowledge in the field of statistical data processing.

The results of the test received during the experiment held in two Ukrainian universities show us the effectiveness of the method.

We believe this paper is only a small step towards this direction, both on the methodological and the practical aspects for creation quality tests for e-Learning. In perspective it is suggested to consider other working parameters with distance courses, which will allow to evaluate the quality of training in whole, on the basis of analyzing different activities of students, not only performance at tests.

References

1. Miller, R., Shapiro, H., Hilding-Hamann, K. E.: School's Over: Learning Spaces in Europe in 2020: An Imagining Exercise on the Future of Learning. Office for Official Publications of the European Communities, online: <http://ftp.jrc.es/EURdoc/JRC47412.pdf> (2008)
2. Johnson, L., Adams Becker, S., Estrada, V., Freeman, A.: NMC Horizon Report: 2014 Higher Education Edition, New Media Consortium, online: <http://cdn.nmc.org/media/2014-nmc-horizon-report-he-EN-SC.pdf> (2014)
3. Johnson, L., Adams Becker, S., Estrada, V., Freeman, A.: NMC Horizon Report: 2015 Higher Education Edition, New Media Consortium, online: <http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf> (2015)
4. Moore, J. L., Dickson-Deane, C., Galyen, K.: E-Learning, online learning, and distance learning environments: Are they the same? *Internet and Higher Education*, 14(2), 129—135 (2011)
5. Hongjiang, X., Mahenthiran, S., Smith, K.: Effective Use of a Learning Management System to Influence On-Line Learning. In: Proc. 11th Int. Conf. on Cognition and Exploratory Learning in Digital Age (CELDA), Porto, Portugal, Oct 25-27, 2014. online: <http://files.eric.ed.gov/fulltext/ED557395.pdf> (2014)
6. Dias, S. B., Diniz, J. A.: Towards an Enhanced Learning Management System for Blended Learning in Higher Education Incorporating Distinct Learners' Profiles. *Educational Technology & Society*, 17, 307--319 (2014)
7. Kim, V. S.: Testirovanie uchebnykh dostizheniy. Monografiya. Izd-vo UGPI, Ussuriysk, Russian Federation (2007). (in Russian)
8. Crocker, L., Algina, J.: Introduction to Classical and Modern Test Theory. Cengage Learning Pub., Ohio, USA: (2006)
9. Mazorchuk, M., Dobryak, S. S., Bondarenko, E. O.: Otsenka kachestva testov na osnove analiza distraktorov po metodu porogovykh grupp. *Radioelektronni i komp'yuterni sistemi*, 62 (3). 39--44 (2013) (in Russian)
10. Lisova, T.V.: Modeli ta metody suchasnoyi teoriiy testiv: Navchal'no-metodychnyy posibnyk Vydavets' PP Lysenko M.M., Nizhyn, Ukraine (2012) (in Ukrainian)
11. Baker, F.B.: The Basics of Item Response Theory. ERIC Clearing house on Assessment and Evaluation, USA (2001)
12. Anisimov, A. M.: Rabota v sisteme distantsionnogo obucheniya Moodle. HNAGH, Harkov, Ukraine (2008) (in Russian)
13. Rizopoulos, D.: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5). 1—25 (2006)

A Review of Teaching Methods for Life-Long Learning

Vira Liubchenko¹

¹ Odessa National Polytechnic University, 1 Shevchenko av.,
65044 Odessa, Ukraine

lvv@edu.opu.ua

Abstract. In the paper, there were analyzed six popular teaching methods – case method, enquiry-based learning, spiral learning, problem-based learning, project-organized learning, and b-learning – for the purpose of selecting and providing relevant teaching methods for LLL courses. There were evaluated the usefulness of each teaching method for LLL courses and identified their relevance for life-long learning. The paper provides some recommendation on teaching methods choosing.

Keywords. Teaching method, LLL course, case method, enquiry-based learning, spiral learning, problem-based learning, project-organized learning, b-learning

Key Terms. Academia, Didactics, TeachingMethodology

1 Introduction

When we mention the life-long learning (LLL), we say first the learning of adult students. To base on the experience of students there are recommended to provide an individual trajectory of learning, adaptive learning strategy and design, and so on. However, there are one more means for adaptation, which is important to enhance students' activity and to discover their abilities. They are teaching methods for the course.

The purpose of LLL teaching methods is to provide relevant teaching methods for LLL courses, which will enhance their usefulness to support the development of continuing and further education, meet the demands of the changing labor market and hence contribute to social and economic growth in the country.

The purpose of this paper is an exploration of existing teaching methods on requirements of LLL courses to formulate recommendations concerning teaching methods choice.

The rest of the paper is organized as follows. Section 2 briefly describes the most popular and known teaching methods. Section 3 presents some recommendation for choosing teaching methods and preparing the courses based on them.

2 Teaching Methods Relevant for LLL

In the life-long learning, there have been exploiting many different teaching methods fitted to the adult students. In this paper, we analyze the methods, which are the most popular now.

2.1 Case Method

A case method is a form of instructor-guided, discussion-based learning. It introduces complex and often ambiguous real-world scenarios (cases) into the classroom, typically through a case study with a protagonist facing an important decision. Since it presents authentic cases and real world problems, case method fully complies with LLL courses and supports the development of some essential skills for life, like communication, social, interpersonal, higher-level reasoning, problem-solving, and decision-making skills. The case method shifts instructor-centred model of education towards a participant-centred one in which students play a lead role in their own and each other's learning.

Teaching with cases is a very suitable method to use when lectures are the primary teaching form in a course. Making the students' work with cases in a structured way contributes to activating them. Thereby they are made to use their knowledge in the field of subject actively and work actively with the present course content. The method was developed at Harvard Business School, and it is very well-known [1].

In this method, the construction and formulation of the cases are crucial. They must cover the topics of the lecture, and they must provide a suitable challenge for the students. It may be a real life problem or a constructed task the students must work with and discuss.

Usually, in the courses practiced case method, the individual oral examination is used as an assessment at the end of the course.

A case method is practiced in onsite teaching as well as in online one. The resources needed to provide the course with encapsulated cases are listed in Table 1.

Table 1. Resources required for case method

Onsite Teaching & Learning	Online Teaching & Learning
Working spaces for group work of students	Online repository to make case descriptions available
Classroom for plenum sessions, preferably with student seats in half-circles	Online communication system to allow for Q&As to/from teacher
Equipment for video recordings of discussions	Virtual forum for students' case discussions Virtual forum for class discussions

2.2 Enquiry-Based Learning

The aim of this method is to give the students a life-long insight into research as a method to create new knowledge and learning. Enquiry-based learning is a student-

centred approach, which focuses on the development of students' higher-order thinking skills. This approach implies a complicated process where students formulate questions, investigate to find answers, build new understandings, meanings, and knowledge, and then communicate their learnings to others [2].

Enquiry-based learning can enhance the effectiveness of LLL courses since it emphasizes the importance of developing and fostering enquiring minds and attitudes in students; it enables them to continue the quest for knowledge throughout life.

In enquiry-based learning, the students are responsible for defining the central parts in the curricula the teacher has chosen as a subject for the course. They must find, evaluate, and use the sources of information available for the subject in question. The students start with one fundamental question, and while they are expanding their knowledge, the teacher inspires them to formulate and answer more detailed questions.

A strong emphasis is placed on reporting achieved results and documentation of the progress in the students' learning. The students themselves do this. The students conduct an evaluation of the results through self- and peer-assessment. Active involving students to knowledge mining process is a central part of the method.

To evaluate students work there are often used the learning portfolios, project reports with accurate descriptions of the problem, work process, results and solutions, and traditional examinations.

The resources needed to organize the onsite or online enquiry-based learning are listed in Table 2.

Table 2. Resources required for enquiry-based learning

Onsite Teaching & Learning	Online Teaching & Learning
Working spaces for group work of students	Online databases and search engines for finding and research information on the topic
Classroom for plenum sessions, preferably with student seats in circles or half-circles	Online communication system to allow Q&As to/from teacher
Equipment for presentations	Virtual forum for students' discussions
	Virtual forum for class discussions

2.3 Spiral Learning

The fundamental principle of the spiral learning is to minimize course risk by breaking a course into smaller topics. Students meet the same topics more than once during the course, with each encounter increasing in complexity and reinforcing previous learning. Spiral learning can be considered as one of the important and relevant method for LLL course.

The teaching sequence is divided into several (2-4) sections or "windings" in a spiral. In each winding the course subject is approached both from the top (overview) and bottom (details).

In the first winding, the students work with the course subject in a very basic way. Elements from all of the course topics are included, and a general overview of the course subject is given – i.e. the topics are not dealt with one at a time like the chapters of a typical textbook. The students are given group assignments to solve based on their present knowledge – maybe only in a qualitative way the first time depending on the subject.

In the next windings, the students work with the subject in increasingly advanced ways. It might be the same problem they are given in each winding, but the solution becomes more and more sophisticated and at a higher level of competency. In the last winding, the final level of competence (learning objectives) is reached.

Usually, the progress assessment is based on the assignments with teacher's feedback and a final evaluation.

The resources needed to provide a course on spiral learning method are listed in Table 3.

Table 3. Resources required for spiral learning

Onsite Teaching & Learning	Online Teaching & Learning
Working spaces for group or individual work of students	Online communication system to allow Q&As to/from teacher
Classroom for plenum sessions	Online for searching the relevant information
Libraries for getting the relevant information	Virtual forum for students discussions
Equipment for presentations	Virtual forum for class discussions

2.4 Problem-Based Learning

Problem-based learning is an active learning approach in which individuals gain knowledge and skills through problem-solving. Students learn the content as they try to address the problem. Students learn both thinking strategies and domain knowledge. The aim of problem-based learning is to develop the students' flexible knowledge, effective problem-solving skills, self-directed learning ability, effective collaboration skills and intrinsic motivation [4]. This teaching method fully complies with LLL demands since it provides a model for life-long learning, and supports building skills and abilities that are so valuable for today's world of constant change.

The leading principle of problem-based learning is that the students learn by relating their knowledge to a given problem or case which the teacher prepares. Through dialogue and discussions, the students in the groups try to solve the problem by using their previous knowledge and the new knowledge in the curricula of the course. A course typically starts with an overview of the course and the learning objectives. Then follows an introduction to the first problem the students are going to work with. The students meet in their groups and work with the problem based on a very structured method, repeated every time the teams meet.

Formulating the problems is crucial for course success. The problems should combine elements from all topics of the course, i.e. they do not deal with one topic at a time like the chapters of a typical textbook.

The progress assessment is based on the individual assessment of learning outcome typically using the triple jump evaluation methodology.

The resources needed to provide problem-based learning are listed in Table 4.

Table 4. Resources required for problem-based learning

Onsite Teaching & Learning	Online Teaching & Learning
Working spaces for group work of students	Online databases and search engines for finding and research information on the topic
Classroom for plenum sessions	Online communication system to allow for Q&As to/from teacher
Equipment for presenting the group work	Virtual forum for students to discuss the assigned case/problem, to provide review and feedback
Laboratories for group experiments	Virtual forum for class discussions
Library for researching relevant information to resolve the problem.	Online management system for group negotiation and planning

2.5 Project-Organised Learning

The idea behind project-based learning is that projects have a multidisciplinary character in which groups of students work while developing life-long learning skills. With a project organized curriculum, each course is structured around the big project. The course is planned around a project case covering all core elements in the course [5].

The use of project-organized learning has several benefits for LLL courses. It can be used with students of all ages. It can be flexible: does not necessarily mean working in a traditional classroom setting, and students can meet, interact and work wherever and whenever it is convenient for them. It provides some key (social, communication, interpersonal, decision-making, problem-solving, leadership, trust-building) skills for individuals to meet the demands of today's constantly changing world.

The objective is to solve the problem, i.e. the outcome includes a product. The students groups are working independently, doing their planning and taking charge of their process. Lectures can be given as a supplement to the group work and to support the course theme.

Outcome and process are assessed based on group project reports with oral presentations and some individual assessment for individual grading.

The resources needed to provide project-organizes learning are listed in Table 5.

Table 5. Resources required for project-organized learning

Onsite Teaching & Learning	Online Teaching & Learning
Working spaces for group work of students	Online repository for upload project descriptions, reports, and presentations
Classroom for plenum sessions, preferably with student seats in half-circles	Online communication system to allow for Q&As to/from teacher
Equipment for making PowerPoint presentation about the project	Virtual forum for students projects discussions
Labs for conducting experiments	Virtual forum for class discussions
Libraries for researching relevant information to fulfill the project work	Online management system for planning and negotiating project work

2.6 b- and e-Learning

Blended learning (b-learning) focuses on offering students choices about when, where and how they learn. It provides opportunities to improve the student knowledge and skills through flexibility in time, space, place (physical, virtual, on-campus, off-campus), mode of study (print-based, face-to-face, blended, online), teaching approach (collaborative, independent), forms of assessment and staffing [6]. This way of organizing university teaching goes far beyond the traditional way of thinking about teaching and learning. e-Learning is a tool often used in b-learning because of its independence of time and place.

Due to its flexible nature b-learning meets the demands of LLL courses and especially needs of adult students. Emphasis is on meeting students' needs, recognizing that all students have unique, complex needs, and this method gives students greater influence on what they need to learn, where, when and how. The method provides greater flexibility for students when it comes to where and when to study because of mix on-campus teaching and distant learning (at the workplace or home), and offer more flexible forms of access, entry, and exit.

In b-learning, very often on-line test are used as an assignment, but more traditional assessment methods can be utilized depending on the context in the course.

The resources needed to provide b-learning are listed in Table 6.

3 Recommendation on Teaching Methods Choice

Because of all teaching methods listed above are analyzed in the context of LLL, we should point their focus (except case method) on the learning to learn, and to take responsibility for own learning.

Case method could be recommended for courses which outcome is learning to deal with complex problems and problem-solving in general. The teacher should take the role of expert and facilitator of the class discussions. He/she organizes the work of the individuals or small group with cases – small, well-defined real-world problems.

Table 6. Resources required for b-learning

Onsite Teaching & Learning	Online Teaching & Learning
Working spaces for group work of students	Online repository to make instructional materials available
Classroom for plenum sessions	Online communication system to allow Q&As to/from teacher
Equipment for video conferencing; making presentations; using audio and video study materials	Virtual forum for students discussions
Libraries and Labs for conducting research and experiments	Virtual forum for class discussions
Alternative format learning resources and specialized equipment for students with disabilities (like, Braille, audio, and e-text for blind and visually impaired students)	Online system for the web-conferencing; web-casting
	Online systems to provide access to instructional materials for students with disabilities

If the outcome of the course is learning to learn, spiral learning and b-learning could be recommended. Spiral learning needs close involving the teacher as an expert and tutor. He/she monitors the work of medium students groups on authentic cases constructed with progression.

In b-learning teacher gets very special duties of student learning facilitator and consultant. We should point that b-learning in the context of LLL is the tool for supporting other teaching methods rather than the separate teaching method. Also, b-learning is the best means to support the individual learning trajectory.

Other three methods are oriented not only on learning to learn and learning to deal with problems but also on getting to know about information resources on the subject. Enquiry-based learning fits the last outcome the most. It requests teacher takes the role of expert, prepares real, ill-defined and open-ended, research-driven case, and supervises the work of small students groups.

Problem-based learning can be recommended as “easier” for teacher alternative of enquire-based learning. The teacher should take the role of facilitator of the group work and guide to the learning process. He/she provides sources of information and structure relevant problems and planning for progression towards the learning objectives. There is no need in one research-driven case for the course. The set of small cases covered the core elements of the course are also suitable. The cases should be realistic and interdisciplinary, ill-defined and open to brainstorm and elaboration with knowledge. Students work on them in large groups.

Project-organized learning can be recommended for courses built around one large, real-life, ill-defined and open case. Students are working on the case in permanent, medium-size groups. This method is the most demanding to the teacher, who works as a tutor (process and group facilitator) and as a consultant (expert).

4 Conclusion

The main point taken into account while developing the LLL course is the elaboration of new students' knowledge and skills based on their previous experience. It requires using of active teaching methods that enhance the students' to work actively on mining new knowledge and training new skills.

On this needs we explored the most known active teaching methods on requirements of LLL courses and formulated recommendations concerning teaching methods choice. Our analysis has a significant restriction: we analyzed the teaching methods separately. While developing the LLL course, each teacher should use different methods to make their course more fruitful.

Acknowledgment

This paper is based on the materials elaborated in the framework of Tempus IV project "A Network for Developing Life-Long Learning in Armenia, Georgia, and Ukraine" (543839-TEMPUS-1-2013-1-SE-TEMPUS-SMHES).

References

1. Case Method in Practice, <http://www.hbs.edu/teaching/case-method-in-practice/>
2. Hutchings, W.: Enquiry-Based Learning: Definitions and Rationale, http://www.ceebl.manchester.ac.uk/resources/papers/hutchings2007_definingeb.pdf (2007)
3. Wehlburg, C. M.: Closing the Feedback Loop Is Not Enough: The Assessment Spiral. Assessment Update, vol. 19(2), pp. 1--2 (2007)
4. Problem-Based Learning. Speaking of Teaching Newsletter, Vol. 11, No.1, http://web.stanford.edu/dept/CTL/cgi-bin/docs/newsletter/problem_based_learning.pdf (2001)
5. Jensen, L.P., Helbo, J., Knudsen, M., Rokkjær, O.: Project-Organized Problem-Based Learning in Distance Education. The International Journal of Engineering Education. Vol. 19, No. 5, pp. 696--700 (2003)
6. The Higher Education Academy. Inspiring Teaching, Transforming Learning: Flexible Learning, <https://www.heacademy.ac.uk/workstreams-research/themes/flexible-learning>

Professional Development of Teachers Using Cloud Services During Non-formal Education

Svitlana Lytvynova, Oksana Melnyk

Institute of Information Technologies and Learning Tools of NAES of Ukraine,
9 Maksyma Berlyns'koho St, 04060 Kyiv, Ukraine

s_litvinova@i.ua, ok_melnyk@ukr.net

Abstract. The rapid development of cloud services and their implementation in secondary education require an increase in the IC-competence of teachers during non-formal education. The implementation of cloud services will make it possible to create some conditions for learning mobility of all participants of teaching and learning activities. This article deals with the organization of teachers' training during non-formal learning. The paper analyzes the main forms of the organization of teachers' learning (workshops, trainings and summer schools). The special features of the non-formal teachers' ICT training include the availability of high-speed Internet and some computer equipment. The obtained basic and additional services allow teachers to make the extensive use of cloud services for different activities, namely the organization of students' group work and "inverted learning"; team-work on projects, assistance during homework; preparation for contests; conducting web-quests.

Keywords. Cloud-oriented, learning environment, non-formal education, cloud services, ICT, IC-competence

Key Terms. TeachingMethodology, TeachingProcess, KnowledgeManagementMethodology, QualityAssuranceMethodology, StandardizationProcess

1 Introduction

The draft of The Concept of Education in Ukraine for 2015-2025 noted that in the period of independency it was accumulated a lot of systematic problems in the Ukrainian education sector. They include the following: reducing of the logistical support of educational institutions; teachers' aging; lack of young specialists; inefficient, over-centralized, out-of-date management and financing of the system; growing inequality in access to education; decline of students' knowledge and skills quality; obsolescence of technology and teaching methods; lack of effective system for education quality control; reduction in textbooks' quality and lack of advanced technolo-

gies. In addition, there are unresolved issues: the unsatisfactory state of computer equipment; the poor quality and accessibility of the Internet, the lack of systematic updating of ICT applications in the educational process.

In the current information age, the transition from education when a teacher is a central figure for information transmission to person-oriented education is one of the most significant educational change [6].

Most of the mentioned problems can be solved by the introduction of cloud services in education, which will provide educational mobility of educational process participants and the virtualization of organizational and methodological components of educational process [8, p.209].

A detailed analysis of the foreign projects in Russia, Germany, Czech Republic, Australia, China, Israel, Africa, Singapore, Brazil, Colombia, Azerbaijan and the United States showed that the cloud-oriented learning environments are used by the teachers and students of foreign countries to improve the educational process, the access to training materials, schedules of work, curriculums for the revitalization of students' activities, to make easily the educational process during a quarantine period, the process of receiving homework and distance learning [8, p. 27].

Thus, cloud-oriented learning environments have a number of advantages for educational institutions in the organization of educational process and the use of learning technologies [9, p. 13].

Therefore, a teachers' training is crucial for an application of cloud technologies. It can be done through non-formal education.

The educational trends in the global space are characterized by considerable support and development of non-formal education. The term "non-formal education" is widely used in the scientific world community today. The understanding of "non-formal education" at the European level is gaining its significance for Ukraine during the current period of integration into the EU.

Zinchenko S. offers the following characteristics of non-formal education: social character of leadership; variability of educational programs and terms; combination of scientific and applied knowledge; voluntary of education; systematic character of education; goal-oriented activity of learners; orientation to satisfaction the educational needs of specific social and professional groups; creating comfortable learning environment for adults' communication; possibility of psychological protection during some social changes [15].

According to the research of N. Terjokhina, non-formal education in Ukraine covers the following spheres: non-school education; postgraduate education and adult education; school and student self-government (through the acquiring of organizational, communication, management skills); educational initiatives to develop additional skills and abilities (language and computer courses, hobby clubs, etc.); civic education (various activities of non-governmental organizations); folk high schools [14].

Summarizing the organizational structure of adult education in current Ukraine, Sighajeva L. identifies three forms of non-formal education: courses; hobby clubs; non-governmental organizations [13].

We consider that it is very important for the further development of non-formal education in Ukraine to take into account the recommendations for the development of non-formal education for the EU members and countries, which are going to be its

members, proposed by the Committee of Ministers of the Council of Europe. They include:

1. The development of effective standards, which recognize non-formal education as an integral part of general education, which are quality criteria of the educational process during non-formal education.
2. Providing information and examples of best practices, teaching methods, lists of skills, experience and knowledge, obtained through non-formal education should be developed.

2 Research Methods

Analysis, theoretical systematization of data; synthesis and evaluation of experimental work; the factor-criteria model to determine the level of teachers' IC-competence on the use of cloud services and the experiment, which was held among 190 subject teachers of Ukrainian secondary schools.

It was determined the ranking coefficient for each criterion to calculate the teachers' IC-competence level in the use of cloud services (Table 1).

Table 1. Factors and Ranking Coefficients

Factors	Ranking coefficients
Understanding of the role and use of cloud services in education	0.158
Basic knowledge about cloud services	0.16
Use of cloud services in professional activities	0.17
Ability for collaboration and self-education	0.174
Use of basic cloud services	0.172
Use of different forms of learning activity	0.178

3 Research Findings

There is a widespread adoption of cloud services in general secondary education. They are the main tools for the efficient organization of cooperation and cooperative work between students and teachers now. That is why the organization of teachers' training during non-formal education is necessary to increase their level of competence in the sphere of cloud services' application.

We can define the following forms of non-formal education: workshops, trainings, summer schools.

Teachers should know theoretical basics and the capabilities of the services during the training seminars, where they will be offered short practical work on the use of several services for the improvement of educational process or the organization of cooperation with students and to drill the material.

The purpose of the training is to improve the skills of working with certain services, so it gives attention on some problematic issues and extending teachers' knowledge.

During the summer schools teachers are given an opportunity not only to combine the two forms of competence development mentioned above, but also unlimited communication with colleagues, team-work on projects, familiarization with the latest innovations of IT industry and pedagogy.

The high-speed Internet and a gadget (laptop, tablet, etc.) are the two components that can be attributed to the special features of the organization of the mentioned training.

The training can be carried out on the basis of services in the cloud-oriented learning environment designed for teachers' training or in schools.

Cloud services include programs that are provided on demand in on-line mode, such as basic and additional ones. The basic programs include calendar, corporate email and document storage, basic office software (Word, Excel, PowerPoint and Forms). The additional cloud services are OneNote, Sway, GeoGebra, OneNote Classroom, the corporate electronic network Yammer, Power BI, Delve, Video, Project etc.

The pedagogical experiment was conducted in secondary schools of Ukraine, working within the framework of the experimental work on topic "Cloud services in education" (order of the Ministry of Education and Science of Ukraine of 21.05.2014 №629, scientific guide Lytvynova S.), namely in 18 experimental schools (Kyiv, Lugansk, Khmelnytsky, Dnipropetrovsk, Zhytomyr, Sumy, Vinnytsia, Ternopil regions). To collect data and determine the levels of teachers' IC-competence, it was developed the portal "The Research System" (<http://expert.obolon365.net>). 190 subject teachers took place in the experiment.

The teacher training was conducted during non-formal education by three stages.

The first stage "Designing of cloud-oriented learning environment" was conducted in the form of a training workshop (8 hours), during which each teacher designed his "cloud space".

On the second stage the teachers were offered intensive implementation of services during the summer school (40 hours). An educational project was developed in the result of the work.

At the third stage the teachers presented their best practices of the use of cloud services for educational purposes during a workshop.

Let us examine the advantages, specific features and disadvantages of the teachers' trainings within non-formal education on the use of cloud services (Table 2).

As a result of the experiment, it was found that the volume of use of cloud services by the teachers has almost doubled, and the level of IC-competence respectively: teachers' understanding of the cloud services increased by 41%, basic knowledge about cloud services - by 45%, the use of cloud services in professional work - by 46%, ability to cooperate with the help of cloud services - by 43%, the use of basic services in students' education - by 45%, the use of various forms of training activities with the help of cloud services - by 44% (Fig.1).

Table 2. Advantages, Features and Disadvantages

Type of non-formal education	Advantages	Features	Disadvantages
Workshops	Demonstration of practical application of services in the real-world	Duration up to 4 hours	An additional training of school teachers to present the use of services is required
Trainings	Presentation or training on the use of new services for educational purposes	Duration 4-6 hours	No more than 3-4 questions are considered, an additional training of trainers is required
Summer schools	Improving the knowledge on the use of certain services, consulting, communication with experts	Duration 40 hours	Public events with 60-120 participants; a special accommodation for training teachers, a broadband Internet connection and an additional training of trainers are required

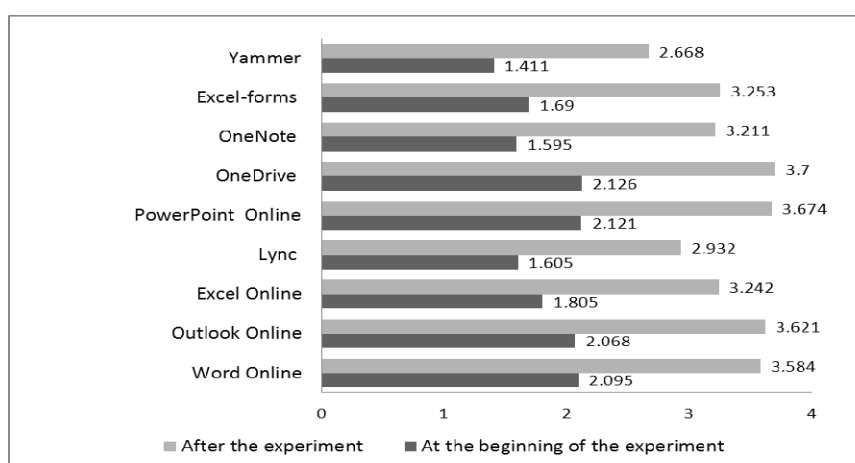


Fig. 1. The volume of using basic cloud services by the teachers

The teachers' training of new information and communication technologies, in particular, cloud services during non-formal education, made it possible to summarize the following: teachers actively improve their ICT qualification; there is no problem of a lack of gadgets; the systematic training in three forms gives significant results in the development of teachers' IC-competence; cloud services become the effective tools in the organization of cooperation and cooperative work between teachers and pupils.

4 Discussion

In the process of the selection of cloud services efficiency criteria, the parameters for an assessment of teachers' educational activity provoked a discussion. Some experts have focused their attention on the fact that cloud services should be used for the coordination of training projects and students' self- preparation for lessons. However, most experts emphasized the need for improving the educational process and defined such criteria as the organization of personalized learning, working in small groups, on-line communication with students.

Teachers' broad practical experience with gifted students became a barrier for the use of cloud services during the students' preparation for school contests (Olympiads) and coordination of scientific works of Minor Academy of Sciences (MAS) members. The results of the teachers' survey are shown in Fig. 2.

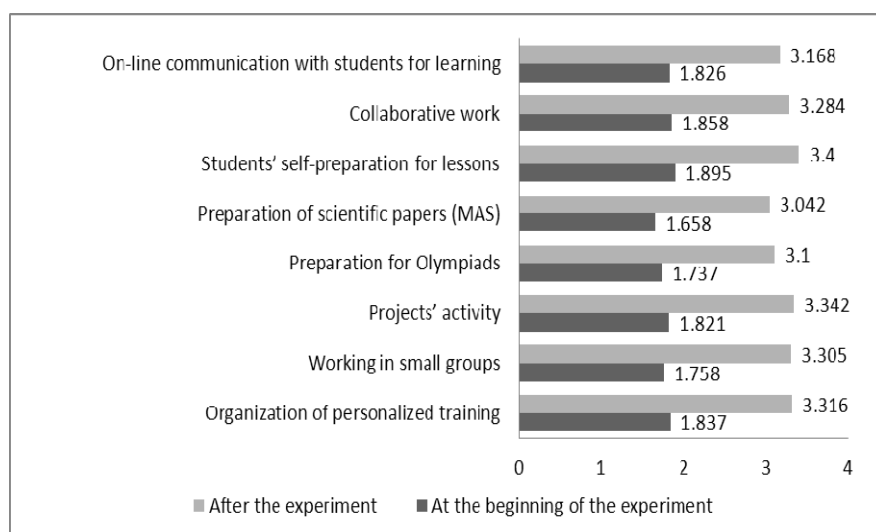


Fig. 2. The results of the teachers' survey

Analyzing the obtained results, it can be stated that teachers' lifelong learning is becoming a necessity. The pupils of the XXI century require new technologies in the organization of teaching process, due to the requirements to the quality of secondary

education. The emergence of new technologies, such as cloud services, gave an impulse for an active development of teachers' non-formal education.

5 Conclusions and Prospects of Further Research

Cloud services contribute to the improvement of the educational process. The development of these services can be arranged through non-formal education of three forms: practical workshops, trainings, summer schools.

The implementing of person-oriented education is a prerequisite for learning with the help of cloud services. There is no need for the absolute traditional teaching methods. The situations, when a student adapt to a teacher, could be avoided [8].

During the active use of cloud services the level of teachers' IC-competence increased from the beginner to average (Table 3).

The efficiency of enhancing teachers' IC-competence in the sphere of learning new technologies is 40% and more. This allows the teachers to make extensive use of cloud services for different activities, namely the organization of students group work, team-work on projects, "inverted learning"; assistance during homework; preparation for contests; conducting web-quests.

Table 3. Criteria and levels of teachers' IC- competence

Criteria	At the beginning of the experiment	After the ex- periment
Technological literacy	beginner level	average level
Extending of knowledge	beginner level	average level
ICT-creativity	beginner level	average level

So, the organizing of teachers' training on the use of cloud services in order to improve their ICT-competence in collaboration, global communication and cooperative work is a new direction in teachers' training.

On the other hand, the mechanism of teachers' training within non-formal education on the use of cloud services is not perfect yet. Therefore, the further research of this issue is required.

References

1. Ablameyko, S. V., Vorotnytsky, Y. I., Listopad, N. I.: "Clouds" technology in education. : Electronics INFO, 9, 30--34 (2013)
2. Barboura, M. K., Reeves, T. C.: The reality of virtual schools: A review of the literature. Computers & Education, 52(2), 402--416 (2009)
3. Bykov, V.: Models of Organizational Systems of Open Education. Atika, Kyiv (2009) (in Ukrainian)
4. Golitsyno, I. N.: Using of cloud computing in educational process. International Electronic Journal "Technology and Society (Educational Technology & Society)", 17, 450--459 online: ifets.ieee.org/russian/depositary/v17_i2/pdf/10.pdf (2014)
5. Kulyutkyn, Yu., Tarasov, S.: Educational environment and development of personality In: Mykhalchuk, M. P., Syvashynskaya, E. F. (eds.) Educational environment as a means of socialization: Proc. IX region. Sci.-Pract. Conf. Pupils and Students Youth, Brest, 16 March

- 2012, Brest A. S. Pushkin State Univ. (2013)
6. Lee, B. C., Yoon, J. O., Lee, I.: Learners' acceptance of e-learning in South Korea: Theories and results. *Computers & Education*, 53 (4), 1320--1329 (2009)
 7. Lubkov, R. V.: Didactic potential of virtual educational environment. Thesis (Cand. Ped. Sci.), 13.00.01, Samara, Russian Federation (2007)
 8. Lytvynova, S. G.: Cloud oriented learning environment, virtualization, mobility - the main lines of XXI century general secondary education. In: *Pedagogy of Higher and Secondary schools: Coll. scientific works*, vol. 40, Copenhagen: CPI State University "NUC", pp. 206-213 (2014)
 9. Lytvynova, S. G.: Design of cloud oriented learning environments of secondary schools. Foreign experience. In: *Information Technologies and Learning Tools: Electronic Scientific Specialized Edition*, 3(41), 10—27, online: http://journal.iitta.gov.ua/index.php/itlt/article/view/1052/810#U7LD9ZR_toE (2014) (in Ukrainian)
 10. Lytvynova, S. G.: The concept and basic characteristics of cloud oriented learning environment of secondary school. In: *Information Technologies and Learning Tools: Electronic Scientific Specialized Edition*, 2(40), 26—41, online: http://journal.iitta.gov.ua/index.php/itlt/article/view/970/756#U2aW6IF_vzA (2014) (in Ukrainian)
 11. Seydametova, Z. S., Seytvelyeva, S. N.: Cloud services in education. *Information Technologies in Education*, 9, 105—111, online: http://ite.kspu.edu/webfm_send/211 (2011) (in Ukrainian)
 12. Shcherbakova, T. N.: On the structure of the educational environment of educational institutions. *Molodyi Vchenyi*, 5, pp. 545-548 (2012) (in Ukrainian)
 13. Sighajeva, L. Je.: Description of the structure of adult education in current Ukraine. *Bulletin of Zhitomir State University*, 59, 38--42 (2011)
 14. Terjokhina, N.: Non-formal education as an important component of adult education system. *Comparative Teaching Studios*, 6, 109--114 (2014)
 15. Zinchenko, S.: Non-formal adult education: psychological context. In: *Pedagogical innovations: ideas, reality, prospects. Coll. scientific works*, vol. 12, pp. 95--101 (2014) (in Ukrainian)

Collaboration between Research Institutions and University Sector Using Cloud-based Environment

Yuliya Nosenko¹, Mariya Shyshkina¹, Vasyl Oleksiuk²

¹ Institute of Information Technologies and Learning Tools of NAES of Ukraine,
9 M.Berlynskoho St., Kyiv, Ukraine

{nosenko, shyshkina}@iitlt.gov.ua

² Volodymyr Gnatiuk National Pedagogical University of Ternopil,
2 Maxyma Kryvonosa St., Ternopil, Ukraine

oleksyuk@fizmat.tnpu.edu.ua

Abstract. The high-tech cloud-based university environment formation, which would provide collaboration between the research institutions and the university sector contributes to the experience exchange; organization of joint events on sectorial, national, and international levels; improvement of the research results' implementation. It would unite the learning resources, the tools of scientific exploration support, and cover different levels of training, that could promote overcoming the gap between the research process, and the implementation and use of its results. This approach is to contribute to the development of university autonomy, which is the aim of the current reformatations in higher education in Ukraine. The article represents the main results of the joint research laboratories activity that is held between the Institute of Information Technologies and Learning Tools of NAES of Ukraine and several universities being organized via cloud technology usage and devoted to its development.

Keywords. Cloud computing, e-learning, open education, research institution, university sector, pedagogical university, cloud-based learning and research environment

Key Terms. Academia, educational process, ICT Environment, methodology research

1 Introduction

During 15years the Institute of Information technologies and Learning Tools of the National Academy of Educational Sciences of Ukraine (IITLT of NAES of Ukraine) carries out research work aimed at solving the problems of development, implementation and use of new learning tools and information technologies in education construction and development of computer-based learning environment of open education

and e-learning, electronic educational resources, management and support of the researches, exploration of cloud technology. A number of important theoretical and applied results were achieved in recent years.

The relevance and the urgent need for these studies is caused mostly by the need to implement a modern educational paradigm, which consists of providing equal access to high-quality education for everyone who needs to learn who has desire and realizes need for life long learning.

Theoretical results and practical orientation of the researches conducted at the Institute are mainly subordinated to the specified educational paradigm, aimed at the development of scientific and methodological foundation of implementing the principles of open education. According to numerous studies, the implementation of the principles of open education is the backbone of the formation of global education systems, development and upgrading of learning and research environments, perspective way for national education system development.

Ukraine has achieved significant results in the study of theoretical and methodological basis of modeling and designing of information environment of open education (V. Bykov, G. Kravtsov, V. Kukhareno, A. Manako, L. Panchenko, S. Semerikov, A. Spivakovskiy, M. Zhaldak, and others). In particular, in V. Bykov's significant work the models of organizational systems of open education are designed; the models of information educational environment are proposed (methodical systems of e-learning, model of educational management in its different organizational levels, modern informatics teachers' training in higher educational institutions and others) [6]. The named works are aimed to create the methodological basis for further researches in this area, given that the cloud oriented environment is a new step in the development of open education.

General trends of implementation of promising ICT, including cloud technology, into organization of educational systems were studied in the works of V. Bykov, V. Kumar, S. Litvinova, N. Morze, V. Oleksyuk, S. Semerikov, Z. Seydametova, M. Shyshkina, A. Stryuk, M. Zhaldak, Yu. Zaporozhchenko (Nosenko) [15, 16], and others.

Problems of creation of cloud oriented learning environment in higher educational institution were studied in works of V. Bykov, A. Fox, R. Griffith, V. Kukhareno, V. Oleksyuk, L. Panchenko, S. Semerikov, M. Shyshkina, K. Slovak, A. Spivakovskiy, A. Stryuk, K. Subramanian, N. Sultan, Yu. Tryus, M. Zhaldak, and others.

In Z. Seydametova's paper [21] the analysis of the use of cloud services in higher school for educational and scientific purposes, and their systematization is proposed.

The main types of service models that reflect the possible ways of ICT-outsourcing use for software and computing resources access provision are considered in M. Shyshkina's paper [23].

The issues of determining the most appropriate technological solutions for corporate network of the cloud-based learning environment of educational institutions are examined by Ukrainian and foreign researchers [24].

S. Semerikov in his paper [12] analyses the term "cloud services" and its evolution, which outlines the prospects and trends of these services and their application to support the processes of research activities. The use of cloud services in maintaining the

processes of research activities is investigated by O. Spivakovsky [32], O. Spirin [31], A. Stryuk [33].

The study of A. Alkhansa, A.A. Shakeabubakor, E. Sundararajan, and A. Hamdan [1] is devoted to issues of productivity of university research processes, and it is proposed to use cloud services as a tool to increase efficiency and enhance the research activities.

Indeed, the cloud technology implementation will provide each educational institution with the following advantages and capabilities: to use the modern, systematically renewed IT infrastructure, software and services at no additional cost; reduce the cost of development and support of local information networks; engage the education participants' personal computer devices into the learning process; ensure the consolidation of electronic educational resources in a single electronic repository; implement mobility of educational process participants, support the distant, blended learning and other innovative forms and technologies [1], [7], [22], [27], [28].

Currently, the potential of cloud technology is engaged not enough in the practice of Ukrainian education. There is a growing discrepancy between the level of our local schools IT-infrastructure and modern European ones. These contradictions might cause a significant technological gap in ICT-equipment of higher educational institutions and even lead to a significant gap between the levels of their equipment in comparison with the similar ICT structures in foreign universities.

Therefore, an important task is to develop the learning and research environment of educational institutions in Ukraine considering the latest developments in the field of scientific and technical progress.

One of the possible ways for this implementation is the organized cooperation among research institutions and university sector, providing the joint research laboratories within the cooperation agreements to implement the results of scientific studies, involving representatives of educational institutions to scientific researches, improvement of teaching staff training, encouragement of participation in the development of joint projects on sectorial, national, and international levels.

The **purpose** of the article is to describe the state of the art of the cloud-based learning and research environment formation in Ukrainian university sector and to expose the research results on its deployment within collaboration between the research institution and the university sector.

2 The Presentation of Main Results

2.1 The Activity of the Institute of Information Technologies and Learning Tools of NAES of Ukraine in Education ICT Support

The National Academy of Educational Sciences of Ukraine (NAES of Ukraine) is the main organization on the national level which provides methodological and scientific support, legal and normative basics for the development of concepts and approaches to education improvement on different levels.

The Institute of Information Technologies and Learning Tools of NAES of Ukraine (<http://iitlt.gov.ua/>) is the principal research institution on scientific and methodological support of formation and development of electronic educational space in the field of general, special, vocational and teachers training and adult education and also on

ICT support of NAES of Ukraine. The Institute carries out computer and technological support of electronic resources of the research and educational institutions within NAES structure, and provides representation of these institutions via Internet [3].

Now the computerization covers all the basic functions of NAES, such as scientific, educational, administrative, informational. The corporate automated information system of NAES of Ukraine is developed systematically. It is connected to URAN, and therefore to GEANT. We should mention that NAES of Ukraine is its founding member since 2006 (<http://www.uran.net.ua/~ukr/uran-statut.htm>). And the Institute of Information Technologies and Learning Tools is a member of the Association of users of Ukrainian research and Academic network "URAN" (URAN - Ukrainian Research and Academic Network) since April 12, 2007 (<http://www.uran.net.ua/~ukr/uran-members.htm>).

The corporate automotive and information system of NAES forms a single computer and technological foundation of information support of NAES, and its principal components are: scientific projects; computer technologies; information, personnel, organizational, managerial, financial and economic resources. Tools and technologies that support corporate networks and electronic information resources of NAES such as sites of its scientific and educational institutions, various computerized databases, information portals, e-learning courses, etc. provide automated systems of research, electronic communications in the Internet space and maintain information process of creating electronic textbooks, manuals and other scientific and educational electronic editions [4], [5], [7].

There is a range of information resources and systems developed at the Institute: the official website of IITLN of NAES of Ukraine (<http://iitlt.gov.ua/>), a peer-reviewed journal in educational sphere, which publishes full-text article online bi-monthly with immediate open-access (<http://journal.iitta.gov.ua>), Digital Library of NAES of Ukraine (<http://lib.iitta.gov.ua/>), conference supporting system "Edu-conference" (<http://conf.iitlt.gov.ua/>), and others.

We can trace trends in the implementation and use of web-oriented open corporate information systems in NAES of Ukraine through the activities of IITLT of NAES of Ukraine during recent years. Thus, the electronic scientific specialized journal "Information Technologies and Learning Tools" (<http://journal.iitta.gov.ua/index.php/itlt/index>) was launched in December 2006, added to the list of electronic scientific professional editions in pedagogical sciences of Ukraine in April 2007. It is bimonthly open-access peer-reviewed e-journal that includes full-text articles in theory, methods and practice of using ICT in education. Since 2011 the edition began functioning on Open Journal Systems (OJS), developed by the Public Knowledge Project (Canada), which is an open information-technological platform for the deployment of scientific journals. Since 2012 the journal is indexed in scientometric databases, including such as Google Scholar, Index-Copernicus, Universal Impact Factor (UIF) et al., and numerous abstracts databases (the Directory of Open Access Journals (Sweden), Academic Journals database (Switzerland), National abstracts database "Ukrainikanaukova" (Ukraine) and many others). Currently the journal holds the 8th place according to "The rating of scientific journals of Ukraine", compiled by The Vernadsky National Library of Ukraine. According to Google Scholar as of April 1, 2016 the number of journal's articles citations is 2615, h-index – 21, i10-index – 82 [8].

The Digital Library NAES of Ukraine (<http://lib.iitta.gov.ua/>) is another significant project of the Institute, and it was founded during 2009-2011, developed on the EPrints platform. Using statistical module IRStatsit is possible to obtain the data on the quantity and quality indicators over the Library for downloads of all scientific productions or separate publications distributed within the specific collections (subdivisions by scientific institutions and their divisions, author name, period of time, topic of research, etc.) [8].

Since 2012 the web-translation tools for distant participation and conducting different scientific events (conferences, workshops, etc.) are used, particularly via Skype, Wiziq platform and others.

In 2014 the institutional system for conference support “Edu-conference” was projected and implemented[30]. By means of this platform the support of submission and reviewing the materials of participants of Ukrainian Scientific Conference for Young Scientists “NaukovaMolod” and annual reports for the conferences of IITLT of NAES of Ukraine is held.

Development of experimental activity of IITLN of NAES of Ukraine, which provides practical implementing of scientific studies into the educational process, is carried out within joint research laboratories, research projects at the regional and national levels, pedagogical experiments, etc.[22],[24,25,26].

2.2 The Expediency of Cooperative Work between a Research Institution and University Sector

Formation of a high-tech learning environment based on cloud technology, which would unite educational resources for learning purposes, support of scientific research, and cover different learning levels, could promote overcoming the gap between the research process, and the implementation and use of its results(that is a significant problem in context of Ukrainian education and science interaction) [27].

One of the possible ways to develop learning and research environment of educational institutions of Ukraine considering the latest developments in the field of scientific and technical progress is the cooperation among the research institutions and university sectorial, providing the joint research laboratories to implement the results of scientific studies, involving representatives of educational institutions to scientific researches, improvement of teaching staff training, encouragement of participation in the development of joint projects on sectorial, national, and international levels [27].

Establishment of joint research laboratories contributes to the spreading of the experience of activity management and its results; organization of joint events on sectorial, national, and international levels; improvement of the organizational techniques of research results’ implementation; overcoming the potential gaps between conducting scientific researches and the impractical use.

This approach can contribute to the development of university autonomy, which is the aim of current reformations in higher education and is consistent with the adoption of the new Ukrainian Law on Higher Education.

The research areas of scientific and educational institutions are coordinated by developing the links with the joint laboratories, and thus the focus is made at those pedagogical issues, being currently at most need in educational space. By virtue of this the mechanisms of the research results’ implementation are improved. The proc-

esses of teaching staff training occur in close cooperation between researchers and teachers, in interrelation of educational process and scientific research.

On the basis of modern network technologies the possibility to appeal to the remote educational resources online appears. For example, it can be realized with the use of virtual laboratories and laboratory systems of remote access, university resource rooms and laboratories for conduction of demonstration experiments [27].

Trends associated with the processes of integration of universities educational environments, suggest their participation in regional clusters formation. Clusters are a form of cooperation in the field of scientific, research and innovative activity and are formed through the merger of companies and organizations relevant to a particular kind of industrial activity [10]. Cooperation can take the form of information exchange, resource sharing, pooling in terms of staff training and employment. In particular, one of the advantages of university clusters is the transfer of non-core functions of organizing and maintaining of the university ICT infrastructure functioning to the experts. To realize this purpose a separate IT department is created in the cluster [10], [27]. Thus the functioning of the high-tech infrastructure is conducted from a single center through outsourcing, i.e. ICT services required by a system are implemented through another external system.

The trend towards consolidation of higher educational institutions has gained acceptance abroad [13], [19], and is manifested in Ukraine. It consists of creating regional universities, which may contain several higher educational institutions. Implementing shared technological platform of the regional educational institution functioning on the basis of cloud computing is a way to solve a number of problems that arise when combining infrastructural ICT into a unite framework. In its turn it allows to access to the best examples of electronic educational tools and resources by those institutions that don't possess the necessary financial support and powerful ICT units [27].

Besides, within the interaction network the university cooperation with research institutions and business structures can be realized, the processes of training, skills development, international projects implementation, linkages between schools and universities can be conducted.

This is consistent with the prospect of creation of integrated (sectorial, national) databases, data collections, resources that are made available to various educational institutions [9], [25]. In order to take advantage of such collections in full, it is also useful to implement tools of cloud computing [27].

2.3 The Joint Laboratory of the Institute of Information Technologies and Learning Tools of NAES of Ukraine, and Kryvyi Rih National University

For the aim of coordination and further development of the researches the Joint laboratory "Cloud Computing in Education" (CCELab) was created on the basis of Kryvyi Rih National University and IITLT of NAES of Ukraine (2012).

The main goal of the CCELab is methodological and experimental research of emerging e-learning technologies and exploration of different aspects of cloud computing application for education and personnel training.

So, main objectives of the CCELab are:

- coordination of research and development on the problems of using cloud technology in education, carried out at IITLT of NAES of Ukraine, National University of Kryvyi Rig, and some other national universities and institutions;
- development, testing, implementation and experimental approve of cloud cloud-based e-learning platforms, tools and services;
- investigations on cloud-based learning environments' modeling and application;
- experimental study of cloud-based e-learning infrastructures for education and training of professionals;
- research for fundamental, methodical, technological, organizational and other use features, presupposition of an introduction and perspective ways of using cloud technology in education;
- publication and discuss of results of scientific and experimental study of the laboratory;
- participation in the organization of international and national scholarly conferences, workshops, etc.

Within the joint activities a range of electronic resources was developed: the site of joint research laboratory (<http://cc.ktu.edu.ua/>); scientific and educational cloud of the Department of Cloud Oriented Systems of Education Informatization of IITLN of NAES of Ukraine, which is used to organize joint activities and resources exchange; cloud oriented educational and scientific laboratory environment. A unique experience is the use of OwnCloud – software with open source, designed for development, modification of specific applications required for the organization of scientific and methodological studies.

In addition to the above, an annual international seminar "Cloud technologies in education" was founded within joint laboratory activity. The seminar is attended by representatives of more than 50 educational institutions from more than 20 cities of Ukraine and other countries – Russian Federation, Australia, USA, etc. The proceedings of the seminar are available in open access (<http://cc.ktu.edu.ua/report.html>). Also, the annual surveys on the development and use of cloud computing services in schools are held.

To show the state of the art of the cloud-based learning environment development and the rate of cloud-based services use by educational personnel in Ukraine the survey was made within the framework of the International internet-seminar "Cloud Technologies in Education" that is held annually since December, 2012. More than 120 members of 54 educational institutions from 22 cities and 18 regions of Ukraine took part in this seminar. The participants were concerned with the problems of cloud technology, well acquainted with the modern trends of technological development, and their organizations were well equipped and oriented for using advanced ICT. So, we involved them into appropriate interrogation.

The responses to the question: "What type of the cloud-based services is mainly used in your educational institution?" showed that 78% of participants had chose SaaS (Software-as-a-Servise), 29,3% – PaaS (Platform-as-a-Service), 22% – IaaS (Infrastructure-as-a-Service); and 31,7% – DaaS (Desktop-as-a-Service). The results are presented in Fig. 1 (the entries are not excluded).

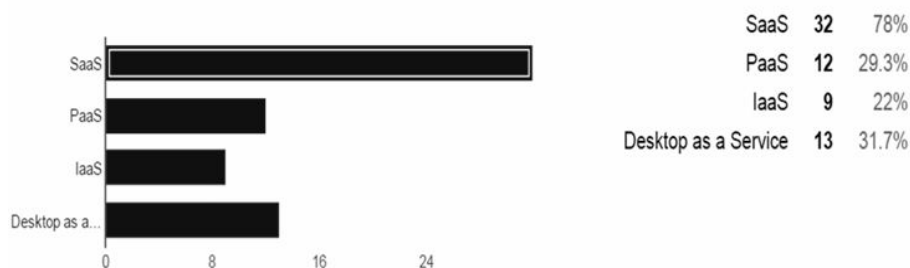


Fig.1. The results of the survey on cloud technology use at the institutions of higher education in Ukraine (2015)

To the question: “What type of the cloud-based platform or services are used in you institution?” we got the next results: Google Apps for Education – 76%;Microsoft Office 365 – 51,1%; Special Software (for example, the SageMathCloud) – 17%; 10,6% use public cloud services, such as Amazon, Microsoft Azure, etc.); 6,4% use services of the private cloud (Microsoft Azure, Xen, WMWare, etc.). These results are presented at Fig. 2 (the entries are not excluding).

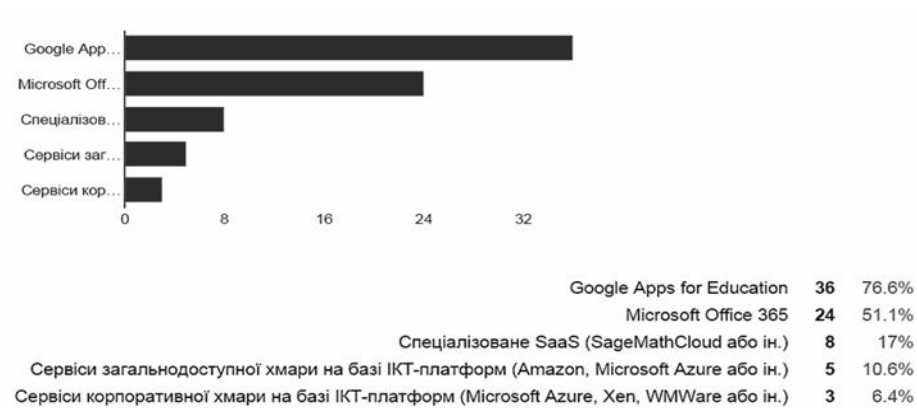


Fig. 2. Application of cloud-based platforms and services at the educational institutions in Ukraine (2015)

As it appears from the study, the cloud-based services are widely used in Ukrainian educational institutions. But still their use is to be more systematic, to be organized into united system, to be consciously and purposely oriented at pedagogical aims. So there is a current need for the upgrading of ICT competence of educational personnel, mainly those ones who engage in providing educational systems with emerging ICT, in particular, public administration employees [28].

2.4 The Joint Laboratory of the Institute of Information Technologies and Learning Tools of NAES of Ukraine and Kherson State University

In 2011 our joint research laboratory on issues of using ICT in control of learning quality was created at Kherson State University. In 2013 the University was approved as the experimental base for research work of the regional level: “Computer-oriented quality management system of electronic educational resources in schools” (2013-2015). The experiment involved four pilot schools from Kherson region.

The aim of the experiment is to determine and verify experimentally the requirements and methods of evaluating the quality of electronic educational resources (EER) in pilot schools learning process. As a result of the experiment the didactic requirements and methodology for evaluating the quality of EER were determined and tested [26]. The significant part of the experiment was the development of the performance indicators for evaluating the quality of cloud oriented components and methods forevaluating the quality of electronic learning resources in the cloudoriented environment. In 2014 within the experiment the cloud storage of open informational resources via Google Drive was created: <https://drive.google.com/folderview?id=0BwItlyWdtLeIbU5pUWR6c0NiWTQ&usp=sharing>. Thus, the experts have opportunity of sharable documents processing, editing, and updating, relating to expert evaluation activity.

The evaluation of the quality of EER in the cloud-based learning environment is a separate trend of the Laboratory activity. In this case there are different approaches and indicators. The access organization has been changed, so the models of learning activity have been changed also. That is a problem: what features and properties have to be checked to measure the pedagogical effect of the cloud-based approach? How to assess the pedagogical innovation, what are the factors of its influence for pedagogical systems, its structure and organization? Were the learning results improved due to the cloud-based models? In this context just the quality of EER is a criterion to estimate the level of organization and functioning of the cloud-based learning environment [22].

The perspective way of estimating the quality of EER is to be made by means of the cloud-based environment. As the resources are in collective access so there is a way to involve experts into the learning process so they could observe and research the functioning. This is a way to make the process of quality estimation more convenient, flexible and quick. The process of estimation becomes anticipatory and timely. The estimation may be obtained at ones with the process of EER elaboration, and this is very important to facilitate the process [22], [26].

The cloud-based learning component used in the experiment has undergone the quality estimation. The method of estimating the quality of learning resources developed in the joint laboratory of educational quality management via ICT [11] was used and adapted for this study.

The advantage of the approach is in possibility to compare the different ways of resources implementation with regard to the learning infrastructure set. There is a perspective of future research in this direction as for taking into consideration different types of resources and environments.

2.5 The Joint Laboratory of the Institute of Information Technologies and Learning Tools of NAES of Ukraine and Ternopil Volodymyr Hnatiuk National Pedagogical University

The joint laboratory was created for consolidating and coordinating the research work of scientists of IITLT of NAES of Ukraine and the Department of Informatics and methods of its teaching of Ternopil Volodymyr Hnatiuk National Pedagogical University. The general purpose of laboratory is implementation of cloud technology in the educational process of higher and secondary education institutions.

The objectives of lab researches are:

- integration of cloud-based and traditional applications into IT-infrastructure of universities;
- implementation and use of public clouds for learning process support;
- analyzing, testing, deployment, implementation and experimental approve of the private cloud platforms in higher education institutions;
- creating virtual laboratories based on corporate clouds, and their technical and pedagogical support;
- development of methodology of virtual labs application in the learning process.

We consider clouds as tools for solving the learning problems. Therefore, researchers help in deploying school clouds based on public platforms and obtaining academic licenses for them.

Scientists of joint laboratory researched the integration of cloud services Google Apps into information and educational space of higher or senior educational institution. They have offered that integration has provide unification authentication and access to learning objects. The unified authentication was designed on the base LDAP-directory. We have adjusted programs for synchronization of the LDAP-directory and public clouds (Google Apps, Microsoft Office 365) [17].

The participants of laboratory investigated models deployment of cloud technology in IT-infrastructure. The hybrid model have recognized most suitable to IT-infrastructure of higher educational institution. Open source platform for organizing enterprise clouds analyzed and tested during last year. We have identified IaaS model as most functional for enterprise clouds. The model of private academic cloud was developed.

The enterprise cloud of Department of Physics and Mathematics of Ternopil V. Hnatiuk National Pedagogical University deployed (<http://cloud.fizmat.tnpu.edu.ua> – available only from university network or through VPN). We have used Apache CloudStack as platform for this cloud [18]. Deployed cloud infrastructure contains such elements: 1 zone, 1 pod, 1 cluster, 3 hosts, 3 primary storages and 1 secondary storage.

In our cloud, Apache CloudStack provides:

- run above 100 virtual machines (VMs);
- connect VMs through simple and advanced networks;
- access to VMs through web-interface and standard protocol such as RDP and SSH;
- distribution of computing resources for VMs;
- create template and snapshot of VMs;

- authentication through LDAP protocol.

We had deployed in cloud infrastructure several virtual networks. We can associate it with physical networks of IT-infrastructure. This allows to model routing process in Internet.

Scientists of joint laboratory are developing virtual laboratories for learning computer sciences, it is described in detail in the article[22].

Three joint research laboratories mentioned in the article were chosen to be the experimental base for the implementation of research on the topic: "Methodology of formation of a cloud-based learning and research environment of educational institution" (Decision of the Scientific Council of the Institute of Information Technologies and Learning Tools of NAES of Ukraine, №4 of March 04, 2015).

Within this research the overall investigation of the ICT competence of the students and lecturers was undertaken. In the focus of investigation there was research and educational community which was formed around the cloud-based trend of research. There were those engaged in research and training actions, seminars, webinars, conferences other events conducted by the joint laboratories involving the vast number of educational institutions. Besides three institutions were joint laboratories were established other institutions where collaboration was carried out under agreements on cooperation with IITLT of NAES of Ukraine also were included.

Among them there were:

- National Pedagogical University, named after M.P. Dragomanov;
- Drohobych State Pedagogical University, named after I. Franko;
- Cherkasy State Technological University.

While forming the control and experimental groups of the experiment the next conditions were taken into consideration:

- In the control group there were those participants of the experiment who had taken part in scientific and methodological events of the Institute and had shown significant interest in the use of cloud technology in their research and educational activity; but they had not be engaged into special training of research projects of the Institute, devoted to the cloud computing tools application – lecturers, researchers, post graduate students and others (58 participants);
- In the experimental group there were those members who also had belonged to the scientific and educational community concerned with the cloud computing applications still they had been members of joint research laboratories and had taken part in special training devoted to CC application and also had been involved into research work in which such training had been supposed (60 participants).

Analysis of the results of pedagogical experiment formative stage showed that the distribution of ICT competencies in experimental and control groups is characterized by statistically significant differences due to the use of cloud services for educational and scientific purposes and also use of the appropriate methods for their application in the process of scientific and pedagogical staff training (Fig. 3).

Another stage of the experiment was related to a survey of students. The experimental group consisted of students who had studied math or informatics with the use

of cloud services due to the specially developed technique (overall by all experimental base – 50 students).

The control group consisted of students who had studied the same subjects and had been familiar with cloud services but used it without a specially developed method (120 students). Survey results are showed in Fig. 4. The distribution of ICT competences had statistically significant differences due to the use of the special technique.

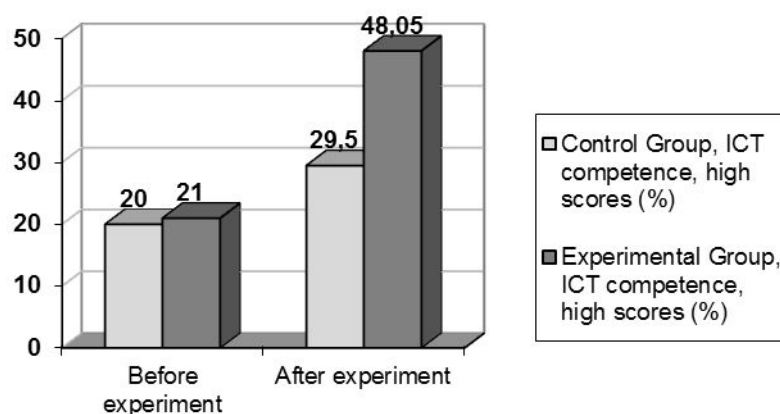


Fig. 3. Measuring the level of lecturers' ICT competence before and after the experiment

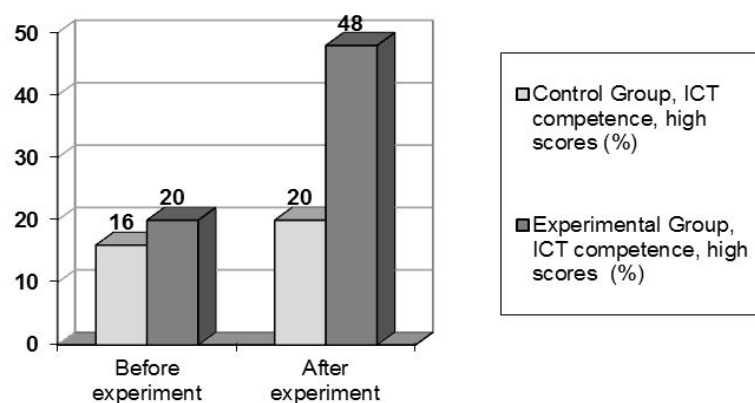


Fig. 4. Measuring the level of students' ICT competence before and after the experiment

Thus, the analysis of the pedagogical experiment results showed that there had been a growth of ICT competencies of students and pedagogical stuff of educational institutions, that had had statistically significant differences, indicating the efficiency

of the cloud-based learning and research environment formation and relevance of the special technique that was used in the training of scientific and pedagogical staff.

3 Conclusions

Within the research the most important scientific and methodological problems which concern the development of learning and research environment of educational institutions in Ukraine were revealed. Among them there are: the pedagogical and technological parameters of the learning environment modernization determining; the methods of research and pedagogical personnel training development with the use of the cloud-based tools; the learning content improvement due to the cloud computing systems application for different learning disciplines; elaboration of methodical guides for the use of the cloud-based learning tools for IT-disciplines.

The specific complex of problems is connected to the use of cloud computing in the field of inclusive education which is one of research priorities of European scientific and educational community as it is reflected in Horizon 2020 international program for research and innovations.

Among the possible ways of improving learning and research environment of higher educational institutions there are:

- wider use of emerging ICT tools and network technologies in the scientific and educational practice as its capacities essentially increased due to the use of cloud technology;
- more active involvement of international network information systems and infrastructures into scientific and educational activity of universities, so as to promote more openness and integration of scientific research and applications;
- organization of international cooperation, coordination and harmonization of national and international strategies for scientific and technological development, broader integration into the European educational and research space;
- expansion of cooperation among educational and research institutions; establishment of corporate structures of various nature (affiliated schools in manufacturing, training and engineering centers, research laboratories etc.) aimed at the development and implementation of advanced ICTs in various spheres of social activity, in particular in education;
- defining the prioritized and relevant researches in IT areas, particularly concerning educational technology and software for learning purposes to attract the attention of society, public sector and industry to this issues.

A promising solution for modernization of the learning and research environment of educational institutions is the development of cooperation through the organization of joint research laboratories that may contribute to the spreading the research results; organization of joint events on sectorial, national, and international levels; improvement of the organizational techniques of research results' implementation; clearing the research priorities; overcoming the potential gaps between conducting scientific researches and their practical use, and so on.

References

1. Alkhansa, A., Shakeabubakor, A., Sundararajan, E., Hamdan, A.R.: Cloud Computing Services and Applications to Improve Productivity of University Researchers. In: J. International Journal of Information and Electronics Engineering, vol. 5, No. 2, 153–157 (2015)
2. Bykov, V.Yu., Mushka, I.V.: Electronic Pedagogics and Modern Instruments of the Open Education. In: J. Information Technologies and Learning Tools, vol. 13, № 5, <http://www.nbu.gov.ua/e-journals/ITZN/em2/emg.html> (2009) (in Ukrainian)
3. Bykov, V.Yu., Oliynyk, V.V.: Information Networks of Open Learning Environment. In: J. PislidyplomnaOsvita v Ukrayini, vol. 1, 54–63 (2008) (in Ukrainian)
4. Bykov, V.Yu.: Increasing Importance of ICT in Education of Ukraine. In: Kremen', V.G. (eds.) PedagogikaiPsikhologiya, pp. 29–33. PedagogichnaPressa, Kyiv (2009) (in Ukrainian)
5. Bykov, V.Yu.: Learning Environment of Modern Pedagogical Systems. In: J. ProfesiynaOsvita: PedagogikaiPsikhologiya, vol. IV, 59–80 (2004) (in Ukrainian)
6. Bykov, V.Yu.: Models of the Open Education Organizational Systems. Atika, Kyiv (2008) (in Ukrainian)
7. Bykov, V.Yu., Prokopenko, I.Ph., Rakov, S.A.: Towards the Issue of Informatization of Higher Education Pedagogical Institutions. In: J. Compyuter u Shkoli ta Simyi, vol. 22, № 4, 8–13 (2002) (in Ukrainian)
8. Bykov, V.Yu., Spirin, O.M., Luparenko, L.A.: Open WEB-based Systems of Scientific and Educational Research Implementation Monitoring. In: J. Teoriya i Praktyka Upravlinia Sotsial'nyimi Systemami, vol. 1, 3–25 (2014) (in Ukrainian)
9. Galimov, A.M., Kashapov, N.Ph., Makhank'o, A.V.: Management of Innovative Activity in Higher Education: Problems and Prospects. In: J. Obrazovatel'nyie Tekhnologii i Obshchestvo, vol. 15, № 4, 392–413, http://ifets.ieee.org/russian/depositary/v15_i4/html/18.html (2012) (in Russian)
10. Gudkova, A.A., Batkilina, Yu.M.: Formation and Development of Regional Innovation Technology Clusters. In: 6th Russian Conference on Scientific, Expert-analytical and Informational Support of National Strategic Planning, Innovation and Technological Development of Russia, pp. 190–193. Moscow (2010) (in Russian)
11. Kravtsov, H.M.: Methods and Technologies for the Quality Monitoring of Electronic Educational Resources. In: 11th Int. Conf. ICTERI, pp. 311–325. CEUR Workshop Proceedings, Lvov (2015)
12. Markova, O.M., Semerikov, S.O.: Cloud Technologies for Education: beginnings. In: J. Information Technologies and Learning Tools, vol. 46, № 2, 29–44, http://nbuv.gov.ua/j-pdf/ITZN_2015_46_2_6.pdf (2014) (in Ukrainian)
13. Mukhamedyarov, A.M., Divayeva, E.A.: Regional Innovation System: the development, operation, evaluation, efficiency. Gilem, Ufa (2010) (in Russian)
14. Nosenko, Yu.G., Shyshkina, M.P.: Recent Trends of Development of Cloud-Based Learning and Scientific Environment of Pedagogical System: From the Experience of the Institute of Information Technologies and Learning Tools of NAPS of Ukraine. In: J. Naukovy-iChasopys M.P. Dragomanov NP, series 2 “Computer Oriented Learning Systems”, vol. 23, № 16, 153–158 (2015) (in Ukrainian)
15. Nosenko, Yu.G.: Cloud Technology in Open Education Space. In: Litvinova, S.G. (eds) The Modeling and Integration of Cloud Oriented Services of Learning Environment, pp. 24–34. Comprint, Kyiv (2015) (in Ukrainian)

16. Nosenko, Yu.G.: The Evolution of Cloud Computing as a Contemporary Learning Tool. In: *J. Informatyka ta InformatsiyniTekhnologiyi v Navchal'nykh Zakladakh*, vol. 5, 16–21 (2015) (in Ukrainian)
17. Oleksyuk, V.: Experience of the Integration Cloud Services Google Apps into Information and Educational Space of Higher Educational Institution. In: *J. Information Technologies and Learning Tools*, vol. 35, № 3, <http://journal.iitta.gov.ua/index.php/itlt/article/view/824/63> (2013) (in Ukrainian)
18. Oleksyuk, V.: Implementation of Cloud Computing as a Component of the University IT Infrastructure. In: *J. Information Technologies and Learning Tools*, vol. 41, № 3, <http://journal.iitta.gov.ua/index.php/itlt/article/view/1042> (2014) (in Ukrainian)
19. Philipov, I.: Higher Educational Clusters in Practice. In: *J. Intelligent Enterprise*, vol. 238, № 4, <http://www.iemag.ru/analitics/detail.php?ID=25820> (2012) (in Russian)
20. Pinchuk, O.P., Demyanenko, V.M., Nosenko, Yu.G., Shyshkina, M.P.: Scientific and Experimental Activities of the Institute of Information Technologies and Learning Tools of NAPS of Ukraine on the Basis of Educational Institutions of Different Levels. In: *J. Computer u Shkoli ta Simyi*, vol. 125, № 5, 18–23 (2015) (in Ukrainian)
21. Seydametova, Z.S., Syeytveliyeva, S.N.: Cloud Services in Educaiton. In: *J. InformatsiyniTekhnologiyi v Obrazovanii*, № 9, 105–111 (2011) (in Russian)
22. Shyshkina, M.: The Hybrid Service Model of Electronic Resources Access in the Cloud-Based Learning Environment. In: *11th Int. Conf. ICTERI*, pp. 295–310. CEUR Workshop Proceedings, Lvov (2015)
23. Shyshkina, M.P. Models of Providing Access to Software in the Cloud-Oriented Educational Environment. In: *J.InformatsiyniTekhnologii v Osviti*, vol. 22, 120–129 (2015) (in Ukrainian)
24. Shyshkina, M.P., Kogut, U.P., Popel', M.V.: Systems of Computer Mathematics in the Cloud-Oriented Educational Environment of the Institution. In: *J. Science and Education a New Dimension. Pedagogy and Psychology*, vol. II, №14, Issue: 27, 75–78 (2014)
25. Shyshkina, M.P., Spirin, O.M., Zaporozhchenko, Yu.G.: Problems of Informatization of Education in Ukraine in the Context of Development of Research of ICT-Based Tools Quality Estimation. In: *J. Information Technologies and Learning Tools*, vol. 27, № 1, <http://journal.iitta.gov.ua/index.php/itlt/article/view/632/483> (2012) (in Ukrainian)
26. Shyshkina, M.P., Zaporozhchenko, Y.G., Kravtsov, H.M.: Prospects of the Development of the Modern Educational Institutions' Learning and Research Environment: to the 15th Anniversary of the Institute of Information Technologies and Learning Tools of NAPS of Ukraine. In: *J. Information Technologies in Education*, vol. 19, 62–70 (2014)
27. Shyshkina, M.P.: Innovative Technologies in the Development of Educational and Research Environment of the Educational Institution. In: *J. Obrazovatel'nyietekhnologiiobshchestvo*, vol. 16, № 1, 599–608 (2013) (in Russian)
28. Shyshkina, M.: Emerging Technologies for Training of ICT-skilled Educational Personnel. In: *J. Communications in Computer and Information Science*, Berlin-Heidelberg, Springer-Verlag, 274–284 (2013)
29. Shyshkina, M.P., Kohut, U.P., Bezverbnyy, I.A.: Formation of Professional Competence of Computer Science Bachelors in the Cloud Based Environment of the Pedagogical University. In: *J. Problems of Modern Teacher Preparation*, vol. 9, part 2, 136–146 (2014) (in Ukrainian)
30. Slovins'ka, O.D.: Organizational Aspects and the Implementation of Video Conferencing in the Educational Process of Distance Education. In: *1st Ukrainian Scientific Practical Conference for Young Scientists "NaukovaMolod' – 2013"*, pp. 83–87. IITLT of NAES of Ukraine, Kyiv (2013) (in Ukrainian)

31. Spirin, O.M.: ICT for Monitoring the Implementation of the Research Results. In: J. Information Technologies and Learning Tools, vol. 36, № 4, <http://journal.iitta.gov.ua/index.php/itlt/article/view/890/655> (2013) (in Ukrainian)
32. Spivakovsky, A., Vinnik, M., Tarasich, Yu.: Using ICT in Training Scientific Personnel in Ukraine: Status and Perspectives. In: 11th Int. Conf. ICTERI, pp. 5–20. CEUR Workshop Proceedings, Lvov (2015)
33. Stryuk, A.M., Rassovyts'ka, M.V.: The System of Cloud Oriented Learning Tools as an Element of Educational and Scientific Environment of Higher School. In: J. Information Technologies and Learning Tools, vol. 42, № 4, <http://journal.iitta.gov.ua/index.php/itlt/article/view/1087/829> (2014) (in Ukrainian)

Mobile Learning Technologies for Learning English

Nataliya Osipova, Olga Gnedkova, Denis Ushakov

Kherson State University, 27, 40 rokiv Zhovtnya St., 73000, Kherson, Ukraine

{Natalie, Gnedkova}@ksu.ks.ua, Coldcloudu@yandex.ua

Abstract. The priority of education development is the introduction of modern information and communication technologies (ICT), in particular - technologies of mobile learning, providing improvement of teaching and educational process, accessibility and effectiveness of education for personal development according to individual inclinations, abilities, based on lifelong learning. Due to rapid expansion, proliferation and increasing of functionality of mobile technologies we should use their potential to improve and facilitate learning, ensuring accessibility, equity, individualization and flexibility. The analysis of the basic features and benefits of mobile technologies for lifelong learning was conducted. The model using mobile learning technologies for learning English phonetics is proposed in the article. The mobile application allows organizing the learning process of English practical phonetics. This article describes the methodical implementation of the proposed model for learning of English phonetics. Modern education in many countries, including Ukraine, foresees the necessity of learning of foreign languages. The proposed mobile learning technologies help to achieve the main aim of learning foreign languages – forming the general and professional communicative competence of specialists in the globalized society.

Keywords. Mobile learning, environment, mobile learning environment; ICT, ICT Tool, mobile application

Key Terms. Mobile learning, ICT Tool, Technology

1 Introduction

This article describes the processes of developing and implementing the model of using mobile learning technologies for English learning using the mobile app, which is designed to teach practical English phonetics. Its main purpose is to increase English communicative competence and professional competence of specialists in a globalized society based on lifelong learning. As stated in the Project of the National Strategy of the Development of Education in Ukraine for 2012-2021 years, the aims of the National Strategy of Education are: increasing access to quality, competitive education for citizens of Ukraine in accordance with the requirements of innovative development of society, economy of the citizen; providing personal development according to the individual inclinations, abilities, needs-based on lifelong learning [5].

2 Related Work

Analysis of recent researches has shown that the priority of education development is the introduction of modern information and communication technologies (ICT), in particular - mobile learning technologies that ensure the improvement of the educational process, accessibility and effectiveness of education, young generation training in the information society. It should be noted, mobile learning is a new educational paradigm, where a new learning environment is created. Students can access to learning materials at anytime and anywhere, making the learning process comprehensive and motivation for continuing education and lifelong learning [4].

According to UNESCO, today there are more than six billion mobile phone subscribers worldwide. Thus, each person who accesses the Internet from a computer accounts two people, making it from the mobile device. Due to the widespread and rapid expansion of functionality of mobile technologies, UNESCO enthusiastically uses its potential to improve and facilitate learning, accessibility, justice, individualization and flexibility for students. The European Commission is funding a number of successful R & D projects focused on mobile learning. Over the years, these projects have been gradually evolved from the development of hardware and software that is targeting primarily at processes, to approach focused on the individual, in which mobile technology plays an important role, along with a fixed technologies to support learning in any anytime and anywhere [9].

In the definition of «mobile learning» given by S. Geddes [2] it is also mentioned the possibility of mobile learning influences on students' behavior and intellectual.

Theory and practice of usage of mobile devices and mobile educational resources in education is actively discussed at various scientific events. Since 2002, the International Conference «MLearnCon» is held, which focuses on applying mobile technologies in the context of learning and performance support, the strategies for integrating these technologies into the training mix, and the best practices for designing, developing, and delivering mobile content [10]. The International Conference of Mobile Learning, which was held since 2005 года, sought for providing a forum for the presentation and discussion of mobile learning research which illustrate developments in the field[6]. There are interesting results of the project «Mobile Technologies in Lifelong Learning: best practices» (MOTILL). The key concepts in MOTILL are Lifelong Learning and Mobile Technologies. The MOTILL project investigates how these technologies may impact on the diffusion of a social model where learning and knowledge are accessible to all, regardless of social and economic background, age, gender, religion, ethnicity or disability [6].

3 Problem Setting

Mobile learning is a part of a new pattern of education, created by technology that supports flexible, affordable, individual training. Daily use of students of mobile phones and other devices (tablets, digital assistants, MP3 players, flash drives, electronic-book readers and smartphones), which can be used in education, is now the main incentive of mass distribution of mobile learning. It is important to note that mobile technologies can help in the provision of quality education in the development

of children, youth and adults around the world, that is stated in the aims of the UNESCO program «Education for All») [9].

The purpose of the article is the design and implementation of the model of usage mobile learning technologies as a means for formation of foreign language communicative competence.

Tasks:

- Analyze of main features and benefits of mobile technologies for lifelong learning
- Find the attitudes of students to use mobile learning technologies
- Design of the model of usage mobile learning in English learning
- Mobile Application of Development for English articulation phonetics
- Description of methodical implementation of the mobile learning model

4 Main Features and Benefits of Mobile Technologies for Lifelong Learning

It is widely recognized that the idea of mobile learning is based on the training opportunities offered by mobile technologies. This method is most relevant when the student is not in the predefined place. Student uses the situational approach and available resources. Mobile learning also allows students to easily change places and conditions and combine training in several schools. Mobile education implies the emergence of a range of new methods of teaching and learning based on the belief that cooperation in terms of the traditional classes are often not as effective as it hoped [7].

Mobile education is closely connected life with training and this type of activity is not associated exclusively with the school or university [9].

Mobile learning is applied best as students' queries support, communities and social networks for training at the working place and etc. Mobile technologies allow fixing students achievement, promoting social inclusion and supporting lifelong learning (MOTILL) [6].

Advantages of Mobile learning for lifelong learning:

- Improving of access to learning resources regardless of time or location;
- Using of inexpensive everyday technologies (hardware and software);
- Choose the own learning rhythm;
- Supporting for popular ways of interaction (social networks, e-mail, phone calls, SMS, forums, chat rooms);
- Fast and quality assessment and diagnosis of possible learning problems.

The adaptation of education to the changing needs of students, encouraging of education continuing to renew and expand the received knowledge, creating a culture of continuous learning, when students do not receive formal education, but also they get to use the technologies of personal use for getting information and extension of the scope of their knowledge are the powerful mobile learning tools to support continuing education [8].

5 Analysis of the Technical and Psychological Readiness of Students for Mobile Learning

We conducted the questionnaire of students of the Kherson State University to determine their technical and psychological readiness for use of mobile learning technologies. 160 students of the Department of Physics, Mathematics and Computer Science were questioned. The questionnaire showed that 99,4% of students have mobile telephones, 82% of which have smartphones, 33,1% have tablets, and 72,5% of students – notebooks.

Figure 1 shows the percentage of students' phones that are equipped with the following technical features: 1. Internet access; 2. 3G Internet access; 3. Play of MP3-files; 4. Recorder; 5. Calculator; 6. Access to Java-applications (games, e-books, etc.); 7. Camera.

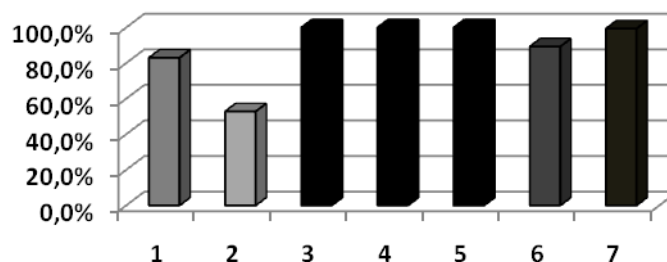


Fig. 1. The technical equipment of students' mobile phones

The technical equipment of mobile phones allows students more than 80% of them go online, use e-books, dictionaries and reference books, play audio files.

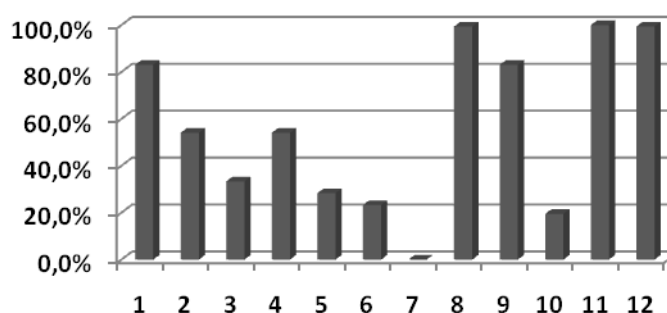


Fig. 2. Use of mobile phone applications by the students

Fig.2 shows which applications of mobile phones used by students, the numbers indicated by the following features: 1. standard browser to view web pages; 2. programs

for viewing of e-mail; 3. ICQ / QIP / JIM; 4. programs for reading e-books; 5. electronic dictionaries; 6. office software (Word analogs, Excel, etc.); 7. training programs; 8. music player; 9. video; 10. voice recorder; 11. calculator; 12. games.

Analyzing the answers, we can conclude that the most commonly used functions such as students are mobile phones, the use of browsers for browsing the web, calculator, audio and video players, receiving and sending e-mail, instant messaging. All other technical phone features are used is significantly lower. Low uses a voice recorder, office software for mobile phones.

Thus, students are not using their own mobile phones learning opportunities, despite the rather high level of technical equipment. But whether they are prepared psychologically to the use of mobile phones in the training? To answer this question, we have offered the students to answer the questions: «Would you like to download on your mobile phone all the necessary books and manuals for learning?»; «Would you like to use Internet resources for learning?»; "Would you like to use mobile applications at preparing for classes to form the necessary skills?».

Analysis of the responses showed that 82% of students would like to upload to mobile phone all the necessary books and manuals for learning, and 18% of the students answered this question in the negative. In answering the second question, opinions were divided as follows: 96% of students wanted to use Internet resources for learning, others answered in negative. Only 58% of students would like to use mobile applications at preparing for classes to form the necessary skills, others do not consider it necessary. At the same time the need for the use of mobile technologies is increasing among students, depending on the year of study. Students were asked to choose the discipline that would be convenient to study with the help of mobile technologies. More than 80% of students have chosen the foreign language. This choice is defined by the need to form foreign language skills for future career, and for this purpose mobile technologies offered: dictionaries, electronic reference books, listening to audio files, watching videos, communicating with native speakers, etc. Thus, the students have sufficient technical capabilities and high level of preparedness for the use of mobile technologies in education.

Obviously, for the use of the new possibilities of mobile learning in the educational process it requires organizational, research and methodological work on the implementation of modern policies, forms and methods of mobile learning in the educational process.

6 Mobile Learning Model of English during the Life

Mobile learning model of foreign languages, based on the basic functions of teaching and ICT opportunities; it includes a set of purposeful and orderly, and the sequence of actions of the teacher and the student through joint and individual study of structured training resources [3]. Thus mobile learning in vocational education should be based on the principle of interactive self-managed learning that will reduce their destructive impact of information and communication technologies on the social and cognitive learning activities [4].

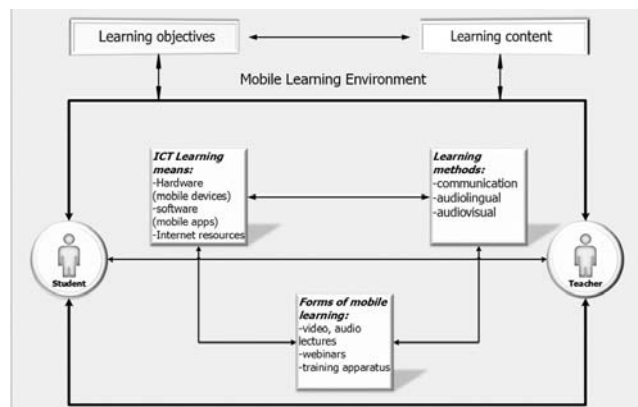


Fig. 3. Mobile Learning Model of English during the Life

7 Development of Mobile Application for English Articulation Phonetics

Mobile Assisted Language Learning (MALL) is technology of language learning using handheld mobile devices such as mobile phones (cell phones), MP3- and MP4-players, PDA, iPhone or iPad, and more. MALL is a subset of mobile learning (m-learning) and Computer Assisted Language Learning (CALL). Taking the results of the analysis of existing developments, it has been designed the mobile app English Sounds to study English phonetics. The application has a simple and user-friendly design and intuitive interface. The mobile app English Sounds can be used by pupils, students, teachers, university professors and for self-study. The main aim of the mobile application is open the opportunity for the students quickly and easily learn the material in the articulation phonetics. It gets teachers the new pedagogical opportunity in education. The program includes systematic data in the pronunciation of English sounds. It will not only improve existing skills of pronunciation, but also improve the communicative and professional competence of future specialists. Home page (Fig.4) consists of a matrix of elements - sounds. Information is presented structurally, so the division is made not only at vowels and consonants, but there is division on the subcategory (short/long, sonorous/ voiceless sound, etc.). It makes the navigation easier and search the sound faster.

The number of icons horizontally and their size was determined based on the maximum convenience, easy and accurately pressing the buttons on the devices of different screen resolutions. The matrix is organized in such way that all its elements (vertical) fit on the two screens.

After clicking on the sound icon at the main page, there is a redirect at the page with information about it (Fig. 4).

YouTube tab has view of the embedded video, which is already included the word examples and comments about the correct formulation of the lips and tongue in the pronunciation. There is need the Internet connection to view it (Wi-Fi, etc.).

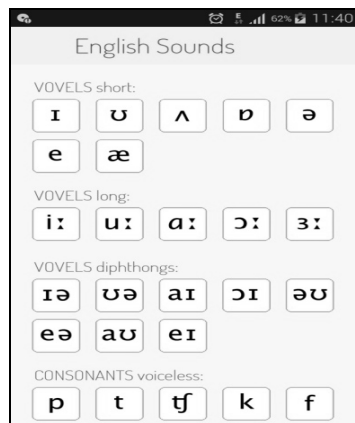


Fig. 4. Home page

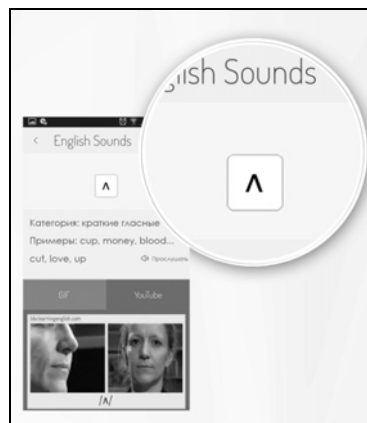


Fig. 5. Sound Page

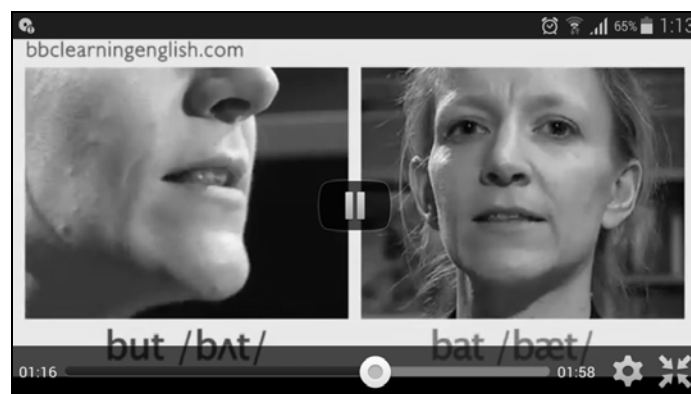


Fig. 6. View of the video

The application is implemented with mobile devices of OS Android version 4, 5, 6. In addition to Java-programming, there are tools for software development: system of automatically assembly Gradle, built on Apache Ant principles and Apache Maven; IDE Android Studio; Custom TextBox with Typeface of fonts ofDosis; YouTube API; component for playback SoundPool.

In particular, the mobile application was developed according to technical requirements that apply to educational software. Thus, according to the international standard ISO / IEC 9126 there is list of basic requirements:• Functionality;• Reliability;• Usability;• Efficiency;• Maintainability;• Portability.

Compliance with requirements at designing software allows creating a really high quality product. Its usage in the educational process will allow implementing an effective learning.

8 Conclusions and Outlook

Most of today's pupils and students are technically and psychologically prepared for the use of mobile technology in education. Mobile learning technologies can be a good addition to the traditional ones thanks to its advantages: availability, efficiency, individualization, flexibility [1]. The result of our research is design and implementation of the model using mobile learning technologies for learning English using the mobile application designed to learn practical English phonetics. The main purpose of it is to improve general communication and professional communication competence of specialists in the globalized society based on lifelong learning.

For effective use of the potential of mobile learning the mutual efforts on the part of educational leaders, software developers, trainers and teachers are required.

9 References

1. Bikov, V. Ju.: Models of Organizational Systems of Open Education. Atika, Kyiv (2008) (In Ukrainian)
2. Geddes, S.: Mobile Learning in the 21st Century: Benefit for Learners. online: <http://knowledgetree.flexiblelearning.net.au/edition06/download/geddes.pdf> (2004)
3. Gnedkova, O., Lyakutin, V.: Mobile Learning Model Designing in Distance Learning System "Kherson Virtual University". Information Technologies in Education, 24, 107--118 (2015) (In Ukrainian)
4. Fine, M.: Mobile Learning in the Educational Process: Foreign Experience. Modern Scientific Research and Innovation, No 1, online: <http://web.snauka.ru/issues/2015/01/43006> (2015)
5. Project of National Strategy for the Development of Education in Ukraine for 2012-2021. online: http://iitzo.gov.ua/files/proekt_rozvitku_osviti_2012_2021_.doc (2012)
6. MOTILL – Mobile technologies in lifelong learning: recommendations. online: <http://www.motill.eu>
7. Rashevskaya, N.: Mobile Software Training. Information Technologies and Learning Tools, 21(1), online: <http://journal.iitta.gov.ua/index.php/itlt/article/view/369> (2011)
8. Spivakovska, E., Osipova, N., Vinnik, M., Tarasich, Y.: Information Competence of University Students in Ukraine: Development Status and Prospects. In: Ermolayev, V., Mayr, H.C., Nikitchenko, M., Spivakovsky, A., Zholtkevych, G. (eds.) Information and Communication Technologies in Education, Research and Industrial Applications. CCIS, vol. 469, pp. 194-216. Springer, Heidelberg (2014)
9. UNESCO Policy Guidelines for Mobile Learning. online: <http://unesdoc.unesco.org/images/0021/002196/219641E.pdf>
10. International Conference on Mobile Learning. online: <http://mlearning-conf.org/>

Perspective Analysis of the Use of Electronic Social Networks in a Learning Environment

Olga Pinchuk¹

¹Institute of Information Technologies and Learning Tools of the National Academy of Educational Sciences of Ukraine, 9 M. Brlyns'koho st., 04060, Kyiv, Ukraine

Opinchu100@gmail.com

Abstract. A continuation of the retrospective analysis of research results of social networks and their use in modern education prospective analysis of using electronic social networks in the learning environment is carried out. Possible changes in teaching methodology when new objects - electronic social networking services, appear in the system of education are studied. Attention is paid to emphasis change from the network communication to productive discussion structure as well as from collaboration to cooperative learning methods for students. The problem of improving information and communication competence of all partners of learning process is described.

Keywords. Electronic social networks, informational and educational environment, learning

Key Terms. ICT, networks, learning

1 Introduction

A noticeable increase in the number of Internet social networks and global involving people in their use is one the forms of expression and development of communication network, that appears to be the basis of the information-oriented society.

Nowadays, electronic social networks (ESN) are the instruments by which a large number of users of the global network get more opportunities for communication, accumulation and transfer of knowledge, use of their own creativity in educational, scientific and socially important problem solving, generation ideas, etc.

It is our opinion that, the use of electronic social networking in education can have a synergistic effect, related, in particular, to the fact, that combined use of several mutually agreed pedagogical strategies appears to be more useful than implementation of any isolated one.

2 Related Work

In [1] we have analyzed a current state of educational research of ESN and formation of practical experience in their use. Retrospective analysis of social networks studies is made taking into account characteristics of their formation in terms of expanding communication space of globalized society; transformation of the old and the emergence of new practices of social interaction in different spheres of society. Based on the comparison of ESN potential and educational outcomes, as benchmarks of learning, we have defined pedagogically appropriate practical training tasks. Organizational forms of training, in which the use of ESN comes to be the most efficient, are also specified.

It necessary to point out that in recent years the formation of practical experience, activation of attempts of social networking services use in teaching practice has been observed in activities of teachers-bloggers, members of popular interest groups on Facebook (e.g. ICT - learning of Ukrainian teachers, Learn with Google, Educational technologists, PC in physical experiment) and professional communities on Google+. But this process is not accompanied by didactic and psycho-pedagogical substantiation.

Researchers, involving experience of foreign educators [2; 3], have identified a score of psychological, social and pedagogical arguments in favor of their application. Hereby we mention only a few which, in our opinion, are the most convincing.

1. ESN provide free use of a server for storing digital data.
2. ESN are popular among young people. This is a comfortable, convenient, positively disposed, familiar environment for the student.
3. With ESN as a teaching tool, students improve ability and create skills: properly and creatively use data to solve problems jointly create learning content, engage others and engage themselves in projects through various forms of communication (wikis, forums, polls, voting, comments, personal messages, chat, etc.), schedule (events, appointments, reminders of important dates) observe and coordinate their work.
4. A classroom discussion can be continued in the social network. Learning takes features of continuity.
5. A virtual learning group, set up in ESN, is always available if using mobile Internet.

3 Perspective Analysis in Pedagogical Research

Equally important than a retrospective analysis of the phenomena is a perspective analysis of educational innovations. Prospective analysis of pedagogical innovations is equally important. Results of this analysis are probabilistic in nature, but without ones, it is impossible to prove both projected growth of education in general and specific subjects teaching techniques in particular. It is necessary to identify the factors that will make a significant impact on results of studies with the use of ESN, as well as degree of this impact due to the establishment and compliance with certain psychological and pedagogical conditions. In prospective analysis we understand studies of educational system by parameters, which determine its future status. The peculiarity

of such analysis is in projection of the past and present state of the object in the future, focus on the selection of specific behavioral strategies of learning process with many alternatives and forming an integral conception of development of educational system. It is assumed that due to changes in state of real learning environment it is possible to correct a strategic plan of behavior of individuals of educational process. It is also important to identify the causes and factors that can negatively influence results of learning activities and elaboration of precautions. In a prospective analysis and conceptual foresight, as a rule, qualitative changes appear to be important aspects when quantitative ones play a supporting role. For prospective analysis of electronic social networks use in the learning environment we offer the following indicators: the intensity of communication, indicator of thematic communication, self-activity indicator, the amount of interpersonal interactions, the set of competencies or academic performance.

Much can be debated about the pros and cons of ESN use. However, their dynamic use has become an active part of modern life. We are of the opinion that it is more effectively to focus on overcoming significant conservatism of teachers, which takes a form of opponency to any innovations, and on specific recommendations as for use of ESN in training and education. An active resistance to change can be overcome if: there is awareness of the essence of ESN, there are no restrictions in accessibility of ESN resources, there is an understanding of the negative consequences of neglect of student's safety in network.

4 Research Methodology

Conceptual theses of national and foreign social philosophy, sociology and psychology, made while researching development and functioning of social media in modern society, have made up theoretical basis of our research. We also used the theoretical conclusions of scientists regarding informatization of education (V. Bykov, R. Gurevich, M. Zhaldak, A. Gurzhiy, etc.) and scientific and education principles of formation and use of information learning environments (V. Bykov, Yu. Zhuk, V. Olijnyk, Ye. Polat, etc.).

Several theoretical methods: analysis of research problems in scientific publications; study of the experience of using electronic social networks in the learning process, methods of comparative analysis, methods of mathematical statistics for processing quantitative characteristics of phenomena under research are used in the study.

Some conclusions of our study also found evidence in thesis work results of national and foreign educators and researchers (T. Arhipova, Ben Romdan Sami, H. Kuchakovska, N. Tverezovska, I. Vylegzhanina and others).

5 Results and Discussion

For 20 years the study of the structure and functioning of learning environments with the use of information and communication technologies (ICT) has been urgent. Experts of various disciplines: technical, physics and mathematics, pedagogical and

psychological, take part in them. Innovations in pedagogy can not be considered in isolation from changes in society in general.

Environmental factors are objective conditions that occur independently from the institution, student and teaching staff, influencing it.

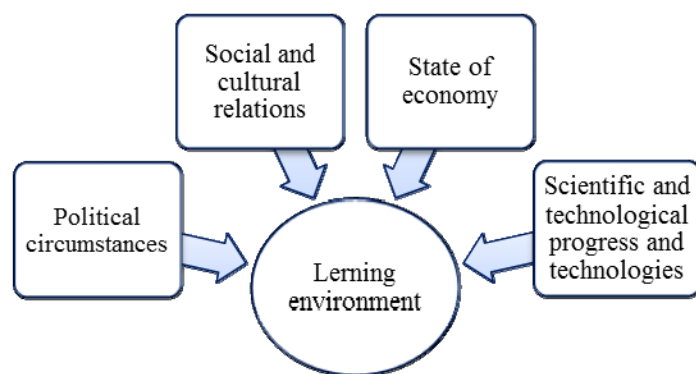


Fig. 1. Factors of indirect effect on the development of modern learning environment

Electronic social networks are widely used today in the life of many people, their popularity is growing rapidly, that is indicated by numerous statistical reports of Internet companies, including site «Alexa's digital marketing tools» (<http://www.alexa.com/>).

We conducted a survey (without use of computer devices and the Internet). The survey involved students and teachers both urban and rural schools (Table 1).

Table 1. A number of respondents according to the residence

A number of residents in community	Number of respondents
over 1 million	79
from 300 000 to 1 million	78
from 100 000 to 300 000	48
from 50 000 to 100 000	8
from 10 000 to 50 000	31
from 2 000 to 10 000	109
from 700 to 2 000	40
under 700	21

It emerged that among students aged 14 to 18 using social networks (such as Facebook, VKontakte and Odnoklassniki) there were 95.2% of respondents. So, in our opinion, not to try using this tool for the purpose of studies would be wrong. At the same time, provided that only 2.2% of teachers do not have access to the Internet, only 66.3% use social networks when working with students and / or parents. We have found that teachers in their professional activity are more inclined to use ESN to communicate with colleagues (61%), to advise students on the subject (41.5%), to get

information on additional resources on subject topics (39.8%), to communicate with students as class master (38.1%) and to put online homework (36.4%). Much less attention is paid to such opportunities as setting up an informal communication on the subjects content (23.7%) and designing of joint educational projects (21.2%).

By creating a "learning situations", which are focused on the use of ESN, a teacher can promote universal learning activities (personal, regulatory, cognitive, communicative) that will provide the development of skills of independent acquisition of new knowledge and skills, formation and development of critical thinking, development of communicative skills of students. Fig. 2 outlines a design process of learning situation in conditions of ESN use by a teacher.

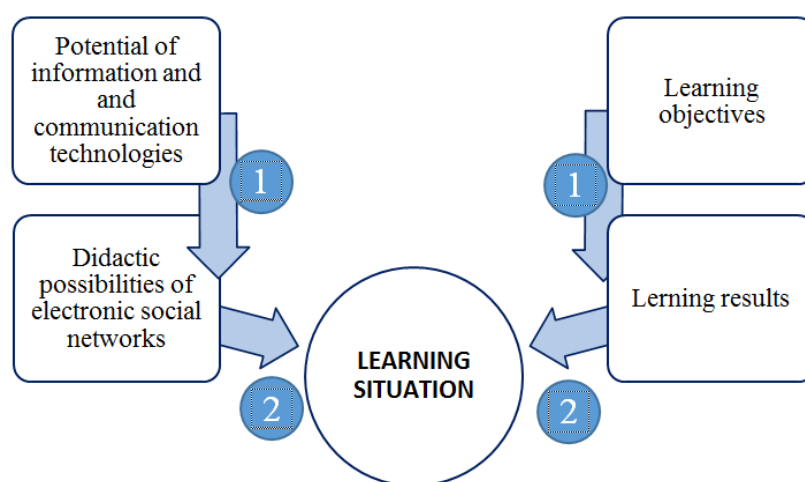


Fig. 2. A design process of learning situation (1 – specification; 2 – shaping)

With ESN services one can arrange a rap session in the form of discussion when development of students' subjectively new knowledge happens by expressing their own thoughts and by comparison of opponents' views on the issue. The situation where the emotional and intellectual stimulus encourages to active thinking is created. Leaders manages the process of discussion. Usually, that is a teacher. It is he, who has to define the topic of discussion, the main question (up to five), to determine the course of the discussion, choose basic training and other materials for participants. Keeping rules of ethical behavior is necessary condition, which everyone should agree.

One of the latest trends in business success is a process of joint activities in intellectual sphere of individuals or organizations to achieve common goals at which the exchange of knowledge, learning, and as they say, agreement occur. If collaborative learning (learning in cooperation, teamwork) is perceived as a certain theory of interaction in the learning process, determining its overall direction, the cooperative learning is a means of implementing collaboration.

Cooperative learning is focused on the use of quantitative methods that take into account achievement, that is learning outcomes. Cooperative learning is a structured, systematic training strategy in which small purposefully selected group of 3-5 stu-

dents work together on a common goal, creating a final product that has semantic specificity. The group inhomogeneous. The group consists of students of different levels of success, different abilities and skills. Thus, every student responsible individually for results of work, and a teacher acts as an advisor of group learning process. All students in the group are responsible for the work (there is no a leader). Cooperative learning is more focused than the system of collaborative learning activities, and more centered on a teacher. Cooperative work in a group includes both successful promotion in the cognitive process of each student and breaking ice, as well as maintaining good working relationships between team members. A number of social skills: ability to listen, respect an opinion of an opponent, expressing criticism to correct errors is necessary for successful cooperative work.

Collaborative learning includes such formats as group projects, joint development and so on. In the context of e-learning, collaborative learning gained a new interpretation (computer-supported collaborative learning). First of all, it is connected with the use of Web 2.0 services, social networking apps, that support mutual activity, virtual communities for the purpose of learning.

The methodological approaches, mentioned above, require "new literacy" [4] of the teacher, a higher level of information and communication competence (IC-competence). On the other hand, active use of ESN and other means of ICT in educational practice leads to gradual development of IC-competence. The teacher possesses methods and styles of information learning activity, that is adequate to the situation, that occur in the process of educational information environment development; forms skills of effective network interaction of educational teams in a global information and education space; develops competencies, needed to shape effective learning programs in open information and educational environment; learns modern tools of conceptual foresight, analytics and diagnostics of educational results.

— Formation of subjectively new knowledge

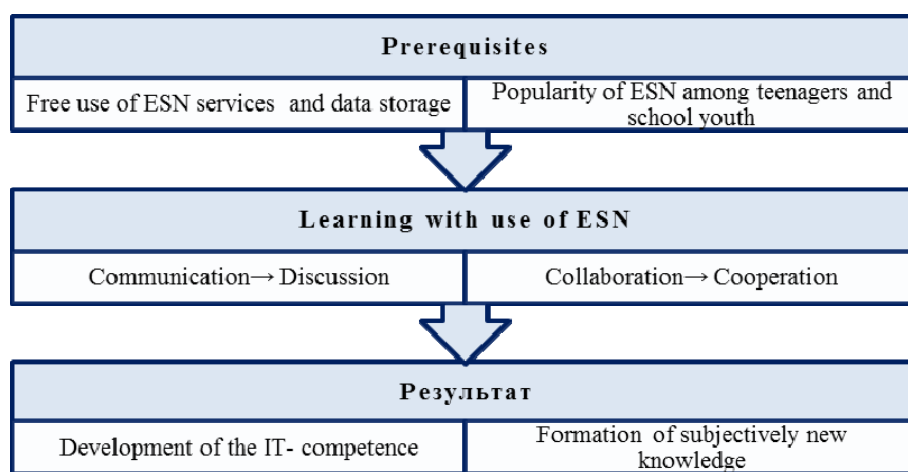


Fig. 3. Particular issues associated with ESN use learning

6 Concluding Remarks and Future Work

In our view, the readiness of a teacher to changes that are focused on active and effective use of ESN in his T and E activity; capacity for effective cooperation and experience exchange with teams of educators in their professional field; knowledge about new teaching methods that provide flexibility and adequacy of the implementation of new ICT in the learning process could not emerge spontaneously, without purposeful pedagogical influence. Creation of appropriate learning courses for teachers is actual problem that requires a solution. We offer to create guidelines for teachers around the following ideas. 1. The effectiveness of group interaction. 2. A need for self-development as a component of cognitive creativity of senior students. 3. A research of features of instructional design of information and educational learning environment. 4. Use of electronic social networks towards the problems solution of various educational formats synchronization and in order to build an integrated trajectory of individual learning.

References

1. Pinchuk, O. P.: Historical and analytical review of social networking technologies and prospects of their use in training. In: Information Technologies and Learning Tools, 4(48), 14--34, (2015)
2. The Facebook Classroom: 25 Facebook Apps That Are Perfect for Online Education. <http://www.collegedegree.com/library/college-life/15-facebook-apps-perfect-for-online-education> (2015)
3. 50 Reasons to Invite Facebook Into Your Classroom. <http://www.onlinecollege.org/2011/07/18/50-reasons-to-invite-facebook-into-your-classroom> (2015)
4. Ovcharuk, O. V., Soroko, N. V.: Using of the teachers' information and communication competency assessment instruments in European countries. Information Technologies and Learning, 2(52), 133--143 (2016)

The Introduction of the Competence-based Approach in Educational Process of Training of Skippers

Tatyana Zaytseva

Kherson State Marine Academy, 20, Ushakova St., 73000, Kherson, Ukraine

sunny@ksu.ks.ua

Abstract. The competence-based professional education is the objective phenomenon in education brought to life by social and economic, political and educational and pedagogical factors. One's competitiveness at the modern job market usually depends on ability to operate new technologies, ability to continuous self-education and fast adaptation to various working conditions. Modern job market imposes the whole layer of new requirements on an employee which are taken into account on inadequate level or aren't taken into account at all in the training syllabus for specialist degree in different subject areas and in maritime education in particular. The main idea of the competence-based approach is that education has to provide not isolated knowledge and skills but to develop students' ability and readiness for future professional activity in various social and working conditions. The competence-based education, which has been introduced in many countries, is nevertheless a new way of organizing educational process in Ukrainian higher educational institutions. Because maritime education is of international nature, the introduction of the competence-based approach into cadets training program is of immediate importance nowadays. Search for effective forms of the educational process organization, which will allow combining the academic, practical and simulator trainings of specialists in this area, is an integral part of teacher's work at higher education institution.

Keywords. Competency, competence-based approach, competence-based education

Key Terms. Teaching Process, Information Communication Technology

1 The general problem statement and its actuality

The decree of President of Ukraine "On measures for priority development of education in Ukraine" dated by 30 September 2010 № 926 determined the number of measures for introduction the regulations in Ukraine, aimed at coordination the national system of education quality assurance with the general European system [4].

Based on modern society requirements, the education quality assurance should be based on the formation such competencies and skills in future specialists that enable to use practically knowledge and skills for the benefit of all Europe.

In the National Standard System of Higher Education the requirements and qualifications are clearly stated, the list of socially and professionally important knowledge, skills and competencies is provided, which are required from the graduate of high institution not only by the national labor market, but also by the European Community.

The National System of Qualifications should be the basis for the introduction of competence-based approach in higher education, including its components – the National and Branch qualifications limits. The National qualifications limits is implemented in order to:

- introduction of European standards and principles of education quality assurance with the requirements of the labor market competencies to professionals;
- ensure harmonization of standards of legislation in the field of education and social, labor relations;
- promote national and international recognition of qualifications acquired in Ukraine;
- establish the effective cooperation of sphere of educational services and the labor market. [3]

Conformity of the quality of graduates' training of high school with the requirements of branch standard of higher education is determined by social and personal (SPC), general scientific (GSC), instrumental (IC) and professional competencies.

The idea of competence-based approach in teaching was originated in the early 80th of the last century, the article by W.de Landshyeer "The concept of "minimum competency" was published in the journal "Perspectives. Question of Education"[8]. Initially, it was not the approach, it was the professional competence of the person as the aim and the result of education.

The competence in a broad sense is understood as "in-depth knowledge of the subject or mastered skill." In due course, there was expansion of the size and content of the concept. Since the end of the last century the scientists began to speak about competence-based approach in education (V.A.Kalney, A.M.Novikov, V.V. Serikov, S.E.Shishov, B.D.Elkonin etc.). [6, 7 10]. Today there are different approaches for understanding the core competencies. Some authors emphasize the personal properties, others - on the knowledge and skills that can be transferred to various conditions.

We will proceed from the fact that competence is the capability and readiness for the implementation of certain actions or functions and competence-based approach in education - a target orientation of educational process on the formation of competencies defined by branch standard, either socio-personal or professional.

The competency-oriented professional education - is an objective phenomenon in education, inspired by socio-economic, political, educational and pedagogical preconditions. At first, it is the reaction of professional education on the changes in the socio-economic sphere, the processes that have emerged with market economies. The

market puts forward the whole layer of demands to the modern specialist that are insufficiently included or not included in the programs of specialists training.

These new requirements are not hard connected with some discipline, they are interdisciplinary and universal. Its formation requires not only new subject matter, but other educational technologies. The competence-based approach allows to:

- coordinate the purpose of study, which the teacher puts with students' goals;
- unload students not by reducing the content, and by improving the part of individual self-education;
- prepare students for conscious and responsible learning, the necessity of constant self-education;
- and most importantly, in our view, provide the labor market by competitive specialists.

The object of the study is to determine the conditions and methods to formation of information competence for students that will promote assimilation of its professional skills.

Based on the objectives of the study were identified the ways of its solution.

The first phase of the study included the following types of work: bring the work programs in compliance with the STCW Code, IMO model course 7.02 and 7.04 and establishing the nature of interdisciplinary connections. In the second stage was done processing teaching methods of disciplines, also reviewed the content requirements for laboratory work in accordance with the competence approach and to implement interdisciplinary courses.

The third stage was dedicated to the order of evaluation procedure results obtained knowledge of students according to the content learned discipline, executed complex laboratory work to ensure appropriate skills

2 Solving basic problems

In our opinion, the introduction of competence-based approach in specialists training system is most important in marine industry, which is international in its essence. Specialists of maritime industry should comply with International Maritime Organization (IMO) irrespective of the country in which they got the education. The list of these requirements is in the International Standard of Training, Certification and Watchkeeping of Seafarers (STCW). The competence-based approach to multilevel training of marine specialists is the base of this document.

For example, we consider the STCW Code for marine electricians, there are two levels of training - electrician and electrical engineer. The document provides the basic functions for each level of specialty (e.g. at operation or support level) and competences are marked for each function, which should have the specialist. The higher level of specialist training, the greater the number of competences he should have. There is an example of part of STCW Code (Table 1).

Table 1. Specifications competencies specialty for marine electricians

№	Competence	National code competence	Requirements convention STCW	Level
Professional competence				
Specialized and professional competence				
1.	Safe use of electrical equipment	SPC-01	A-III/6	At support level
The general science competence				
1.	Basic knowledge of the fundamental branches of mathematics , to the extent necessary to own mathematical tools relevant industry knowledge, ability to use mathematical methods in their chosen profession.	GSC-2		At the level of exploitation
2.	Basic knowledge in computer science and information technologies; use software skills and skills in computer networks , the ability to create databases and use online resources.	GSC -3		At the level of exploitation

Today, the labor market dictates to education system what level of specialists' knowledge it needs. And competence-based approach is an attempt to bring in line the professional education and the requirements of employers. Competence-based approach suggests that the significant results of education are recognized outside the education system.

From the 2014-2015 school year in the Kherson State Maritime Academy the experimental research work in the integration of educational process on the basis of competence-based approach to the state education policy of Ukraine was started. [3]

All training departments of the academy are actively involved in the project. The teachers of department of information technologies, computer systems and networks have developed the new working program of courses in which the disciplines are considered as the mean of mastering the certain competencies within these disciplines.

Competence-based approach fixes and sets up the subordination of knowledge and skill sets. An important role in this process is computer science as a science and a subject, as competencies, which are formed during the study of the subject can be transferred to the study of other objects to create the integrated information space of cadets' knowledge.

To show the interdisciplinary connections the teachers of the department acquainted not only with the work programs of other disciplines, but also discussed with the leadership of faculties and departments the level of information culture should have the cadet for successful studying the special disciplines. As a result of this work in the discipline "Information Technologies", that studied on first year, new themes were added and the content of course "Computer Science for Skippers" has been revised and has acquired the applied nature.

The content of the course "Information Technologies" has two components:

1. Theoretical Computer Science, which is currently one of the fundamental areas of scientific knowledge, it forms the system-information approach to the analysis of the environment.
2. Information Technologies, which represent the methods and means of obtaining, transformation, transmission, storage and use of information. This component has extremely the important practical importance; it takes the social order of society to prepare the future professionals of the field of marine branch in the information world society.

The main purpose of the discipline "Information Technologies" is developing the information and communication competencies. The information and communication competence can be seen as a comprehensive ability to search, select the information, analyze, organize, represent, pass it; simulate and design objects and processes, implement projects, including in the area of individual and group activities. [5]

Most courses in computer science can be realized by means of competence-base approach. After analyzing the content of the courses we have identified the following competencies and described them from the point of view of the subject and with the aim to expand on other disciplines (Table 2).

Table 2. The list of competencies

The competence		Projection common objective competencies to curriculum subject
Key	Generalized (common objective)	
	Work with different media.	Skills of work with the applied software, multimedia reference books, electronic books, Internet resources.
	Use of information and telecommunication technologies.	The application of information and telecommunication technologies for solving a wide class of

The competence		Projection common objective competencies to curriculum subject
Key	Generalized (common objective)	
Information Competence		academic and applied problems.
	Job information (search, converting, storing, systematization, analysis and selection of information).	Information search in directories, search engines; conversion information (with graphics in text, with analog-digital); database design; work with different types of sorting; filtering, structuring the file system;
	The wording of the purpose, plan of action, awareness of the presence of certain requirements to the product of its activity.	Formulation of the goal in the study of certain subjects subject when creating the project, the choice of the theme report; writing the plan development and implementation of the project; understanding the software requirements, framework operation of databases; the formulation and description of its limitations.
	Possession of stylistic techniques	Create text documents, using the rules of presenting information in a presentation
Communicative competence	Dialogue the "man-machine"	Familiarity with the principles of constructing different types of interface (dialog boxes, menus and toolbars)
	Group work	Work on projects together, and in the group, performing different roles in a group interact in networks.
	Tolerance	Work in a network society, the use of telecommunication resources
Educational-cognitive competence	The ability to propose hypotheses, ask questions to the observed facts and phenomena, evaluate the initial data and planned result	Owning the technology problem solving with the help of a computer, computer modeling, formalization, numerous methods of problem solving, conducting computer experiment

The competence		Projection common objective competencies to curriculum subject
Key	Generalized (common objective)	
	The ability to issue the results of its activities, submit it at the present level	Knowledge of ISO requirements for registration statements, build charts and graphs, create presentations
General cultural competence	Understanding the place of science in the system of other sciences	Awareness of the availability of certain software requirements, analysis of the advantages and disadvantages of information technology
	The ethics of labor and social relations	License for software, information security, legal responsibility for violation of legislation.
	Creation of conditions for self-realization, self-knowledge	The computer as a tool self-reflection (use test simulators), create your own information space), publishing, receiving authority in network society

The main task of the competence-base approach is to find out and to include in the educational trajectory that, without what the professional training can be happened, what is necessary and sufficient to know and be able to do the future professionals in the maritime industry. For this purpose we adjusted the content of educational material of disciplines by means of the implementation the tasks of following types:

- Tasks containing a large amount of text data and information presented in tables, charts, graphs, drawings, diagrams (transforming of information, work with different types of information).
- Tasks, in which it isn't clear to which area of knowledge it is necessary address to determine the method of action or the information (selection of the necessary software, the use of competencies in practice).
- Tasks, with a large number of tasks of different topics and different formats that require different algorithms for solving, forms of records of the answer.
- Tasks to optimize solutions.

Preferably it is necessary to use not formalized, but meaningful statement of the problem, so that the most difficult but important for getting the experience for solving problems, stage of formalization will be completed by the cadet.

For example, at learning the optimization, classic transportation tasks and use the capabilities of MS Excel for solving the tasks, formulas of specific functions and constraints are offered. But it is much more effective to give the applied task, for the solution of which the student should reasonable formulate the objectives, describe the

task and its limits in terms of selected technologies, create the information model of the task.

To generate activity algorithm for solving the tasks it is necessary to train cadets of the implementation of sequence of the following stages:

- Formation of the needs and intentions of a particular activity;
- Selection of required software and the way of activity;
- Planning of activities, description of restrictions;
- Implementation of actions;
- Analysis of the results.

Competence-base approach focuses on the use of knowledge and skills in the so-called extra-curricular, life situations, including in solving practical professional problems. The basis for the formation of competencies in students of 1 course is their experience, gained earlier in life and learning situations. But we must take into account the different levels of school knowledge in first-year student at updating knowledge in training lessons.

Therefore, the offering the individual tasks, tasks of different complexity, participation of students in project activities is the necessary ways to individualization of the learning process and encourage them to creative self-dependent scientific and practical work.

During the laboratory practical course "Information Technologies" in the first semester, the tasks of different difficulty levels can be presented and cadets can choose the tasks. This method will allow the cadets objectively determine their level of subject competence, well-prepared students will be able to realize the high potential, and students who has the poor basic training, avoid getting psychological trauma and motivate them to deepen knowledge. At the initial stage of training, the level of conformity cadets with the requirements prescribed for them is presented and it allows the teacher to adjust the methodical forms of teaching classes and forms of control according the results.

If we analyze the content of the discipline "Computer Science for Skippers" studied by the cadets of 2nd year of training, during the laboratory work, the subject competencies are formed that associated with the use of means of data processing presented in tabular form; using the computing techniques, mathematical and information modeling, business graphics can be purchased and secured in the study of this discipline. The content of the subject has interdisciplinary connections with many subjects, particularly with the disciplines: Ship computers and computer networks, Navigation information systems, Descriptive geometry and engineering graphics; Theory and construction of vessel, and others.

The basis for the formation of subject competencies is put at first year during studying the discipline Information Technologies. The study of functionality of spreadsheet (for example, Microsoft Excel) is intended, to explore the possibility of systematic presentation of information in tabular form and performance calculation works of any complexity, to demonstrate practical importance of informatics and implementation of interdisciplinary connections.

Further, the subjects of some laboratory works of the discipline "Computer Science for Skippers" are given:

Task 1. Calculation of the latitude variation, the longitude variation and the coordinates of points of departure and arrival of vessel

Task 2. Calculation of directions as to the geographic meridian and diametric plane of the vessel.

Task 3. Calculation of the arc's parameter of a great circle for mapping using MS Excel.

Task 4. Calculation and construction of static stability diagram using MS Excel (Figure 1).

Formulation of the problem. With Excel spreadsheet to perform calculations static stability shoulder and construct a diagram of static stability and dynamic stability diagram, constructed in the same coordinate axes.

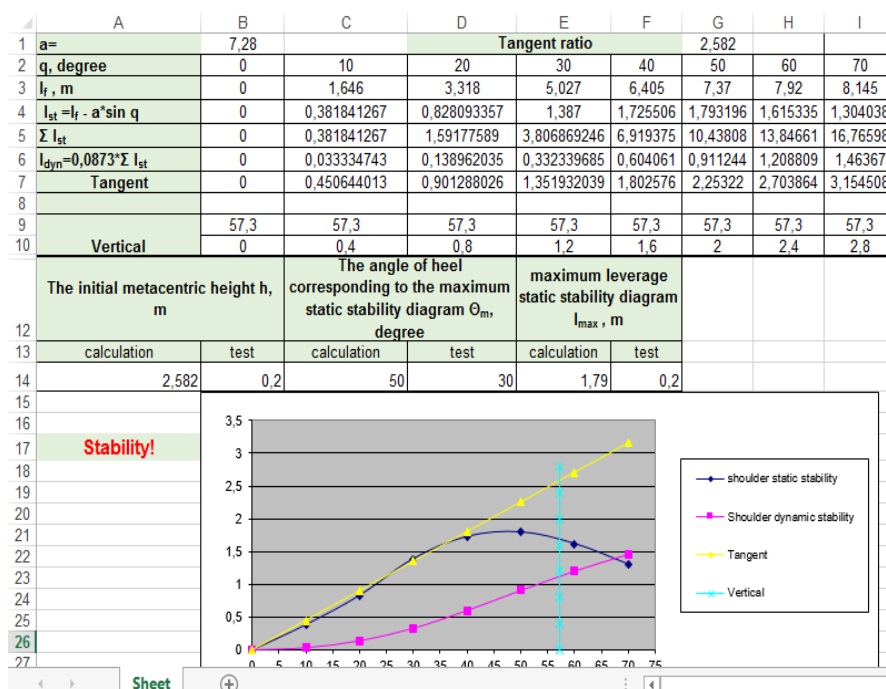


Fig. 1. Example of solving the cadets with MS Excel

Task 5. Calculation the course at navigating on LDC (large diameter circle).

Task 6. Calculation and construction of curve of total inertial error of gyrocompass, arises in resulting maneuvering.

Task 7. Calculation of coordinates of the vessel by the direct analytical method (by two measured heights) with the help of MS Excel.

Task 8. Calculation the assessment of the accuracy of vessel place by three equilateral bearings using MS Excel.

Task 9. Work with hydro meteorological data in MS Excel. Calculating the actual wind.

Task 10. Conducting the navigation calculations at planning of vessel voyage using MS Excel (Fig. 2).

Formulation of the problem. Given the coordinates of waypoints, you need to calculate the exchange rate to move in the next point, the length of each section of the route. It is also necessary to determine the total length of the route, as well as build a circuit-moving vessel.

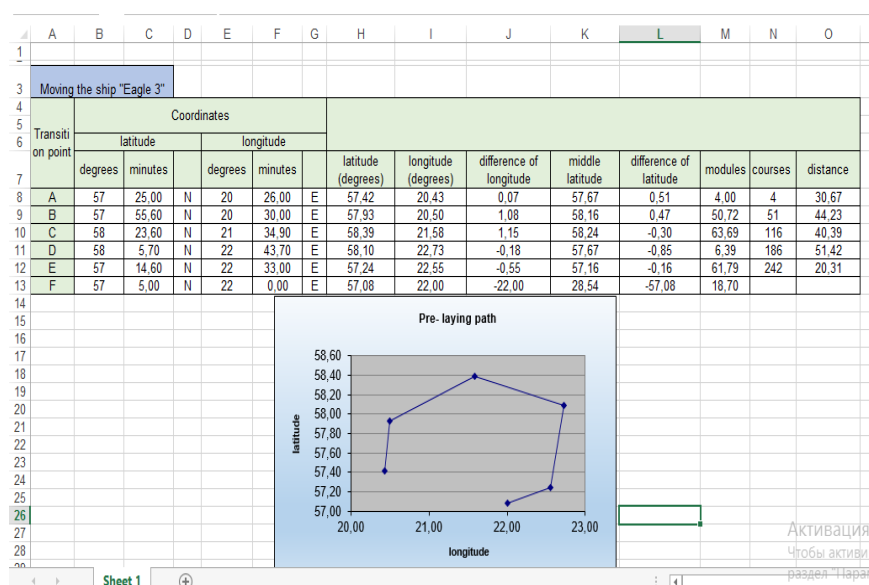


Fig. 2. Example of solving the cadets with MS Excel

As it turned out, the cadets have difficulties in solving meaningful problems, but solving them, they are acquiring invaluable experience of professional competence.

It should be changed not only the working programs of disciplines, content and methods of education, but also it should be realized that personality of the teacher, who uses competence-base approach, should meet certain requirements [9]:

- Set the goals and estimate the level of its achievement together with the cadets.
- Evaluate the achievements of students not only by the mark but by the meaningful description.
- Connect the investigating material with professional direction, everyday life and cadets' interests.
- Plan the lessons with all variety of forms and methods of educational work.
- Strengthen the knowledge and skills in educational and extracurricular practice.
- The ability personally to orient in the situation in the labor market.

- Take into consideration the cadets' ideas.
- Successfully solve own problems.

3 Conclusions and directions for further research

Competence-base approach makes the cadet a major participant of the educational process with his individual goals and objectives. This approach allows directing the educational activities to involve students in active, conscious activity, the development of information, communicative, educational, cognitive competencies and the development of the personal cadet's potential, forming self-appraisal, self-control and teacher's reflection that allows to achieve the better results in education.

It is the complex issue to connect the learning results and competencies, to which it should be paid much attention. The orientation on the results of education is now the urgent issue for Ukrainian high school, and it requires the integration of academic and professional education, recognition of the qualifications, getting in the process of higher education, the development of the education during whole life. Society should get used to the situation when the description of the results of the education will be provided in the language of competencies.

As a result, we note that the essence of the new paradigm of education can be characterized by the following factors:

1. The displacement of the main emphasis from mastering the large amounts of information to master the methods of continuous acquisition of new knowledge and the ability to learn independently;
2. The mastering of skills to work with any information, with mixed, contradictory data, forming the skills of independent, critical way of thinking;
3. The gradual change of the traditional principle "form of knowledge and skills" to the principle "to form professional competence."

Today there are already the first results of the implementation of competence-base approach in the educational process. One of the positive issues is the changes that have occurred in the redistribution of motivational aspects of cognitive and training activity of cadets.

During two academic years we observed the changes in the motivational component of the educational process, namely, we were interesting how the content of the course "Computer Science for skippers", the strengthening of applied direction of the discipline and the interdisciplinary connections of the courses affect on the interest of students, increase their motivation level. The following types of motives were selected:

- motives for avoiding troubles (traditionally they had the higher percentage of dominance);
- motives of the content of educational activity;
- motives of attitude to the learning process.

Questionnaires of the cadets showed what types of motives are dominant. The motives of the content of educational activity were put at the first place.

The data are given in the table (see. Table. 3) and represented in the diagram (Fig. 3).

Table 3. Motives of educational activity

Study year	Motives avoid trouble	Motives of educational content	Motives related to learning
2014-2015	44%	36%	20%
2015-2016	40%	42%	18%

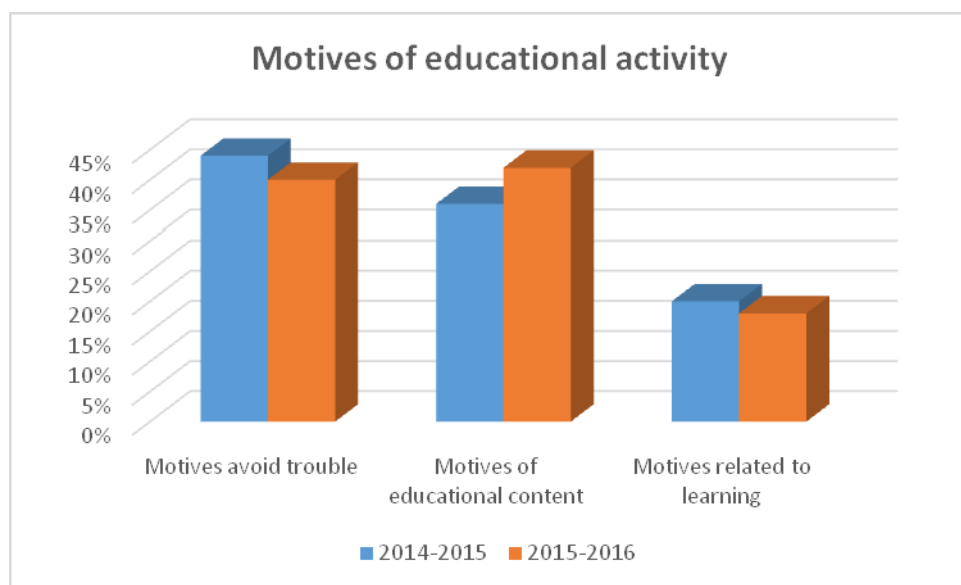


Fig. 3. Diagram of distribution motives of educational activity

In terms of implementation of competence-based approach into practice the training of specialists maritime industry, information literacy is the basis of formation of professional competence of students, promotes the informed use of information technologies for solving applied problems in finding and making decisions in predicting and analyzing the results of solving problems.

At this stage in the KSMA the experiment in the introduction of competence-base approach in the educational process is conducting.

Today we have the task to analyze the results of the experiment to further adjustment of educational material and forms of learning.

References

1. Complex of normative documents for the system development standards of a higher education. Annex 1 to order of Ministry of education of Ukraine dated 31.07.1998, No 285, with changes and additions. Bulletin "Higher education", 10 (2003)
2. The national classifier of Ukraine: Classifier of professions DK 003-10. Kiev (2010)
3. The Order of the Ministry of education and science of Ukraine № 1148 from 7.10.2014. Bulletin, 2 (2014)
4. The order of the President of Ukraine "on measures for the priority development of education in Ukraine", No. 926 (2010)
5. Baidenko, V. I.: Competence approach to the design of state educational standards of higher professional education (methodological and methodical questions): Methodical grant. Research center of quality problems of training of specialists, Moscow (2005)
6. Winter, I. A.: Key competences – new paradigm of result of education. Higher education today, 5, 14--20 (2003)
7. Competence in education: experience design, <http://projects.ict.usc.edu/dlxxi/materials>
8. Landseer, V.: The concept of "minimal competence. Prospects. Education. 1, 28--34 (1988)
9. Lebedev, O. E.: Competence approach in education. School technologies, 5, 19-23 (2004)
10. What is the competence approach in modern education? <http://festival.1september.ru/articles>

Author Index

A

Abdelhady, Mohamed	168
Alexeyev, Alexander	137
Atamanyuk, Igor	168, 184, 292
Atamanyuk, Igor P.	458
Azimi, Vahid	168

B

Balkova, Yulia	57
Barto, Taylor	168
Bespalov, Yuri	57
Bezobrazov, Sergei	534
Bisikalo, Oleg	4
Bogach, Ilona	4
Bulba, Yevgeniy	308
Burov, Oleksandr	608
Bykovyy, Pavlo	384

C

Chernenko, Irina	215
Chernyshov, Mykhaylo	384
Chornomaz, Bogdan	13
Corcho, Oscar	3

D

Davies, John	1
de Keijzer, Ander	273
Doroshenko, Anatoliy	111
Dorozhinsky, Vladimir	469
Drozd, Alex	322
Drozd, Miroslav	322
Dzikovska, Yuliia	148

E

Ebeigbe, Donald	168
-----------------	-----

F

Fakoorian, Seyed Abolfazl	168
Farzad, Babak	33
Fedasyuk, Dmytro	148
Fesenko, Herman	384

G

Gamzayev, Rustam	41
Glazunova, Olena	157
Gnedkova, Olga	620, 672
Golovko, Vladimir	534

H

Hladchenko, Lidiia	477
Hots, Nataliya	148
Hovorushchenko, Tetiana	87, 332

I

Ivanchenko, Oleg	349
Ivanov, Vitalii	73
Ivasiuk, Aleksandr	308

K

Karpenko, Oksana	284
Khademi, Gholamreza	168
Khadikov, Arsen	469
Kharchenko, Vyacheslav	349, 384, 414, 427
Kobets, Vitaliy	485, 502, 518
Kochan, Volodymyr	384
Kolomiychenko, Vadim	57
Komar, Myroslav	534
Kondratenko, Yuriy	168, 184, 292
Kondratenko, Yuriy P.	458
Konovalova, Nataliya	137
Korol, Elena	137
Kostolny, Jozef	441
Koval, Vasyl	384
Kozlovskiy, Evgen	215
Krasiy, Andriy	87
Kravets, Tetyana	550
Kravtsov, Hennadiy	620
Krupnyk, Oleksandr	284
Kuchma, Olesya	257
Kushnir, Nataliya	200
Kuts, Viktor	148
Kuzminska, Olena	632
Kuznietsov, Mykola	322
Kvassay, Miroslav	361, 441

L

Levashenko, Vitaly	361, 441
Liashenko, Olena	550
Litvinenko, Olena	200
Liubchenko, Vira	640
Lozova, Kateryna	137
Lvov, Michael	215
Lyaletski, Alexander	25
Lyaletsky, Alexandre	25
Lysenko, Sergii	375
Lytvynova, Svitlana	648

M	
Mazorchuk, Mariia	632
Melnyk, Oksana	648
Moroz, Olha	558
Morze, Nataliia	231
Mulyak, Oleksandr	399
Mykyychuk, Mykola	148
N	
Nicheporuk, Andrii	375
Nosenko, Yuliya	656
Nosov, Konstantin	57
O	
Oleksiuk, Vasyl	656
Osipova, Nataliya	200, 672
P	
Paientko, Tetiana	566
Panchenko, Taras	119
Paskevich, Andrei	25
Pichugina, Oksana	33
Pinchuck, Olga	680
Poltoratskiy, Maksym	103, 485
Pomorova, Oksana	332, 375
Ponochovny, Yuriy	308
R	
Rak, Larisa	57
Reisig, Wolfgang	2
Rong, Yiming	73
Roshanineshat, Arash	168
Roshchupkin, Oleksiy	384
Roth, Hubert	384
Rukkas, Kirill	13
S	
Sachenko, Anatoliy	384, 534
Savenko, Oleg	375
Shebanin, Vyacheslav	292
Shishko, Ludmila	215
Shyshkina, Mariya	241, 656
Simon, Dan	168, 184
Sirenko, Natalia N.	458
Sklyar, Vladimir	129, 308
Spivakovska, Evgeniya	257
Spivakovsky, Aleksandr	103
Stincic Clarke, Sandra	1
T	
Tarasich, Yulia	103, 257

Tkachuk, Mykola	41
Troino, Kseniia	13
Tulika, Eugene	111
U	
Ushakov, Denis	672
Ushakova, Iryna	575
V	
Valko, Nataliya	200
van der Stappen, Esther	273
van Moergestel, Leo	273
Varchenko-Trotsenko, Liliia	231
Vashchenko, Svitlana	73
Vekshyn, Oleksii	41
Verezubova, Tatsiana	566
Vinnyk, Maksym	103, 257
Volochiy, Bogdan	399
Voloshyna, Tetyana	157
Vysotskaya, Elena	57
W	
Weissblut, Alexander	502
Y	
Yakovyna, Vitaliy	399
Yanovskaya, Olga	414
Yanovsky, Max	414
Yaremchuk, Svitlana	427
Yashyna, Kseniia	284
Yatsenko, Valeria	518
Yatsenko, Viktoria	591
Yatskiv, Vasyl	384
Z	
Zaitseva, Elena	361, 441
Zaytseva, Tatyana	687
Zhereb, Kostiantyn	111
Zholtkevych, Grygoriy	57
Zholtkevych, Nina	469