



# INFORMATION TECHNOLOGY & INTERACTIONS (SATELLITE)

04 December, 2020  
Kyiv, Ukraine

ISBN 978-966-2399-61-5



9 789662 399615

TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV  
(FACULTY OF INFORMATION TECHNOLOGY,  
FACULTY OF COMPUTER SCIENCE AND CYBERNETICS)  
NATIONAL TECHNICAL UNIVERSITY OF UKRAINE “IGOR SIKORSKY KYIV  
POLYTECHNIC INSTITUTE”  
VIKTOR GLUSHKOV INSTITUTE OF CYBERNETICS OF THE NAS OF UKRAINE  
INSTITUTE OF INFORMATION TECHNOLOGY AND LEARNING TOOLS OF THE NAES OF  
UKRAINE  
INSTITUTE OF INFORMATION REGISTRATION PROBLEMS OF THE NAS OF UKRAINE  
INSTITUTE OF SOFTWARE SYSTEMS OF THE NAS OF UKRAINE  
THE COUNCIL OF YOUNG SCIENTISTS OF THE FACULTY OF COMPUTER SCIENCE AND  
CYBERNETICS AND THE FACULTY OF INFORMATION TECHNOLOGY OF  
TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV

## **VII INTERNATIONAL CONFERENCE**

### **Information Technology and Interactions (Satellite)**

04 December, 2020

Conference Proceedings



UDC 004(082)

I-60

*Volume editor:* Vitaliy Snytyuk, Dr.Sc., Prof.

*Program Committee:* Anatoly Anisimov (co-chair), Vitaliy Snytyuk (co-chair), Aldrich Chris, Andreas Pester, Frederic Mallet, Hiroshi Tanaka, Iurii Krak, Karsten Henke, Mykola Nikitchenko, Oleg Chertov, Oleksandr Marchenko, Sándor Bozóki, Vitaliy Tsyganok, Vladimir Vovk.

*Organizing Committee:* Anatolii Pashko, Andrii Biloshchytskyi, Kateryna Kolesnikova, Mykola Nikitchenko, Nataliia Lukova-Chuiko, Oleh Ilarionov, Oleksandr Marchenko, Oleksii Bychkov, Serhii Toliupa, Taras Panchenko, Valentyna Pleskach, Viktor Morozov, Volodymyr Zaslavsky, Iurii Krak, Yurii Samokhvalov, Yurii Kravchenko.

*Conference Secretary:* Hryhorii Hnatiienko.

I-60 Information Technology and Interactions (Satellite): Conference Proceedings, December 04, 2020, Kyiv, Ukraine / Taras Shevchenko National University of Kyiv and [etc]; Vitaliy Snytyuk (Editor). – Kyiv: Stylos, 2020. – 388 p.

ISBN 978-966-2399-61-5

This book includes abstracts of the 7th International Conference "Information Technology and Interactions (Satellite) – 2020". Philosophical, theoretical and applied aspects which describe the results, problems and prospects of the creation and use of intelligent computing methods and creating of information systems and technology on their basis are reviewing.

Main tracks of the conference are: Artificial Intelligence Technologies, Cyberspace Protection Technologies, Data Analytics, Digital Project Management Technologies, E-commerce, E-government and E-learning Technologies, Mathematical Foundations of Information Technology, Network and Internet Technologies.

UDC 004(082)

ISBN 978-966-2399-61-5

© Authors of abstracts, 2020

© Стилос, видання

## CONTENTS

### **ARTIFICIAL INTELLIGENCE TECHNOLOGIES**

<i>Abduramanov Z., Seidametova Z., Valiieva N.</i> Color Recognition Deep Learning Model	13
<i>Antonevych M., Didyk A., Snytyuk V.</i> Choice of Better Parameters for Method of Deformed Stars in N-Dimensional Case	17
<i>Astakhov A., Ilarionov O.</i> Analysis of Speech Emotion Recognition Methods	21
<i>Bondar T., Hnatienko H.</i> Video Registration and Face Recognition Technology on Stream Video	25
<i>Derevianchenko O., Nikolaiev A.</i> Implementation of Artificial Intelligence Module for Learning Purposes	27
<i>Hlavcheva Y., Glavchev M., Bobicev V., Kanishcheva O.</i> Language-Independent Features for Authorship Attribution on Ukrainian Texts	29
<i>Kadomskyi K.</i> Evaluating Deep Learning Models for Anomaly Detection in an Industrial Transporting System	31
<i>Nazarchuk I., Krasovska, H., Ilarionov O.</i> Intellectual Agent for Sentiment Analysis on Movie Reviews	33
<i>Neskorodieva T., Fedorov E.</i> Automatic Analysis Method of Audit Data Based on Neural Network Mapping	36
<i>Samokhvalov Y., Bondarenko B.</i> Use of Neural Networks in Information Retrieval Systems	40
<i>Semerikov S., Kucherova H., Los V., Ocheretin D.</i> Neural Network Analytics and Forecasting the Country's Business Climate in Conditions of the Coronavirus Disease (Covid-19)	42
<i>Sharkadi M.</i> Neuro-Fuzzy Modeling of Level Assessment in the System of Financial-Economic Security	46
<i>Soroka P., Krasnovidov S.</i> Business Analytics Information Technologies for Analysis of the Activity of a Commercial Organization	49
<i>Soroka P., Savchenko R.</i> Machine Learning Methods for Sport Result Prediction	51
<i>Sus B., Revenchuk I., Bauzha O.</i> Model of Implementation Virtual Laboratory Work for Supporting Educational Process	53
<i>Tmienova N., Dulich O.</i> Automatic Question Generation System for Ukrainian-Language Texts	57
<i>Yakymenko D., Tregubenko I.</i> Modified Method of Construction of Information Image of Electronic Text Documents By Means of Intellectual Data Analysis	59

## CYBERSPACE PROTECTION TECHNOLOGIES

<i>Buchyk S., Andrushchenko Y.</i> Searching for a Potential Criminal Using Wireless Internet Networks as one of the Targets of State Security	65
<i>Buchyk S., Symonychenko Y., Symonychenko A.</i> The Method of Detection of Hidden Information Using Steganographic Methods	68
<i>Kashtalian A.</i> Honeypots Models in Computer Networks According to Malicious Attacks Types	71
<i>Koltsov D., Parkhomenko I.</i> Traversal Utilities For Nat	74
<i>Lukova-Chuiko N., Bystrov A.</i> Advice on Selecting an Intrusion Detection System for Small and Medium-Sized Businesses	77
<i>Lukova-Chuiko N., Fesenko A., Papirna H., Gnatyuk S.</i> Threat Hunting as a Method of Protection Against Cyber Threats	79
<i>Lukova-Chuiko N., Klochko V.</i> Collective Defense of Corporate Networks Against Computer Attacks	83
<i>Nakonechnyi V., Bondarenko M.</i> Application of Biometric Methods of User Identification in Information and Communication Systems	86
<i>Nakonechnyi V., Voitenko I.</i> Comparative Characteristics of Algorithms to Improve Spam Prevention Mechanism	88
<i>Nicheporuk A., Savenko O., Kazantsev A.</i> The Architecture of CNN Model for Android Malware Detection	92
<i>Ponomarov S., Lukova-Chuiko N.</i> Breach and Attack Simulation as a new vector of information security	95
<i>Rusyn V., Sambas A.</i> Simple Autonomous Security System Based on the Fingerprint Scanner Module and Arduino Platform: a Study Case	97
<i>Shved A., Buchyk S.</i> Basic Approaches to Personal Data Protection in Client Relationship Management System	99
<i>Slipachuk L., Toliupa S.</i> Synthesis Features of Functional Model of Integrated Industry Management System of National Cybersecurity	102
<i>Stetsiuk M., Nicheporuk A., Savenko B.</i> Ensuring the Fault Tolerance And Survivability of Specialized Information Technologies in Corporate Computer Networks Under the Influence of Malicious Software	105
<i>Toliupa S., Brailovskiy M., Parkhomenko I., Zhurakovskiy B.</i> Safety of Critical Functions Infrastructure	107
<i>Toliupa S., Buchyk S., Shestak Y., Kulko A.</i> Cyberattack Detection Systems Based on the Signature Method	110

*Toliupa S., Nakonechnyi V., Kotov M., Solodovnyk V.* Signals Encryption in Wireless Data Input Devices 113

*Tukalo S., Kostiv O., Shpur O., Buhyl B.* Methods Development to Protect IoT From Botnets 115

## DATA ANALYTICS

*Bokan V., Tsykun V., Khlevnyi A.* Information Analysis of Methods for Forecasting the Population of Ukraine 121

*Bura Y., Khlevna I.* House Price Modeling by Machine Learning 124

*Burmistenko O., Bila T., Statsenko V., Statsenko D.* Information Analysis of the Bulk Materials Continuous Dosing Process 126

*Dolgikh S., Mulesa O.* Covid-19 Epidemiological Factor Analysis: Identifying Principal Factors with Machine Learning 128

*Dvoretskyi M., Dvoretska S., Horban H., Nezdoliy Y.* Using the Analytic Hierarchy Process for Optimization the Database Structure of a Distributed Corporate Information System Node 131

*Fedorchenko I., Oliinyk A., Stepanenko A., Kharchenko A., Saman M.* Research and Development of a Genetic Algorithm for Diagnosing the Strength of the Blade Structure in Gas Turbine Engines 135

*Horban H., Kandyba I., Dvoretskyi M., Boiko A.* Principles of Searching for a Variety of Types of Associative Rules in OLAP Cubes 139

*Khlevnyi A., Koval B., Shabatskaya S.* Development of a Fraud Detection System in Payment Services Using CRISP-DM Methodology 143

*Kiktev N., Lendiel T., Osypenko V.* Application of the Internet of Things Technology in the Automation of the Production of Compound Feed and Premixes 145

*Kondruk N.* Segmentation of Data Sets by Different Types of Clusters 148

*Koval B., Khlevna I.* Fraud Detection Technology in Payment Systems 150

*Linder Y., Veres M., Kuzminova K.* Modeling and Prediction of Covid-19 Using Hybrid Dynamic Model Based on Seird With Arima Corrections 153

*Mikhieiev V., Mezentseva O.* Analysis And Forecasting Of Environmental Pollution By Carbon Dioxide 157

*Minaeva J.* Processing of Multidimensional and Multi-Aspect Data in Conditions of Uncertainty 159

*Mudra A., Mezentseva O.* Examination of the Dependence Between Criminal's Appearance and His Offense Using Machine Learning 161

<i>Orlovskyi D., Kopp A.</i> A Business Intelligence Dashboard Design Approach to Improve Data Analytics and Decision Making	163
<i>Rudenko V., Mezentseva O.</i> Influence Analysis Of Different Management Methodologies On The Result Of Big Data Projects	165
<i>Shelest T., Yeremieieva V.</i> Analysis Of The Possibility Of Using Vr Technologies In Environmental Awareness Projects	168
<i>Shtovba S., Petrychko M.</i> An Informetric Assessment of Various Research Fields Interactions on Base of Categorized Papers in Dimensions	170
<i>Taborovskyi A., Kolesnikova K., Khlevnyi A.</i> The Impact of Automated and Robotic Warehouses on the Scope of Supply Chain Process	172
<i>Tereshchenkova O., Kondrashov K.</i> Informational Expert System For Minimizing The Time For Searching Of Failures Of Ship Electrical Equipment	174
<i>Vavilenkova A.</i> Ragularity of Context Units Identification in Electronic Text Documents	178
<i>Yefremov H., Kolesnikova K.</i> Opinion Mining Methodology in Market Research	181
<i>Yeshchenkov V., Mezentseva O.</i> Identification of the Main Problems of Collection and Analysis of Speech Data Using Machine Learning	184
<i>Zhovtukhin D., Yehorchenkov O.</i> Classification of Bottles Images Using Convolutional Neural Networks	186

## DIGITAL PROJECT MANAGEMENT TECHNOLOGIES

<i>Dehtiarova Y., Morozov V.</i> Practical Implication of Digital Project Management Technologies	191
<i>Gamotska S., Soroka P.</i> Choice of Method of Quantitative Risk Assessment in Risk Management Tasks of IT Projects	194
<i>Kambur M., Yehorchenkov O.</i> Smart Kitchen Development Project Management	197
<i>Kovalenko A., Ivanov I., Morozov V.</i> Research of Methods of Formation of the Initial Description of the Project of Creation And Start-Up of the Enterprise on Production of Street Furniture Made of Recycled Materials	200
<i>Latysheva T., Smishchenko D.</i> Process of Effective Project Management of Developing Mobile Application for Carsharing	202
<i>Loik O., Triska M., Lub P., Sharybura A.</i> Information Technology in Project Management of the Agriculture Technological Systems Development	204
<i>Morozov V.</i> Use of Machine Learning Methods in Data Analysis for Digital Project Management	206
<i>Morozov V., Proskurin M.</i> Analysis of the Prospects for Applying Methods for Customer Churn Prediction Using Machine Learning in Innovative Startup	208

## Projects

<i>Naumenko A., Kolomiets A.</i> Specific Characteristics of Project Management in the Banking Sector	212
<i>Oberemok I., Oberemok N.</i> Priority Of Values Of Project Stakeholders	214
<i>Raichuk I.</i> Models of Digitalization of Business Processes of Project-Oriented Organizations Based on Artificial Neural Networks	217
<i>Samonenko A., Yehorchenkov O.</i> Peculiarity of RPA Projects	221
<i>Sazonov A., Yehorchenkova N.</i> Concept of Organization of Portfolio of Projects and Programs of Financial Companies	223
<i>Shelest T., Rudenko A.</i> Analysis of Prerequisites for the Application of IT Projects in Conscious Consumption Management	225
<i>Steshenko G., Buhrov A., Horban D., Timrova Y.</i> Basic Metrics of Startup Evaluating	227
<i>Suprun O., Klimenkova N.</i> It Audit as a Key Component of Information Systems Effectiveness and Data Security	229
<i>Timinsky A., Kerdun N.</i> MS Project as a Digitalisation Tool of Project Management System for Project Oriented Companies	231
<i>Timinsky A., Patsyuk M.</i> Team Management Models of SEO-Optimization Start Up Projects	233
<i>Yas V., Kolomiets A.</i> Implementation of Projects in the Medical Field Using Big Data and Waterfall Methodology	237
<i>Zharikova A., Morozov V.</i> Project Management of Development Business Messenger for Communication With Foreign Clients	239
<i>Zubets D., Steshenko G.</i> Business Analysis In Ukraine	241

## E-COMMERCE, E-GOVERNMENT AND E-LEARNING TECHNOLOGIES

<i>Bezlutska O., Leshchenko A., Yurzhenko A., Paziak A.</i> Informational Visualization on E-Courses of Higher Maritime Educational Institutions	247
<i>Domanetska I., Ilarionov O., Fedusenko O., Vlasenko O.</i> Dynamic Analysis Of The Quiz Complexity In Moodle	251
<i>Gradinari O.</i> Analysis Of Existing Models Of Information Competence	253
<i>Horbas I.</i> “A State in a Smartphone” Concept by Ukrainian government	256
<i>Makhachashvili R., Semenist I., Bakhtina A.</i> Ict Tools for Final Qualification Assessment Survey Study for European and Oriental Languages Programs	260

<i>Mironova V., Pyroh M., Harko I.</i> Methodology of Building Agile-Education Processes in Higher Education Institutions	262
<i>Morze N., Makhachashvili R.</i> Digital Competence In E-Governance Education: A Survey Study	264
<i>Morze N., Strutynska O.</i> Development of the Digital Transformation Model for Higher Educational Institutions	266
<i>Ponomarenko R.</i> Knowledge Test Systems Based on Type 2 Takagi-Sugeno Fuzzy Inference	270
<i>Provotar A., Veres M., Samoilenco M.</i> Using Educational IoT System	272
<i>Riabov O., Khlevna I.</i> Recommendation System Design in Python by Methods of Emotional Analysis and Machine Learning	274
<i>Selivanova A., Pursky O., Yurchenko Y., Samoylenko H., Dubovyk T.</i> Agent Modeling of Online Store Activities	276
<i>Yurchenko A., Semenikhina O., Shamonia V., Khvorostina Y.</i> Open Educational Resources in IT Sphere	278
<i>Zagorodnyuk S., Sus B., Bauzha O.</i> The Application of Network Communication for Organizing a Laboratory Work	281
<i>Zinchenko V., Kyrpa A., Stepanenko O.</i> Information and Communication Technologies While Forming Non-Philological Students' Professional Language and Speech Competences	284

## **MATHEMATICAL FOUNDATIONS OF INFORMATION TECHNOLOGY**

<i>Bychkov O., Ivanchenko O., Merkulova K., Zhabska Y.</i> Mathematical Methods for Information Technology of Biometric Identification in Conditions of Incomplete Data	289
<i>Hnatienko H., Rimek A.</i> Use of Algebraic Approach When Evaluating the Correct Sequence of the Present List Elements in Testing Tasks	292
<i>Klyushin D.</i> Randomness: Old And New Ideas	295
<i>Kovalenko I., Davydenko Y., Shved A.</i> Structuring of Group Expert Judgments Formed Under Complex Forms of Ignorance	297
<i>Kredentser B., Mogylevych D., Kononova I., Mohylevych V.</i> Analytical Model with Interruption of Service of Short-Term Objects with Temporary Reservation	301
<i>Makarenko A.</i> Cellular Automata Models With Riemann Surfaces	304
<i>Polishchuk V., Malyar M., Polishchuk A.</i> The Technology for Determining the Level of Process Control in Complex Systems	306

<i>Rusyn V., Sambas A.</i> Circuit Realization of the Pulse Transformation of the Analog Nonlinear Signals Based on Chua's Generator	309
<i>Semenov V., Vedel Y.</i> Convergence of Adaptive Methods for Equilibrium Problems in Hadamard Spaces	311
<i>Sobchuk V., Olimpiyeva Y., Musienko A., Sobchuk A.</i> Ensuring the Properties of Functional Stability of Manufacturing Processes Based on the Application of Neural Networks	314
<i>Solomko M., Zubyk L., Zubyk Y., Ivanytska A.</i> Modified Algorithm for Transformation of Boolean Functions	317
<i>Vostrov G., Khrinenko A.</i> Mathematical Models of Pseudorandom Processes Behavior for Nonlinear Dynamical Systems	320

## NETWORK AND INTERNET TECHNOLOGIES

<i>Barannik V., Babenko Y., Shulgin S., Parkhomenko M.</i> Video Encoding to Increase Video Availability in Telecommunication Systems	323
<i>Belfer R.</i> The Architecture of The Layered Peer to Peer Network	325
<i>Buchyk S., Palageychenko D.</i> Information Technologies in Ukrainian Judicial System	327
<i>Cherevatenko A., Paliy S.</i> The Use of Artificial Intelligence in the Internet of Things System	329
<i>Dudnik A., Kobylchuk M., Pokutnia D.</i> Analysis of the Current State of Technology "Smart Home"	332
<i>Gladka M., Lisnevskyi R., Kostikov M.</i> Using the Internet of Things When Introducing CRM Systems in the Banking Sector	335
<i>Hnatienko H., Kudin V., Ilarionov O., Vlasenko O.</i> Fuzzy Definition of Relative Estimates of Alternatives Based on Pairwise Comparisons Using Pseudobasic Matrices	339
<i>Kondratiuk I., Vlasiuk S., Paliy S.</i> Current Problems of Information Security of IoT Systems	344
<i>Kovbas Y., Izmailova O.</i> Scenario Formation Construction of a Local Corporate Network of the Enterprise	346
<i>Kravchenko Y., Dakhno N., Leshchenko O., Tolstokorova A.</i> Machine Learning Algorithms for Predicting the Results of Covid-19 Coronavirus Infection	350
<i>Kucherenko R., Kravchenko O.</i> IoT Solutions System for Climate Control Process of Making Cheese	352
<i>Kudin V., Onyshchenko A., Ilarionov O.</i> Modeling of Dynamic Ecological-Economic Interaction	354

<i>Leshchenko O., Dakhno N., Herasymenko O., Lavrinovich V.</i> Application Peculiarities of Gradient Descent Algorithms in Neural Networks	357
<i>Myroshnychenko Y., Paliy S.</i> Road Traffic Optimization By IoT	360
<i>Nakonechnyi V., Pliushch O., Bielikov A.</i> Development and Analysis of Algorithms for Recognizing Moving Objects in the Data Stream	363
<i>Nemchenko K., Paliy S.</i> Statement of the Task of Building an Adaptive System of Energy-Efficient Lighting for Administrative Buildings Based on the Internet of Things	365
<i>Nikolyuk P., Neskorodieva T., Fedorov E., Chioma E.</i> Intellectual Algorithm Implementation for Megacity Traffic Management	367
<i>Paiuk V., Heidarova O.</i> Detecting Software Malicious Implant Based On Anomalies Research On Local Area Networks	369
<i>Ponomarenko R., Tkachenko R.</i> Method of Processing Complex Objects Based on Object-Oriented Proxy System	371
<i>Sakharov D., Kravchenko O.</i> IoT Seismological Situation Monitoring System Development With one of the Regions of Ukraine as an Example	373
<i>Selivorstova T., Kyrychenko S., Brodskyi V., Tarkovska N.</i> Research of Application Metrics Deployed in Monolithic and Microservice Architectures	375
<i>Turovsky O., Kozlovskyi V., Balaniuk Y., Boiko Y.</i> Minimization of Phase Error Dispersion in Closed Type Phase Synchronization Systems in Carrier Frequency Tracking Mode	378
<i>Tymoshchuk S., Ponomarenko R.</i> The Research and Development of the Software to Support the Educational Process in Higher Education Institutions	380
<b>AUTHORS</b>	382

# **ARTIFICIAL INTELLIGENCE TECHNOLOGIES**



**<sup>1</sup> Zinnur Abduramanov**

Senior Lecturer

**<sup>2</sup> Zarema Seidametova**

Professor

**<sup>3</sup> Niiare Valiieva**

Lecturer

<sup>1-3</sup> Crimean Engineering-Pedagogical University

## COLOR RECOGNITION DEEP LEARNING MODEL

Hardware or software systems of certain human activities (for example, the recreation of intelligent reasoning and actions using software systems and devices, or the imitation of a device of intellectual behavior inherent in humans). Machine Learning (ML) and Deep Learning (DL) are AI subsets. In ML we create models (algorithms) which are first trained on the basis of available data (datasets), and then allow making predictions based on the data. DL uses an artificial neural network structure imitating the neurons located in the human brain. The term “deep” is used to refer to multiple layers in an artificial neural network. The basic idea is that the network of artificial neurons, constructed from interconnected switches can learn to recognize patterns in the same way as is done by the brain and the nervous system of animals. Deep learning can be defined as neural networks with many parameters and layers. One of the largest neural networks “The Sparsely-Gated Mixture-of-Experts Layer” is described in the paper [1]. Authors of the paper [1] proposed a method to increase the capacity of the model without linearly increasing the number of calculations.

Classic machine learning (ML\DL) algorithms, which are not subset of DL algorithms, train and make predictions much faster than deep learning (DL) algorithms [2] – one or more processors are enough to train a classical model. Deep learning models require additional hardware for training as well as for large-scale deployment of software infrastructure; without this, it takes a lot of time to train the model.

There are problems for which classical machine learning algorithms build a “good enough” model. But there are problems for which classical machine learning algorithms don’t work very well. For example, deep learning algorithms are used for natural language processing tasks (text translation, discourse analysis, morphological segmentation, object recognition, natural language generation, natural language understanding, mood analysis, and speech recognition). Deep learning algorithms also are used in the fields where image classification is required (image classification with localization, object detection, object segmentation, transfer of image styles, coloring, reconstruction, super-resolution and image synthesis) [3]. In addition, deep learning is used in pharmaceutical industry in the development of new drugs (to predict how molecules will interact; to search for subatomic particles, etc.).

An interdisciplinary group of experts within the framework of an independent initiative of the Human-Centered Artificial Intelligence Institute at Stanford University prepared and published the study [4] on the impact of artificial intelligence on various aspects of social development. Authors of the papers [5, 6, 7, 8] presents examples of

using the frameworks, programming language and libraries for solving artificial intelligence problems and building deep learning models. The papers [9], [10], [11] describe libraries and tools for machine and deep learning (OpenAI Gym, TensorFlow, Keras, Scikit-Learn).

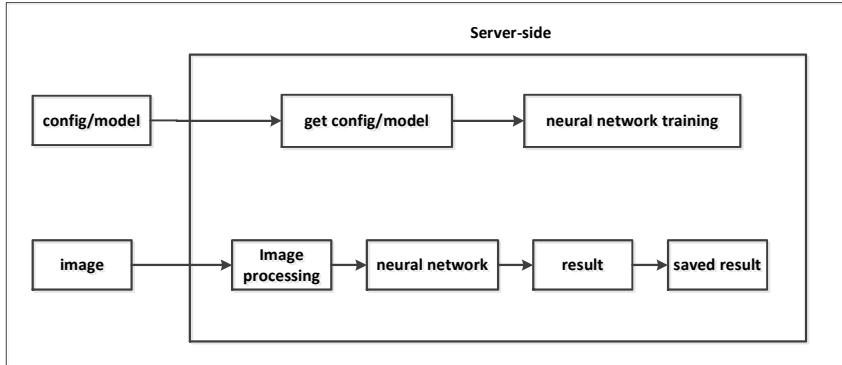


Figure 1 – Schematic DLMCR client-server application architecture

After the user has sent the image to the server, it is necessary to obtain its mathematical representation. We need to get the number of the image pixels and their representation in the three channels of R, G, B. Thus, we get two models of the input image (Fig. 2).

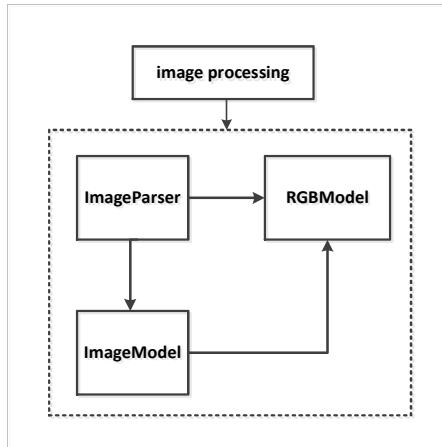


Figure 2 – Input image representation in the format required for neural network

In order to create a deep learning model, it is important to prepare a dataset that stores each color value of the RGB model and the value of the color opacity. The obtained data should be normalized, this is due to the nature of all machine and deep

learning algorithms, including neural networks. Data that is different in physical meaning and arrives at the input of a neural network can lead to its incorrect operation: incorrect predictions, non-existent data, it can also slow down the learning process and the modeling process.

After normalization, all the numerical values coming to the input of the neural network are in one narrow intermediate interval (in our case, an interval from 0 to 1).

After receiving data from the input signals, they are transmitted to the subsequent (hidden) layers of the neural network. The first such layer in the entire hierarchy determines the number of all colors in the image and assigns its own class to each. Then delving deeper into the layers of the network, each color is responsible for its own layer of the network. It means that some layer of the network can only be responsible for the defined color and it will give a result about whether the given image has this color or not. The result from each deep (hidden) layer of the network is analyzed and summarized at the output of the external, global neural network. It can give results in the form of what percentage a given color is in the image.

The number of colors neural network knows depends on the dataset on which network was trained and trained. The more colors it was given for training, the higher the accuracy of the result produced by the network.

After training, the neural network will know unique colors. The array of output values is an array of ones and zeros: [0,0,1,0,0,0,0,0,0,0,0,0], where 0 is no this color. Each color in this array has its own index. Then this dataset transforms into a json file that transferred it to train the neural network.

To implement a deep learning color recognition model, we choose the multi-paradigm JavaScript programming language that allows to cover a large number of areas for solving diverse problems: development and support of direct server services for receiving and sending data; access to the content of files, and transformation of the content into the necessary data structures; deployment and implementation of artificial intelligence systems for image analysis; transforming the output data into a human-readable format; use of the received output data for visual presentation in the client part.

Neural networks can solve almost any problem posed to a person in any area of the activity, they do it faster, more accurately and more efficiently. The goal of the project was to design and develop a deep learning model for building a color recognition system, which can be applied in the work of the designers, photographers, medical workers and in the agricultural industry.

For the analysis of existing models and architectures of neural networks, libraries that allow to implement neural networks we use Synaptic.js as the main library for implementing a neural network.

#### References:

1. N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, J. Dean, Outrageously large neural networks: The sparsely-gated mixture-of-experts layer, arXiv preprint arXiv:1701.06538. 2017. URL: <https://arxiv.org/abs/1701.06538>.
2. I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep learning (Vol. 1, p. 2). Cambridge: MIT press. 2016.

3. I. Arel, D.C. Rose, T.P. Karnowski, Deep Machine Learning – a New Frontier in Artificial Intelligence Research. Computational Intelligence Magazine, IEEE. 2010. Vol. 5, No. 4. P. 13-18. DOI: 10.1109/mci.2010.938364
4. Artificial Intelligence Index Report 2019 / Human-Centered Artificial Intelligence Institute, Stanford University. URL: <https://hai.stanford.edu/ai-index/2019>
5. P. Bezak, Building recognition system based on deep learning. In 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR) (pp. 1-5). IEEE. 2016.
6. F. Marra, G. Poggi, C. Sansone, L. Verdoliva, A deep learning approach for iris sensor model identification. Pattern Recognition Letters, 113, pp.46-53. 2018.
7. M. Zhang, P. Wang, X. Zhang, Vehicle Color Recognition Using Deep Convolutional Neural Networks. In Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science. pp. 236-238. 2019.
8. A. Tang, K. Lu, Y. Wang, J. Huang, H. Li, A real-time hand posture recognition system using deep neural networks. ACM Transactions on Intelligent Systems and Technology (TIST), 6(2), pp.1-23. 2015.
9. T. Hope, Y.S. Resheff, I. Lieder, Learning tensorflow: A guide to building deep learning systems. "O'Reilly Media, Inc.". 2017.
10. Building a Color Recognizer in Python. URL: <https://towardsdatascience.com/building-a-color-recognizer-in-python-4783dfc72456>
11. K. Wongsuphasawat, D. Smilkov, J. Wexler, J. Wilson, D. Mane, D. Fritz, D. Krishnan, F.B. Viégas, M. Wattenberg, Visualizing dataflow graphs of deep learning models in tensorflow. IEEE transactions on visualization and computer graphics, 24(1), pp.1-12. 2017.

<sup>1</sup> **Maryna Antonevych**

Student of the Faculty of Information Technology

<sup>2</sup> **Anna Didyk**

Student of the Faculty of Information Technology

<sup>3</sup> **Vitaliy Snytyuk**

Dr. of. Sci., Professor, Dean of the Faculty of Information Technology

<sup>1,2,3</sup> Taras Shevchenko National University of Kyiv

## CHOICE OF BETTER PARAMETERS FOR METHOD OF DEFORMED STARS IN N-DIMENSIONAL CASE

**Abstract.** This paper is devoted to the problem of optimization of a function in  $n$ -dimensional space, which in the general case is polyextreme and undifferentiated. In contrast to the classical method of deformed stars [4], we obtained a method that solves problems in  $n$ -dimensional space, where the population of solutions consists of 3-, 4-, and 5-point groups. The advantages of the developed method over genetic algorithm [1], differential evolution [2] and evolutionary strategy [3] as the most typical evolutionary algorithms are shown. Testing was performed to investigate the best configuration of method of deformed stars parameters.

**Keywords:** Function, optimization, method of deformed stars (MODS),  $n$ -dimensional space.

**Introduction.** A large number of modern practical problems belong to the class of constraint satisfaction problems (CSPs). Stochastic search, combinatorial optimization methods, and evolutionary algorithms are used to solve such tasks. Exactly, the use of evolutionary algorithms does not require strict target functions constraints and also does not guarantee the finding of a global optimum, although according to certain conditions there is a probability convergence.

### General algorithm of the method.

Step 1. We initialize the parameters of the algorithm, let  $t = 0$ .

Step 2. Generate  $m$  potential solutions in domain  $D$  (population  $P_t$ ).

Step 3. Form  $w$  figures (triangles, rectangles or pentagons)  $F_i, i = \overline{1, w}$ .

Step 4. For each  $F_i$  find the vertex in which the function  $f$  takes the best value and consider it the best vertex.

Step 5. Find the center of each figure as the average of all its vertices.

Step 6. For each figure  $F_i$  find a compressed figure  $T_i$  in which the best point is transferred along the line connecting the center of the figure and the best point, and all the others are transferred to it.

Step 7. For each figure  $F_i$  find a compressed figure  $U_i$ , in which all points are compressed to the best vertex.

Step 8. For each figure  $F_i$  find the figure  $Q_i$ , obtained by rotating  $F_i$  around the best vertex.

Step 9. For each figure  $F_i$  find the figure  $B_i$ , obtained by rotating  $F_i$  around the center of the figure.

Step 10. For each figure  $F_i$  find a modified figure  $R_i$ .

Step 11. Form a general population  $P_t$ , which will contain all the new points created in the previous steps. Thus,  $P_t = P_t + T_i + U_i + Q_i + B_i + R_i, i = \overline{1, w}$ .

Step 12. For all potential solutions from  $P_t$  find the value of the function  $f$  and sort the potential solutions from the best to the worst.

Step 13. Leave in  $P_t$  only  $m$  best solutions and check the fulfillment of the stop criterion.

Step 14. If the stop condition is not met, go to step 3. Otherwise, complete the algorithm and the best element in the population will be considered the best solution.

Note that the stop criterion may be:

- a given number of iterations;
- the worst value of fitness function in neighboring populations is less than specified;
- the average value of fitness function in neighboring populations is less than specified, etc.

**The experiment results.** Table 1 shows the results for the Schwefel 2.20 function in 10-dimensional space. The comparative graph of results is presented in fig. 1. It shows how well-known methods and MODS found a solution during the execution with the stop condition by iterations. As we can see, MODS found a solution on the initial iterations, in contrast to other known methods.

Table 1.  
Comparison of the results of methods under three conditions of completion

Test function	Global minimum	1000 iterations, 10 launches, accuracy = $10^{-5}$								
		GA		ES		DE		MODS-n3	MODS-n4	MODS-n5
		$f$	$f$	$f$	$f$	$f$	$f$	$f$	$f$	
Schwefel 2.20	0	0.143	194.1793	79.7148	0.0	0.0	0.0			
Worst fitness-function between populations, epsilon = $1*10^{-10}$										
Schwefel 2.20	0	8.4094	191.1157	283.9126	0.0	0.0	0.0			
Average fitness-function between populations, epsilon = $1*10^{-10}$										
Schwefel 2.20	0	3.6056	244.8562	170.2587	0.0	0.0	0.0			

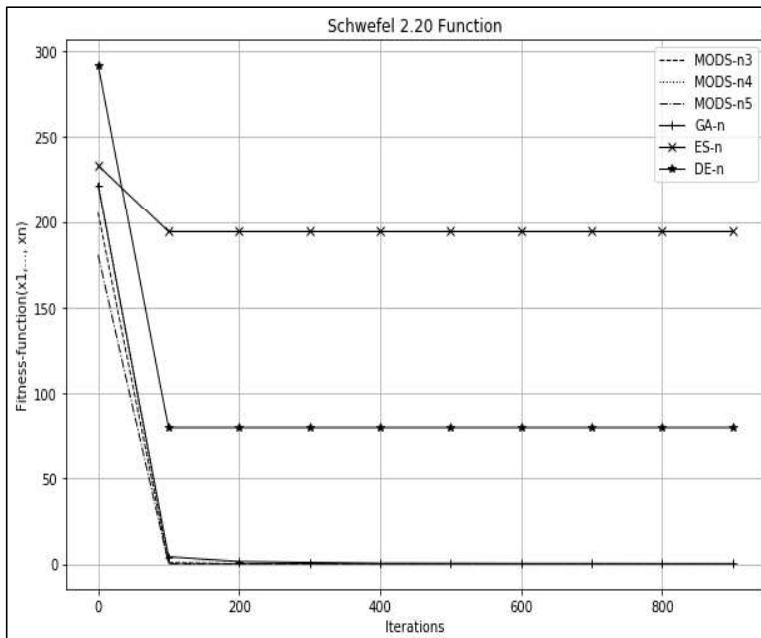


Figure 1 - Convergence plots for the Schwefel 2.20 function, the condition by iterations, coefficient = 2

**Choice of better parameters for MODS.** As we can see, first experiments were made with the parameters, which are used to perform transformations in MODS populations, equal to 2, for all stop conditions.

And MODS with these parameters was used for comparison with well-known methods in Table 1.

To investigate the best configuration of the MODS, it was decided to conduct testing, considering the different values of the input parameters of the method. For the study, it was decided to set the values of all parameters, which are used to perform transformations in MODS populations, equal to 1.5.

After conducting experiments, it was found that even with the change of parameters, MODS was able to find the correct extremes of the tested functions.

Thus, we can say that when changing the parameters of the method, the solution will still be found.

However, it is important to note, that for coefficients = 1.5 the method worked faster. Therefore, the change in the value of the parameter affected the speed of finding the result.

The comparative graph of results is presented in fig. 2. It shows how well-known methods and MODS (but, in this case, with parameters, that are equal to 1.5) found a solution with the execution of the stop condition by iterations.

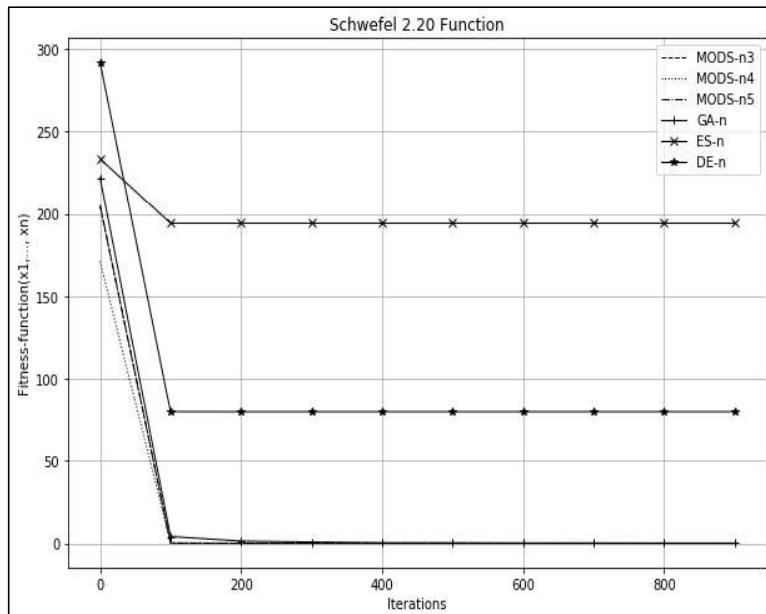


Figure 2 - Convergence plots for the Schwefel 2.20 function, the condition by iterations, coefficients = 1.5

**Conclusion.** In implementing the method of deformed stars, significantly fewer steps performed in the wrong direction, in contrast to genetic algorithm, the method of differential evolution and evolutionary strategy as representatives of classic evolutionary paradigm. The accuracy of the obtained solutions is, on average, higher than that of competing algorithms due to a deeper study of the solution search area. It is also important to note that in the course of research it was found that when changing the parameters of the MODS solution will still be found.

#### References:

1. J.H. Holland (1975) Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, Michigan; re-issued by MIT Press (1992).
2. Storn, R., Price, K. Differential Evolution – a Simple and Efficient Heuristic for Global Optimization over Continuous. Journal of Global Optimization 11, 341-359 (1997).
3. I. Rechenberg (1973) Evolutions strategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution, Frommann-Holzboog Verlag, Stuttgart (2nd edition 1993).
4. V. Snytyuk, “Method of Deformed Stars for Multi-extremal Optimization. One- and Two-Dimensional Cases”, in International Conf. Mathematical Modeling and Simulation of Systems. MODS 2019, Advances in Intelligent Systems and Computing, vol 1019. Springer, Cham.
5. Antonevych M., Didyk A., Snytyuk V. Optimization of functions of two variables by deformed stars method // In Proc. 2019 IEEE International Conference on Advanced Trends in Information Theory ATIT, Kyiv, Ukraine.

<sup>1</sup> **Anton Astakhov**

Master's student

<sup>2</sup> **Oleh Ilarionov**

PhD in Engineering Science, Associate Professor

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## ANALYSIS OF SPEECH EMOTION RECOGNITION METHODS

**Abstract.** Human beings can naturally recognize emotions from speech, but building automated speech emotion recognition systems is a challenging and relevant problem. Such systems could be used for interactive voice-based assistants and customer satisfaction analysis.

**Keywords:** Speech emotion recognition; Speech Datasets; Classification; Emotion Recognition

Speech is the most common way for humans to communicate with each other. People can not only understand the meaning of words, but can also extract a lot of useful information from the speaker's emotions.

Speech emotion recognition (SER) — is the process of identifying human emotions embedded in their speech [1]. However, human beings can naturally solve this problem, it is still a challenging task to automatically recognize human emotions. Therefore, SER is a relevant topic for scientific research. We have examined the structure of SER systems and identified the crucial parts of them, such as datasets and classifiers.

Speech emotion recognition system usually consists of three main components: signal acquisition, feature extraction, emotion recognition. General framework is shown in Figure 1 [2].

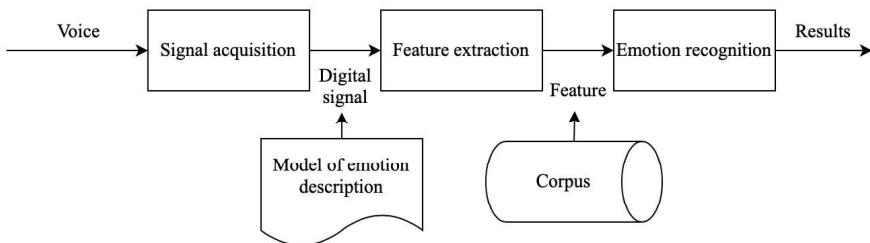


Figure 1 — SER system block diagram

One of the key stages to building SER systems is to find correct and complete dataset. Audio recordings should be performed by diverse groups of people (men and

women of different ages with different accents), who can express the whole range of emotions (anger, happiness, sadness, disgust, surprise, fear, etc.). Usually, this kind of audio recordings are made by professional actors [3]. Moreover, there are some other criteria to match when choosing datasets, such as English language and free to use license (at least for scientific purposes). Datasets, that match all of the criteria, are presented in Table 1.

Table 1 — Review of SER datasets

<b>Dataset</b>	<b>Emotions</b>	<b>Size</b>
Surrey Audio-Visual Expressed Emotion (SAVEE) (Surrey audio- visual expressed emotion database, 2019)	Anger, disgust, fear, happiness, sadness, surprise, neutral, common	14 speakers (male), 120 utterances
Toronto Emotional Speech Database (TESS) (Toronto emotional speech database, 2019)	Anger, disgust, neutral, fear, happiness, sadness, pleasant, surprise	2 speakers (female), 2800 utterances
eINTERFACE'05 Audio- Visual Emotion Database (Martin et al., 2006)	Anger, disgust, fear, happiness, sadness, surprise	42 speakers (34 male, 8 female) from 14 nationalities, 1116 video sequences
SAMAIN Database (McKeown et al., 2011)	Valence, activation, power, expectation, overall emotional intensity	150 speakers, 959 conversations
TUM AVIC Database (Schuller et al., 2009)	Five level of interest; 5 non- linguistic vocalizations (breathing, consent, garbage, hesitation, laughter)	21 speakers (11 male, 10 female), 3901 utterances
AFEW Database (Kossaifi et al., 2017)	Anger, disgust, surprise, fear, happiness, neutral, sadness	330 speakers, 1426 utterances from movies, TV-shows

Another key component of building a SER system is to choose the right classifier, which is a method to classify underlying emotions for a given utterance [4]. There are several different options to choose from, such as machine learning, traditional classifiers, deep learning algorithms. However, there is no consensus on what should be the ultimate classifier to solve the problem. Classifiers used in literature are presented in Table 2.

Table 2 — Review of classifiers, studies, datasets, and results

	<b>Number of studies</b>	<b>Datasets</b>	<b>Results</b>
HMM	7 [5, 6]	Berlin Emo DB	79%
GMM	4 [5, 7]	Berlin Emo DB, EPSAT, EMA, GES, SES, WSJ	75%
SVM	10 [5, 8]	LDC, Berlin Emo DB, AIBO DB, ABC DB, SUSAS DB, EMA DB, VAM I-II DBs, VAM DB	74%
MLP	5 [5, 9]	Berlin Emo DB	72%
kNN	2 [9, 10]	One natural and one acted speech corpora in Mandarin	66%
Decision Tree	3 [5, 8]	AIBO DBUSC IEMOCAP DB	58%
Rule Based Fuzzy Estimator	1 [11]	EMA DB, VAM I-II DBs	70%
DNN	2 [12]	IEMOCAP DB	54%
CNN	5 [5, 13]	SAVEE DB, Berlin EMO DB, DES DB, MES DB, RECOLA DB	70%
RNN	6 [5, 14]	IEMOCAP DB	63%

Among the most common classifiers are Hidden Markov models, Support vector machines, and Recurrent Neural Networks. The most precise are Hidden Markov models and Gaussian Mixture Models. The most underrated are Convolutional Neural Networks and Rule Based Fuzzy Estimators.

#### References:

1. Mehmet Berkehan Akçay, «Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers», Department of Software Engineering, Izmir University of Economics, Izmir, Turkey, 13.12.2019.
2. Chenchen Huang, Wei Gong, Wenlong Fu, and Dongyu Feng, «A Research of Speech Emotion Recognition Based on Deep Belief Network and SVM», 12.08.2016.
3. Kerkeni L, Serrestou Y, Mbarki M, Raoof K, Mahjoub MA. A review on speech emotion recognition: Case of pedagogical interaction in classroom. In: 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). IEEE; 2017. pp. 1-7.
4. Helen Chan, Travis Ebisu, Caleb Fujimori, «Speech Emotion Detection and Analysis», COEN296: Natural Language Processing, Department of Computer Engineering Santa Clara University, 04.12.2018.

5. Babak Basharirad and Mohammadreza Moradhaseli, «Speech emotion recognition methods: A literature review» (2017)
6. B. Schuller, G. Rigoll, M. Lang, «Hidden markov model-based speech emotion recognition», Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP'03). 2003 IEEE International Conference on, 2, IEEE (2003), pp. II-1.
7. O.-W. Kwon, K. Chan, J. Hao, T.-W. Lee «Emotion recognition by speech signals», Eighth European Conference on Speech Communication and Technology (2003).
8. M. Borchert, A. Dusterhoft, «Emotions in speech-experiments with prosody and quality features in speech for use in categorical and dimensional emotion recognition environments», Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE'05. Proceedings of 2005 IEEE International Conference on, IEEE (2005).
9. B. Schuller, S. Reiter, R. Muller, M. Al-Hames, M. Lang, G. Rigoll, «Speaker independent speech emotion recognition by ensemble classification», Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on, IEEE (2005).
10. J. Rong, G. Li, Y.-P.P. Chen, «Acoustic feature selection for automatic emotion recognition from speech», Inform. Process. Manag., 45 (3) (2009).
11. M. Grimm, K. Kroschel, E. Mower Provost, S. Narayanan, «Primitives-based evaluation and estimation of emotions in speech», Speech Commun., 49 (2007).
12. K. Han, D. Yu, I. Tashev, «Speech emotion recognition using deep neural network and extreme learning machine», Fifteenth annual conference of the international speech communication association (2014).
13. Q. Mao, M. Dong, Z. Huang, Y. Zhan, «Learning salient features for speech emotion recognition using convolutional neural networks», IEEE Trans. Multimed., 16 (8) (2014).
14. S. Mirsamadi, E. Barsoum, C. Zhang, «Automatic speech emotion recognition using recurrent neural networks with local attention», 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE (2017).

<sup>1</sup> Tamara Bondar

Bachelor of Computer Science

<sup>2</sup> Hryhorii Hnatienko

PhD

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## VIDEO REGISTRATION AND FACE RECOGNITION TECHNOLOGY ON STREAM VIDEO

**Abstract.** In this paper, the analysis of modern methods of face recognition on streaming video in real-time is carried out and the system analysis for the development of the intelligent system and the user interface for work with it is carried out.

**Keywords:** video registration, recognition, database, streaming video.

One of the systems that can increase security in public places and educational institutions is the face recognition system. Currently, many systems of this type are used in the world, the most famous of which are SkyNet [1], FacePro [2]. Based on the analysis of existing systems [3], it was decided to develop and adapt a similar system for education.

During the analytical review of existing solutions, a general description of the process of visitor identification was formed in the form of an IDEF0 diagram (Fig. 1).

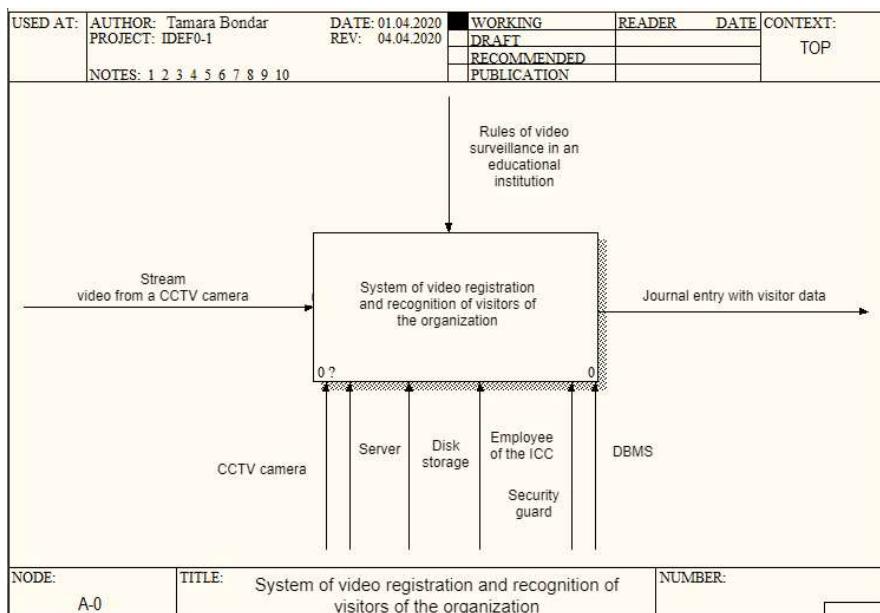


Figure 1 - IDEF0 diagram of the process of identifying visitors to the organization

At the entrance, the system must receive a streaming video, which will be broadcast from a video surveillance camera located in the direction of the flow of visitors to the organization. The result of the system should be an entry with the visitor's data in the electronic journal.

The operation of the video recording and visitor recognition system of the organization includes three processes:

- search and highlight faces on streaming video;
- search for a visitor in the database;
- making a record of the visitor in the electronic journal.

The intelligent system of video registration and identification of visitors of the organization must receive streaming video at the entrance and interpret the information received in it based on the results of personal identification. Comparative analyzes described in the sources [4-5] prove that convolutional networks will fully satisfy the requirements.

The main feature of the developed network: it was based on the MobileFaceNet architecture [6], which uses the ReLU activation function, alternation of deep cores, and bottleneck structures. This type of architecture is ideal for the limited capacity of the technical resources of the organization.

In order to determine the level of similarity of facial characteristics, it is necessary to find the distance from the characteristic removed by the convolutional neural network to the nearest similar to it among the characteristics of identified visitors stored in the database. The Euclidean distance square is used to give more weight to those class implementations that are significantly separated from each other. This feature is ideal for face recognition, as it allows to more accurately separate the facial features of visitors from each other.

#### **References:**

1. Weida Li, China updates Skynet system with facial recognition, 2018. URL: <https://bit.ly/3onLqCQ>
2. FacePRO: Panasonic Facial Recognition System, Panasonic Security System, 2019. URL: <https://bit.ly/35rIfBs>.
3. Tamara Kopchyk, Analysis of intelligent search and face recognition systems on streaming video. Information technologies and interactions: materials VI International. scientific-practical conf., Kyiv, December 20, 2019 / Ministry of Education and Science of Ukraine, "National Taras Shevchenko University of Kyiv", and [other] - p. 232-233.
4. Nicolas Delbiaggio, A comparison of facial recognition's algorithms, Degree Programme in Business Information Technology, Bachelor's Thesis, 2017. URL: <https://bit.ly/31CSH8g>.
5. Thai Hoang Le, Applying Artificial Neural Networks for Face Recognition / Thai Hoang Le. – Ho Chi Minh City, 2011. – p. 16. – (Department of Computer Science, Ho Chi Minh University of Science).
6. X.Gao, Y. Liu, S. Chen, Z. Han, MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices URL: <https://arxiv.org/abs/1804.07573>.

<sup>1</sup> Oleksandr Derevianchenko

PhD., Associate Professor Faculty of Computer Science and Cybernetics

<sup>2</sup> Andrii Nikolaiev

M.S. CS

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE MODULE FOR LEARNING PURPOSES

**Abstract.** Games are a great environment to explore different methods and algorithms. Developers often rely in their research on the ability to train AI with the help of games. It is also known that children acquire complex skills by learning and applying different patterns of behavior with a low level of risk while playing games. The purpose of the work is to implement AI module using modern machine learning methods and show some benefits for learning purposes.

**Keywords:** gaming artificial intelligence, reinforcement learning, gamification.

For the base algorithm was taken the idea of combining a tree-based search with a learned model which is presented in the [1] by Google DeepMind (2019). It is a reinforcement learning algorithm which could master games without having any knowledge about the game rules and environment dynamics making it better than state-of-the-art model-based and tree-based domain algorithms respectively.

According to the algorithm idea, there are provided four components that run simultaneously in the one loop. The shared storage holds the latest neural network weights, the self-play uses the weights to generate self-play games and place them in the replay buffer. The played games are used to train a network and store the weights in the shared storage (Figure 1).

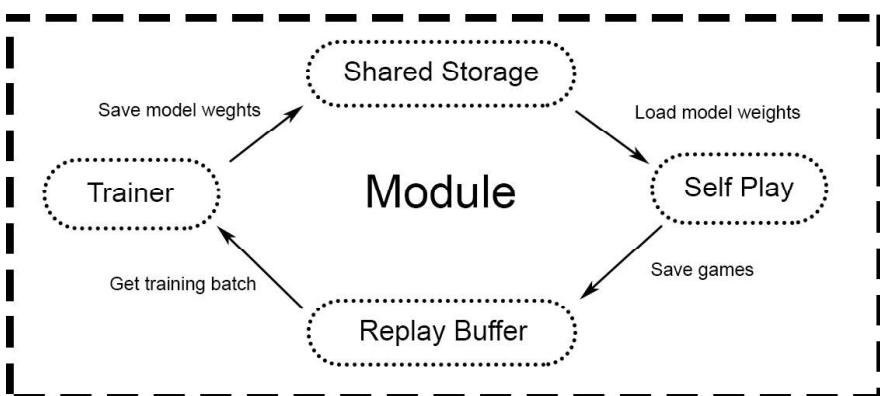


Figure 1 – How the module works.

Google Colaboratory was used for the model training as it has some powerful tools for machine learning in cloud computing [2]. The model was making about  $T_{steps} = 100$  training steps per minute during the training what makes the module competitive in such game as Connect 4 less than a few hours (Figure 2).

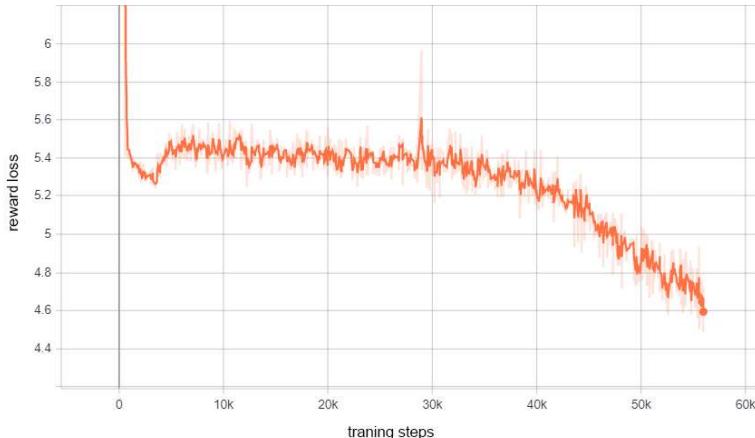


Figure 2 – Reward loss for the game Connect 4.

After reaching the certain level of playing, the developed module can be applied for the learning purposes: e.g. in the games position analysis; for searching the best move at the given position; to provide an expert evaluation for professional or beginner players (in chess and other board games).

The idea could be also applied for some educational projects as a gamification tool which could help students to master some of the board games and develop their own critical and logical thinking. Such method is highly approved as players tend to have more motivation to learn [3].

The strong point about the work is that the module could be used for different game domains (including more visually complex ones) which could bring higher variety of environments, where such learning skills could be developed.

#### **References:**

1. Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., Lillicrap, T., & Silver, D. (2019). Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. ArXiv, abs/1911.08265.
2. Google Colaboratory. URL: <https://research.google.com/colaboratory/faq.html>.
3. Burgers, C., Eden, A., van Engelenburg, M. (2015) How feedback boosts motivation and play in a brain-training game. Computers in Human Behavior 48: 94–103.

<sup>1</sup> **Yuliia Hlavcheva**

Deputy Director of the Library

<sup>2</sup> **Maksym Glavchey**

PhD in Economics, Associate Professor

<sup>3</sup> **Victoria Bobicev**

PhD in Computer Science, Associate Professor

<sup>3</sup> **Olga Kanishcheva**

PhD in Computer Science, Associate Professor

<sup>1,2,4</sup> *National Technical University “Kharkiv Polytechnic Institute”, Ukraine*

<sup>3</sup> *Technical University of Moldova, Republic of Moldova*

## LANGUAGE-INDEPENDENT FEATURES FOR AUTHORSHIP ATTRIBUTION ON UKRAINIAN TEXTS

**Abstract.** Authorship attribution is the natural language processing task of identifying the author of an input text. The main goal of this task is to define salient characteristics of documents that capture the writing author's style. In this paper, we analyze language-independent features for authorship attribution. For the experiments we used ML methods. The experimental results on scientific text classification showed that DT method outperforms most other ML methods, and these language-independent features are appropriate for the Ukrainian scientific documents authorship attribution.

**Keywords:** Writing Style, Language-Independent Features, Authorship Attribution, Text Classification, Machine Learning Methods.

The task of authorship identification is not new. The results of authorship detection studies are actively used in various spheres of human life. Authorship research can be divided into three main areas: authorship identification, authorship characterization, similarity detection [1]. Similarity detection is most often used to identify potential academic plagiarism [2]. This topic is relevant and important.

This paper focuses on the third direction of similarity identification. The decision is based on the group of properties that reflect the author's style measurement and comparison. Stylometric properties by which the author's style can be identified make up a list of style markers. All style markers are equally effective when used for different languages [3]. We distribute the stylometric characters into two groups: language-dependent features and language-independent features. This paper focuses on the study of the statistical characteristics of scientific texts in the Ukrainian language, which can be attributed to language-independent stylometric properties.

For the experiments we used our own preprocessed text corpus. The text corpus consists of individual scientific publications in Ukrainian. For stylometric properties we used only the paper main text, which best reflects the author's written style. For each author, a collection of paper fragments is formed. In our case, stylometric properties are determined for each fragment of the paper separately.

For stylometric properties, we used only the paper main text. We created two subsets for our experiments, one of them consist of 8 classes (8 authors, 1019

fragments), other – 32 classes (32 authors, 2633 fragments). These classes were selected randomly but sets are balance.

Groups of properties, which refer to the text statistical parameters and allow to determine the author's style with high accuracy are described in [3]. The authors of this paper created their own list of text statistical properties of Ukrainian, which are divided into 5 groups: average number of words in a sentence, average word length, average word frequency, punctuation (5 indicators); the number of words with length from 1 to 20 characters, the number of words with a word frequency from 1 to 8 times (28 indicators); frequency of using letters of the Ukrainian alphabet (33 indicators); frequency of using stopwords and pronouns (65 indicators); coefficients of language diversity (5 indicators) [4].

For our experiments of text classification, we took our corpus (2 subsets) and five groups of features: for separate groups and their combination. Weka software was used for classification task. Bayes Based Algorithms (Naive Bayes Multinomial, NBM), Support Vector Machine (SMO), Decision Trees (LMT, J48) were used as classification methods with the cross-validation parameter – 10 folds.

We conducted experiments for 32 authors (32 classes) and 8 authors (8 classes) and compared the results. Experiments showed that for 1-3, 1-4 and 1-5 groups of properties, the classification indicators are similar, despite the increase in the number of features. The best result (F-measure) of 32 classes we received for the SMO method (0.586) and LTM (0.614) for 1-5 groups of properties. The best result (F-measure) of 8 classes we received for the SMO method (0.794) and LTM (0.806) for 1-5 groups of properties.

According to the experiments, we obtained for 8 classes an average increase in the value of Correctly Classified Instances – 20%, MIN increase in the value Correctly Classified Instances – 15%, MAX increase in the value Correctly Classified Instances – 28%. The result of the experiments demonstrated the usefulness of the proposed language-independent stylometric properties indicators for text authorship attribution.

### **References:**

1. B. Alhijawi, S. Hriez, A. Awajan, Text-based authorship identification - A survey. Paper presented at the 5th International Symposium on Innovation in Information and Communication Technology, ISIICT 2018. 2018, pp. 1-7. doi:10.1109/ISIICT.2018.8613287.
2. M. AlSallal, R. Iqbal, V. Palade, S. Amin, V. Chang, An integrated approach for intrinsic plagiarism detection. Future Generation Computer Systems, Vol. 96., 2019 pp. 700-712. doi:10.1016/j.future.2017.11.023.
3. S. Adamovic, V. Miskovic, M. Milosavljevic, M. Sarac, M. Veinovic, Automated language-independent authorship verification (for Indo-European languages). Journal of the Association for Information Science and Technology, Vol. 70.8, 2019, pp. 858-871. doi:10.1002/asi.24163.
4. V. A. Vysotska, V. V. Pasichnyk, Yu. M. Shcherbyna, T. V. Shestakevych. Matematychna lingvistyka. Knyha 1. Kvantytatyvna linhvistyka, Lviv, Novyi Svit–2000, 2012.

**Kyrylo Kadomskyi**

Assistant Professor

*Taras Shevchenko National University of Kyiv*

## EVALUATING DEEP LEARNING MODELS FOR ANOMALY DETECTION IN AN INDUSTRIAL TRANSPORTING SYSTEM

**Abstract.** In Cyber-Physical Production Systems (CPPS), the task of anomaly detection is commonly solved by model-based methods. While these methods have proven effective in some use cases, their performance can drop dramatically in other systems. In this study the problem of representativeness of evaluation of such methods is addressed. The CPPS data is used, on which existing models have proven ineffective. The perspective of applying deep learning approach to constructing a process model in such systems is investigated.

**Keywords:** anomaly detection, autoencoder, model evaluation, cyber-physical production systems, industrial IoT.

Modern industrial plants demonstrate both increasing pressure for efficiency, and new possibilities to use growing set of sensors to facilitate automation. In this context, diagnosis of complex production processes has gained new attention due to research agendas such as Cyberphysical Production Systems (CPPS) [1]. One of the most important goals is self-diagnosis, which includes identification of anomalous system behavior, suboptimal energy consumption, or wear in CPPS.

The model-based diagnosis is a commonly used approach, in which a dynamic process model captures spatio-temporal features of the system's behavior. Considering the challenges of CPPS agendas, precise mathematical or expert modeling is infeasible in most cases. Thus, novel dynamic modelling techniques are being developed for learning the model from system observations [1, 2].

Applying these techniques in CPPS poses two main challenges:

1. While showing good results in certain applications, existing models yield relatively poor performance in other similar use cases [2, 3, 4]. The hypothesis is that this effect is due to limited nature and fixed structure of spatio-temporal features learned by the model, which are imposed by the structure of the model itself. Then the informativeness of learned features will vary in different physical systems, which can explain the observed effect. To meet this challenge, more generalized models are required.
2. As previous studies suggest, results of model evaluation in some systems may not be representative [2]. Thus, new evaluation criteria are required for representative comparison and benchmarking of the models.

In this study the two mentioned challenges are addressed. Deep Learning models, such as autoencoders, are applied to remove the first limitation by automatically selecting the most relevant features and structure to represent the data. To address the second problem, two robustness criteria are proposed for representative model evaluation: reconstructed variation rate and reconstruction's sensitivity to anomalies.

These criteria are assessed from the statistical distributions of model's response to normal and anomalous data.

A set of autoencoder architectures were modelled and evaluated on the HRSS dataset [5], which has proven challenging for applying novel dynamic models [2, 4]. Proposed robustness criteria were used in conjunction with traditional performance metrics, aiming for accurate benchmarking of the two approaches. This in turn provides the possibility to assess the limits of model-based anomaly detection in given class of CPPS.

Each model is trained in unsupervised manner to reconstruct normal time series, targeting for minimal reconstruction loss. Then the trained model is used to reconstruct unseen time series with anomalies, where the reconstruction error is expected to peak at anomalous intervals. To evaluate the model, the distributions of reconstruction error in normal and anomalous intervals are analyzed for being statistically distinguishable. Finally, from these error distributions, a decision-rule classifier for anomaly detection is built in a supervised mode. This method detects anomalies with time step precision, and most of evaluated models can be applied in real time.

6 LSTM and 3 ConvNet architectures were empirically tested. A grid search approach was applied to each architecture in order to select optimal model's hyperparameters. It was shown that increasing model complexity, both in LSTM and convolution-based models, allows to increase anomaly detection performance, but has a strong robustness tradeoff. This indicates that model evaluation in CPPS of this class cannot rely completely on performance metrics. Considering both performance measure and proposed robustness criteria, a single LSTM model is selected for HRSS data. Comparing to the baseline efficiency [2, 4], an increase by 102% in anomaly detection score and an increase by 121% in recall are achieved.

#### References:

1. O. Niggemann, C. Frey, Data-driven anomaly detection in cyber-physical production systems, AT – Automatisierungstechnik, 2015, vol. 63, issue 10. doi: 10.1515/auto-2015-0060.
2. A. von Birgelen, O. Niggemann, Using self-organizing maps to learn hybrid timed automata in absence of discrete events, in: Proceedings of the 22nd IEEE international conference on Emerging Technologies and Factory Automation, ETFA, Limassol, Cyprus, 2017, pp. 1–8.
3. N. Hranisavljevic, O. Niggemann, A. Maier, A novel anomaly detection algorithm for hybrid production systems based on deep learning and timed automata, in: Proceedings of the 27th international workshop on Principles of Diagnosis, DX-2016, Denver, Colorado, 2016.
4. M. Cerliani. Predictive maintenance with LSTM siamese network, 2019. URL: <https://towardsdatascience.com/predictive-maintenance-with-lstm-siamese-network-51ee7df29767>.
5. Physical factory / demonstrators IMPROVE, 2016. URL: <http://improve-vfof.eu/background/physical-factory-demonstrators>.

<sup>1</sup> **Iryna Nazarchuk**

Student

<sup>2</sup> **Hanna Krasovska**

PhD in Engineering Science, Associate Professor

<sup>3</sup> **Oleh Ilarionov**

PhD in Engineering Science, Associate Professor

<sup>1,2,3</sup> *Taras Shevchenko National University of Kyiv*

## INTELLECTUAL AGENT FOR SENTIMENT ANALYSIS ON MOVIE REVIEWS

**Abstract.** Movie reviews collection from IMDb website was used as the main dataset and evaluation of the text sentiment by negative/positive attitude was performed. The research proposes bag-of-words and tf-idf models. The core of the intellectual agent system model was trained using such classification algorithms - Logistic Regression, SGD, Random Forest and a Deep Learning model in a form of feedforward neural network. The comparative analysis of the accuracy scores was held and model with the highest AUC rate was chosen for the system integration.

**Keywords:** Motion pictures, Sentiment analysis, Feature extraction, Classification algorithms, Data mining, Opinion mining, Movie review, Information Retrieval, Deep learning, Neural networks, Classifier.

The modern stage of human development involves a rapid increase of information generation. With the spread of Internet and web platforms, where every user is given an opportunity to express their opinions regarding any type of product or service, as well as a movie or book, it is an urgent and essential task to handle these huge amounts of data to determine the attitude of users towards a particular object.

In this paper, the definition of the sentiment or polarity of reviews for films is being examined. By constructing an intellectual agent that automatically extracts and identifies opinion within the text, given as an input, in particular, movie reviews (comments or posts in social networks) in order to determine the mood of the audience in relation to a particular movie.

In the process of analyzing already existing systems used for automated classification of the text sentiment such systems as RapidMiner [1], GATE [2], Google Cloud Prediction API [3] revealed the instability of the accuracy while determining the emotional expression of the submitted movie reviews. These systems are currently in high demand and gain huge popularity in market researches to analyze reviews for commercial products [4]. The results of the semantic analysis using the above-mentioned systems have shown that the accuracy rate of sentiment classification for input text fragment is 87% when using the RapidMiner system, 92% utilizing the Google Cloud Prediction API service, 90% with the help of GATE software application. Also, while using the above-mentioned systems in multiple iterations, the accuracy of the classification has undergone significant fluctuations of classification metrics scores. According to the author, the main disadvantage of these systems is their

general-purpose focus, which does not take into account the semantic subtleties inherent in such a genre as movie reviews.

The purpose of this study is to develop a software product to increase the effectiveness of the sentiment analysis of the opinion expressed in text and ensure the stability of classification metrics, regardless of the nature of the input information.

To build the classification model, a collection of movie reviews data from the IMDb web site [5] was used for machine learning model training. The polarity of the reviews was based on the type of review (positive, negative and partly positive or partially negative) chosen by the author. In order to apply these data as an input parameter for the classification algorithms, the pre-processing phase was applied, which included such procedures as stemming, tokenization, stop words elimination and lemmatization. In order to form a training dataset for machine learning models from the texts of movie reviews, a set of features was constructed, having a numerical value, that is, the procedure of vectorizing the text was carried out. As a way to represent vectorized text in a form of features, models bag-of-words [6] and tf-idf [7] were used.

For the purpose of training, three classification algorithms were selected: Logistic Regression [8, 9], SGD [10], and Random Forest [11]. Alternatively, to find the best model that performs a sentiment analysis to determine the polarity of the movie review with high accuracy, training of the feedforward neural network with a different configuration of hyperparameters and layered architecture was performed.

To compare the results of the trained models' accuracy of the classification, the AUC [12] metric was used, in the assessment of the ROC curves for each model (fig. 1).

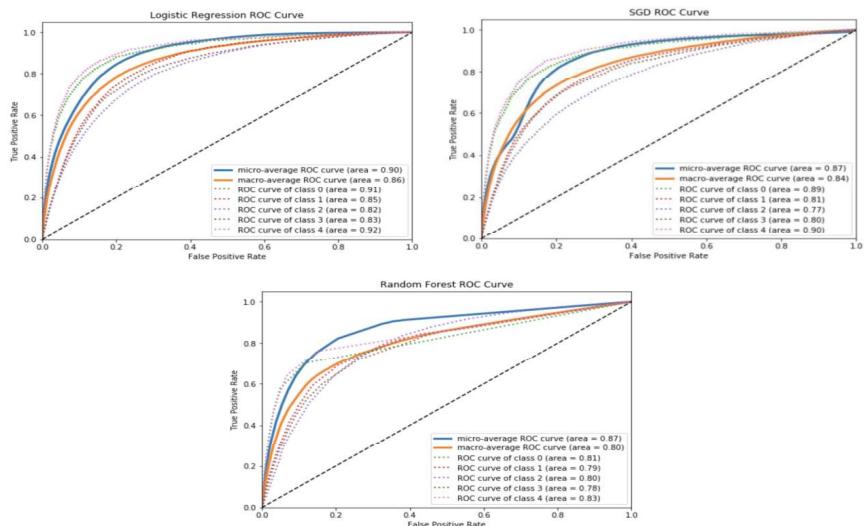


Figure 1 – ROC-curve graphs for logistic regression, stochastic gradient and random forest models

The received results showed that the highest value of the algorithmic accuracy was held by the neural network model with the AUC metric equal to 0.95, that is, the accuracy rate of the input movie review sentiment classification is 95%, using this software. On the basis of this, it can be concluded that the prototype system developed in the framework of this study is the best option for application with the purpose of sentiment analysis of movie reviews.

The prototype, developed in this research can be used as a basis for creating commercial software or as integration into existing systems. The novelty of this work is to use the ensemble of Machine Learning methods to achieve high accuracy of the text data classification.

Further improvement of the algorithm for analyzing the sentiment of the text is possible using deeper levels of natural language processing (syntactic and morphological). Also, it is necessary to investigate the processing of bipolar words, double objections, irony and sarcasm in the text.

#### References:

1. RapidMiner. Lightning Fast Data Science for Teams. Retrieved May 10, 2019, from <https://rapidminer.com/>.
2. GATE: a full-lifecycle open source solution for text processing. Retrieved May 1, 2019, from <https://gate.ac.uk/overview.html>.
3. Cloud machine learning engine. Retrieved April 29, 2019, from <https://cloud.google.com/ml-engine/>.
4. Pascual F. (2018, August 2). Why and How Companies Should Use Sentiment Analysis. Retrieved from <https://www.northeastern.edu/levelblog/2018/08/02/companies-use-sentiment-analysis/>.
5. IMDb datasets. Retrieved May 1, 2019, from <https://www.imdb.com/interfaces/>.
6. Wallach H.M. Topic modeling: beyond bag-of-words. Proc. 23rd Intern. Conf. on Machine learning. ACM, 2006, pp. 977–984.
7. Salton, G. and McGill, M. J. 1983 Introduction to modern information retrieval. McGraw-Hill, ISBN 0-07-054484-0.
8. Kleinbaum D.G., Logistic regression. A self-learning text, Springer-Verlag, 1994.
9. Hosmer DW, Lemeshow S. Applied Logistic Regression. 2nd ed. N.-Y.: Wiley; 2000. 375 p.
10. Roy R. (2018, April 10). ML | Stochastic Gradient Descent (SGD). Retrieved May 10, 2019, from <https://www.geeksforgeeks.org/ml-stochastic-gradient-descent-sgd/>.
11. Biau G. Analysis of a random forests model. The Journal of Machine Learning Research, 98888:1063–1095, 2012.
12. Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers. Kluwer Acad. Publ; 2004. 38 p.

**<sup>1</sup> Tatiana Neskorodieva**

Candidate of Technical Sciences (Ph.D.), Head of the Computer Science and Information Technology Department, Associate Professor

**<sup>2</sup> Eugene Fedorov**

Doctor of Technical Sciences, Professor of the Department of Robotics and Specialized Computer Systems, Associate Professor

<sup>1</sup> Vasyl' Stus Donetsk National University

<sup>2</sup> Cherkasy State Technological University

## **AUTOMATIC ANALYSIS METHOD OF AUDIT DATA BASED ON NEURAL NETWORK MAPPING**

**Abstract.** The urgent task of increasing the audit efficiency was solved by automating the mapping of audit indicators by forward-only counterpropagating neural network. A learning algorithm based on k-means has been created, intended for implementation on a GPU using CUDA technology, which increases the speed of identifying parameters of a neural network model.

**Keywords:** audit, mapping by neural network, forward-only counterpropagating neural network, sequences of payment and supply of raw materials.

Currently, the analytical procedures used during the audit are based on data mining techniques [1]. The aim of the work is to increase the efficiency of automatic data analysis in the audit DSS by means of a neural network mapping of sets of audit indicators in order to identify systematic misstatements that lead to misstatement of reporting. It is assumed that the audit indicators are noisy with Gaussian noise, which in turn simulates random accounting errors (as opposed to systematic ones).

For the achievement of the aim it is necessary to solve the following tasks:

- generate vectors of indicators for objects of sequences of payment and supply of raw materials;
- choose a neural network model for mapping audit indicators (which are noisy with Gaussian noise, which in turn simulates random accounting errors (as opposed to systematic ones, which lead to distortion of reporting));
- choose a criterion for evaluating the effectiveness of a neural network model;
- propose a method for training a neural network model in batch mode;
- propose an algorithm for training a neural network model in batch mode for implementation on a GPU;
- perform numerical studies

**Choosing a neural network model for mapping audit sets.** In the work, the Forward-only Counterpropagating Neural Network (FOCPNN), which is a non-recurrent static two-layer ANN, was chosen as a neural network [2]. FOCPNN output is linear.

FOCPNN advantages:

1. Unlike most ANNs are used to reconstruct another sample using hetero-associative memory.
2. Unlike bidirectional associative memory and the Boltzmann machine, it works with real data.
3. Unlike a full counterpropagating neural network, it has less computational complexity (it does not perform additional reconstruction of the original sample).

FOCPNN model performing mapping of each input sample  $\mathbf{x} = (x_1, \dots, x_{N^x})$  to output sample  $\mathbf{y} = (w_{i^*1}^{(2)}, \dots, w_{i^*N^y}^{(2)})$ , is represented as

$$i^* = \arg \min_i z_i, z_i = \sqrt{\sum_{k=1}^{N^x} (x_k - w_{ki}^{(1)})^2}, i \in \overline{1, N^{(1)}}, \quad (1)$$

where  $w_{ki}^{(1)}$  – connection weight from the  $k$ -th element of the input sample to the  $i$ -th neuron,

$w_{i^*j}^{(2)}$  – connection weight from the neuron-winner  $i^*$  to  $j$ -th element of output sample,

$N^{(1)}$  – the number of neurons in the hidden layer.

**Criterion choice for assessing the effectiveness of a neural network model for mapping audit sets.** In this work for training model FOCPNN was chosen target function, that indicates selection of the vector of parameter values  $W = (w_{11}^{(1)}, \dots, w_{N^x N^{(1)}}^{(1)}, w_{11}^{(2)}, \dots, w_{N^{(1)} N^y}^{(2)})$ , which deliver the minimum mean square error (difference between the model sample and the test sample)

$$F = \frac{1}{P_N Y} \sum_{\mu=1}^P \| \mathbf{y}_\mu - \mathbf{d}_\mu \|_W^2 \rightarrow \min, \quad (2)$$

where  $\mathbf{y}_\mu$  –  $\mu$ -th output sample according to the model

$\mathbf{d}_\mu$  –  $\mu$ -th test output sample.

**Training method for neural network model in batch mode.** The disadvantage of FOCPNN is that it does not have a batch learning mode, which leads to reducing of the learning speed. For FOCPNN was used concurrent training (combination of training with and without a teacher). This work proposes training FOCPNN in batch mode.

First phase (training of the hidden layer) (steps 1-6).

The first phase allows you to calculate the weights of the hidden layer  $w_{ki}^{(1)}$  and consists of the following blocks (Fig 1).

Second phase (training the output layer) (steps 7-12). The second phase allows you to calculate the weights of the output layer  $w_{ij}^{(2)}$  and consists of the following blocks (Figure 2).

**Algorithm for training neuron network model in batch mode for implementation on GPU.** For the proposed method of training FOCPNN on audit data example, examines the algorithm for implementation on a GPU with usage of CUDA parallel processing technology.

**Numerical research.** The results of the comparison of the proposed method using GPU and the traditional FOCPNN training method are presented in Table 1.

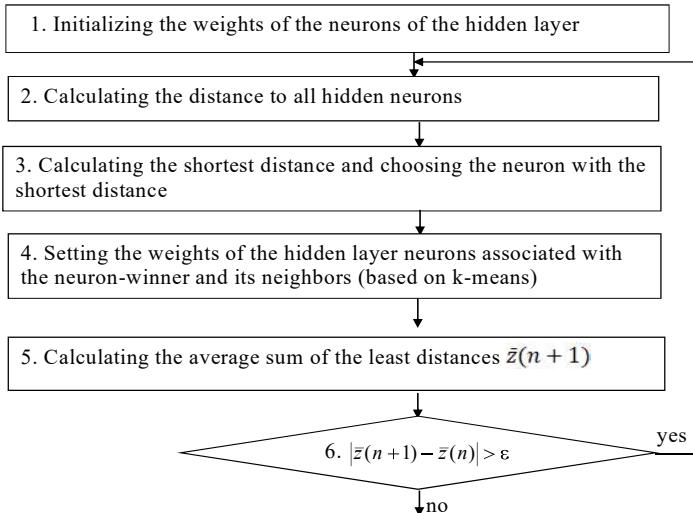


Figure 1. The sequence of steps in training method of FOCPNN in batch mode (the first phase)

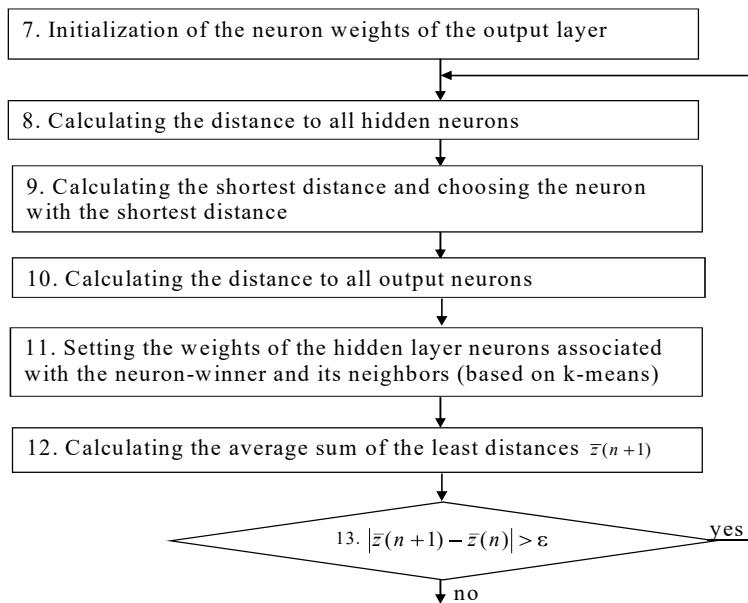


Figure 2. Sequence of procedures for the FOCPNN training method in batch mode (second phase)

Table 1

Comparison of the computational complexity of the proposed and traditional training methods of FOCPNN

Feature	Method	
	proposed	traditional
Computational complexity	$O(n_1^{max} + n_2^{max})$	$O(PN^{(1)}n_1^{max} + (PN^{(1)} + P)n_2^{max})$

Evaluation of computational complexity of the proposed method using the GPU, and the traditional method of teaching FOCPNN were based on the number of calculation distances, computing of which is the most consuming part of method. Moreover,  $n_1^{max}$  – the maximum number of iterations of the first training phase,  $n_2^{max}$  – the maximum number of iterations of the second training phase,  $N^{(1)}$  – the number of neurons in the hidden layer,  $P$  – the power of the training set.

**Discussion.** The traditional FOCPNN learning method does not provide support for batch mode, which increases computational complexity (Table 1). Proposed method eliminates this flaw and allows for approximate increase of learning rate in  $PN^{(1)}$ .

### Conclusion

1. The urgent task of increasing the effectiveness of audit in the context of large volumes of analyzed data and limited verification time was solved by automating the formation of generalized features of audit sets and their mapping by means of a forward-only counterpropagating neural network.

2. For increased learning rate of forward-only counterpropagating neural network, was developed a method based on the k-means rule for training in batch mode. The proposed method provides: approximately increase learning rate in  $PN^{(1)}$ , where  $N^{(1)}$  is the number of neurons in the hidden layer and  $P$  is the power of the learning set.

3. Created a learning algorithm based on  $k$ -means, intended for implementation on a GPU using CUDA technology.

4. The proposed method of training based on the  $k$ -means rule can be used to intellectualize the DSS audit.

Prospects for further research is the study of the proposed method for a wide class of artificial intelligence tasks, as well as the creation of a method for mapping audit features to solve audit problems.

### References:

1. T.V. Neskorodieva. Postanovka elementarnykh zadach audytu peredumovy polozhen bukhhalterskoho obliku v informatsiinii tekhnologii systemy pidtrymky rishen (Formulation of elementary tasks of audit subsystems of accounting provisions precondition IT DSS). Modern information systems. 3(1), 48–54, 2019. doi:10.20998/2522-9052.2019.1.08
2. A. Fischer, C. Igel. Training Restricted Boltzmann Machines: An Introduction Pattern Recognition, vol. 47. pp. 25-39, 2014.

**<sup>1</sup>Yuriii Samokhvalov**

Doctor of Technical Sciences, Professor, Professor of the Department of Intellectual Technologies.

**<sup>2</sup>Bohdan Bondarenko**

Master student of the Intellectual Technologies Department

<sup>1, 2</sup> Taras Shevchenko National University of Kyiv

## USE OF NEURAL NETWORKS IN INFORMATION RETRIEVAL SYSTEMS

**Abstract.** The issues of using neural networks in modern information retrieval systems are considered. A comparison of classical methods of information retrieval and neural networks is performed. The efficiency of neural networks application for information retrieval is substantiated.

**Keywords:** information retrieval systems, neural networks, information retrieval, DSSM network, semantic search.

Machine learning plays an important role in many modern information retrieval systems. Recently, such systems have begun to use deep learning. The rapid pace of modern deep learning research has given rise to many different approaches in information retrieval. The use of neural networks has greatly increased the efficiency of information retrieval systems.

Many experts, such as Tom Kenter, Christophe Van Gysel and others, have dealt with the use of neural networks for information retrieval. A large number of international conferences, seminars and round tables are devoted to this topic. However, the use of neural networks in information retrieval systems needs further development.

Information retrieval systems evaluated the document's relevance to the query based on word matches. As the amount of information increased, there was a need to sort it by relevance. Classic algorithms for estimating the importance of a word in the context of a document, such as TF-IDF and Okapi BM25, have appeared [1].

Classical methods of information retrieval are able to study complex relationships in text, but they use only a fixed set of features designed for search queries and documents [1]. This shortcoming is absent in neural networks, the deep hierarchical structure of which allows to analyze data and generate connections without human intervention.

Neural networks are actively used in image search. Using a network with a specially constructed architecture, image is converted into a vector in N-dimensional space. The query, which can be either text or image, is also converted to a vector. To calculate relevance, the two resulting vectors are compared with each other. The closer one vector is to another, the more the image matches the query.

Using neural networks to search for text is a complex and resource-intensive process, but it justifies itself by showing higher efficiency than with classical methods. The network must find hidden semantic relations to process the request for documents at the semantic level, while keyword-based comparisons often fail. It has to represent

the text of the request and the text of the document title in the form of vectors, scalar multiplication of which would reflect the relevance of the document to the request. To do this, the network has to generate similar vectors for semantically close texts.

A significant breakthrough in the use of neural networks in information retrieval was made by experts from Microsoft Research. They introduced their network for modeling semantic similarity between two text strings, which was called DSSM (Deep Semantic Similarity Model) [2]. The text of the request and the document title are fed to the model input. They are divided into trigrams. This allows representing the text as a vector of several thousand elements that have a value of 0 or 1. Trigrams present in the input text take the value of 1, otherwise - 0. The input vectors are processed by three following hidden layers that have a size of 300 - 300 - 128 neurons respectively. Thus, hidden layers convert the input vector into a vector of hidden semantic features [3]. The output of the model is the result of scalar multiplication of these vectors for the query and the title of the document. The goal of training in this case is to maximize the output value of the similarity function.

The results of the DSSM neural network were compared with other methods of information retrieval according to the NDCG metric, which is used to measure the quality of data ordering in search systems. The neural network has showed the best result in this comparison [2]. At the same time, the performance of the network with such architecture can be significantly improved, for example, by adding to the input layer not only trigrams, but whole words and phrases.

Besides DSSM, other neural network architectures exist and are being developed to search for semantic relations in a text, for instance, MT-DNN, UniLM, and others [4]. An improved version of DSSM is used in various search engines and shows high performance results.

To sum up, the ability of neural networks to find deep semantic connections allows us to build high quality data models. Deep neural networks can learn to process images and are being used in image search. Simple similarity functions, such as the cosine of similarity, can be applied to data generated by neural networks to detect semantically similar words, sentences, paragraphs, and so on. In addition, neural networks can optimize an information retrieval system.

### References:

1. Thiago Akio Nakamura, Pedro H. Calais, Davi de Castro Reis, André Paim Lemos. An anatomy for neural search engines. *Information Sciences*, Elsevier, vol. 480. April 2019
2. Po-Sen Huang, Xiaodong He, Jianfeng Gao. Learning Deep Structured Semantic Models for Web Search using Clickthrough Data. At ACM International CIKM. October 2013.
3. Nguyen, Gia-Hung and Tamine, Lynda and Soulier, Laure and Souf, Nathalie. Toward a Deep Neural Approach for Knowledge-Based IR. 2016.
4. Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Jianfeng Gao. Unified Language Model Pre-training for Natural Language Understanding and Generation. NeurIPS 2019. 2019.

**<sup>1</sup> Serhiy Semerikov**

Doctor in pedagogy, Professor (Full)

**<sup>2</sup> Hanna Kucherova**

Doctor in Economics, Professor

**<sup>3</sup> Vita Los**

PhD in Economical Sciences, Associate professor

**<sup>4</sup> Dmytro Ocheretin**

PhD in Economical Sciences, Associate professor

<sup>1</sup> *Kryvyi Rih State Pedagogical University*

<sup>2</sup> *Classic Private University*

<sup>3,4</sup> *Zaporizhzhia National University*

## **NEURAL NETWORK ANALYTICS AND FORECASTING THE COUNTRY'S BUSINESS CLIMATE IN CONDITIONS OF THE CORONAVIRUS DISEASE (COVID-19)**

**Abstract.** The paper proposes an approach to modeling the business climate of the country, which is based on the principles of information transparency, and makes it possible to assess the development trends of the studied indicator in conditions of the COVID-19. This approach has been tested on the example of Ukraine. The results obtained make it possible to analyze the cyclical development of the country's economy with high accuracy and reliability even under quarantine restrictions.

**Keywords:** Business climate, Business confidence index, Correlation analysis, Socio-economic indicators, Taxonomic model, Neural network model, COVID-19.

The dynamism of changes in the business climate of the countries of the world is accompanied by the increasing uncertainty of the external environment and internal disturbances of socio-economic systems. This is a reaction to new conditions of functioning and development, the emergence of which is due to the global pandemic and quarantine restrictions. The sensitivity of the business climate to such changes is high, therefore, the trends in the indicators that characterize it require system monitoring, thorough and multidimensional data analysis, and increased forecasting accuracy without time delay. This ensures that proactive management decisions are made on time in the context of the impact of COVID-19, which determines the goal and the task of this research.

One of the key indicators that determine the country's business climate is the business confidence index (BCI). The assessment of the business climate is based on the results of generalizing the opinions of business entities about their expectations of the dynamics of changes in production, demand, reserves, the general socio-economic state in the country. Therefore, the results of surveys of business entities, which underlie the formation of the BCI, determine the subjectivity, vagueness, and poor structuredness of the constructed index, which well-known researchers are trying to overcome. Despite the obvious subjectivity of the methodic approach to assessing the business climate of countries, scientists have repeatedly proved the close relationship of

the series of its values with the dynamics of macroeconomic indicators.

To solve the problem of predicting trends in the business climate of countries as a tool for strategic analysis, a wide range of forecasting tools is actively used. The paper proposes an approach to modeling the business climate of the country, which is based on the principles of information transparency, and makes it possible to assess the development trends of the studied indicator in conditions of the COVID-19.

The authors' previous research was based on statistical methods, however, the popularity and efficiency of neural network technologies proved the expediency of their application to solving problems of forecasting the country's business climate. The authors proposed to predict the business confidence index (BCI) using a methodological approach, which includes the step-by-step construction of taxonomic and neural network models.

As a result of using the methodological approach, a time series of quarterly values of the business confidence index in Ukraine was predicted for the period 2008-2020. The forecast was based on socio-economic indicators selected by their closeness to the business confidence index, namely: Retail sales, Industrial production, Steel production, Export, Imports and GDP annual growth rate. The forecast value of the composite index of business activity is obtained as follows:

$$\begin{aligned} \overline{BCI}_i = & W_1 \cdot RS_i + W_2 \cdot IP_i + W_3 \cdot SP_i + W_4 \times \\ & \times Exports_i + W_5 \cdot Im\ ports_i + W_6 \times \\ & \times GDP\_AGR_i = 0,218 \cdot RS_i + 0,176 \times \\ & \times IP_i + 0,128 \cdot SP_i + 0,096 \cdot Exports_i + \\ & + 0,096 \cdot Im\ ports_i + 0,286 \cdot GDP\_AGR_i \end{aligned} \quad (1)$$

The quarterly values of socio-economic indicators for the past thirteen years (2008-2020) were taken as input data. The results of taxonomic analysis established that the GDP annual growth rate and retail sales have the greatest impact on the business confidence index. A forecast has been built for the trend of changes in the business confidence index (forecast accuracy of 89.38%), which proves the similarity of development trends in the country's business climate.

In addition, the most important thing is that the tendency of the studied indicators is identical, in particular, during the period of the emergence of crisis phenomena (beginning of 2009, end of 2014, beginning of 2015, period of the COVID-19 in second quarter of 2020), the decrease in the level of indicators is similar, which suggests that there is a real possibility of using the alternative to business confidence index, which calculated by the taxonomic method of in order to predict the business climate in conditions of limited information transparency.

Having determined the predicted value of the business confidence index (BCI) using a taxonomic model in accordance with the proposed methodology, the next step is forecasting using neural network technologies. An artificial neural network consists of one hidden layer, which contains two neurons, and one output layer (business confidence index). The number of variables in the input layer corresponds to the

number of selected economic indicators for modelling, i.e. six. Thus, to predict business confidence index, used the neural network of the type [6–2–1].

The activation function of the hidden layer is the sigmoid function. This type of function is often used for modeling and the outgoing values of such a function continuously fill the range from 0 to 1. The learning algorithm is the back-propagation error algorithm (Back-Propagation) with a learning rate of 0.1. The difference between the reference and the real output of the network is less than 0.05 (learning rate). The number of learning iterations is 10000.

Formation and analyzing a neural network model were carried out on the basis of the analytical platform Deducor Studio Academic 5.3, which allows you to perform all the steps of data mining from their loading and visualization to building and evaluating the quality of finished models. The time period for analysis is 50 values (first quarter of 2008 - second quarter of 2020). The training set consists of 88% of the data (44 values, time period between first quarter of 2008 and fourth quarter of 2018), and the test set – 12% of data (6 values, time period between first quarter of 2019 and second quarter of 2020).

The constructed neural network model with training capabilities showed the best results in the accuracy and quality of the forecast (forecast accuracy of 96.22%). A decrease in the business confidence index is predicted in third quarter 2020 (will be 87.65). The sharp decrease in the dynamics of the indicator in the studied forecast period is also explained by the influence of the negative consequences of COVID-19 and the introduction of quarantine restrictions in the country and the world.

The article examines the risks of deteriorating the business climate in Ukraine, as a result of which such preconditions as: the weakness of the judicial system, corruption, political and economic instability, the growth of tax pressure, changes in legislation, the slowdown and curtailment of reforms are identified. The situation due to the introduction of prolonged restrictive measures due to COVID-19 was worsened. Insufficient attention has been established in Ukraine to the issues of the ecological system's influence on the formation of the country's business climate, which requires a separate research.

The results obtained make it possible to analyze the cyclical development of the country's economy with high accuracy and reliability even under quarantine restrictions.

The effectiveness of the proposed alternative approach is manifested in saving costs for generating input data for assessing the country's business climate by using official statistics instead of survey results, the subjectivity of which is much higher. In general, the implemented alternative approach is unified, can serve as the basis for further deepening the methodological provisions for studying the business climate of countries with high accuracy and reliability of the results. The prospect of the research is to determine the impact of COVID-19 and introduction of quarantine restrictions on the value and dynamics of the business climate in other countries.

The problem for the implementation of an alternative approach remains limited access to key statistics, which is the result of a policy of ensuring information transparency in different countries.

**References:**

1. S. Arslankaya, V. Öz. Sakarya, Time Series Analysis on Sales Quantity in an Automotive Company and Estimation by Artificial Neural Networks. University Journal of Science 22, 1482-1492 (2018). doi: 10.16984/saufenbilder.456518.
2. D. Ocheretin, V. Los, H. Kucherova, O. Bilska, An alternative approach to modeling the country's business climate in conditions of limited information. E3SWC 166 (2020): 13024. URL: [https://www.e3s-conferences.org/articles/e3sconf/abs/2020/26/e3sconf\\_icsf2020\\_13024/e3sconf\\_icsf2020\\_13024.html](https://www.e3s-conferences.org/articles/e3sconf/abs/2020/26/e3sconf_icsf2020_13024/e3sconf_icsf2020_13024.html).
3. L.A. El'shin, Mechanisms for the identification of business cycles of regional economic systems based on cross-correlation analysis. Regional Economics: Theory and Practice 15(8), 1540-1551 (2017). doi: 10.24891/re.15.8.1540.
4. V. Los, D. Ocheretin, H. Kucherova, O. Bilska, Neural network technology forecasting the country's business climate, in: Hryhoruk, P., Khrushch, N. (eds.), Proceedings of the 6th International Conference on Strategies, Models and Technologies of Economic Systems Management (SMTESM 2019) 95, pp. 320-324. Atlantis Press (2019). DOI: 10.2991/smtesm-19.2019.62.
5. M. R. Safiullin, L.A. El'shin, A.I. Shakirova, Evaluation of business and economic activity as a short-term forecasting tool. Herald of the Russian Academy of Sciences 82(4), 623-627 (2012). doi: 10.1134/S1019331612040053
6. S. Feuerriegela, J. Gordon, News-based forecasts of macroeconomic indicators: a semantic path model for interpretable predictions. European Journal of Operational Research 272(1), 162-175 (2019). doi: 10.1016/j.ejor.2018.05.068.
7. H. F. Mendonca, A. F. G. Almeida, Importance of credibility for business confidence: evidence from an emerging economy. Empirical Economics (2018). doi: 10.1007/s00181-018-1533-5.
8. H. Sakaji, R. Kuramoto, H. Matsushima, K. Izumi, T. Shimada, K. Sunakawa, Financial Text Data Analytics Framework for Business Confidence Indices and Inter-Industry Relations, in: Proceedings of the First Workshop on Financial Technology and Natural Language Processing (FinNLP@IJCAI 2019), pp. 40-46. Macao, China (2019).
9. V. Los, D. Ocheretin, Construction of business confidence index based on a system of economic indicators, in: Semerikov, S., Soloviev, V., Kibalnyk, L., Chernyak, O., Danylchuk, H. (eds.) SHS Web of Conference. The 8th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2 2019), vol.65, pp. 1-6. SHS Web of Conferences (2019). doi: 10.1051/shsconf/20196506003.

**Marianna Sharkadi**

PhD, associated professor

*Uzhhorod National University*

## **NEURO-FUZZY MODELING OF LEVEL ASSESSMENT IN THE SYSTEM OF FINANCIAL-ECONOMIC SECURITY**

**Abstract.** The solution of the actual problem of determining the level of financial-economic security for companies through the prism of neuro-phase modeling is presented. In this study, it is proposed to use a multilayer neural network, each layer of which solves a number of problems. The proposed approach will make it possible to determine the level of financial security of the company at different times of its operation. The developed model allows each company to use its own set of financial indicators to determine the level of security. Each layer of the neural network is an autonomous unit that allows you to develop a network.

**Keywords:** security level, neural network, fuzzy modeling.

### **Introduction**

The use of information technology in various fields of human activity is accompanied by the development of intelligent systems that use the connection of knowledge in the general case with the outside world. The solution to any problem is related to specific subject areas, which are usually badly or poorly structured. During the design and development of an intelligent system, knowledge undergoes a similar transformation of data - from more generalized sets to narrower, specific to a given subject area. In the development of intelligent systems, knowledge of the specific subject area for which the system is developed is rarely complete and reliable.

One of the most promising and active areas of applied research in the field of management and decision-making in poorly structured systems is fuzzy modeling. The fuzzy modeling methodology specifies the methodology of system modeling in relation to the process of construction and application of fuzzy models of complex systems. Every year, the range of fuzzy models and methods expands to cover various new areas. The essence of fuzzy mathematical modeling is that the elements of the study are not numbers, but some fuzzy sets or combinations thereof. At the heart of this approach is not traditional logic, but logic with fuzzy truth, fuzzy connections and fuzzy inference rules.

A significant number of important problems in supporting management decisions that arise in various areas of human activity, is reduced to the task of assessing various kinds of phenomena and processes. When designing and managing a complex socio-economic system, a problem arises when a person is unable to give accurate and then practical values of judgments about their behavior.

The existence of any state in today's globalized world depends on its economic security, which is one of the important components of national security as a whole. One of the main segments of economic security, which significantly affects its level, is the financial segment, which is the set of financial indicators of the economic entity, which are combined into a global indicator. Forecasting this indicator is a complex analytical

and computational process and requires a detailed study of development trends and prediction of the impact of the studied factor's components on the level of the state's economic security.

The urgency of the work lies in the development and study of models and methods for obtaining multicriteria assessment using neuro-fuzzy technologies, which is currently undisclosed sufficiently.

### Formulation of the problem

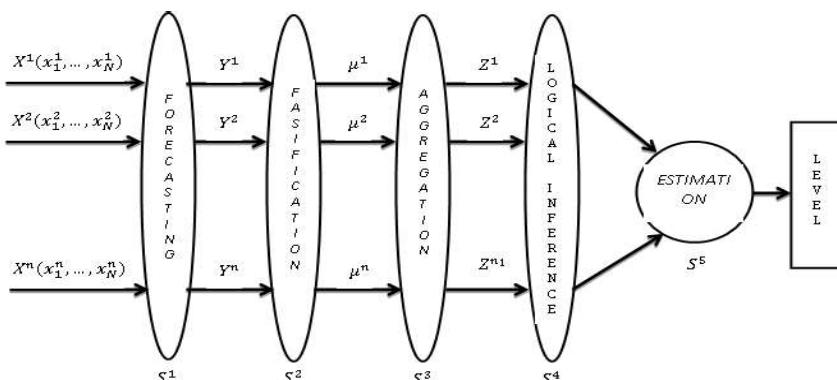
For an economic entity, economic security is considered as a state in which its economic development and stable activity is ensured, the protection of its financial and material resources is guaranteed. Ensuring financial security involves planning, forecasting and anticipating many factors of the internal and external environment. At the same time, a systematic, comprehensive approach based on the effective use of appropriate information and analytical support, logic and modeling to involve the modern mathematical apparatus is extremely important.

*The general statement of the problem (task) can be presented as follows.* Let a set of quantitative and qualitative indicators of its functioning be known for a certain subject of economic management, as well as the history of these indicators for certain periods of time is known. There is a task to provide an assessment of the entity's economic security level.

We formulate the statement of the evaluation problem as follows. Suppose we have at the entrance some object of study, which is evaluated by many indicators  $K = (K_1, K_2, \dots, K_m)$ . Indicators  $K$  can be a whole system of criteria and models. Each indicator is a quantitative estimate, which can be obtained, for example, using financial reporting models [1].

Based on a set of estimates  $K = (K_1, K_2, \dots, K_m)$ , it is necessary to establish the level of financial security of the object.

To solve the formulated problem, a model of a neuro-fuzzy network is proposed (Figure 1), which consists of a set of successive layers, each of which solves a number of specific classes of problems[2].



**Figure1:** The structure of the neuro-fuzzy network.

## Conclusions

A study of the current task of determining the assessment of the level of economic security for economic entities of the socio-economic system and the state as a whole, taking into account key management indicators. The following results were obtained for the first time.

The structure of a multilayer neural artificial network with a fuzzy mathematical model for determining the security level assessment in the financial and economic system is proposed. On the basis of this network the information technology is developed, which allows to forecast step by step on the basis of input statistical information on financial indicators the values of these indicators for future periods, to determine their efficiency through membership functions, to obtain aggregated fuzzy estimates on certain groups of criteria [3]. Conclusion to obtain a fuzzy integrated assessment of the safety level, which with the help of defasification methods reduce to a clear value.

## References

1. Walsh Karyan. Key Management Indicators: A Complete Guide to Working with Critical Numbers Managing Your Business, 4th ed., Companion Group, Kyiv, K, 2010.
2. Yu. P. Zaichenko. Fuzzy models and methods in intelligent systems [Text]: textbook. Manual, Slovo, Kyiv, K, 2008.
3. M.M. Malyar. Models and methods of multicriteria limited-rational choice: Monograph, RA "OUTDOOR-SHARK", Uzhhorod, 2016.

<sup>1</sup>**Petro Soroka**

PhD in Physical and Mathematical Sciences, Associate Professor,  
Associate Professor of the Intellectual Technologies Department

<sup>2</sup>**Serhii Krasnovidov**

Master student of the Intellectual Technologies Department

<sup>1,2</sup>*Taras Shevchenko National University of Kyiv*

## BUSINESS ANALYTICS INFORMATION TECHNOLOGIES FOR ANALYSIS OF THE ACTIVITY OF A COMMERCIAL ORGANIZATION

**Abstract.** This paper considers modern business analytics system approaches for analysis of a commercial organization's activity, including data mining and market basket analysis.

**Keywords:** business analysis, business analytics, data mining, market basket analysis.

Current economic situation, given the global crisis, does not allow large and small enterprises to operate at full capacity. In such conditions, it's important for each commercial organization to maintain its profitability and competitiveness in the market. Business analytics systems help to cope with this task allowing managers to make informed decisions reducing significant amount of their time in finding and analyzing the necessary information.

Nowadays, one of the most effective tools for managing a company is business analysis. It allows to get a full view of your organization and the prospects for its development. Business analysis is a set of tasks and techniques used both to understand the activities of the organization as a whole and to obtain effective management decisions [1].

In our time, the amount of data has reached such a mark that even a group of people is not able to analyze them on their own. However, large raw data sets often contain knowledge that can be used in decision making. For their processing and analysis Data Mining technology is used extracting previously unknown, non-trivial, practically useful and interpretable knowledge necessary for decision-making [2].

Consider some methods of Data Mining: sequence, classification, regression, association. The sequence is to find a temporal pattern between events, i.e. such a relationship that if event X occurs, then after some time event Y will occur too. For example, after purchasing a car, the driver will most likely take out an insurance policy, then buy a first aid kit, spare wheel etc.

Classification helps to identify features that define a group of certain objects. The purpose of regression is to find a function of given variables that would determine the range of valid values. A prerequisite for such an analysis is the relationship between variables [3].

The task of the association is to identify patterns between related events, i.e. rules of the type «event Y follows event X», or association rules. Analysis of purchases and goods sold together is often called market basket analysis.

Market basket analysis focuses on the study of a large set of information in search of trends, patterns and relationships that contribute to effective decision making [4]. A market basket is a set of goods purchased by a customer. The most popular algorithm that solves this problem is still the Apriori algorithm, which has been repeatedly improved by various researchers. There are now a large number of software products and commercial technologies, such as Oracle Data Miner, Tableau and Deductor Studio, that effectively solve this problem.

Deductor Academic software was used in the research to form a data warehouse, analyze the activities of a commercial organization – a shop selling auto parts, and generate its annual reports. Initially, the data cleansing removed the records of accidental transactions (there were about 20 such records) and completed some records replacing ‘auto part code’ nulls with a position code from the same department of similar price. After that, an OLAP cube was formed by product groups. The products that bring 80% of the profits are automotive oils, automotive chemicals, cosmetics and small spare parts. Friday was found as the busiest day of week in the store.

To identify sales increasing techniques, there were also found the association rules of the most popular pairs of goods pre-configured with a minimum support of 10% and a minimum probability of 80%. As a result, 4 rules were found: 2 of them had already been implemented in the form of respective departments’ planning (trivial rules), the other 2 were recommended to the store administration.

*Conclusions.* Modern business analysis technologies allow a commercial organization to consolidate knowledge about its activities, make OLAP-slices of sales, manage store load and design its own data warehouses. Association rules allow to determine its most popular sets of goods and see the options of redistribution of less popular goods. Therefore, based on the analysis, the company can optimize its business processes, which will lead to increasing sales.

## References

1. Franks, B. (2014). *Taming The Big Data Tidal Wave*. Hoboken, NJ: John Wiley & Sons, Inc.
2. BABOK Guide 2.0. Retrieved from: <https://www.iiba.org>
3. Rouaud, M. (2013). *Probability, Statistics and Estimation (Short Edition)*. Retrieved from: <http://www.incertitudes.fr/book.pdf>
4. Babin D. V. (2004). *Genetic algorithm for solving the problem of market basket analysis* (Institute for Artificial Intelligence Problems). Retrieved from [http://www.iai.dn.ua/public/JournalAI\\_2006\\_4/Razdel3/06\\_Babin.pdf](http://www.iai.dn.ua/public/JournalAI_2006_4/Razdel3/06_Babin.pdf)

**<sup>1</sup>Petro Soroka**

PhD in Physical and Mathematical Sciences, Associate Professor,  
Associate Professor of the Intellectual Technologies Department

**<sup>2</sup>Roman Savchenko**

Master student of the Intellectual Technologies Department

<sup>1,2</sup>Taras Shevchenko National University of Kyiv

## MACHINE LEARNING METHODS FOR SPORT RESULT PREDICTION

**Abstract.** The purpose of this paper is to consider various approaches of utilizing machine learning methods for sport result prediction.

**Keywords:** machine learning, API, betting market, hierarchical scoring structure, match, set, game, point.

As a result of exponentially growing computing capabilities and data availability machine learning now is being applied to almost all areas of life, from financial institutions and medicine to self-driving vehicles. At the same time, its utilizing for sport result prediction and the related betting market is given relatively less attention. More traditional statistical approaches still dominate this area. In addition, one of the main areas of investigation has been the football market, and tennis has been in the background, although it is one of the most popular sports. The potential profit, as well as scientific interest, encourages to search for effective methods of predicting result of tennis matches.

Most modern approaches to tennis prediction use a hierarchical scoring structure to determine the probability that a player would win a match. Assuming the points have independent uniform distribution, the expressions only need the probability that two players will win a point on their own serve. From these basic statistics, it becomes possible to deduce the probability that a player would win a game, then a set, and finally a match.

Barnett [1], O'malley [2], and Knottenbelt [3] in their researches defined hierarchical models for calculating the probability of winning a point on serve using only matches with players' common opponents instead of all previous opponents. This reduces bias due to differences in the level of opponents. Madurskaya [4] improved the model by using different winning probabilities in different sets and allowing the model to reflect how a player's performance changes during a match.

This mathematical approach represents the level of players using only one value (points won), it does not take into account a large number of extremely important parameters. For example, a player's susceptibility to a particular opponent's game strategy, time after the last injury or fatigue accumulated from a previous match. In addition, there are such important characteristics of the match itself as location,

weather, surface, etc. Given the huge amount of diverse historical data, an alternative approach to predicting the results of tennis matches can be based on machine learning. Player parameters and features of the match itself, combined with its outcome, can form a dataset for machine learning application. Supervised learning algorithms can be applied to determine the function of predicting the results of future matches.

The paper examines not common approach to predicting tennis matches - predicting the amount of total games played in the second set based on the result of the first set. For this purpose, a set of historical data was gathered and pre-processed. The goal was to classify matches into two classes: with the total of games in the second set less than 9.5 or more than 9.5 games. To solve it, the following machine learning methods were used: k-nearest neighbors, support vector machine, multilayer perceptron. The return on investment when tested on real data ranged from -17.16% to 18.67%, which is close to the results of existing researches.

For the convenience of applying prediction result in practice an automatic notification system was developed. Console app based on C#.NET in combination with the Telegram API allows to collect information about current matches from open resources in real time, analyze it and send notifications about potentially profitable matches to the Telegram messenger.

*Conclusion.* In this paper, machine learning methods were applied to predicting the result of tennis matches. Also were found ways to improve result by using more complete and informative data for training as well as optimizing the configuration of the neural network. This indicates the significant potential of machine learning for predicting the results of sport competitions.

## References

1. T. Barnett and S. R. Clarke. Combining player statistics to predict outcomes of tennis matches. *IMA Journal of Management Mathematics*, 16:113–120, 2005.
2. J. A. O’Malley. Probability Formulas and Statistical Analysis in Tennis. *Journal of Quantitative Analysis in Sports*, 4(2), 2008.
3. W. J. Knottenbelt, D. Spanias, and A. M. Madurska. A common-opponent stochastic model for predicting the outcome of professional tennis matches. *Computers and Mathematics with Applications*, 64:3820–3827, 2012.
4. A. M. Madurska. A Set-By-Set Analysis Method for Predicting the Outcome of Professional Singles Tennis Matches. Technical report, Imperial College London, London, 2012.

**<sup>1</sup> Bohdan Sus**

PhD, Assist. Prof. of Nanophysics of Condensed Matter Dept.

**<sup>2</sup> Ilona Revenchuk**

PhD, Assoc. Prof. of Software Engineering Dept.

**<sup>3</sup> Oleksandr Bauzha**

PhD, Assoc. Prof. of Computer Engineering Dept.

<sup>1,3</sup> Taras Shevchenko National University of Kyiv

<sup>2</sup> Kharkiv National University of Radio Electronics

## **MODEL OF IMPLEMENTATION VIRTUAL LABORATORY WORK FOR SUPPORTING EDUCATIONAL PROCESS**

**Abstract.** Due with the current pandemic, online education is becoming very popular and it is one of the good ways to get access to educational resources without a threat to health. The article discusses Model of Implementation Online Educational Systems for Supporting Educational Process at the organization (schools, universities etc.). E-learning laboratory work are mainly remote and should be integrated with Learning Management System for more effective learning process.

**Keywords:** Learning Management System, E-learning, Virtual Laboratory Work

The proliferation of new technologies and internet tools is fundamentally changing the way we live and work. The lifelong learning sector is no exception with technology having a major impact on teaching and learning. This in turn is affecting the skills needs of the learning delivery workforce. From Moodle to Edmodo to Inquisiq3, there's a lot of tools we can use to manage, track, and deliver educational courses and training programs in our universities.

All Supporting Technology that could be used to deliver educational courses and programs could be classified to several groups by their role in the educational process:

- Online Education Systems - OES (Online Education, E-learning, OES, Integrated OES, Standards Specifications);
- Content Creation Tools -CCT (CCT, Authoring Tools, Assessment Tools, Learning Content Management Systems (LCMS), Learning Objects);
- Learning Management System - LMS (LMS, Learning Platform, Virtual Learning Environment (VLE), Learning Service, Provider (LSP));
- Management System (MS) (Student Management System, Enterprise Resource Planning System, Human Resource Information System, Knowledge Management System, Competency Management System).

There are many terms for online education. Some of them are: virtual education, Internet based education, web-based education, and education via computer-mediated communication. The Web-edu project uses a definition of online education that is based on [1] definition of distance education. In general, online education is characterized by:

- the separation of teachers and learners which distinguishes it from face-to-face education;

- the influence of an educational organization which distinguishes it from self-study and private tutoring;
- the use of a computer network to present or distribute some educational content;
- the provision of two-way communication via a computer network so that students may benefit from communication with each other, teachers, and staff.

E-learning is here defined [1] as interactive learning in which the learning content is available online and provides automatic feedback to the student's learning activities. Online communication with real people may or may not be included, but the focus of e-learning is usually more on the learning content than on communication between learners and tutors.

Very often the term e-learning is used as a synonym for online education. E-learning includes a set of applications and processes, as instance Web-based learning, computer-based learning, virtual classrooms and laboratory, digital collaboration. It includes the delivery of information or content via Internet.

Therefore online education is covering all systems that support it.

Beside content, integrated learning system - ILS includes a tools for making assessments, notes, report creation, and user files that help to evaluate and identify a monitoring learning progress via students learning needs and competences.

CCT are the tools that course designers and teachers use to create the content in online education courses. The CCT are used to develop learning material (text, slides, graphics, pictures, animations, simulations, assessments, audio, video etc.)

Authoring tools could be regarded as a subset of CCT. A software application, used by non-programmers, that utilizes a metaphor (book, or flow chart) to create online courses [2].

The LCMS is a computer application provide procedures to manage workflow in a collaborative in common environment such as for creation text and publishing, editing and modifying content, organizing, deleting as well as maintenance from a central interface. It use for creation teaching contents and materials for educational process in online or as blended learning.

The learning object is a reusable, media-independent chunk of information used as a modular building block for e-learning content. Learning objects are most effective when organized by a meta data classification system and stored in a data repository such as an LCMS.

The content-management systems focused in general more attention to the creation, developing, and managing content for online courses, and less - to the control of students experience.

Institutions use LMS software for support of organization of learning process, namely: to plan educational process, implement it, the evaluation educational process, support of access for the students and teachers, monitor the student learning.

Beside this, software provides course preparation, educational content and resources according to the standards or templates. It centralizes the delivery and tracking of student activities during education, as instance discussion, collaboration, to

do tasks, assessments. All these activities will control by the virtual environment of LMS that provides a user data protection.

Each LMS is different and it give to users different possibilities for implementation such as content-oriented, activity-oriented, network-oriented, linear, and branching. Some systems realize asynchronous instruction, while others are providing synchronous instruction [3].

There are many different types of LMS, or LCMS, for manage learning process and course content delivery. That why one is the important question is how to choose effective LMS or LCMS according to the organization needs. The main rule is to know how LMS will be deliver training materials to students, and then compare your organization needs with LMS opportunities by the LMS functionality.

There are many projects for the development of virtual laboratory work (VLW) in natural science. Laboratory work for cloud electronic learning environments with algorithms and methods for protecting information between devices using combined communication channels and embedded systems is discussed in [4-6].

To develop methods of virtual laboratory work wide use of interactivity with the implementation of interdisciplinary methods to learning is required. In the case of laboratory work, the essence of these approaches is to expand ways of virtual laboratory work on every step, the presence of self-control and methods of evaluation of outcomes. During laboratory work, students have to perform a task of conscious choice and means of experiment process. This may be the choice from the possible list of virtual devices and conditions of the experiment. An adaptive model of the virtual electronic laboratory work is presented on fig.1.

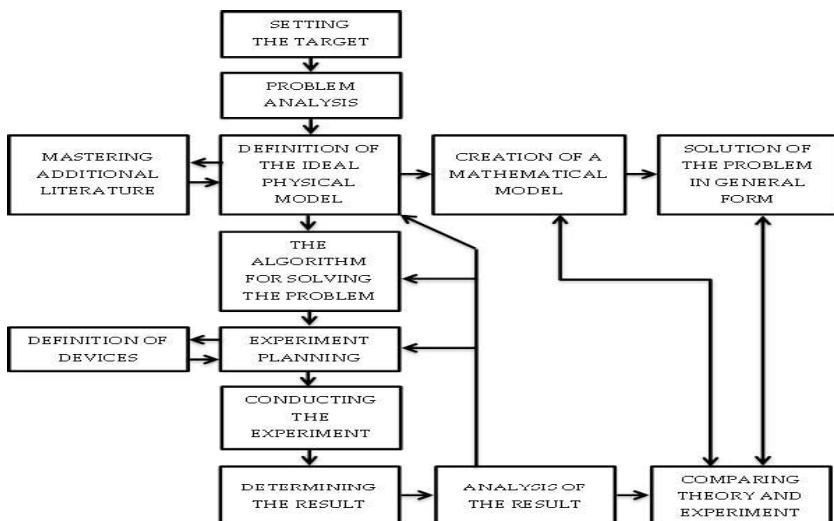


Figure 1 - Block diagram of the algorithm of functioning of electronic laboratory work

Methods of visual intercommunication, methods of modeling and data acquisition have been also realized for virtual laboratory development. Tasks are performed in real-time, with the illustration of complex processes and interactive analysis of additional data.

At any step of the process, the student can investigate the data from the current set of measurements.

Each element of the virtual device can interchange data with other elements of the VLW and external programs by the means of network technologies.

VLW can also be a good tool to get introduced with original laboratory devices with limited access.

*Conclusion.* VLW as a as custom eLearning implementation and design solution for virtual class are based on real experimental setups. Integration process of Moodle with VLW and variety of other programs with functional together as a system in further very useful to meet various needs of a website for eLearning, including developing courses using authoring tools and make it available through Moodle with the internal course booking system.

#### **References:**

1. Desmond J. Concepts: Problems in defining the field of distance education, 2009. URL: <https://www.tandfonline.com/doi/abs/10.1080/08923648809526619>.
2. D. Stewart, D. Keegan, B. Holmberg, Distance education: International perspectives, 2020. URL: [https://books.google.com.ua/books?hl=en&lr=&id=OujyDwAAQBAJ&coi=fnd&pg=PT7&ots=TL2D99Oz\\_J&sig=PCh7apz6xYcx4INPJegCFm9eiZE&redir\\_esc=y#v=onepage&q&f=false](https://books.google.com.ua/books?hl=en&lr=&id=OujyDwAAQBAJ&coi=fnd&pg=PT7&ots=TL2D99Oz_J&sig=PCh7apz6xYcx4INPJegCFm9eiZE&redir_esc=y#v=onepage&q&f=false)
3. LMS Selection Guide 2020. URL: <https://mindflash.com/resource-center?postType=literature>
4. Tmienova, N., Sus, B.: Hardware data encryption complex based on programmable microcontrollers. In: CEUR Workshop Proceedings, pp. 199–208 (2018). URL:<http://ceur-ws.org/Vol-2318/paper17.pdf>.
5. Bauzha, O., Sus, B., Zagorodnyuk, S., Stuchynska, N.: Electrocardiogram Measurement Complex Based on Microcontrollers and Wireless Networks. In: International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T, pp. 345-349. (2019)
6. Sus, B., Tmienova, N., Revenchuk, I., Vialkova, V.: Development of Virtual Laboratory Works for Technical and Computer Sciences. In: Damaševičius, R. and Vasiljevičienė, G. (eds.) Information and Software Technologies, pp. 383–394. Springer International Publishing, Cham (2019). URL:[https://doi.org/10.1007/978-3-030-30275-7\\_29](https://doi.org/10.1007/978-3-030-30275-7_29).

<sup>1</sup>**Nataliia Tmienova**

PhD in Physical and Mathematical Sciences, Associate Professor of the Intelligent Technologies Department, Associate Professor

<sup>2</sup>**Oleksandra Dulich**

Master student of the Intelligent Technologies Department

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## AUTOMATIC QUESTION GENERATION SYSTEM FOR UKRAINIAN-LANGUAGE TEXTS

**Abstract.** The problem of automatic generation of questions for Ukrainian-language texts is considered. The problems that arise when generating questions for Ukrainian-language sentences are presented, the analysis and comparison of the used methods are given. The main result is a software module for automatic question generation.

**Keywords:** natural language processing, text processing, tokenization, stemming, POS-tagging, automatic question generation, question generation methods, rule-based methods, text corpus.

We live in a time when the amount of information produced by humanity is greater than ever, and the amount of this data is growing every day. However, significant benefits can only be obtained from this information if this data is properly processed and analyzed. On the other hand, the number of tasks that humanity trusts computers to solve is growing at an incredible tempo. More and more processes are being automated. Thus, there are problems with computer analysis and natural language synthesis, and the need to improve natural language processing methods is growing and enjoys inexhaustible interest. This is especially relevant for Ukrainian realities, since, unfortunately, there is a lack of tools for processing the Ukrainian language, such as libraries for programming languages, marked corpora, dictionaries, and thesaurus.

Automatic question generation in natural language processing is one of the most urgent tasks of Computational Linguistics [1]. Systems with similar functionality are usually used in the field of education to test students' knowledge, namely when compiling questions on theoretical material [2]. Thus, the implementation of a system for automatic question generation is a promising area of work and can be used in the development of chatbots, compiling tests for online courses or distance learning, which are now in high demand during the pandemic.

The main problem of automatic generation of questions for Ukrainian-language texts is that before the stage of direct question generation, you need to perform a number of transformations with the input text in order for the computer to understand human speech at first, and only then generate a question. Performing these actions for the English language is easier due to the availability of appropriate tools for processing the English language. Meanwhile for the Ukrainian language, even at the preliminary stage of processing, a few problems appear, because the quality of understanding the language depends on many factors.

Thus, there is a need for pre-processing of the Ukrainian-language text: it is necessary to have the text processed at the semantic and syntactic levels, and for this, it is urgent to implement such tasks of graphemic and morphological analysis as determining the boundaries of sentences, words tokenization, lemmatization or stemming, stop words extraction and parts of speech tagging. It should also be noted that sometimes there is a problem in finding new ways to define key parts of a sentence when well-known algorithms for determining keywords.

Existing tools for preprocessing Ukrainian-language text, such as sentence boundary detection, tokenization, and parts of speech tagging, did not meet the expected results. So, to solve the problem of sentence boundary detection and tokenization, it was decided to resort to the NLTK library [3], namely, tools created for English and Russian, which were partially improved and adapted for the correct processing of Ukrainian-language texts. This library also provides the ability to add Ukrainian stop words, the step of getting rid of which is also quite important when analyzing text.

The morphological analyzer pymorphy2 was used for parts of speech tagging [4]. Work on the tagging ordinary words is performed by pymorphy2 tools based on the Russian-language OpenCorpora corpus, so it is not always possible to get the correct processing result for the Ukrainian language texts.

It was decided to use a text corpus-based generation method for question generation for the implementation of the software module. Unlike other methods, this one has no restrictions on the number of question types and format of text data. It can be divided into 3 stages: text preprocessing, sentence filtering, and question generation itself, which has a rule-based implementation. This approach is based on the development and extension of existing rules, so the system does not require a massive training corpus compared to the machine learning approach.

The result of implementing this method is a software module for automatic question generation, which receives text for input, performs text preprocessing, determines the keywords and key sentences for which questions are generated. The used rule-based approach is flexible and easy to debug. It also does not require a massive training corpus and shows high accuracy. It should be noted that adapting text processing tools of other languages to Ukrainian-language texts raises the possibility of errors during analysis, which is a disadvantage of this module, along with the need to improve and expand the existing rules for generating questions.

#### References

1. A.M. Kurtasov, A. N. Shvetsov. Program for generating educational tests based on a semantic approach // Proceedings of the International scientific and methodological conference "Informatization of engineering education" - INFORINO-2012. Moscow: publishing house of MEI. P. 71-74 (in Russian).
2. I. L. Bratchikov. Generation of test tasks in expert training systems // Bulletin of the peoples' friendship University of Russia. Series: Informatization of education. 2012. № 2. P. 47-60 (in Russian).
3. Bird, Steven, Edward Loper and Ewan Klein, Natural Language Processing with Python. O'Reilly Media Inc., 2009.
4. Korobov M.: Morphological Analyzer and Generator for Russian and Ukrainian Languages // Analysis of Images, Social Networks and Texts, pp 320-332, 2015.

<sup>1</sup> Dmytro Yakymenko

Postgraduate student (applicant)

<sup>2</sup> Iryna Tregubenko

PhD, Associate Professor

<sup>1,2</sup> Cherkasy State Technological University

## MODIFIED METHOD OF CONSTRUCTION OF INFORMATION IMAGE OF ELECTRONIC TEXT DOCUMENTS BY MEANS OF INTELLECTUAL DATA ANALYSIS

**Abstract.** The problems of analysis and processing of electronic documents are investigated in the work. The existing methods of analysis of text documents are analyzed. The study identified ways to improve existing methods. The article presents a mathematical model of the method.

**Keywords:** data mining, Kohonen maps, information retrieval image of documents, TF-IDF, array of values, cluster.

### 1. Formulation of the problem

The rapid increase in the number of electronic documents that is currently observed clearly shows that traditional mechanisms for processing electronic documents are not able to cope with the needs of users. This trend is noticeable both on the Internet and in corporate networks.

That is why there is a need for methods that will provide a quick and easy distribution of documents by category or by keyword.

Thus, we can identify the main problems associated with increasing the amount of information:

— the rapid growth of information contained on the Internet is the cause of more and more difficulties in finding the necessary documents and organizing them in the form of structured repositories;

— most technologies for working with text documents are focused on the organization of convenient work with information for humans, but there are virtually no opportunities to convey the semantic content of the text, ie there is no semantic indexing;

— unstructured information is a significant part of modern electronic text documents.

Therefore, there is a need to develop new modern methodologies for data processing and analysis. Data mining has become such a new methodology data mining [1].

Data mining is the processing of information and the identification of patterns and trends that help make decisions. The principles of data mining have been known for many years, but with the advent of large amounts of data, they have become even more widespread. Large amounts of information have led to an explosive increase in the popularity of broader methods of data mining, because information has become much more, and it by its very nature and content is becoming more diverse and

extensive. When working with large data sets, relatively simple and straightforward statistics are no longer enough.

The reasons for the popularity of IAD are as follows:

- rapid accumulation of data (the account is already on exabytes);
- general computerization of business processes;
- penetration of the Internet into all spheres of activity;
- progress in the field of information technologies: improvement of DBMS and data warehouse;
- progress in the field of production technologies: rapid growth of computer productivity, storage volumes [2].

## **2. Presenting main material**

After analyzing the analogue, it was decided that there is a problem in the method that needs to be solved.

The method proposes to modify the method of data clustering, based on the Kohonen method. It is planned to modify the method of selecting the data to be analyzed.

To increase productivity, it is proposed to introduce normalization of data sampling, within the specified limits. This will ensure faster data sampling, as data for analysis will be more carefully selected.

The developed algorithm for forming images of documents is based on a statistical approach to the analysis of texts in natural language. It is proposed to form the image of each document in the form of a multidimensional vector of normalized and weighted single words (signs) found in the text of this document. The dimension of such a vector will be equal to the number of unique features in the document collection.

The proposed method of forming images of PD consists of the following main stages:

$$\Phi D = \langle \Phi P, \Phi DP, \Phi R \rangle \quad (1)$$

where  $\Phi P$  - a way to remove features from the texts of documents;  $\Phi DP$  - a way to display documents in the space of their features;  $\Phi R$  - algorithm for reducing the space of document features [3].

The method of removing the signs of  $\Phi P$  is to perform the following operations: lexical analysis of the text (removal of markup, punctuation, numbers, conversion of all letters to uppercase, etc.), removal of stop words, ie commonly used words that do not have an independent meaning , for example, prepositions, conjunctions, particles and pronouns; morphological analysis.

We propose to use such an approach to morphological analysis as the selection of pseudo-bases of words. As a result of this analysis, words from the text are reduced to a special type, and in the future, words that have the same special form (pseudo-basis) are considered as one feature. As a result of extracting features by the  $\Phi P$  method, it is possible to obtain  $\Phi D$  - a dimensional set of features (pseudo-words) of the document collection  $P$ , which is also called the general dictionary of features of the document collection.

The method of displaying documents in the space of their features FDP is based on the procedure of weighing features. Weighing of features of documents is offered to carry out by means of traditional technique tf \* idf which is independent of existence of a training set, considers frequency of occurrence of a term, both in the separate document, and in all collection as a whole.

The need to develop an algorithm for reducing the feature space of FR documents is due to the fact that high-dimensional and sparse document vectors in the feature space are not sufficiently clear orientation so that automatic methods by calculating the distance between them could make an unambiguous conclusion about their relationship or difference. To solve this problem in modern information retrieval systems, forced reduction of the feature space by the DF criterion is used. The algorithm of forced reduction according to the DF criterion removes from the general dictionary of features P of the document collection all those features whose document frequency is above the threshold value DF  $\max\tau$  and below the threshold value DF  $\min\tau$ .

The input of the neural network is fed a set of document vectors in the form of a matrix of size N by M, where N is the number of documents that are clustered, and M is the number of unique terms in the collection of documents that are clustered. At the intersection of columns and rows are the weights of the j-th term in the i-th document, calculated by the method tf \* idf.

The basic algorithm for learning the Kohonen network is as follows:

Step 1 Initialize the weight matrix with small random values (on the interval [0, 1]).

Step 2 Randomly select a vector from the source set.

Step 3 For each output neuron j calculate the distance between its weight vector  $w_i$  and the output vector x:

$$d_j = \sqrt{\sum_{i=1}^n (w_{ij} - x_i)^2} \quad (2)$$

Step 4 Find the original winning neuron  $j_{\min}$  with the minimum distance  $\min(d_j)$ .

Step 5 For the original winning neuron  $j_{\min}$  and for its neighbors in the vicinity, the weight vectors are updated as a rule:

$$W_{ij}(t+1) = W_{ij}(t) + e(t) * h(t, j, m) * (x_i - W_{ij}(t)) \quad (3)$$

where  $w_{ij}(t)$  is the value of the weighting factor of the input neuron i and the output neuron j at time t;  $h(t, j, m)$  - the value of the neighborhood function with the central neuron of the source layer m for the neuron of the source layer j at time t;  $e(t)$  is the coefficient of learning speed at time t;  $x_i$  is the output of the neuron of the first layer numbered i.

Step 6 Repeat the steps from step 2 for all elements of the source set.

The training cycle lasts until the system reaches the desired state. The following can be used as criteria for stopping the learning process:

- topological ordering of the feature map (weight matrix);
- weight changes become insignificant;
- the network output is stabilized, ie the output vectors do not pass between

cluster elements;

- the limit value of the error on the map has been reached;
- passed a given number of epochs.

## Conclusions

As a result of the research, a modified method of forming an information retrieval image of an electronic document was developed. The modified method provides faster and better data selection for the formation of IPO. The method of constructing an information retrieval image of electronic documents by means of data mining provides stable and fast work with documents.

In the future it is planned to refine the method and make changes to it. Since this topic is currently in high demand, it is possible to make amendments to the method that will improve and speed up the work of this method.

## References:

1. Korneev VV Databases. Intelligent information processing. / Korneev VV, Gareev AF, Vasyltin SV, Reich VV -M : Nolidzh, 2001. 496 p.
2. Barsstyan AA Data analysis technologies: Data Mining, Visual Mining, OLAP./ Barsgyan AA, Kupriyanov MS, Stepanenko VV, Kholod II - СПб: БХВ-Петербург, 2007. 275 c.
3. Oksanich IG Intellectual analysis of an array of text documents based on Text Mining technology / IG Oksanich, DM Piskunov, DP Chernysh // Information processing systems. - 2013. - Vip. 2. - P. 139-143.

# **CYBERSPACE PROTECTION TECHNOLOGIES**



**<sup>1</sup>Serhii Buchyk**

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection

**<sup>2</sup>Yaroslav Andrushchenko**

Student

<sup>1,2</sup>Taras Shevchenko National University of Kyiv

## **SEARCHING FOR A POTENTIAL CRIMINAL USING WIRELESS INTERNET NETWORKS AS ONE OF THE TARGETS OF STATE SECURITY**

**Abstract.** The article discusses one of the possible ways of finding potential criminals using wireless Internet networks in order to strengthen the security of facilities of particular importance and the country in general. It is also described how modern digital and network technologies can be used to develop a comprehensive search system and control a person's cell phone.

**Keywords:** wireless network security, ‘man-in-the-middle’, cybercriminals, cyber special service, data interception, deauthers.

In the context of the hybrid war, in which Ukraine is currently engaged, the question of ensuring security and integrity of the country is beyond doubt. Many countries, for their territorial integrity, are investing not only in the armed forces, in the traditional sense of this term, but also in cyber-weapons. The advent of cyber-weapons also makes it necessary to develop means of countering these weapons. That is why governments fund and support projects that protect the country in cyberspace.

Detection and prevention are the main objectives of the country's special services. That is why they must use modern equipment and technology, by means of which, they can monitor and, if necessary, influence the work of information systems or individual devices. But how to set up this solution? And what would that take? It's impossible to track every single person in the country. People generate a lot of information, and there are physically not enough specialists to process that information. And furthermore, the Constitution protects the human rights of privacy. Special services have the right to interfere in the life of a person only if suspicion has been made and a court order has been executed. And only after that you can legally interfere with a certain person's life.

You can't defend yourself if you can't attack. That is why it is not uncommon for special services to use the help of cybercriminals to solve cybercrime. They also train their specialists, the so-called "white hackers" who work for the country. However, such specialists are not of much use without special tools. The authorities seek help from developers who create tools to search for and deanonymize criminals.

One of those tools that could help special services is a tool that can scan mobile phones with an enabled Wi-Fi adapter. This type of solution should consist of two parts. The first part is a server that processes, stores and transmits information, and the second part is an access point that emits the work of a Wi-Fi router. Such Wi-Fi points are placed in strategically important and lively places such as airports, bus stations,

train stations and so on. They're also placed in shopping and entertainment centers - all the places where there are a lot of people. In such public places, there is usually free Wi-Fi that people gladly use. Also, these places are visited by criminals, as they can be "invisible" among the crowd. And it is this solution that would allow us to detect the presence of such attackers.

As mentioned earlier, the solution should consist of a server and an access point. However, there should be two access points, so that a person can be found by x and y coordinates. Each point runs at frequencies of 2.4 Ghz and 5.0 Ghz. These are the open frequencies on which the Wi-Fi protocol works according to IEEE 802.11[1] standard. These points are different from others because they have a much higher power, their signal power reaches several kilometers. The points cover 360 degrees around, but their antennas have a clear direction, which is why they have to be positioned in the direction of these antennas. Otherwise, their efficiency will be reduced. These antennas must be reprogrammed with special software that is based on the Linux kernel with long-term support distributions such as Debian, or its descendant - Ubuntu. However, the specific user (operator) "shell" should be developed which is oriented to the Cisco command-line interface (CLI). These access points should be connected to the Internet and directly to the mini-server on which the scanned information will be stored and the database will be supplemented.

Next up is the server. There should be two types of server. The first one is a cell phone - a small computer that will be connected directly to the point which stores, processes and transmits the information to the data center, where a more powerful server, operated by the information security administrator will already be. The mobile server will also update and configure the access point.

For easy operation, a graphical interface (GUI), operated by a cybersecurity specialist, should be developed. With this interface, the operator should be able to process all incoming information. Scanning information should be displayed on his/her monitor and have certain filters to make the work easier. He/she should also be able to control, configure and reconfigure the version of the access point. In case of sensor problems, the operator should be able to connect to it directly, and with debugging commands that are specially designed to test the work of the sensor, reconfigure it or gather information which will be forwarded to the developers for further software correction and updating.

The question then is: What is this all about? With such a solution, the operator will monitor and, if necessary, conduct an attack on the mobile phone in order to capture the attacker's phone. In that case, two phases of intervention would be carried out. The first one is known as the Man-in-the-Middle attack[2]. The essence of the attack is that it interferes with the transmission of information in such a way, that from the target victim the information passes through the attacker and then to the router and vice versa. When the victim requests the information, the router first directs it to the attacker because it considers him to be the end-user and then the attacker redirects it to the user.

This is a description of the classical method, but the solution will use a more effective method. The information, that is most often transmitted, is encrypted and

cannot be retrieved. In case of using this solution, the public router, which is located in the mall, will be replaced by the router of the cybersecurity operator. This all will be done thanks to these sensors. That is why these access points will work exclusively with the cell phone of the criminal and not with the cell phone of ordinary citizens. The operator chooses his target; the system first calculates the frequency at which it operates - 2.4Ghz or 5.0Ghz, then finds the exact channel using the phone. The sensor then sends deauthers[4]. Deauthers are images that are sent to the target's phone and report that it has been disabled. At this time, the point is fixed to a specific channel at a certain frequency and creates an access point with an exact copy of the SSID[3]. When the criminal's phone tries to reconnect the Wi-Fi, the access point is no longer public, but artificially created by the system. The victim won't notice any change in work, because it happens pretty quickly. And once the target is caught, we can analyze and alter the traffic. That's how phase one ends. Next, phase number two - infection of the target's phone and retrieval of information from the device - begins. The solution starts sending packets and switching the phone so that it can "tell" about itself. This enables the installation of malicious software for spying purposes, the ability to redirect the phone to phishing sites in order to obtain confidential data or download programs.

All actions described above are illegal and can be carried out exclusively by special services, the ones that have the permission of the court. The operator's target won't be random. The MAC address of the target's phone must be approved by the court, and permission for obtaining the information from the target phone must be granted. Otherwise, all obtained information in the course of an investigation cannot be considered in court, because it has been obtained illegally.

The theoretical material discussed above is all about improving cybercrime investigations and the work of the special services of the country.

#### **References:**

1. IEEE 802.11-2012 - IEEE Standard for Information technology--Telecommunications and information exchange between systems Local and metropolitan area networks--Specific requirements: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications
2. Man in the middle (MITM) attack. - <https://www.imperva.com/learn/application-security/man-in-the-middle-attack-mitm/>
3. What Is an SSID, or Service Set Identifier? - <https://www.howtogeek.com/334935/what-is-an-ssid-or-service-set-identifier/>
4. Wi-Fi deauthentication attack. - [https://en.wikipedia.org/wiki/Wi-Fi\\_deauthentication\\_attack](https://en.wikipedia.org/wiki/Wi-Fi_deauthentication_attack)

**<sup>1</sup> Serhii Buchyk**

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

**<sup>2</sup> Yaroslav Symonchenko**

Graduate student at the Department of Telecommunication and Radio Electronic Systems, Faculty of Aeronautics, Electronics and Telecommunications

**<sup>3</sup> Anna Symonchenko**

Student at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

<sup>1,3</sup> Taras Shevchenko National University of Kyiv

<sup>2</sup> National Aviation University

## THE METHOD OF DETECTION OF HIDDEN INFORMATION USING STEGANOGRAPHIC METHODS

**Abstract.** In this work it is proposed the method of detection of hidden information using steganographic methods. Abstracts focus on the method of detection of data in a digital image that is hidden using steganographic methods. The proposed method of analysis is based on detecting presence of hidden data embedded in a digital image using bit modification methods of color components of an image based on the RGB color model.

**Keywords:** steganographic system, stegocontainer, steganographic analysis, detection of hidden information, software steganographic tools, information protection, linguistic group, linguistic ratio parameter.

Currently, the development of methods of computer steganographic analysis is an urgent task.

The realization of steganographic analysis allows to conduct research of digital messages for the purpose to find the fact of presence of hidden data and possible decoding of a hidden message [1].

The aim of this work is the realization of the detection method of presence of hidden message using a digital image while saving it as a digital image file.

This method analyzes the distribution of the number of single bits in conditional blocks of bit planes of color components of a digital image with the purpose to detect a hidden message.

This method provides the ability to determine the type of hidden data, namely, belonging to a linguistic group, if a message is text information.

This approach is carried out by comparing the distribution of the number of single bits in conditional blocks of bit planes of color components of an image and symbols codes of text message at their binary representation.

Most often, steganographic tools are as steganographic means that can be used to hide and decode the hidden data. So, given the functionality of modern steganographic software [2]:

- the most common type of supported computer files is files with graphic information (an image file of BMP format with a digital image);
- the most common type of hidden data is text information;
- as the method of hiding data it is used computer steganographic methods that

support the direct replacement of least significant parts of an image (extra information) with bits of hidden data.

Considering that the most common type of hidden data is text information, we analyzed the distribution of number of single bits (further— DNB) for codes of message symbols of texts in English, Russian, Ukrainian.

To conduct this research we follow the steps according to the developed algorithm [3]:

Stage 1. Make representation of each message symbol in proper binary code (8 bits each).

Stage 2. Make the division of the formed bit representation into the blocks that corresponds to the bit digit of each binary representation (from 0 to 7).

Stage 3. Make calculation of the bit number with a value «1» in each bit block.

We chose 3 messages of different length in each language and make the representation of their DNB (distribution of number of single bits).

We used 24-bit digital bitmap images to research the detection method of hidden data in a digital image.

An image was embedded by the method of modification of a lower bit in the blue color component when using the RGB color model.

We conduct the realization of the proposed method of analysis of detection of hidden data by following these steps:

Stage 1. Decode each color component of the researched image using the RGB color model.

Stage 2. Convert the matrix of the blue color image component into a vector-column formed from the color gradation values of the blue component, and convert the value of the vector into binary.

Stage 3. Split each column of bit plane into eight conditional blocks that are formed by counting bits with a value «1» as follows: each conditional block (0...7) of size  $1 \times 8$  contains the values of the number of single bits the proper column of the binary representation matrix (0...7) with a shift step «+8».

Stage 4. Get a matrix of conditional blocks (further MCB), of size  $8 \times 8$ , the columns correspond to the number of a bit plane of the image components, and rows—to the conditional block number with the count of distribution of the number of single bits in each block.

It is a graphic representation of the values of the formed matrix of conditional blocks.

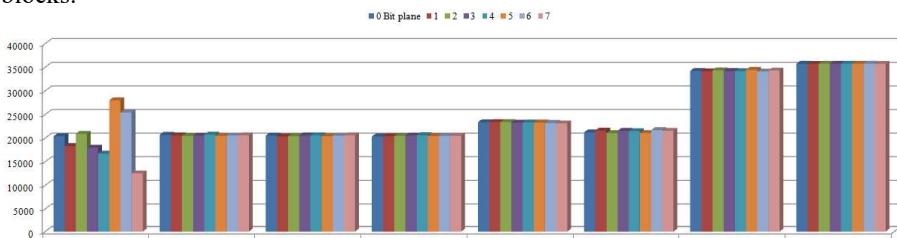


Figure 1 - Graphic representation of the values of the formed matrix of conditional blocks

To evaluate and compare DNB matching in bit image planes of an image and codes of symbols of a text message we compare them based on the Pearson correlation coefficient indicator.

Make the determination of the Pearson correlation coefficient DNB of 5 bit planes of blue color image component and symbols codes of the text message.

We used 24-bit digital bitmap images of the following classes to research the detection method of hidden data in a digital image [4-6]:

class 1 – images with few colors (4-16) and large areas filled with the same color;

class 2 – images with smooth color transitions built on a computer;

class 3 – photorealistic images;

class 4 – photorealistic images with overlay business graphics.

The degree of image modification when embedding an English text message to the blue color component of each image.

For more detailed research we embed the English message into each image. Determine the value of Pearson correlation coefficient of DNB of zero bit planes of blue color component of certain images and symbols codes of text message.

We proposed the method of steganalysis of the detection of presence of hidden data formed using steganographic system. This method allows to research a digital image for presence of embedded hidden text message using the method of modification of lower bit of color component of RGB model and determine language of a hidden message.

The detection of presence of hidden data in the image is conducted by comparing DNB MCB of the bit planes of the image color components and symbol codes of the message at their binary representation.

#### **References:**

1. Barannik, V. and Shulgin, S., (2016), The method of increasing accessibility of the dynamic video information resource. Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET): 13th Intern. conf., (Lviv-Slavsk, Ukraine, febr. 23–26, 2016). Lviv-Slavsk: 2016. pp. 621-623.
2. O. Yudin, Y. Symonychenko and A. Symonychenko, "The Method of Detection of Hidden Information in a Digital Image Using Steganographic Methods of Analysis", 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 2019, pp. 262-266, doi: 10.1109/ATIT49449.2019.9030479.
3. A. N. Alimpiev, V. V. Barannik, and S. A. Sidchenko, "The method of cryptocompression presentation of videoinformation resources in a generalized structurally positioned space," Telecommunications and Radio Engineering, vol. 76, no. 6, 2017, pp. 521–534.
4. V. Barannik, A. Bekirov, A. Lekakh, and D. Barannik, "A steganographic method based on the modification of regions of the image with different saturation," Proceeding of the XIVth International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), Lviv-Slavsk, Ukraine, 2018, pp. 542-545.
5. M.Grundmann, V.Kwatra, M.Han, I.Essa, "Efficient hierarchical graph- based video segmentation," Preceedings of the IEEE Conference on Computer Vision and Pattern, San Francisco, 2010, pp. 85-91.
6. P.Gurzhiy, ,B.Gorodetsky, ,O.Yudin, ,Y.Rybukha, "The Method of Adaptive Counteraction to Viral Attacks, Taking into Account Their Masking in Infocommunication Systems", Proceedings of the 3rd International Conference on Advanced Information and Communications Technologies (AICT), Lviv, Ukraine, 2019, pp.423-426.

**Antonina Kashtalian**

Ph.D., Associate Professor

*Khmelnitsky National University***HONEYPOTS MODELS IN COMPUTER NETWORKS ACCORDING TO  
MALICIOUS ATTACKS TYPES**

**Abstract.** The paper proposes honeypots models based on typical computer attacks and architectural features of baits developed taking into account the architecture of a distributed system with baits. The main features are configuring different types of honeypots and their integration with other components in multilevel security system. Typical features of honeypots and methods of system analysis and decision theory for solving problems are analyzed with using baits by types of attacks and organizing the interaction of components of a multilevel system.

**Keywords:** honeypot, malicious actions, computer attack detection, forecasting

Computer networks connected to Internet become objects for malicious actions [1, 2]. Detection of malicious actions and protection against them are used in systems of different types [3]. A promising way of computer networks protection is using separated honeypots and honeynets and their integration with other protection systems. The work goal is to develop honeypots models according to their architecture features, utilization features and attacks types on networks. Honeypots perform functions of information collection and analysis about malicious actions in networks [4, 5].

Models characteristics are grouped and generalized for model creating. With this purpose honeypots operation levels are taken into account with its functions: detection, analysis, reaction, execution. The honeypot model, taking into account the functions from which it is formed, is set in the way  $M_P = \langle P, \Omega_1 \rangle$ , where  $P$  is the set of a honeypot functions;  $\Omega_1$  is the set of predicates on the set  $P$ .

It is necessary to specify features of a certain type of an attack for its detection and its type determination. A honeypot should provide: ports and services that are attacked; collection, storage and processing data of network traffic of a honeypot; interaction with a honeynet.

Ports and services deployed on a honeypot depend on attack type or types which a honeypot intercepts. Collection and storage of traffic data may be implemented in the same way for different types of attacks.

Honeypots architectures are considered, taking into account typical attacks in corporate computer networks.

The attack ‘mailbomb’ is aimed at e-mail box or e-mail server. For detection of the attack ‘mailbomb’ it is necessary to develop SMTP-service for simulating e-mail server (fig. 1). Data monitored by the honeypot: IP-address/ IP-prefix of e-mail source; domain name/ URL/ URL type of e-mail source; user or users ID; received e-mail volume. On the basis of these data the analysis of activity on honeypot SMPT service is performed, threshold values of e-mail or e-mails volume and their received speed are

determined. Threshold values are used for defining e-mail traffic as malicious and one that requires analysis.

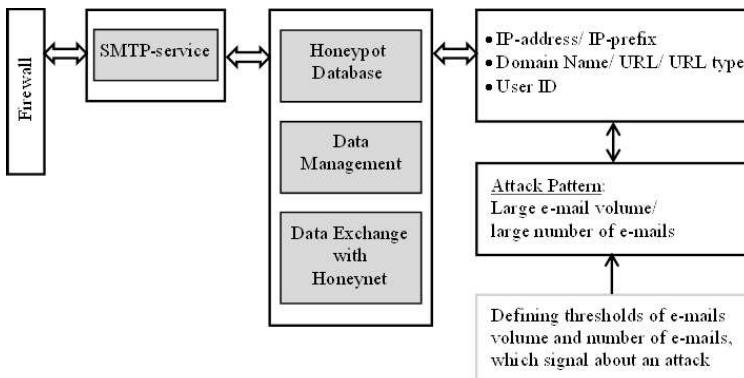


Figure 1 – The honeypot designed for an attack ‘mailbomb’

For detection of the attack ‘neptune’ the honeypot should contain the service, which works by TCP protocol, for example HNNP and FTP services. Data monitored by this type of honeypot: IP-address/ IP-prefix of the request source; domain name/ URL/ URL type of the request source; the number of half-open requests. As the result of analysis the network traffic is classified as not malicious/ malicious and threshold values of requests number and their received speed are defined.

For detecting the attack ‘portsweep’ the honeypot should contain significant number of ports, which may be both physical and virtual and have to correspond to real services ports. Data collected by the honeypot in the case of attack ‘portsweep’: IP-address/ IP-prefix; domain name/ URL/ URL type; existence of ports access. The analysis of time series of ports access provides a way to classify an activity as malicious.

The developed honeypots model for typical attacks can be applied for creating the network of honeypots inside a corporate computer network. This allows them to be used depending on the types of attacks. It is necessary to use the methods of decision theory for the organization of functioning such honeypots network with the purpose of determining a dynamic architecture of honeynet.

The dynamic honeynet allows collecting different types of attacks and analyzing intruders. The analysis provides finding similar intruders, anomaly activity detection, and forecasting intruder activity with traditional statistical and state of the art methods like deep learning.

Finding similar intruders is performed on the basis of clustering their activity, i.e., time series belonged intruders. Clustering is executed with different methods, among them there are clustering on the basis of informational criterions, clustering on the basis of Gaussian mixture models, clustering on the basis of hidden Markov models, clustering on the basis of neural network models etc.

Forecasting intruders activity provides early detection of malicious actions and determination of attack probability and their characteristics on the basis of current state and previously detected patterns. The analysis of intruders actions allows to get the patterns of attacks.

The developed models of honeypots with considering their architectural features and typical attacks features are basis for creating systems of wrong attacks objects integrated in the general security system of corporate computer network. This improves security level including through analysis of information collected in a honeypot. When organizing the detection and interaction of distributed system components, it is necessary to involve the methods of system analysis and decision theory which allow enhancing the result of work of the entire system.

Conducted research of mentioned methods allows separating important ones from them for their use in the developed system. The result of experimental investigation allows performing the creation of the honeypot network on the basis of a distributed multilevel system.

#### **References:**

1. Savenko O.S Research of methods of antiviral diagnostics of computer networks / O.S. Savenko, S.M. Lysenko // Visnyk of Khmelnytsky National University. Technical sciences. - 2007. - № 2, v. 2. - P. 120–126. (in Ukrainian)
2. Savenko O.S., Klots Y.P, Mostoviy S.V. Research and analysis of process blocking in a computer system // Visnyk of Khmelnytsky National University. - 2007. - № 3, Volume 1.- P.248-251. (in Ukrainian)
3. Sidiropoulos S./ Composite Hybrid Techniques for Defending Against Targeted Attacks// S. Sidiropoulos, A.D. Keromytis. Part of the Advanced in Information Security book series (ADIS, volume 27), 2007, 213-229pp.
4. Sochor Tomas/ Study of Internet Threats and Attack Methods Using Honeypots and Honeynets// Tomas Sochor, Matej Zuzcak - Springer International Publishing Switzerland 2014, A. Kwiecień, P. Gaj, and P. Stera (Eds.): CN 2014, CCIS 431, pp. 118–127, 2014.
5. Sokol Pavol/ Data Collection and Data Analysis in Honeypots and Honeynets// Pavol Sokol, Patrik Pekarčík, Tomáš Bajtoš. <http://spi.unob.cz/papers/2015/2015-19.pdf> [Access 18.04.2020].

<sup>1</sup> **Danil Koltsov**

<sup>2</sup> **Ivan Parkhomenko**

PhD in Technical Sciences, Associate Professor Department of Cybersecurity and Information Protection

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## TRAVERSAL UTILITIES FOR NAT

**Abstract.** The question of why you can't connect to the server via NAT is considered. Requirements to traversal NAT. Options for this to be done.

**Keywords:** Traversal NAT, P2P applications, STUN, TURN, ICE.

### 1. INTRODUCTION

A big problem for many applications is the inability to establish a connection at all. This is especially true for P2P applications such as VoIP, messengers, and file sharing, which often need to act as both a client and a server to provide two-way direct communication.

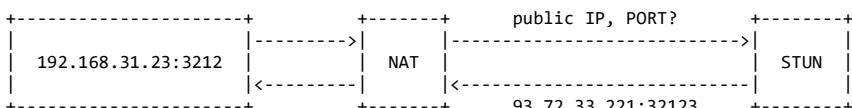
If NAT is available, the internal client does not know about its public IP address. It knows its internal IP address, and NAT devices overwrite the output port and address in each TCP/UDP packet and the output IP address inside the IP packet. However, if the client transmits its private IP address a part of its application data to an equal external environment outside its private network, then the connection will fail.

Therefore, if you want to share peer-to-peer code outside your private network, the application must first detect its public IP address. Another packet that arrives at the public IP address of the Nat device must also have a destination port and an entry in the NAT table that can translate it to the internal IP address of the destination host and a tuple of ports.

To eliminate this discrepancy in NAT, there are methods for bypassing STUN, TURN, and ICE, which are used to establish end-to-end communication between peer members on both sides.

### 2. SESSION TRAVERSAL UTILITIES FOR NAT (RFC 5389)

Session bypass utilities for NAT STUN (RFC 5389) is a protocol that allows the host application to detect the presence of a network address translator on the network and, if available, get a dedicated public IP address and a tuple port for the current connection. To do this, the protocol requires the help of a well-known third-party stun server, which must be located on a public network.



**Figure 1 – STUN request for public IP and Port**

Assuming that the IP address of the stun server is known (via DNS detection or at a manually specified address), the application first sends a request to bind to the stun server. In its turn, the stun server responds with a response that contains the public IP address and client port that is visible from the public network.

This process has several problems.

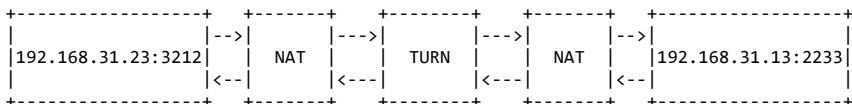
- The app detects its public IP and port packet, and can then use this information as part of its app data when communicating with its members.
- An outgoing binding request to the stun server sets NAT routing records along the path, so that incoming packets arriving at the public IP address and Port tuple can now find their way back to the host application on the internal network.
- The STUN protocol defines a simple ping saving mechanism to avoid waiting times for NAT routing records.

With this mechanism, when two peer-to-peer partners want to communicate with each other, they first send binding requests to their respective STUN servers, and after both parties successfully respond, they can use the established public tuples of IP and ports to exchange data.

### 3. TRAVERSAL USING RELAYS AROUND NAT (RFC 5766)

However, in practice, STUN is not sufficient to work with all NAT topologies and network configurations. In some cases, UDP may be blocked by a firewall or other network device - a common scenario for many corporate networks. To solve this problem when the STUN fails, we can use the relay bypass protocol around Nat (TURN) (RFC 5766) as a backup option that can work over UDP and switch to TCP if all else fails.

The key word in TURN is, of course, "relays". The protocol relies on the availability and availability of a public repeater to transmit data between members.

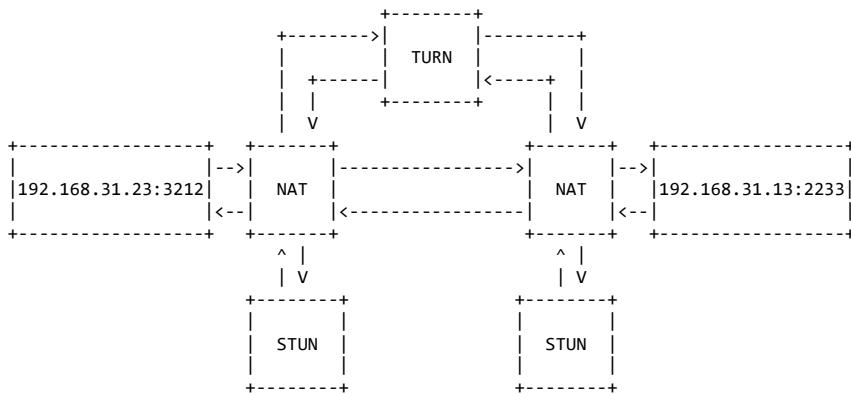


**Figure 2 – TURN**

- Both clients start their connections by sending a distribution request to the same turn server, followed by permission approval.
- Once reconciliation is complete, both peers communicate by sending their data to the TURN server, which then passes it to another peer.

### 4. INTERACTIVE CONNECTIVITY ESTABLISHMENT (RFC 5245)

ICE (RFC 5245) is a protocol and set of methods that aim to establish the most efficient tunnel between participants, provide direct connection where possible, use stun negotiations where necessary, and finally return to TURN if all else fails.



**Figure 3 – ICE**

## 5. References

- Recommendations [1, 2, 3]
- RFC 5389 Session Traversal Utilities for NAT
- RFC 5766 Traversal Using Relays around NAT
- RFC 5245 Interactive Connectivity Establishment

**<sup>1</sup>Nataliia Lukova-Chuiko**

Doctor of Technical Science, Professor of the Department of Cybersecurity and Information Protection

**<sup>2</sup>Alexander Bystrov**

Student

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **ADVICE ON SELECTING AN INTRUSION DETECTION SYSTEM FOR SMALL AND MEDIUM-SIZED BUSINESSES**

**Abstract.** This article provides recommendations for choosing an intrusion detection system for small and medium-sized businesses. These recommendations can be applied in practice by employees of the information security department of the enterprise. These recommendations can significantly increase the level of information security of the enterprise and minimize possible losses in the future.

**Keywords:** Open-Source, SIEM, Intrusion Detection, Monitoring

In the modern information field, there is an acute problem of information security. This problem includes both protection from attacks and their detection at an early stage in order to minimize the consequences of an attack and respond to it in time. One of the main tools to detect attacks is SIEM-systems, which, if properly configured, reduce the response time to an attack to a minimum. Typically, efficient systems are expensive, and their implementation and support are heavily funded, which can be afforded by an Enterprise or Medium-Business companies. For small companies, SIEM implementation and maintenance becomes an impossible burden, as usually the means to purchase an SIEM system are bigger than possible losses from a successful attack, which can paralyze the whole business for an indefinite period of time. The way out of this situation are open source applications that can be adapted to штектгішшт detection systems and perform their functions.

This application can be Osquery, the software was developed in 2014 by Facebook. This application is distributed under MIT license, so it can be used for commercial purposes, without any usage fees [1]. The software uses a client architecture, so the application must be installed on each server from which data is to be received. The essence of this application is that the operating system is perceived as a relational database with tables displaying information about the system. These data can be conveniently obtained with the help of SQL language and record both in the log file and sent to a remote syslog-server, which will accumulate data from different endpoints to present a complete picture of the infrastructure security. One example of using this application is the detection of one of the most popular tactics for launching malware, namely deleting the executable file after the process has already been created, making it difficult to detect the attack. So to search for these processes, you need to form a simple SQL query - "SELECT \* FROM processes WHERE on\_disk = 0", after executing this query, Osquery will display all the processes whose executable files have been removed from the file system. Thus, with the help of this software it is possible to

create many markers that will trigger the presence of malicious software in the system. All settings for information security markers are recorded in a configuration file, which also contains settings for information output (syslog-server, log file, etc.) and the frequency of marker checks. Thus, these configuration files are conveniently distributed between a large number of endpoints using automation tools such as Ansible.

The problem with Osquery is that the application does not provide a user-friendly interface to analyze all incidents that have been collected from different parts of the infrastructure, so to use it effectively you need to implement an application to analyze and display this data. This application is ELK-Stack which is also distributed under MIT license, so it can be used for commercial purposes [2]. This software consists of three components, each of which is responsible for different purposes and tasks. Elasticsearch is the main component of the system, which accumulates data and analyzes them. Kibana - is responsible for managing the components in the web browser interface and building charts, maps and other graphical means of displaying information using data stored in Elasticsearch. The last component used in the system is Logstash. This component is responsible for converting data to provide them in a format understandable to Elasticsearch for further analysis. Together, all these components form a powerful stack of technologies that can be used to analyze and visualize information from different sources. In addition, this software can be used not only for analysis and displaying information security events, but also for its intended purpose, namely as software for analysis and aggregation of log files and collection of metrics for information system operation. It may also be noted that the data is stored in a document-centric database, with effective indexing and mechanisms for managing this data using the API and Index Lifecycle Management. The whole set of tasks for which ELK-Stack can be adapted makes it attractive for implementation in companies with small IT departments and limited financing.

In combination, these software tools provide an opportunity to constantly monitor the entire infrastructure and identify possible attacks, which will minimize losses from them. In addition, the cost of implementing this software package is minimal, as only an information security officer with knowledge of the software is needed. At the same time, this software complex can be developed together with the company that has implemented it and scaled up together with information systems of enterprises.

#### **References:**

1. Osquery license. Available:  
<https://github.com/osquery/osquery/blob/master/LICENSE-Apache-2.0>
2. ELK license. Available:  
<https://github.com/elastic/elasticsearch/blob/master/LICENSE.txt>

**<sup>1</sup> Natalia Lukova-Chuiko**

Doctor of Technical Science, Professor

**<sup>2</sup> Andriy Fesenko**

Candidate of Technical Science, Associate Professor

**<sup>3</sup> Hanna Papirna**

Master's degree student

**<sup>4</sup> Sergiy Gnatyuk**

Doctor of Science in Engineering, Associate Professor

<sup>1,2,3</sup> Taras Shevchenko National University of Kyiv

<sup>4</sup> National Aviation University

## THREAT HUNTING AS A METHOD OF PROTECTION AGAINST CYBER THREATS

**Abstract.** The article presents the structuring of a new approach to countering cyber threats - Threat Hunting. The article proposes a functional diagram for an application of this approach in practice by specialists in the field of cybersecurity.

**Keywords:** Threat Hunting, indicators of compromise, proactive cybersecurity.

### 1. Introduction

To date, most information security threats are known, and can be defended by traditional means of protection such as antivirus, firewalls, and so on. Such threats include spam, denial-of-service attacks, viruses, rootkits, and other classic malware. The remaining minority of threats are unknown and the most dangerous. They are difficult to detect and even more so to protect against them. Examples of such threats are encryption viruses, crypto miners, etc. This situation has led to the development of means of protection against cyber threats in the direction of developing new technology that would be able to counteract the most serious and complex threats.

Proactive threat search or Threat Hunting (hereinafter - TH) is the latest way to counter cyberattacks, which through proactive and iterative search, allows to detect complex threats that traditional means of protection are not even able to notice. It should be noted that TH is not a specific software or hardware product and is not a passive activity. Proactive threat search is, first of all, mainly a manual process with elements of automation, in which the analyst, based on his knowledge and skills, checks large amounts of information for indicators of compromise, according to a predetermined hypothesis of the presence of a threat.

### 2. The order of Threat Hunting conduct

According to the approach of the leading American company in the field of cybersecurity and big data analytics - Sqrrl, in general, the whole process of TH can be reduced to four main stages, which are repeated cyclically [1].

On the first stage of the hypothesis creation begins with asking questions about how an attacker can gain access to an organization's network. Then these questions need to be divided into specific and measurable hypotheses that determine what threats may be present in the network and how they can be identified [2].

Once the observations have led to the development of hypotheses, they should be tested during the stage of research. In general, it is possible to identify four types of techniques that can be used by specialists in TH at this stage: search, clustering, stack counting or accumulation and machine learning [2].

A fairly effective method at this stage is Linked data - a method of publishing structured data that allows to link them and seek confirmation of hypotheses using semantic queries. Related data analysis is particularly effective in presenting the data needed to solve hypotheses in an understandable form, and is therefore an important component of the TH. Linked data can even help prioritize and direct visualization, making it easier to search large datasets and use more powerful analytics. Methods of analysis of both source and related data should be used to identify patterns in disparate data sets, to detect the actions of attackers [1].

Today, there are many information security technologies that can provide assistance in the process of TH. However, the authors of this paper tend to narrow the set of technologies to the next most necessary: SIEM (Security Information and Event Management) systems, EDR (Endpoint Detection and Response) systems and NTA (Network Traffic Analysis) systems. As experts in the field of cybersecurity, the authors of this article note that the above systems are the technical basis for the construction of a modern SOC (Security Operations Center).

SOC is a specialized center for monitoring and prompt response to information security incidents. Such a center is a group of information security experts who are responsible for continuous monitoring and analysis of the security of the organization, using a combination of technological solutions and acting within well-structured processes. It is important to note that most often TH processes seek to implement organizations that already have their own SOC or use such services through outsourcing.

The third stage allows to reveal new harmful patterns of behavior and tactics, techniques and procedures (hereinafter - TTP). The gap in the detection of violations arises from the ability of attackers to evade the mechanisms of detection. As detection capabilities continue to evolve and expand, cybercriminals will find new ways to evade these measures. Thus, over time, TTPs of attackers will evolve to ensure that they can evade detection and act unnoticed in the IT environment.

It is due to this stage of the previous TH procedures that the TTP frameworks are filled and improved.

MITRE ATT&CK is a structure that describes the methodologies used by attackers during cyberattacks. It is presented in the form of a matrix consisting of eleven tactics, each of which contains a list of related techniques [3].

The fourth stage of the cycle forms the basis for informing and enriching automated analytics. This can be done in a variety of ways, including developing a default search for regular execution, creating new analytics or even providing feedback to a controlled machine learning algorithm [1].

One of the main mistakes of organizations initiating the TH process to their overall information security strategy is that they do not define metrics for assessing the TH either because of the difficulty of defining such indicators or because they believe

that because threat detection must be a flexible process, indicators cannot be identified.

However, there are useful metrics that can measure the performance of the TH process to help improve it, as well as help build a business case for further investment (financial and time) in staff training and tools. The following is an approximate set of metrics that can be used [2]:

- graph of trends and / or comparisons: number of incidents detected proactively (compared to reactively); the number of proactively identified vulnerabilities (compared to vulnerability assessments); waiting time (delta) in the detection of incidents proactively (compared to those detected reactively);
- percentage: data coverage (data types and asset coverage).
- pie chart: number of hypotheses on MITRE ATT&CK tactics; number of TH procedures according to MITRE ATT&CK tactics; number of incidents under MITRE ATT&CK tactics.
- service level: the percentage of successful THs that led to a new analytical conclusion or detection rule; sensitivity and specificity of analytics or rules obtained as a result of TH (number of true and false positive results).

### 3. Developing Threat Hunting process diagram

The "Hunting loop" by Sqrrl [1] provides a working and stable cycle of actions for experts. However, in the context of organizations where there are information security departments wishing to implement TH process, the cycle needs to be detailed and supplemented with initial data, as well as a connection with the classic incident management process. The authors of the paper propose the improvement of the original cycle, described on the Figure 1 and its integration in the whole information security management system of organizations.

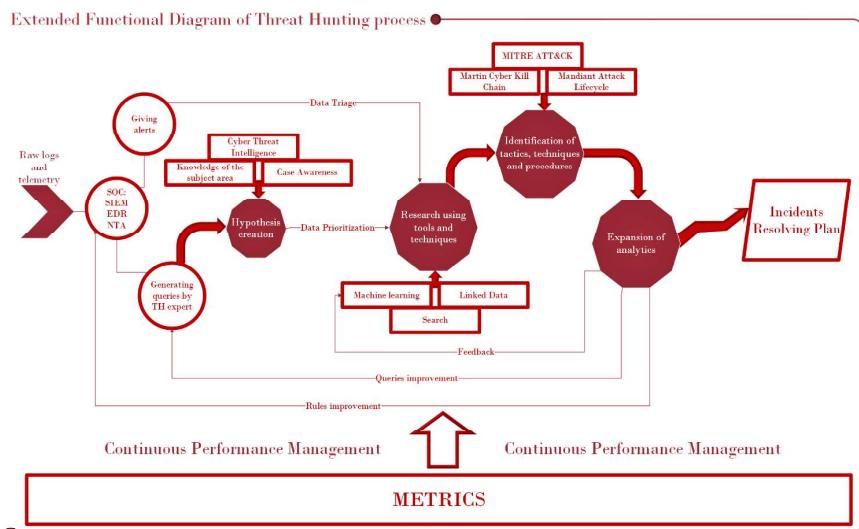


Figure 1 – Extended functional diagram of Threat Hunting process

The proposed extended functional diagram of the TH process includes:

- raw logs and telemetry from network and infrastructure assets of the organization;
- means of the initial analytics and information processing – SOC: SIEM, EDR, NTA;
- alerts and queries conducted by TH expert using initial analytics means;
- stage of the hypothesis creation;
- preliminary data triage and prioritization before conducting the research stage;
- stage of the research using tools and techniques; stage of the identification of the TTPs enriched by the trusted frameworks in the field;
- detailed outputs of the expansion of analytics stage;
- incidents resolving plan as an ultimate goal of the incident management process;
- metrics for the continuous performance management of the TH process.

Existing SOC tools, such as the SIEM platform, in the context of the second stage of TH, can be used to query data, from basic search to more advanced methods, and visualization can help identify anomalies and unusual patterns of behavior.

It can be said that the planning and implementation of the TH procedure in the daily process of information security can afford organizations with a high level of maturity of security processes, which already have established procedures and technologies for threat prevention and are ready to move to a higher level - the level of proactive threat response.

Thus, we can conclude that the TH cycle is a simple but effective process that can radically improve the level of security of the organization. This procedure is most effective when used in conjunction with traditional security systems, complementing the measures and tools to detect cyber threats that already exist in most organizations. The ultimate goal of TH should always be to go through the four-stage loop as efficiently as possible.

### References:

1. White paper: A Framework for Cyber Threat Hunting, Sqrrl Data, 2018. URL: <https://www.threathunting.net/files/framework-for-threat-hunting-whitepaper.pdf>.
2. Detecting the Unknown: A Guide to Threat Hunting, Home Office Digital, Data and Technology, version 2.0, 2019. URL: <https://hodigital.blog.gov.uk/wp-content/uploads/sites/161/2020/03/Detecting-the-Unknown-A-Guide-to-Threat-Hunting-v2.0.pdf>.
3. MITRE ATT&CK Framework, 2020. URL: <https://attack.mitre.org/tactics/enterprise/>.

<sup>1</sup>**Nataliia Lukova-Chuiko**

Doctor of Technical Science, Professor of the Department of Cybersecurity and Information Protection

<sup>2</sup>**Victoria Klochko**

Student

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## COLLECTIVE DEFENSE OF CORPORATE NETWORKS AGAINST COMPUTER ATTACKS

**Abstract.** These days, content analysis of text information is used to prevent threats, along with the analysis of the network traffic characteristics, the behavior of corporate networks and their security policy. Existing systems of text analysis and modeling include different kinds of search engines and information-analytical systems. They are capable of solving such tasks as classification of documents by its subject matter, author identification, detection of plagiarism, modeling representations of the knowledge about the subject area and the content of text.

**Keywords:** corporate network, attacks, SDA, System Monitoring Unit, cybersecurity, collective protection.

In the modern world, problems related to the use and spread of malicious software, information attacks and other types of cyber threats, which have received the general name “cybercrime” are becoming more and more relevant. Sophisticated threats require an innovative approach. Collective defense empowers organization to stay ahead of evolving threats to better defend network through real-time sharing and collaboration across industries and sectors.

Earlier, the main efforts of developers were pointed to create effective detection algorithms. These detection algorithms have used different mathematical basis: statistical methods, methods of automata theory, methods of interacting sequential processes calculus, methods of mathematical logics, neural networks, fuzzy logics, and other formalisms.

Some detection algorithms, in particular algorithms on basis of neural networks have cyberspace-adaptive properties. However, the rapid dynamics of the environment change (the variety of network structures, the variety of types of attacks, etc.) often reduced efforts of designers to zero.

A number of freely distributed and commercial systems of defense from attacks (SDA) was developed and became widely accepted in the field of corporate computer networks building [1].

The analysis of the structure of circulating packages in the corporate network is the essence of the analysis at the network layer of protection in SDA. As a rule, the package flags, the port addresses for network nodes, the time intervals between specific events and so on are analyzed here.

The package contains the information about the sender, which is often represented as a DNS-address. This information is definitely of a great value as it can

clearly point at the source of the attack. However, the truth of address information about the source of the attack is often questionable, since it can be easily corrected by the sender of the package. For some protocols, such as mail, the address of the attacker may also be obviously stated. However, as in the previous case, the address of the sender can easily be changed.

As a result, there is a need to allocate one more level of realization of the protective methods – the level of the global network.

At this level the information, which is contained in the text documents on websites, global network portals, social networks or other legitimate objects of the information space can be analyzed and both the sources of attacks and their information characteristics can be indirectly identified.

The concept of a text document here is multivalued: it is text information from websites and portals, and emails, and program codes that are entered into the computing environment of the victim's computer. In any case, this level is characterized by, on the one hand, methods used in intelligence activities, including business or competitive intelligence, and, on the other hand, methods of text processing.

IT professionals very often have problems with viruses and other malware. Actual threats include spreading spam, phishing, network attacks on enterprise infrastructure, including tar- get and DDoS attacks, where use potentially dangerous software vulnerabilities.

These and other similar examples show a close relationship between cybersecurity systems and word processing systems: when detecting spam, data loss, detecting and tracking potentially dangerous messages, etc.

This field of the research is actively evolving lately. From one side, it is connected with intellectual property protection, from another, it is connected with the necessity of cyber threats prevention, which arises because of the malware usage. In the latter case, it is hard to overestimate the possible damage, which can be caused to control systems by the key infrastructure, including to the military targets. Because there are new kinds of malware being created all over the globe, there is a necessity of the identification of the malicious code creators and bringing them to justice.

Processing, careful analysis and synthesis of information collected from Internet resources is made using content and/or rapid analysis methods, bibliometric and/or cluster analysis, as well as expert and/or situational methods [2].

However, a tight time limit for the search, collection, extraction and processing of information circulating in the global information space of the Internet, its accumulation, classification by certain attributes, further analysis, synthesis, compilation and making it accessible to the concerned users, as well as transformation into synthesized conclusions and recommendations necessitates some arrangements. First, the automation of all measures in the complex of risks monitoring system associated with these processes. Second, the configuration of SDAs subordinate to the System Monitoring Units of corporate networks according to their risk vectors.

The development of a corporate networks protection model with a collective System Monitoring Unit defense module, methods for detecting and identifying computer attacks with help of content analysis of the global information space and the

architecture of SDA, related to it, will provide a basis for the synthesis of a reliable and high-performance adaptive cyber threats detection systems and will shorten the detection time of the computer attacks of the new generation.

Further improvement of the security and stability in functioning of the information and telecommunication systems of corporate networks in the conditions of massive influence of computer attacks requires an increase in the probability of detection of new computer attacks and a decrease in the recognition time for the signs of known attacks [3].

To solve this problem, it is not enough to use only traditional methods that utilize identification characteristics of network traffic and information about the work of corporate networks and security devices. The processing of data sets of the body of network packages, content of Internet pages, information from social networks is very valuable in this area.

Calculations of risks from various attacks require the identification of sources of attacks on indirect grounds, determining their inclinations to attacks or undesirable influences of one kind or another, determining the characteristics of attack activity, calculating predictive activity indicators based on time series analysis, and the like other [4].

This protection becomes possible or by configuring the corporate network SDA to prepare the activation of attack detection algorithms. Given the temporary limitations of the attack detection process, such actions should be performed based on predictions of the activity of potential attack sources, the detection of which is the task of the global network security level of the corporate network.

#### **References:**

1. Chi, S.-D., Park, J. S., Jung, K.-C. & Lee, J.-S., Network security modeling and cyber-attack simulation methodology. Lecture Notes in Computer Science. Springer-Verlag 2001.
2. Martovitsky V., Ruban I., Lukova-Chuiko N., Kortyak E., Kruglikov Y., Model monitoring network infrastructure based on standard FIPA, 2020.
3. Ruban I., Martovitsky V., Lukova-Chuiko N., Approach to Classifying the State of a Network Based on Statistical Parameters for Detecting Anomalies in the Information Structure of a Computing System (2018), 302-309.
4. Shannon, C. E., A mathematical theory of communication. Bell System Technical (1948), Vols. 27.

**<sup>1</sup> Volodymyr Nakonechnyi**

Doctor of Technical Sciences, Professor of the Department of Cybersecurity and Information Protection

**<sup>2</sup> Maksym Bondarenko**

Master's degree, Student

<sup>1,2</sup> Kyiv National University of Taras Shevchenko, Ukraine

## **APPLICATION OF BIOMETRIC METHODS OF USER IDENTIFICATION IN INFORMATION AND COMMUNICATION SYSTEMS**

**Abstract.** Biometrics are physical or behavioral human characteristics to that can be used to digitally identify a person to grant access to systems, devices or data. In this publication it is described why this topic is an important part of information security and corporate information security as well; moreover, using of multimodal biometric identification system was proposed for stronger security.

**Keywords:** biometry, IT, security, cybersecurity, characteristics, error, identification, recognition, reliability.

The relentless expansion of the scope of computer information processing and computer telecommunications is attracting more and more people to the field of information technology, which increases the risks of information threats and their implementation. Despite the extensive technological capabilities of protection, today the number of crimes and fraud is growing with every minute.

One of the most common protection technologies is biometric protection system. It is the most convenient because it does not require storing complex passwords or carrying special identifiers (keys, cards, etc.), and it will be enough to just say the code word, put your finger or hand, or put the face to scan to access.

There is a limited number of characteristic personality attributes that can be used to identify a person [1]:

- what the person owns (identification mark, key or plastic card);
- features of behavior (language, handwriting, the nature of the keyboard);
- some physical characteristics (fingerprints, hand shape, blood vessel pattern).

These properties are used in everyday practice when people communicate with each other to identify visitors, messages, etc. On their basis special automatic devices are created and methods of identification of the person are developed.

A common characteristic that is used to compare different methods and techniques of biometric identification are statistical indicators [2]: error of the first kind (do not let into the system of "local") and second kind error (let into the system of someone else).

It is very difficult to sort and compare statistical and dynamic biometric methods according to the first kind of errors, as they are different for the same methods due to

the equipment for which they are implemented [3].

According to the indicators of errors of the second kind, the general sorting of biometric authentication methods looks like this (from best to worst) [1]:

- DNA.
- Iris, retina.
- Fingerprint, facial thermogram, palm shape.
- The shape of the face, the placement of veins on the palm and hand.
- Signature.
- Keyboard handwriting.
- Voice

Based on the analysis of modern biometric systems of human recognition, it was proposed to use a multimodal (bimodal) system of identification, which consists of two characteristics: face and voice.

Multimodal biometric identification system is a multi-factorial identification of a person, which consists of two main static components [3]:

- 1) identification with the image of a person;
- 2) identification with a passphrase.

Face identification is performed in real time mode at the moment of raising or approaching the device with the camera. Three images are enough for registration and identification.

Voice identification is based on the use of a static password phrase. At the stage of phrase registration it is necessary to repeat several times, in this way the maximum reliability is reached and the variability of the utterance is estimated.

#### **References:**

1. Біометрія як універсальний спосіб ідентифікації людини [Електронний ресурс]. – Режим доступу: <http://bablyukh.clan.su/publ/1-1-0-4>
2. DigitalPersona Fingerprint Identity Solutions for Identity Protection, Security and Compliance [Електронний ресурс]. – Режим доступу: <http://www.digitalpersona.com>.
3. Identix – Protecting and Securing Personal Identities and Assets [Електронний ресурс]. – Режим доступу: <http://www.11id.com/pages/17>.

**<sup>1</sup> Volodymyr Nakonechnyi**

Doctor of Technical Sciences, Professor of the Department of Cybersecurity and Information Protection

**<sup>2</sup> Ivan Voitenko**

Master's degree, Student

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **COMPARATIVE CHARACTERISTICS OF ALGORITHMS TO IMPROVE SPAM PREVENTION MECHANISM**

**Abstract.** E-mail spam is one of the main problems of the modern Internet, causing financial damage to companies and irritates individual users. Among the approaches designed to prevent spam, filtering is essential and popular. [1] This article provides an overview of the state of comprehensive algorithms used in machine learning applications for spam filtering and evaluating and comparing different filtering methods. There is also a brief description of mechanisms that operates to prevent spam.

**Keywords:** IT, security, cybersecurity, characteristics, spam, e-mail, prevention, filter, algorithms.

Spam is any unwanted, unsolicited digital communication, often e-mail, that is sent in bulk. Today, there are about 900,000 active spam servers that send malicious mail to large companies and users. [2] Methods of protection against this type of attack exist, but they may not be sufficient to weed out unwanted e-mails. Therefore, this issue is an urgent problem today. Spam e-mails are classified into the following categories: [3]

- Advertising companies or people who advertise their products by sending e-mails.
- Advertising of illegal products - people or companies that advertise illegal products by sending e-mails.
- Anti-advertising - the type of ads that negatively affects a company's reputation, person or other communities.
- Phishing - a type of fraud in which the letter may have malicious code or a file. Usually, launching the program or going to a link from this type of message is accompanied by an infection of the workstation or withdrawal of funds from the user's bank card.

Bulk spam has a low cost per message for the sender. However, a considerable amount of useless messages causes apparent harm to recipients. First of all, we are talking about the time wasted on filtering out unnecessary mail and looking for necessary individual letters among it. [4] Very often, internet traffic is expensive, and the user has to pay for obviously unnecessary e-mails. The most significant harm comes from the ignorance of spam recipients who open spam e-mails, follow links allegedly sent by their friends, download viruses and, without suspecting that, spread them in society. It is providers who have to waste resources on redundant hardware and anti-spam systems. According to publicly available statistics, at least 80% of forwarded e-

mails are currently spam. [5] Mail servers cut most of it off at the time of receipt. But even the remaining small part is enough to complicate the lives of users. Providers incur additional costs due to the constant need to fight spammers. Since sales letters tend to be very different from regular correspondence, filtering them out of the incoming mail flow has become a standard method of dealing with them. Currently, this method is the main and most widely used one. [6]

Generally, there are two preventing methods from e-mail spam: [7]

- Automatic filtering - There is software for automatically detecting spam. It can be intended for end-users or use on servers. This software takes two main approaches.
  - The first is that the letter's content is analyzed, and a conclusion is made whether it is spam or not. A message classified as spam is separated from other correspondence: it can be marked, moved to another folder, deleted. Such software can run both on the server and the client's computer. In the latter case, the user does not see the filtered spam but continues to incur the costs associated with receiving it since the filtering software receives each message and only then decides whether to show it or not. However, the user risks not receiving a letter perceived by the filter as spam in this case.
  - The second approach uses various methods to identify the sender as a spammer without looking at the letter's body. This software can only run on a server that directly receives messages. With this approach, additional traffic is spent only by the server for communicating with spam mail programs and calls to other servers during the check.
- Manual filtration - many programs and mail services on the Internet allow users to set their filters. Such filters can consist of words or, less often, regular expressions, depending on the presence or absence of which the message gets or does not end up in the trash bin. However, such filtering is time-consuming and inflexible and requires a certain degree of familiarity with using computers. On the other hand, it allows you to filter out some spam effectively, and the user knows strictly which messages will be filtered out and why.

Automated filtering software uses statistical analysis of e-mail content to decide whether it is spam. In practice, Bayesian spam filtering methods are popular. [8] For these methods to work, preliminary "training" of the filters is required by sending them manually sorted letters to reveal periodic messages and spam's statistical features. E-mails have many attributes, such as headers, domain keys, MX records, etc. used by automatic spam filters to analyze them. [9] Therefore, the question of choosing a fast and efficient automated spam filter algorithm is essential. The following charts use the notation for recognition algorithms to be analyzed: [11]

- Bc –Bayes;
- Kn –Correlation;
- Mh – Mahalanobis distance;
- Th – Thermal agitation;
- Kp – Capon.

We described possible algorithms for using automated spam filters to enhance performance. Each of the algorithms has its advantages and disadvantages, among others [11]. The first figure shows the efficiency of algorithms when the number of email attributes is more than 5 (figure 1).

The second figure represents a graph of the efficiency of the algorithm when the number of system errors is more than 5 (figure 2).

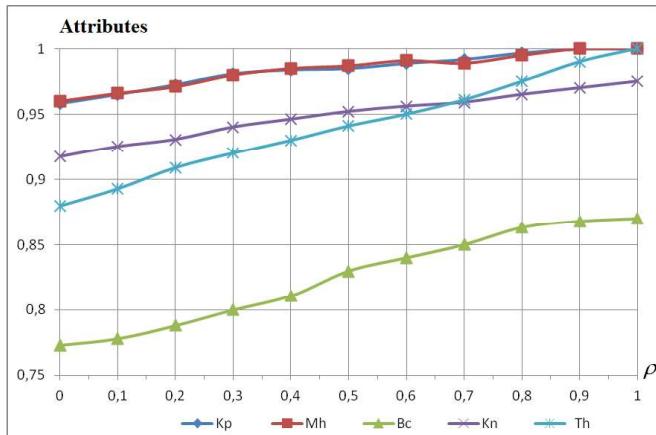


Figure 1 - Dependences on the coefficient with the number of e-mail attributes

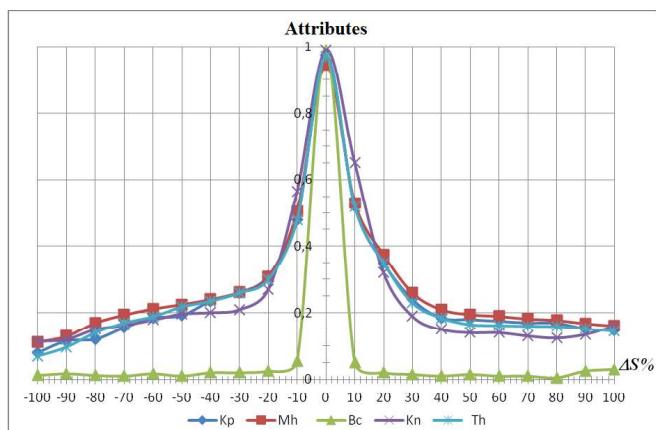


Figure 2 - Dependences on the coefficient with the number of false positive attributes

Mathematical modeling methods analyzed the effectiveness of detecting spam or malicious e-mails by the proposed algorithms. In this article, we confirm that Capon's algorithm is the most effective compared to other methods. Its use in preventing spam messages and improving the performance of the automated spam filter by 5%.

**References:**

1. Agrawal B, Kumar N, Molle M (2005) Controlling spam emails at the routers., Proceedings of the IEEE international conference on communications, ICC 2005, vol 3, pp 1588–1592
2. Albrecht K, Burri N, Wattenhofer R (2005) Spamato—an extendable spam filter system, Proceedings of second conference on email and anti-spam, CEAS'2005
3. Androutsopoulos I, Koutsias J, Chandrinou KV, Spyropoulos CD (2000a) An evaluation of naive bayesian anti-spam filtering, Potamias G, Moustakis V, van Someren M (eds) Proceedings of the workshop on machine learning in the new information age, 11th European conference on machine learning, ECML 2000, pp 9–17
4. Androutsopoulos I, Koutsias J, Chandrinou KV, Spyropoulos CD (2000b) An experimental comparison of naive bayesian and keyword-based anti-spam filtering with personal e-mail messages, Proceedings of the 23rd annual international ACM SIGIR conference on research and development in information retrieval, SIGIR '00. ACM Press, New York, NY, USA, pp 160–167. ISBN 1-58113-226-3. <http://doi.acm.org/10.1145/345508.345569>
5. Androutsopoulos I, Palioras G, Karkaletsis V, Sakkis G, Spyropoulos C, Stamatopoulos P (2000c) Learning to filter spam e-mail: a comparison of a naive bayesian and a memory-based approach, Zaragoza H, Gallinari P, Rajman M (eds) Proceedings of the workshop on machine learning and textual information access, 4th European conference on principles and practice of knowledge discovery in databases, PKDD 2000 pp 1–13
6. Androutsopoulos I, Palioras G, Michelakis E (2004) Learning to filter unsolicited commercial e-mail (Technical Report 2004/2). NCSR “Demokritos”. Revised version
7. Androutsopoulos I, Magirou E, Vassilakis D (2005) A game theoretic model of spam e-mailing, Proceedings of second conference on email and anti-spam, CEAS'2005
8. Aradhye H, Myers G, Herson J (2005) Image analysis for efficient categorization of image-based spam e-mail, Proceedings of eighth international conference on document analysis and recognition, ICDAR 2005, vol 2. IEEE Computer Society, pp 914–918.
9. Blanzieri E, Bryl A (2007) Evaluation of the highest probability svm nearest neighbor classifier with variable relative error cost, Proceedings of fourth conference on email and anti-spam, CEAS'2007. pp 5
10. Enrico Blanzieri & Anton Bryl (2009) - A survey of learning-based techniques of email spam filtering, Artificial Intelligence Review, pp 62-92.
11. Наконечний В.С Методи та засоби підвищення ефективності функціонування радіотехнічних систем розпізнавання багатопозиційного базування / В.С. Наконечний, С.В. Толопа, В.А. Дружинін, Н.В. Лукова-Чуйко, І.І. Пархоменко: монографія - К.: 2019. - 237 с.

**<sup>1</sup> Anastasiia Nicheporuk**

Postgraduate Student at the Computer Engineering & System Programming Department

**<sup>2</sup> Oleg Savenko**

Doctor of Technical Sciences, Full Professor, Computer Engineering & System Programming Department

**<sup>3</sup> Andrii Kazantsev**

Postgraduate Student at the Computer Engineering & System Programming Department

<sup>1-3</sup> Khmelnitsky National University

## THE ARCHITECTURE OF CNN MODEL FOR ANDROID MALWARE DETECTION

Humanity has made a significant leap forward in the development of the information technology industry and, in particular, mobile devices operated by the Android operating system. However, the advent of new mobile devices features has automatically created new vulnerabilities for them and has driven an increase in the amount of malware, which are trying to use them.

Considerable attention today is being paid to the problem of Android malware detection. Several solutions have been developed using an academic approach, utilizing features such as permissions, API calls, opcodes, strings, metadata or intents. Naive Bayes, J48, decision tree, C-means clustering methods are used as machine learning algorithms [1, 2]. However, the presented solutions are characterized by a fairly high level of complexity which prevents their use in real-time systems.

As a result of ours' research architecture of convolutional neural network (CNN) for Android malware detection have been proposed. Involvement of convolution layers creates an analogy with the human brain, allowing the identification of local features that are subsequently fed to the input of fully connected layers to form a membership degree of an input object to one of the predicted classes. In the field of pattern recognition, such features may be, for example, the presence of inclined lines at a certain angle. Another important advantage is that the weights in the convolution layer are locally connected and move throughout the feature map. This leads to involving much less of a number of weights compared to fully connected neural network architectures.

When designing architecture of neural network, we aimed to use it to detect malware. As an input data for convolutional neural network, we used the API method calls and a set of permissions for Android app. Application Program Interface (API) is a set of procedures that represent an intermediate layer for communicating applications between themselves and the Android kernel. In fact, no one high-level action doesn't take place without the participation of API invocation. Thus, by analyzing them, we could represent the behavior of the application through the sequence of API calls. For example, the sequence `getDeviceId()`, `loadLibrary()`, `sendTextMessage()` might be determined as the behavior of receiving and sending information. The detection process may then be

defined as a procedure of search the similarity of the program's behavior with the knowledge about the typical malware behaviors.

Except API calls, no less important attributes that can enhance behavior representation is a set of app's permissions. The permissions mechanism restricts access to certain components or functionalities of the application. All permissions used by the application are specified in the AndroidManifest.xml file. According to the results of previous studies, the distribution of permissions in malware and benign applications is differed. Thus, knowledge of attracting permissions may indicate a set of potential actions that will need to be granted. In order to implement convolutional neural network both type of data was represented in binary form.

The proposed neural network consists of two separate parallel convolutional branches, each of which processes its own type of data (see Fig. 1). As a result of convolution and max-pooling operations, the input data, i.e. behavioral patterns of Android app, is prepared for fully connected layers (FCL). In order to produce nonlinear decision making, there is one hidden layer between the first and third FCL. The result is provided by the last layer consisting of two neurons.

The proposed neural network architecture utilized an approach without convolutional and max-pooling layers alternating. This is due to the fact that after the next pair of CONV + POOL layers, the dimension of the data decreases, which leads to the loss of some information about the input object. In the proposed architecture, in each of the two sub-branch, the two convolution layers  $C_{11}$  and  $C_{12}$ , as well  $C_{21}$  and  $C_{22}$ , are placed one after the other, where the first convolution layers  $C_{11}$  and  $C_{21}$  highlight simple features that will be used by the layers  $C_{12}$  and  $C_{22}$  to represent higher-level behavior patterns. The input matrices with size  $K_{11} \times D_a$  and  $K_{21} \times D_p$ , respectively, are feed to the convolutional layers  $C_{11}$  and  $C_{21}$ . The  $D_a$  and  $D_p$  are the dimension of input feature vector for API call and permissions respectively. For layers  $C_{12}$  and  $C_{22}$ , the size of the convolution kernel is  $K_{12} \times 1$  and  $K_{22} \times 1$ , respectively. Following each pairs of convolution layers there is placed one aggregation layer, which reduce the dimension of each type of feature. In order to transform the data into a one-dimensional vector, each sub-branch uses a Flatten layer. After concatenation the data of both sub-branches, the resulting vector with size  $F_1 + F_2$  is feeds to FCL. The output layer consists of two neurons that accumulate the probability that a suspicious Android application belongs to one of two classes – malware or benign. The quantitative indicators of the proposed convolution network are presented on the Table 1.

For all layers, except last, the ReLu activation function was selected. The neurons of the last layer were activated by a softmax function that simulates the probabilities of belonging suspicious app to one of the two classes. The neural network minimized the cross-entropy loss function. In order to reduce the impact of overfitting of the neural network between fully connected layers dropout regularization was used with parameter  $p = 0.5$  (during testing, the dropout parameter was  $p = 1.0$ ). The learning rate and the batches size were set at 0.001 and 64, respectively.

Table 1 – Quantitative indicators of the proposed convolution network

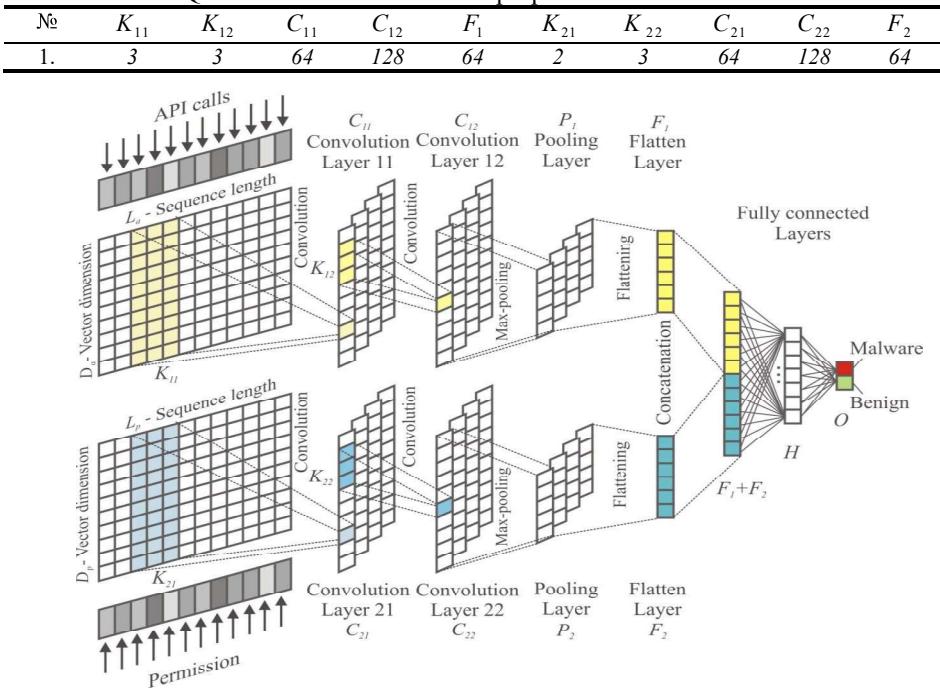


Figure 1 – The architecture of CNN model for Android malware detection

We propose architecture of convolution neural network which is the basis for Android malware detection. The Android malware detection process is based on the involvement of a neural network for training on a test sample, each instance of which is presented in the form of API calls and a set of permissions. The architecture of the proposed neural network consists of two separate parallel convolutional branches, each of which processes its own type of data. The outputs from both branches of the network are combined to form the input for fully connected layers, which determine the probabilities of belonging suspicious app to one of the classes – malware or benign.

#### References:

1. M.K. Alzaylae, S. Yerima, S. Sezer DL-Droid: Deep learning based android malware detection using real devices, Computers & Security (2020) vol. 89 doi: 10.1016/j.cose.2019.101663
2. L. Wen, H. Yu, An Android malware detection system based on machine learning, Proceedings of the International Conference on Green Energy and Sustainable Development, Chongqing City, China, 2017, pp. 1-7 doi:10.1063/1.4992953

<sup>1</sup> **Ponomarov Serhii**

Student.

<sup>2</sup> **Natalia Lukova-Chuiko**

Doctor of Technical Science, Professor of the Department of Cybersecurity and Information Protection.

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## BREACH AND ATTACK SIMULATION AS A NEW VECTOR OF INFORMATION SECURITY

**Abstract.** This article discusses the excellence of the increasingly popular automated testing tool over today's common network security testing tools. This approach allows you to automate the capabilities of the traditional penetration test, keep the condition of protection under constant control and monitor all key vectors of attacks simultaneously.

**Keywords:** BAS, Red Team, Penetration Test, IT, SaaS, cybersecurity

### 1. Introduction

Penetration test is one of the most common methods for assessing the level of protection and security strength, demonstrating methods for attacks and identifying existing security problems.

However, penetration test require a significant portion of human participation and are held with a certain frequency and in a short period of time. The results reflect the static image recorded at the time of the event.

Nowadays Breach and Attack Simulation (BAS) is the most popular product for simulating intrusions and attacks on the new, rapidly growing market of tools for automatic security testing in real time.

### 2. Analysis of how to test network security and their shortcomings

Penetration tests are performed manually by employees of some company or external consultants who try to assess the security of infrastructure of organization by breaking it down safely.

There are some drawbacks of the penetration test:

- The results depend on skills and experience of the tester and do not give a complete image as it is not possible to check all aspects of the system manually.
- A limited testing environment does not allow using of all features that real hackers could have.
- A tester is not able to check all known attack technologies.
- Penetration test results reflect the state of the system in a certain period of time. The high cost of such testing does not allow it to be conducted often.

Red Team testing is an imitation of a targeted cyber attack, which is becoming more and more popular. In addition to identifying critical vulnerabilities and the overall security assessment, a proactive approach provides valuable information about the ability of IT services to detect and block attacks directly during their implementation. [1]

The disadvantages of this approach are:

- Requires trained full-time or part-time staff.
- Imitations should be performed regularly.
- Test results can be difficult to compare as they can be performed under different rules and conditions.
- Significant resources are involved.
- Due to insufficient automation, it is important to repeat testing uniformly.
- It is difficult to assess the impact of changes in the IT environment and track the dynamics of protection effectiveness.

### **3. Breach and Attack Simulation as the way automatic testing**

The BAS platform develops the idea of simulating targeted attacks and assesses the actual readiness of the organization to repel cyber attacks. This method allows to detect critical infrastructure vulnerabilities by conducting cyberattacks from several vectors as it would be done by real attackers. [2,4]

Penetration tests are carried out according to the patterns of real hacker groups, state cyber forces and even on behalf of imaginary unreliable employees.

The SaaS model allows you to run simulations at any time without affecting users or infrastructure[3].

Here are some reasons why BAS is singled out in the cybersecurity services market:

- Allows to automate the verification process.
- Decreases the influence of the human factor.
- IT infrastructure is a living organism and it changes regularly with new vulnerabilities and threats appearing every day, which leads to the need for constant security testing.
- Not every big enterprise can afford continuous penetration testing services or its own Red Team.

### **4. Conclusion**

From the material mentioned above we can make a conclusion that BAS products allow companies to independently and continuously assess their own security, to check security mechanisms by simulating attacks in various directions. Thus, the use of BAS solutions for continuous security assessment can significantly increase the level of actual security of the IT infrastructure.

### **5. References:**

1. List of Adversary Emulation Tools. Available: <https://pentestit.com/adversary-emulation-tools-list/>
2. What is breach and attack simulation. Available: <https://blog.cymulate.com/what-is-breach-attack-simulation>
3. H.A. Kuchuk, A.A. Kovalenko, N.V. Lukova-Chuyko Method to minimize the average delay of packets into virtual connection support network cloud services. Management systems, navigation and communication. - Poltava. National University «Yuri Kondratyuk Poltava Polytechnic», 2017. - Vol. 2 (42). - P. 117-120.
4. Ruban, V. Martovitsky, N. Lukova-Chuiko approach to classification of network condition on the basis of statistical parameters for detection of anomalies in the information structure of the computing system. Cybernetics and Systems Analysis. 2018. Vol. 54, No. 2. P. 142-150.

**<sup>1</sup> Volodymyr Rusyn**

PhD, Assistant Professor of Department of Radio Engineering and Information Security

**<sup>2</sup> Aceng Sambas**

Lecturer of Department of Mechanical Engineering

*<sup>1</sup> Yuriy Fedkovych Chernivtsi National University, Ukraine*

*<sup>2</sup> Universitas Muhammadiyah Tasikmalaya, Indonesia*

## **SIMPLE AUTONOMOUS SECURITY SYSTEM BASED ON THE FINGERPRINT SCANNER MODULE AND ARDUINO PLATFORM: A STUDY CASE**

The basic premise of biometric authentication (the term is derived from the Greek word “bio” meaning life and “metric” meaning to measure) is that every person is unique and each individual can be identified by his or her intrinsic or behavior traits. Biometric technology is able to recognize a person on the basis of the unique features of their face, fingerprint, signature, DNA or iris pattern and then impart a secure and convenient method for authentication purposes [1-7].

Biometrics is therefore the measurement and statistical analysis of a person's physical and behavioral characteristics. For example, voice recognition systems work by measuring the characteristics of a person's speech as air is expelled through their lungs, across the larynx and out through their nose and mouth.

The speech verification software will compare these characteristics with data already stored on the server and if the two voiceprints are sufficiently similar, the biometric security system will then declare it a match. In this paper, we proposed a simple autonomous security system that based on the fingerprint scanner module and Arduino Uno.

Arduino board was designed in the Ivrea Interaction Design Institute intended for students without a background in electronics and programming concept. This board started altering to adapt to new requirements and challenges, separating its present from simple 8-bit boards to products for IoT (Internet of Things) applications, 3D printing, wearable, and embedded surroundings [8]. All boards are entirely open-source, allowing users to build them separately and finally adapt them to their exact needs. Over the years the Arduino boards has been used to build thousands of projects, from daily objects to compound scientific instruments.

The Arduino integrated development environment is an environment in which an Arduino board can be programmed. A written program or code is called SKETCH. In this work, the Arduino IDE is used as an environment in which the Arduino Uno program is written, compiled and uploaded on the Arduino board. The Arduino can be

connected to a computer through the USB port and programmed using a language similar to C++. The connection scheme is quite simple and it's rather difficult to make a mistake (Figure 1).

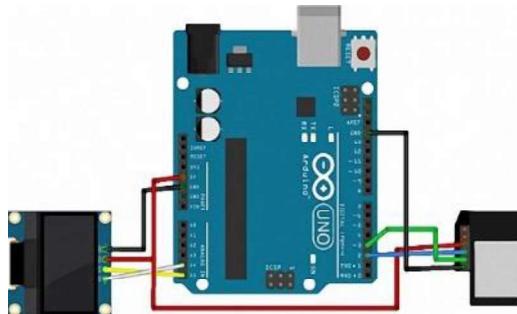


Figure 1 – The connection scheme

#### References:

1. Elmir, Y., Ghazaoui, O., Boukenni F.: Multimodal Biometrics System's Resistance to noise (Fingerprint and Voice). CEUR Workshop Proceedings **942**, 25-28 (2012)
2. Xia S., Liu Y., Yuan G., Zhu M., Wang Z., Indoor fingerprint positioning based on Wi-Fi: an overview, ISPRS Int. J. Geo-Inf. 6, 135, (2017)
3. Suining H., Gary Chan S. H., Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons, IEEE Commun. Surv Tut, 18, 466-490 (2016)
4. G. Li, C. Busch and B. Yang, "A novel approach used for measuring fingerprint orientation of arch fingerprint," 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, 2014, pp. 1309-1314. doi: 10.1109/MIPRO.2014.6859770
5. S. P. Sandip and P. H. Zope, "Selective review of fingerprint enhancement, classification and matching techniques," 2015 IEEE Bombay Section Symposium (IBSS), Mumbai, 2015, pp. 1-6. doi: 10.1109/IBSS.2015.7456656
6. Caso G., De Nardis L., On the Applicability of Multi-wall Multi-floor Propagation Models to WiFi Fingerprinting Indoor Positioning, Future Access Enablers for Ubiquitous and Intelligent Infrastructures: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Springer, (2015)
7. Tomas Trainys, Algimantas Venčkauskas. Encryption Keys Generation Based on Bio-Cryptography Finger Vein Method. CEUR Workshop Proceedings **2145**, 106-111 (2018)
8. Rusyn V, Subbotin S., Sambas A.: Analysis and Experimental Realization of the Logistic Map Using Arduino Pro Mini. CEUR Workshop Proceedings **2608**, 300-310 (2020).

<sup>1</sup> **Anastasiia Shved**

Bachelor's degree, Student

<sup>2</sup> **Sergii Buchyk**

Doctor of Technical Sciences, Professor

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## BASIC APPROACHES TO PERSONAL DATA PROTECTION IN CLIENT RELATIONSHIP MANAGEMENT SYSTEM

**Abstract.** Customer Relationship Management (CRM) systems based on the storage of customers' personal data and their processing in the system. Ensuring proper protection of this data is the most important stage in the development, implementation and enforcement of this system. Modern CRM systems are created using various data protection methods, which will be outlined in this work.

**Keywords:** Customer Relationship Management system, personal data protection, CRM system architecture, logging, access delimitation.

The business processes of any commercial organization are aimed at ensuring its main business goals: making a profit, providing services or goods. It is not a secret that most of a company's profits depend on customers, that why the loss of them means the loss of most of the revenue. Therefore, attracting new customers, awakening existing ones – all this forces companies to spend a lot of resources on their marketing companies. It is also important to know everyone's needs, to be able to offer exactly what person interested. Only this approach will provide an effectively customers' attract. Small businesses, can save the entire customer base, for example, in Excel. For companies with a large amount of customer data and complex business processes, this will not be enough. In addition, maintaining a large customer base in such programs can lead to loss or damage some data, because of fact that they don't provide any protection of information, and all security is assigned to access to the user's workplace.

The Customer Relationship Management (CRM) system allows storing customers' personal data in a convenient way, providing complex connections between them. In addition, modern CRM systems allow automating many processes of the enterprise, which will work out depending on the characteristics of each client and his needs. [1] Such automation can significantly speed up the customer service process – in today's pandemic realities this is a very big advantage – the speed of work minimizes the manifestations of large queues and crowds of people in one room. Great opportunities for the use of CRM systems determine their growing popularity and necessity.

However, usually, no matter how extensive the capabilities of modern CRM systems, companies seek to customize it to their needs. Customers who use CRM systems, in most cases, pay attention to expanding its capabilities, and security issues are left to the developers of the platform. But it should be noted that the storage of customers' personal data in one place is the main advantage, but at the same time the disadvantage of Customer Relationship Management systems – having access to the

system and its main components provides an access to customers' personal data. Therefore, it is necessary to understand the basic capabilities of protection methods in CRM systems in order to build the right vector for the development of personal data protection system in the CRM system.

On the example of the architecture of CRM system Creatio we will consider the main approaches to data protection in the CRM systems. [2]

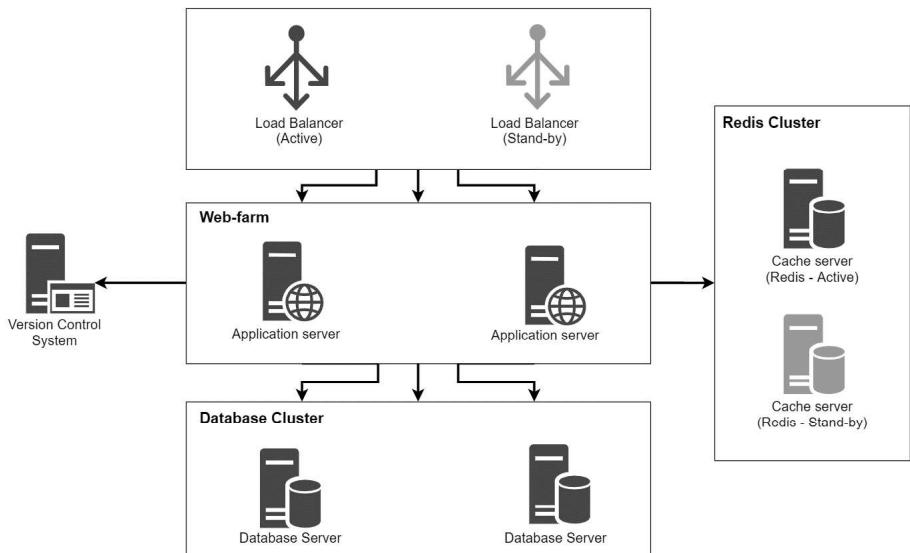


Figure 1 – Customer Relationship Management system architecture

According to the architecture of the presented CRM system, it is necessary to provide protection at the following levels: authorization to the system, work with personal data and security of data transfer between the main components of the system.

Access to any secure system begins with entering login and password – this allows to identify the user who plans to work in the system and to authorize the user in the system. An effective way to protect data at the application server level is to integrate CRM systems with security systems of the database management system (DBMS). With this approach, a separate record is created for each user in the system in the corresponding database table, and all rights are distributed at the database level.

Anyone in the enterprise's local network can get an access to the application by following the link. In case of the databases, only an employee with the role of system administrator, database administrator or security administrator can get an access to database tables, which minimizes the risk of login and password theft. In addition, user passwords are usually encrypted with a 128-bit key and stored in the database in an unreadable form. [3]

As mentioned above, companies seek to expand the functionality of the CRM system to their needs. This uses a variety of web services, which in turn connect using the SSL protocol that allows using public key encryption to authenticate and encrypt client-server connections. [4]

Virtually any CRM implementation project is carried out with the initial configuration of access rights for individual users or groups of users within the system. System administrators can configure access for the entire table, as well as for individual records or fields in the table. Rights are distributed at the database level, so it is important to protect database servers most effectively, because gaining direct access to the database gives an access to all information in the system.

Another good tool for recording system changes is logging, which allows recording the fact of adding, changing or deleting data, the time of these actions and the users who make them. In addition, this tool allows tracking the actions that may preceded the threat. The collection of statistics can prevent the realization of the threat of copying data, or their change: one suspicious action can be explained by human factors, two – by coincidence, three similar actions are the reason to pay more attention to these actions.

After analyzing the approaches to data protection in CRM systems, we can conclude that the basic methods of personal data protection meet the basic needs of the enterprise. However, to achieve a higher level of protection, it is necessary to refine the protection system, using, for example, two-factor authorization by phone number, allocation of Demilitarized Zone, development of additional processes for tracking employees and preventing third parties from accessing system components. All of these methods will be considered in next papers.

### References:

1. Efromeeva Elena Valentinovna, Lelaev Magomed Isaevich, Efromeev Nikolay Maksimovich The relevance of the implementation of CRM systems // Problems of Science. 2016. No. 8 (50)
2. Creatio Customer Relationship Management system documentation. URL: <https://academy.terrasoft.ua/documentation>
3. Shpilevoy DI Regulation of data access in Oracle DBMS // Izvestiya SFU. Technical science. 2008. No. 8
4. Bekker Mikhail Yakovlevich, Terentyev Andrey Olegovich, Gatchin Yuri Armenakovich, Karmanovsky Nikolay Sergeevich Using digital certificates and SSL / TLS protocols for data encryption in cloud computing // Scientific and technical bulletin of information technologies, mechanics and optics. 2011. No. 4 (74)

**<sup>1</sup> Lada Slipachuk**

PhD student

**<sup>2</sup> Serhii Toliupa**

Doctor of technical sciences, professor

*<sup>1,2</sup> Taras Shevchenko National University of Kyiv*

## **SYNTHESIS FEATURES OF FUNCTIONAL MODEL OF INTEGRATED INDUSTRY MANAGEMENT SYSTEM OF NATIONAL CYBERSECURITY**

As of today, a number of issues concerning constructing functional models of industry integrated automated IT, providing general industry management based on integrated automated IT, and ensuring industry integration in functional modeling is still completely uncovered.

The national cybersecurity sector is a strategically important but problematic area for the state. Gradually developing and improving, the national cybersecurity system looks for and builds new trend-perspective mechanisms for its improvement:

- by solving the problems of good governance;
- based on the development and implementation of new management tools and technologies.

Therefore, providing computer-integrated management of the national cybersecurity sector based on an integrated industry-specific information control system is of strategic importance for the subject area of IT [1].

**Setting the research objective.** The problems outlined above determined the objective of our scientific research, which is the substantiation of features of a functional type of model creation of a branch integrated system of management of the sector.

**The topic of the presented research** is quite relevant at the state level because:

- it is linked closely to the common state needs and industry problems;
- it is aimed at the creation of a functional model of a sectoral integrated management system, which is a modern tool for improving the management of the national cybersecurity sector.

**Analysis of recent research and publications, which started the elaboration of the problem.**

Peculiarities of engineering of integrated management systems of the national cybersecurity industry were studied by S. Toliupa, V. Nakonechnyi, L. Slipachuk and other scientists [1-5].

**Methodological apparatus of the research.** In order to build the model a system-integrated approach for structural and functional filling of the model was used.

**Summary of the main research material.**

The format of representation of the genesis essence of the functional model was decided to be fixed by the following argumentation:

1. Due to a pragmatic and rational approach when designing the model, it was decided to dwell on the functional type of model.
2. The functional model type:

- is at most applied, informative, dynamic, effective, and flexible;
- meets the requirements that are advanced to its functional capacity of the system;
- clears up functioning mechanisms [2].

**At the engineering stage, functional modeling of industry-integrated control systems required:**

- an in-depth study of the structural and functional features of the industry macro-object that needs management.
- adequate consideration of the needs, tasks, goals, objectives and specifics of the national cybersecurity sector as a problem area.

**The design process of functional model engineering:**

- provided a high degree of scientific and methodological reasoning for research and design work;
- was focused on significant aspects concerning the specifics of the national cybersecurity sector;
- separated everything secondary, minor and insignificant, which does not affect the achievement of the goal;
- singled and took into account only the functional aspects [3].

Let us consider in more detail the typological features of the model [2].

1. The functional model was based on certain functional aspects that not only provide operational efficiency of the control system but also allow displaying:

- functions performed by each component of IT as a control system;
- processes occurring in the system;
- inputs to the system and outputs from the system;
- the behavior, mode of action, and properties of IT as a control system;
- information flows and control signals;
- closed loops of work cycles;
- functional load of working modules;
- dynamics of everything that occurs within the model;
- all interconnections and interoperability of parts in the functioning of IT as a control system;
- functional properties of a control system;
- functional and industry capabilities embedded in the simulated system [4].

2. Architectural basis of the system is working modules, as software subsystems, united by a single task in the software and technology chain.

3. Working modules as working subsystems have defining characteristics. They:

- perform specific tasks that go beyond goals and objectives;
- implement the requirements that are imposed on them;
- carry multifunctional workloads (manage modes, communications, data exchange, processes, informatization, analytical data processing, managerial decision-making);
- play an applied role as regards the integrated provision of centralized

management of the national cybersecurity sector;

- work according to their own principles of functioning [5].

This feature of the structure allowed to ensure the functional industry and management capacity of the system.

**Obtained scientific result and value of the research.** The presented paper discloses:

- genesis essence of the functional type of the model of the branch integrated system of management of the national cybersecurity sector;
- features of the creation of a functional type of model of branch integrated system of management of the national cybersecurity sector;
- elements of scientific research aimed at solving the problems related to the management of the national cybersecurity system on the basis of a functional model of an integrated sectoral information system for management;
- reasons for choosing a functional type of model of the integrated sectoral information system for management.

**Conclusions.** The materials of this article prove that the functional type of the model of the integrated branch information system of management of the national cybersecurity sector is the most appropriate choice.

**Prospects of further research.** Further research advisable to focus on the creation and implementation of integrated sectoral management information systems and their models for other sectors of the military defense or industrial complexes of the state.

#### **References:**

1. Slipachuk, Lada. (2020). Relevance of the integrated information management system as a subject of management of the national cyber security sector of Ukraine. The synergetic concept. Journal of Physics: Conference Series. 1454. 012002. 10.1088/1742-6596/1454/1/012002.
2. Slipachuk, Lada & Nakonechnyi, Volodymyr. (2019). Typology of the Model of Integrated Sectoral Information System of the National Cyber Security Management. doi: 271-276. 10.1109/ATIT49449.2019.9030595.
3. Lada Slipachuk. Design principles of the functional model of the integrated industrial MIS by the national cybersecurity sector // International science journal "Polish Science Journal" (ISSUE 3(24), 2020) – Warsaw: Sp. z.o. o. "iScience", 2020. – pp. 39-46. - ISBN 978-83-949403-4-8.
4. S. Toliupa, L. Slipachuk Applied aspects of the integrated industry information system management model for the national cybersecurity sector // 2rd International Scientific-Practical Conference on Problems of Cyber Security of Information and Telecommunication Systems (PCSITS) – 2019, pp. 305-308.
5. Lada Slipachuk. Functional principles of the model of the integrated industrial MIS by the national cybersecurity sector // Actual scientific researches in the modern world. Scientific publications Journal – No. 3 (59) Part 1, 2020. - pp. 75-85. - ISSN 2524-0986.

**<sup>1</sup> Mykola Stetsiuk**

Postgraduate Student at the Computer Engineering & System Programming Department

**<sup>2</sup> Andrii Nicheporuk**

Candidate of Technical Sciences (PhD), Associate Professor at the Computer Engineering & System Programming Department

**<sup>3</sup> Bogdan Savenko**

Master Student at the Computer Engineering & System Programming Department

<sup>1-3</sup> Khmelnitsky National University

## **ENSURING THE FAULT TOLERANCE AND SURVIVABILITY OF SPECIALIZED INFORMATION TECHNOLOGIES IN CORPORATE COMPUTER NETWORKS UNDER THE INFLUENCE OF MALICIOUS SOFTWARE**

Fault tolerance and survivability [1] define one goal – to ensure high efficiency of IT, which is achieved in different ways. One of its parameters is the time of unavailability, i.e. the time when the system is unable to perform its functions within the requirements for it. For different systems, this time is different and ranges from zero to a certain, still acceptable value. For specialized IT, which operate in corporate computer networks and perform the function of information support in such a highly specialized subject area as financial and economic activities in various fields of application, this parameter is much higher than zero, but the requirements for such IT are also quite high, especially constant growth of their quantitative parameters of functioning (increase in the number of users, complexity of information flows and volumes of processed data) and work in the conditions of influence of malicious software.

The survivability of the developed IT is provided by: redundancy of the server part of IT with territorial diversity of the main and backup server, the feature of redundancy is that the server function, at a critical moment, takes over the mirror SQL server, which in normal mode provides FTP-server; redundancy of client software, the feature of which is that the reserve is not a dedicated computer, and the performance reserve of individual client computers, which, according to the redundancy plan, is installed software of the client, which in critical the moment will be used as a regular, without losing the functionality of IT.

Fault tolerance of its client part of IT is ensured by performing a set of measures, which includes in addition to traditional hardware redundancy and functional redundancy: organization of automatic updating of system and application software of client PCs by monitoring its relevance with a given frequency; algorithms of procedures that implement critical functions of the client part of IT, with the inclusion of a non-trivial (intelligent) error handling unit, which is performed in parallel with the procedure itself; use of non-trivial data editors, which include in their algorithm an interactive procedure that eliminates uncontrolled manipulation of database data by the operator; implementation of critical for the use of resources, calculation procedures with the ability to quickly select the place of their execution, which prevents overloading of the IT hardware platform.

Maximization of criteria for fault tolerance and survivability in IT configurations based on the client-server architecture can be achieved for each of the parts of the system separately. Then the function  $f_1(S_i), i = 1, 2, \dots, n$  determination of fault tolerance in computer systems in quantitative form will look like:

$$f_1(S_i) = \frac{T_{f_1(S_i),1}}{T_{f_1(S_i),1} - (T_{f_1(S_i),2} + T_{f_1(S_i),3})}, \quad (1)$$

where  $i$  is the number of components of specialized IT,  $i = 1, 2, \dots, n$ ,  $T_{f_1(S_i),1}$  – time between adjacent failures;  $T_{f_1(S_i),2}$  – the time required to detect the failure and find a way around it;  $T_{f_1(S_i),3}$  – time required to recover IT after failure.

The task of structural survivability analysis requires the definition of: the system architecture required to fulfill the purpose of IT operation at some point or time when adverse effects on the system occur. We define the function  $f_2(S_i)$ , in which  $i = 1, 2, \dots, n$ , the definition of survivability in quantitative units in computer networks is presented as follows:

$$f_2(S_i) = \frac{T_{f_2(S_i),1}}{T_{f_2(S_i),1} - (T_{f_2(S_i),2} + T_{f_2(S_i),3})}, \quad (2)$$

where  $T_{f_2(S_i),1}$  – the time of operation of IT processes in standard mode,  $T_{f_2(S_i),2}$  – the time spent on the processes of survivability,  $i = 1, 2, \dots, n$ .

This definition of the survivability function makes it possible to display the standard mode of operation with a unit value, and if there is a need to ensure survivability and in the case of a much longer time than the standard mode of operation, the function value will display a quantitative ordinal value.

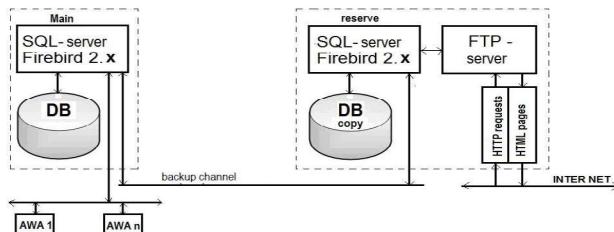


Figure 1 – The scheme of ensuring the survivability and fault tolerance of the server part of the IS.

An important area of further research to improve the effectiveness of IT is to develop a method to ensure effective protection of information directly in the structure of IT and computational processes that take place under the influence of malware.

#### References:

1. DSTU 3396.2-97 Zaxy'st informaciyi. Texnichnyj zaxy'st informaciyi. Terminy' ta vy'znachennya [DSTU 3396.2-97 Information protection. Technical protection of information. Terms and definitions]

**<sup>1</sup>Serhii Toliupa**

Doctor of Technical Sciences, professor

**<sup>2</sup>Mykola Brailovskyi**

Candidate of Technical Sciences, Associate Professor

**<sup>3</sup>Ivan Parkhomenko**

Candidate of Technical Sciences, Associate Professor

**<sup>4</sup>Bohdan Zhurakovskiy**

Doctor of Technical Sciences, professor

<sup>1,2,3</sup>*Taras Shevchenko National University of Kyiv*

<sup>4</sup>*National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"*

## **SAFETY OF CRITICAL FUNCTIONS INFRASTRUCTURE**

**Abstract.** The questions of conceptual and regulatory bases of critical infrastructure's information security are considered, and the critical infrastructure objects cyberpower level assessment under the conditions of external cybernetic influence is carried out. The strategy, that allows us to obtain a quantitative assessment of the critical infrastructure protection level from the risk of external cybernetic influence and to establish requirements for the cybernetic security systems formation with a set of measures aimed at increasing the security level of these objects, is proposed.

**Keywords:** critical infrastructure, objects of critical infrastructure, threats to critical infrastructure, protection of critical infrastructure, cyberspace, cyberattack, cybersecurity, cyberpower.

### **Introduction**

In many countries, the concept of critical infrastructure is being implemented, which allows us to focus on systems, networks, and individual objects, the destruction or disruption of which will have serious negative consequences for national security. As the world experience shows, the process of establishing a legislative framework in the field of critical infrastructure protection is rather laborious and long-lasting. Legislation on the protection of critical infrastructures in different countries is often uncoordinated, in addition, there are certain problems with the mechanisms of assigning objects to critical infrastructure. Each country defines critical infrastructure by considering its specificity, the criticality of individual sectors, and the importance of certain services for the society and the state's security [1].

### **Analysis of known research and problem statement.**

Some countries have tried to define the critical infrastructure and develop a strategy for its protection. The list of vital (critical) infrastructures is different for each country and is determined according to its traditions, social and political beliefs, as well as geographical and historical characteristics of each country [2].

After researching scientific publications and analytical materials related to the international experience of the formation and implementation of the critical infrastructure protection system, it can be concluded that the organization of measures

concerning critical infrastructure protection in different countries is implemented in different ways. In some countries, an organizational model is defined and structured, and measures are focused and systematic, while other countries' models have a non-systematic nature when measures are carried out in an informal manner [3].

### **The purpose and objectives of the research**

To propose a strategy that will provide a quantitative assessment of the critical infrastructure objects protection level from the risk of external cybernetic influence and to establish requirements for the formation of cybernetic security systems with a set of measures aimed at increasing the protection level of these objects.

1. Usually, critical infrastructure includes life support systems (water and heat supply) of megalopolises, high-speed and government communication channels, central authorities, power and transport trunk networks, oil and gas pipelines, seaports, emergency response services, and emergency services to the population, high-tech enterprises and military-industrial complex enterprises [4].

That is why the adoption of measures to formulate a cybernetic security policy at the state level should be a priority task of the political leadership in any country.

2. Cybernetic security has become one of the most important components of any country's national security. The maintenance of a country's optimal state is impossible without the development of a national system based on a tolerant attitude to the norms and international law principles, the protection of the primary values determined by the current legislation, and also the national interests of safeguarding cyberspace.

3. Relying on the experience of scientists, it is impossible to disagree with the fact that the national system of cybernetic security should be the engine of its subjects' interaction in order to unite the special services, state, and law enforcement authorities, which regulate the field of telecommunications and information security.

The main goal of governance in this area should be the development and application of all possible methods for timely detection, cessation, and prevention of cybernetic nature threats.

Tasks such as assessing the protection level of Critical Infrastructure Objects against the risk of third-party cyber impact belong to the multi-criteria class. For their collegial solution under conditions of uncertainty and conflict among the existing methods of mathematical modeling, methods of formation and research of generalized quality indicators using graph analytic and similar approaches, expert methods for solving complex tasks of evaluation and choice of any objects, including special purpose objects, as well as analysis and situations forecast with a large number of significant factors, the most rational and determinant are the expert methods [5]. They provide an opportunity to explore more deeply the phenomena that significantly affect the protection level of both the state as a whole, as well as its individual information and cyberinfrastructure objects against the influence of internal and external cybernetic interventions and threats, to identify the most important and significant in these processes, without omitting those details and interconnections, without which the model of the problem under study cannot be constructed.

## Conclusions

Thus, the proposed strategy will provide an opportunity to get a quantification of the OCI protection level from the risk of the outside cybernetic influence, to set the organizational requirements of their own cybernetic security systems, and to work out measures aimed at increasing their effectiveness. The reason for such actions can be the detection of deviations from the normal mode of IP, IT systems and networks, functioning, as well as respective software and hardware.

## References

1. Slipachuk, L., Toliupa, S., & Nakonechnyi, V. (2019). The Process of the Critical Infrastructure Cyber Security Management using the Integrated System of the National Cyber Security Sector Management in Ukraine. In 2019 3rd International Conference on Advanced Information and Communications Technologies, AICT 2019 - Proceedings (pp. 451–454). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/AIACT.2019.8847877>
2. Toliupa, S., Parkhomenko, I., & Shvedova, H. (2019). Security and regulatory aspects of the critical infrastructure objects functioning and cyberpower level assessment. In 2019 3rd International Conference on Advanced Information and Communications Technologies, AICT 2019 - Proceedings (pp. 463–468). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/AIACT.2019.8847746>.
3. Dovgan, O.D. Critical infrastructure as an object of protection against cybernetic attacks / O. D. Dovgan // Information security: challenges and threats of the present: materials of the scientific and practical conference., April 5, 2013 - K.: P. 17-20.
4. Makhutov N., Petrov V., Reznikov D. Analysis of Critical Facilities and Infrastructures in Russia /2nd International Disaster and Risk Conference, Davos, 2017. P. 118–119.
5. Pederson P., Dudenhoeffer D., Hartley S., Permann M. Critical Infrastructure Interdependency Modelling: A Survey of U.S. and International Research / Idaho National Laboratory Critical Infrastructure Protection Division Idaho Falls, Idaho 83415. 2016.

**<sup>1</sup>Serhii Toliupa**

Doctor of Technical Sciences, professor

**<sup>2</sup>Serhii Buchyk**

Doctor of Technical Sciences, professor

**<sup>3</sup>Yanina Shestak**

Candidate of Technical Sciences

**<sup>4</sup>Andrew Kulko**

Master of Cyber Security

<sup>1,2,3,4</sup>Taras Shevchenko National University of Kyiv

## **CYBERATTACK DETECTION SYSTEMS BASED ON THE SIGNATURE METHOD**

**Abstract.** The presence of important information in the functioning of the systems and objects of critical national infrastructures enables its usage by the negatively-minded elements and groupings for the implementation of unlawful actions in the cyber space by violating the integrity, availability and confidentiality of information, and inflicting damage on information resources and information systems. The purpose of the material is to develop a system for recognizing cyber threats based on signature analysis, which would reduce the time of detection of an attack of a cyber defense system while the number and complexity of cyber attacks are increasing.

**Keywords:** cyberspace, cyber attack, decision-making system, cyber intrusion.

### **Introduction**

Possibilities of the cyberspace, rapid development and implementation of leading-edge information and telecommunication technologies provide the unprecedented opportunities for accumulation of data and its usage. The presence of important information in the functioning of the systems and objects of critical national infrastructures enables its usage by the negatively-minded elements and groupings for the implementation of unlawful actions in the cyber space by violating the integrity, availability and confidentiality of information [1]. The purpose of the is to develop a system for recognizing cyber threats based on signature analysis, which would reduce the time of detection of an attack of a cyber defense system while the number and complexity of cyber attacks are increasing.

### **Main part**

In general, modern systems of intrusion and cyber attacks detection are far from ergonomic and effective solutions, according to the security. But the improvement of efficiency should be considered not only in the sphere of detection of improper activities on the infrastructure of secure information objects, but also according to everyday exploitation of these measures and to the saving of computing power and information resources of an owner of a security system.

The most widespread cyber threats to information resources can be considered as potentially possible cases of natural, technical or human-induced nature, which may lead to unwanted effects on the information system, as well as on the information

stored therein. The emergence of a cyber threat, that is finding the source of actualization of certain events in the threat, is characterized by such an element as vulnerability. By integrating a variety of approaches, as well as suggestions for solving this issue, we believe that the following kinds of cyber threats to information security can be identified: disclosure of information resources; violation of their integrity; failure of the equipment itself [2].

All developers of attack detection systems and organizations that use CADS should understand and study their classification in order to choose the best solutions for information security systems. In the study of various aspects of taxonomy and the application of various options, we can achieve a higher level of security of information systems.

The systems for detecting abnormal behavior are based on the fact that CADS has some features that characterize the correct or permissible behavior of the object of observation.

As the world experience has showed, the most effective methodological approach for constructing of innovative intellectual cyber attack monitoring systems is the way to create a hierarchical multilevel structure of cyber attack detection at the beginning of their implementation [3]. Furthermore, a hierarchical approach allows to solve difficult problems of the information protection process managing from cyberattacks in the distributed information systems (IS) as sequence of local tasks, coordinated with each other.

Let's consider one of the effective methods of detecting intrusions and cyber attacks, which is based on the signature approach. Signatory methods allow you to describe a cyber attack with a set of rules or using a formal model, which can be used as a character string, semantic expression in a special language, etc. The essence of this method is to use a specialized database of templates (signatures) of cyber attacks to find actions which fall under the definition of "cyberattack" [4].

The signature method can protect from a viral or hacker cyber attack when its signature is already known (for example, the unchanged fragment of the body of the virus) and it is included in the database of CADS. If the network is experiencing the first attack from the outside, the first infection is still unknown, and the database simply lacks the signature for its search - the signature method CADS will not be able to signal the danger because it considers the attacking activity to be legitimate.

Most of the existing software products which claim to use the signature method, in fact, realize the most primitive way of signature recognition. In such systems, the signature method is implemented as an algorithm that examines only the dynamics of cyberattack development [5]. And it is based on a state machine to assess the scenario of the developing attack. According to the plan, this approach should allow tracking the dynamics of the development of cyber attacks in accordance with the actions of the intruder, while as the module for data collection even the systems for detecting cyber attacks can be used.

### **Conclusions**

Thus, the effectiveness of the signature CADS is determined by three main factors: the efficiency of refinement of the signature base, its completeness from the

point of view of the determination of the signature of the cyber attack, as well as the presence of intelligent algorithms for reducing the attacking party's actions to some basic steps, within which there is a comparison with the signatures.

In order to implement the chosen method of determination and identification of CADS, models of the signature and statistical analyzers of network traffic are offered, and the fuzzy intellectual system is used to determine the sources of cybermedia and the choice of solutions for their elimination [6].

Most of the existing software products which claim to use the signature method, in fact, realize the most primitive way of signature recognition. In such systems, the signature method is implemented as an algorithm that examines only the dynamics of cyberattack development. And it is based on a state machine to assess the scenario of the developing attack. According to the plan, this approach should allow tracking the dynamics of the development of cyber attacks in accordance with the actions of the intruder, while as the module for data collection even the systems for detecting cyber attacks can be used.

In order to implement the chosen method of determination and identification of CADS, models of the signature and statistical analyzers of network traffic are offered, and the fuzzy intellectual system is used to determine the sources of cybermedia and the choice of solutions for their elimination.

#### **Reference:**

1. Toliupa S., Parkhomenko I Signature and statistical analyzers in the cyber attack detection system. Scientific and Practical Cyber Security Journal (SPCSJ) № 3 (02). c. 47-53
2. Amer, S.H., Hamilton, J.A., "Intrusion Detection Systems, (IDS) Taxonomy – A Short Review," DOD Software Tech News, vol. 13, no. 2, June 2010, DOD Data & Analysis Center for Software, Air Force Research Laboratory, Rome, N.Y., pp. 23 – 30.
3. Toliupa S., Nakonechnyi V., Uspenskyi O. Signature and statistical analyzers in the cyber attack detection system. Information technology and security. Ukrainian research papers collection Volume 7, Issue 1 (12). c. 69-79.
4. 17. IDS / IPS. Netgate Documentation: [website]. Washington: Rubicon Communications LLC, 2017. [Electronic resource]. Online: <https://www.netgate.com/docs/pfsense/ids-ips/>.
5. Ghahramani, Z. An Introduction to hidden Markov models and Bayesian networks / Z. Ghahramani // International Journal of Pattern Recognition and Artificial Intelligence — 2001. — Vol. 15. — P. 9–42.
6. Barbara D. Detecting novel network intrusions using Bayes estimators / D. Barbara, J. Couto, S. Jajodia, N. Wu. // In: Proc. of the 1st SIAM International Conference on Data Mining. — 2001.

<sup>1</sup> **Serhii Toliupa**

PhD in Technical Sciences, Professor at the Department of Cybersecurity

<sup>2</sup> **Volodymyr Nakonechnyi**

PhD in Technical Sciences, Professor at the Department of Cybersecurity

<sup>3</sup> **Maxym Kotov**

3rd year of bachelor's degree student

<sup>4</sup> **Valeria Solodovnyk**

3rd year of bachelor's degree student

<sup>1-4</sup> Taras Shevchenko National University of Kyiv

## SIGNALS ENCRYPTION IN WIRELESS DATA INPUT DEVICES

Cryptography is the science that creates strategies for utilizing complex mathematical changes to transmit data through conveyance channels in a frame that no one but authorized individuals can get. Encryption process is a key object in the field of cryptographic research, it is the method of changing the frame of data which is transmitted through open transmission channels [1-5].

The cipher quality of the encryption is measured by the time that it takes to decrypt the content with brute force which is checking of all conceivable key combinations.

There are two types of symmetric encryption [4, 5]: block encryption and stream encryption. An example of symmetric block encryption algorithm is the AES competition finalist, the American encryption standard - Rijndael [6, 7]. In AES 128 form of the calculation, the cipher key comprises of 128 bits isolated by 16 bytes that are composed to the InputKey matrix. The InputKey comprises 4 columns. Utilizing those columns an arrangement of 44 words ( $w_0 - w_{43}$ ) where each word comprises of 32 bits is shaped. Thus, these words become the round keys.

The AES scheme is shown in Figure 1 [6]:

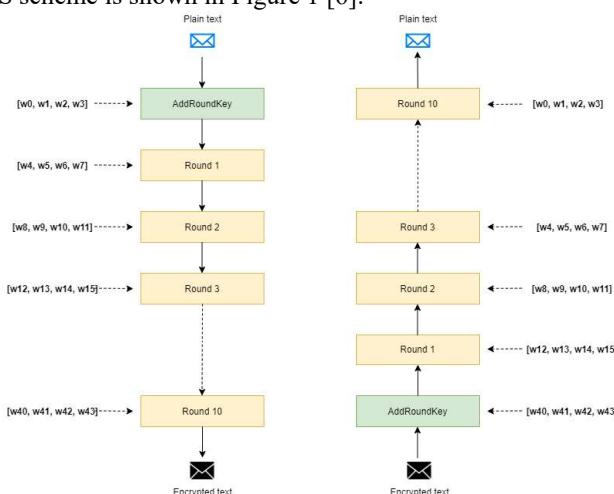


Figure 1 - AES scheme

There are numerous wireless input devices in access at nowadays. As the devices of wireless information input, we consider - the gadget of remote transmission of the entered information to the computer (wireless data input device WDID). Examples of such gadgets are remote consoles or mice, remote headsets, touch input gadgets, etc.

WDID gadgets transmit data through radio waves with a frequency from 27 MHz to 2.4 GHz. WDIDs perform their work with transmitter and a collector. An example of a records acquisition technique is Mousejack.

The essence of this attack is in following stages: a wireless keyboard and a wireless mouse use USB-dongle for signal transmission. a few models of keyboards encrypt those signals, but the majority of mouses do not. In case of keyboards, this works as follows: only USB dongle has the encryption key, which makes it the only item that has the potential to decrypt a signal, an attacker, despite the fact that it intercepts that signal, will no longer be able to decrypt it. But as it was stated most of mouses are not encrypting their broadcast and so the attacker is able to receive unencrypted packets.

The following happens during the first three steps: the user determines the shift (x , y) of the mouse location coordinates, and the transmitter in the mouse transmits the radio signals without encryption to the USB-dongle. At the same time, hackers intercept unencrypted signals using their own personalized USB dongle.

The attacker sends a sequence of requests to connect to the USB dongle during 4-6 stages, then the USB dongle receives the sequence of these requests and connects to the computer of the user.

The attacker sends a series of characters to the user's machine during steps 7-9. If all the stages shown above have succeeded, the hacker has full access to the computer of the victim. This vulnerability can be used on all operating systems since this vulnerability does not apply to operating system vulnerabilities.

### References:

1. Dan Boneh, Victor Shoup, A Graduate Course in Applied Cryptography, version 0.4, Stanford University, September 2017.
2. Bruce Schneier. Applied cryptography. Protocols, algorithms, source texts in the C language.
3. S. Toliupa, L Slipachuk, V. Nakonechnyi. The Process of the Critical Infrastructure Cyber Security Management using the Integrated System of the National Cyber Security Sector Management in Ukraine, 3rd International Conference on Advanced Information and Communications Technologies, AICT 2019 – Proceedings, 2019.
4. William Stallings, Cryptography and Network Security Principles and Practices, Fourth Edition, Prentice Hall, November 16, 2005/
5. Christof Paar, Jan Pelzl, Understanding Cryptography, Springer-Verlag Berlin Heidelberg, 2010
6. Joan Daemen, Vincent Rijmen, AES Proposal: Rijndael, October 1999.
7. Nigel Smart, Cryptography: An Introduction, 3rd edition, McGraw-Hill College, December 30, 2004.

<sup>1</sup> **Sviatoslav Tukalo**

Bachelor, MA Department of Telecommunications

<sup>2</sup> **Orest Kostiv**

Senior Lecturer Department of Telecommunications

<sup>3</sup> **Olga Shpur**

PhD, Assistant of the Department of Telecommunications

<sup>4</sup> **Bohdan Buhyl**

PhD, Assistant of the Department of Telecommunications

<sup>1-4</sup> Lviv Polytechnic National University

## METHODS DEVELOPMENT TO PROTECT IOT FROM BOTNETS

Modern technologies are evolving very fast. Almost every day we can hear news about new developments in a particular field. One of the fastest growing technologies is the Internet of Things. IoT is integrated into almost all spheres of our life. Also it make demands from developers for new requirements of security standards for this technology. Its implementation directly depends on the security of personal data of users.

Telecommunication company Cisco in its annual report [1] notes that almost 90% of all IoT devices are vulnerable to cyberattacks. The number of such gadgets in the world has become larger than population of the Earth, and the approach to cybersecurity almost equal zero, it becomes obvious that cyber attackers are starting to create IoT-based botnets (Fig. 1).

### THE TREND

C&C QUANTITY

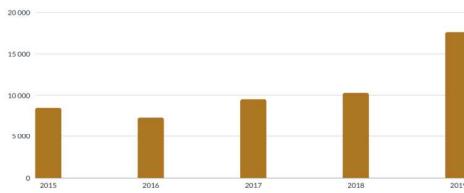


Fig. 1- The trend of increasing C&C servers

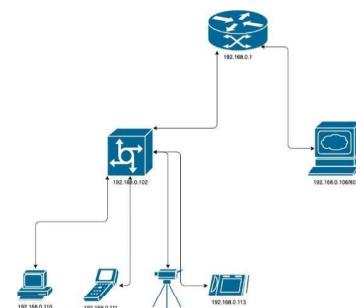


Fig. 2 - Scheme of the experiment

A botnet it's a typical computer network, that includes in its hierarchy such elements like: botmaster, C&C server and bots. Main principles of botnets are: infection, connection, control and pervasion. Investigation shows, that there are three common and typical Botnet attacks – DDoS, Spam and Brute - force.

For this research botnet with Client-Server architecture was developed. The scheme of the experiment is presented on the Figure 2. The operating system of the attacking host was Kali Linux 64 bit 2020.2 Operating system of the victim's host - Debian 10 64 bit.

For this attack was used usual ICMP network protocol from command ping. The default command view looks like “ping [-AaDdfnoQqRrv] [-c count] [-G sweepmaxsize] destination IP”. As this command is standard for all operating systems, it was initially decided that it was necessary to investigate the possible generated traffic when executing this command. It was explored that the network card processed traffic of 1,31 kbit/s. By using the same configuration for the command ping, but sent at the same time from 4 bots, traffic amounted to 5,25 kbit/s. As a result of the system work, it is established that under normal conditions ping request cannot injure the system, because 100 bots will generate 0.13 Mbit/s traffic, but when changing the packet size and packet sending speed, 100 bots will generate traffic exceeding 77 Mbit/s.

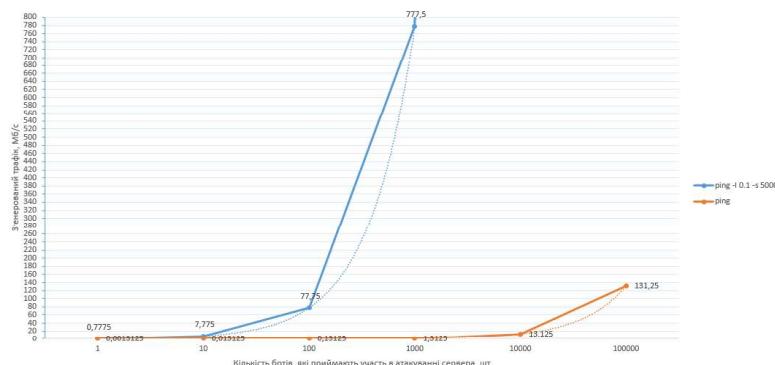


Fig. 3 - Comparative characterization of the generated traffic size using ping commands with parameters and without

By investigating the structure of TCP packet requests, it became clear that the SYN flood can disable the system very quickly, by creating half-open connections. Therefore, a botnet of 100 bots will incapacitate the home WEB server in ~ 3 seconds.

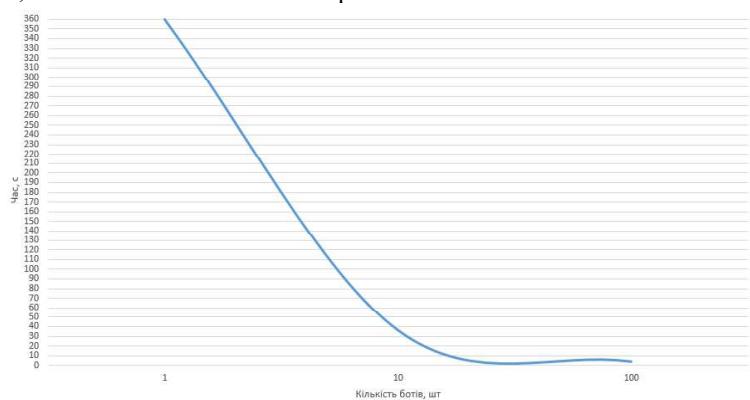


Fig. 4 - Time required for system failure

The basis of the ICMP – flood attack is the system's automatic responses. This means that the first step is ping request disabling.

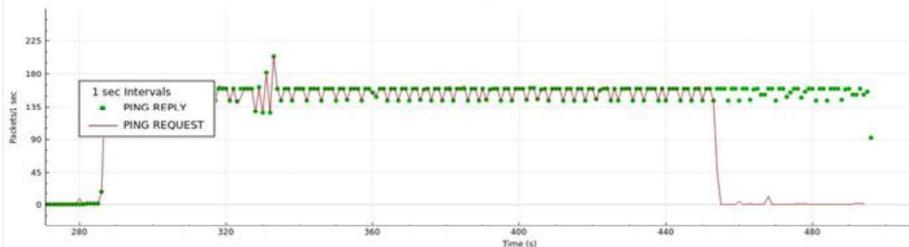


Fig. 5 - The time required to disable the system

Thus, the algorithm for combating the ICMP flood will take the form refer to Figure 6.



Fig. 6 - ICMP flood reflection algorithm

SYN – flood attack can destroy system in a short time. The graph shows that attack with generated traffic 18 kB / s will destroy system in 90 seconds. The danger is that the network card will stop opening new connections and stop directing traffic to existing ones.

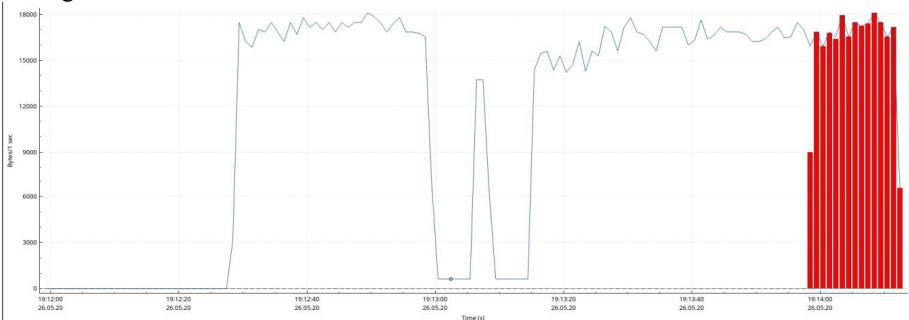


Fig. 7 - SYN - flood traffic generated during the experiment

By changing the system settings (reducing the retention time of half-open connections, syncookies collection and increasing the queue of half-open connections), it was significantly possible to reduce the power of the attack and make it almost safe. Figure 8 shows that the traffic has decreased to 8 kB / s, and there are no more failures to create a new connection.

Thus, the proposed algorithm to struggle this type of attack, turn to Figure 9

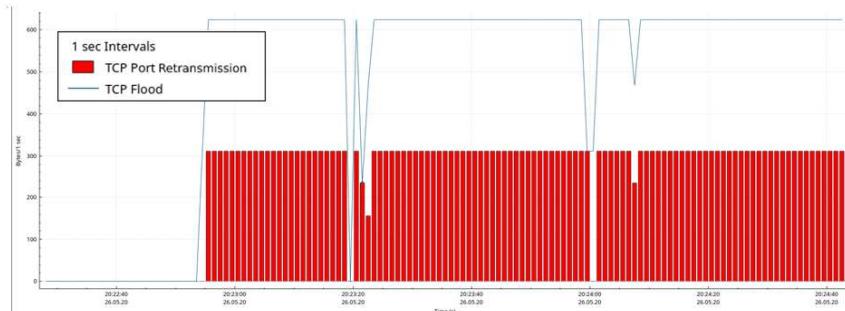


Fig. 8 - SYN - flood traffic after applying the reflection algorithm

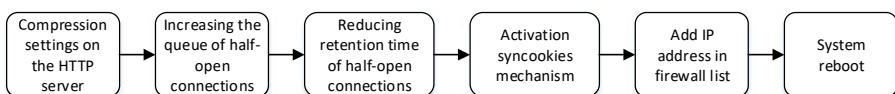


Fig. 9 SYN flood reflection algorithm

**Conclusion.** In this research presents methods for ICMP and TCP flood attacks repelling. This investigation establishes the time required for the botnet to disable the system and the size of generated traffic. This made it possible to understand the real threat that the bot network is now creating. This research improved algorithm reflection SYN flood by reducing the retention time of half-open connections, syncookies collection and increasing the queue of half-open connections. And also the process of data compression on the web server played an important role in this algorithm improving.

#### References:

1. Cisco Inc. (2018). Cisco Cyber Security Report 2018 [Electronic resource]. Access mode: [https://www.cisco.com/c/uk\\_ua/products/security/security-reports.html](https://www.cisco.com/c/uk_ua/products/security/security-reports.html)
2. Botnets and their types [Electronic resource] // EC-Council – Resource access mode: <https://blog.eccouncil.org/botnets-and-their-types/>.
3. Into the Battlefield: A Security Guide to IoT Botnets // Trend Micro. – 2019. – Resource access mode: <https://www.trendmicro.com/vinfo/ph/security/news/internet-of-things/into-the-battlefield-a-security-guide-to-iot-botnets>
4. Water Torture: A Slow Drip DNS DDoS Attack // Secure64 Software Corporation. – 2014. – Resource access mode: <https://secure64.com/2014/02/25/water-torture-slow-drip-dns-ddos-attack/>.
5. Scanlon M. Peer-to-Peer Botnet Investigation: A Review / M. Scanlon, T. Kechadi // Future Information Technology, Application, and Service, LNEE 179, pp. 231–238, DOI: 10.1007/978-94-007-5063-0\_33
6. Almomani A. A proposed framework for Botnet Spam-email Filtering using Neucube // A. Almomani, M. Alauthman, O. Almomani, O. Albala. // Proceeding of The International Arab Conference on Information Technology (Yassmine Hammamet, Tunisia, December 22-24, 2017). – 2017.

# **DATA ANALYTICS**



<sup>1</sup> **Veronika Bokan**

Master of the Department of Management Technology

<sup>2</sup> **Viktoria Tsykun**

Master of the Department of Management Technology

<sup>3</sup> **Andrii Khlevnyi**

PhD, Associate Professor of the Department of Technologies Management

<sup>1-3</sup> Taras Shevchenko National University of Kyiv

## INFORMATION ANALYSIS OF METHODS FOR FORECASTING THE POPULATION OF UKRAINE

**Abstract.** The urgency of population forecasting is given in the work, the methods of such forecasting are analyzed. The parameters on the basis of which it is expedient to forecast the demographic situation in the near future are identified: the establishment of births, deaths and migration. The expediency of application of methods of neural networks is established, parameters are allocated, prospects of further researches are established.

**Keywords:** population, forecasting, neural network.

Demographic forecasting is the basis for scientific planning of indicators of socio-economic development of the country, giant companies and local companies. Accordingly, consumers of such forecasts form the transformation of geopolitical processes in terms of creating plans for the production of goods and services, planning in education and medicine and other important areas of life, infrastructure development and housing, and etc. According to the results [1], it is established that in Ukraine there is a rapid decline in population. This depends on: declining population, high mortality, working age, poor health of adults and children, reduced life expectancy and gradual aging, migration.

From the analyzed approaches [2] of demographic forecasting, namely the separation of forecasting by: time, detailing information about the population of Ukraine, regional coverage, purpose, object, it is established that special attention is needed by time forecasts (short-term, medium-term and long-term). Short-term forecasts are considered to cover the period before reaching the age of. The short-term forecast should correspond to high level accuracy and detail. Forecasts, called medium-term, covering a period of twenty to thirty years. Such forecasts are not always the same as in the short-term, but it is possible to show a deeper picture for demographers, noting the prospect of growth or decline of the population and its individual parts with an accepted level of reliability. Long-term forecasts cover fifty or even a hundred years. It is clear that the results of this type of forecast can be operated as scientific predictions. It is established that in terms of economic prospects for the near future it is advisable to pay attention to the development of models and methods of short-term forecasts. The expediency of such a forecast is to plan and achieve economic and cultural goals of companies, the region and the country as a whole. The main indicators on the basis of which it is expedient to forecast the demographic situation in a certain

area for the near future: the establishment of births, deaths and migration.

Short-term demographic forecasting for certain parameters can be implemented by the following methods: regression, cohort-component, analytical, extrapolation, Markov chains, neural networks.

Methods of regression analysis for short-term demographic forecasting. The essence of this approach is to establish qualitative relationships between the factors that shape their intensity. The assessment is based on historical changes in indicators that are considered and act as factor factors and affect the forecast of demographic phenomena. The result of such modeling is the construction of multidimensional regression models, which are the results of the analysis of correlation and regression sets. As an independent variable in such modeling, it is appropriate to use not time, but a numerically defined material characteristic, which is a factor [3]. Therefore, it is appropriate to use this method to determine the forecast by regional coverage.

Cohort-component method. The essence of the method is to form cohorts and determine their changes according to certain parameters, such as age, sex, life expectancy, and so on. This method is most often used in population forecasting. The advantage is that such a number can be considered from the standpoint of the established cohorts. The input data of such modeling are data on the initial number of cohorts. It is appropriate to implement on the basis of the equation of demographic balance.

Analytical method. The essence of the method is to select a function based on historical data that will most accurately describe it in the future. This method should be used for short-term forecasting.

Extrapolation method. The essence of the method is that the calculations are based on exponential and linear functions. Input data contain data on changes in population, both average and absolute, for a particular period or data on changes in the average annual population growth rate. If we assume that the factors or groups of factors that can affect the process of forecasting using the extrapolation method, is unchanged, in order to allow connecting the population to any required time [4]. However, it is advisable to use for short-term forecasting. The method will not change in certain groups of the population. Therefore, it cannot be used to predict, for example, age, working groups.

Markov chains. convenient to use when there is a need to solve the problems of population transition from one cohort to another.

The disadvantages of these methods can be minimized using the neural network approach. The expediency of their application is that in the system of indicators of demographic status linear methods do not cover all the patterns. The task of the neural network is to learn to solve the problem on the basis of a training sample. Such a neural network is able to determine the relationships between the data that enters the input and requires the output signal [5]. The trained neural network is able to summarize the acquired skills and give a forecast for new values of the input signal.

Based on the analysis, it is established that for short-term forecasting of the population of both the region and the country as a whole, it is advisable to use such promising methods as neural networks. Data for training is proposed to be used from

<http://www.ukrstat.gov.ua/> from 1989 to 2020. To implement the method, the following parameters are proposed according to which forecasting by neural network methods will be carried out:

- to predict the birth rate: permanent population, migratory population growth, average wages, number of unemployed, number of children who applied to preschool educational institutions, number of doctors, infant mortality rate, indicators of the number of family institutions (marriages and divorces), state social assistance for children, the number of families of reproductive age who received housing from the state who purchased housing.

- to predict mortality: permanent population, migration growth, life expectancy, number of pensioners, average wages, number of registered crimes, morbidity, number of road accidents, number of suicides, number of deaths in hospitals.

It is established that for the implementation of forecasting it is advisable to choose the Python programming language for simplicity in syntax, broad support of the programming community and a huge amount of available documentation.

The prospect of further research is to build a neural network architecture, its training, testing. This model will be the basis of information technology aimed at combining the analysis and forecast of the population, its labor potential with the socio-economic processes that take place. It will make it easier for stakeholders to draw conclusions that will be aimed at changing the trend of population decline, increasing the socio-cultural level in general.

#### **References:**

1. Distribution of permanent population by sex, age groups and type of locality. URL:[http://database.ukrcensus.gov.ua/Mult/Dialog/varval.asp?ma=000\\_0204&path=../Database/Population/02/02/&lang=1&multilang=uk](http://database.ukrcensus.gov.ua/Mult/Dialog/varval.asp?ma=000_0204&path=../Database/Population/02/02/&lang=1&multilang=uk)
2. Власенко Н.С., Макарова О.В., Пирожков С.І., та інші. Комплексний демографічний прогноз України на період до 2050 р. / за ред. членкореспондент НАНУ, д.е.н., проф. Е.М. Лібанової. К.: Український центр соціальних реформ, 2006. 138 с.
3. A. Sen, M. Srivastava, Regression Analysis. Theory, Methods, and Applications, Springer-Verlag, Berlin, 2011 (4th printing).
4. George Lindfield, John Penny. Numerical Methods (Fourth Edition), 2019.
5. N.D. Lewis. Neural Networks for Time Series Forecasting, 2017.

<sup>1</sup> Yuliia Bura

Master's degree student Department of technology management

<sup>2</sup> Iulia Khlevna

Doctor of Engineering Science, Associate professor Department of technology management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## HOUSE PRICE MODELING BY MACHINE LEARNING

**Abstract.** The objective of this paper is to present the relevance of house price prediction. During the research, we've established that machine learning methods give a fairly accurate result (with a small RMSE error). The aggregate model gives the smallest prediction error.

**Keywords:** Machine Learning, House Price.

The real estate sector, both in Ukraine and in the world, is one of the sectors with the largest capitalization. Here often appear new profitable projects. These projects play an important role in the formation of capital flows and economic stability. Providing objective information to those who make decisions about certain transactions in the real estate market is an urgent application task. Its solution can be interpreted through the creating of information technology, which is based on the analysis, modeling, and prediction of house prices in the real estate market. The scientific task is the implementation of a rational method of prediction in such technology. This is the task of this article too.

The real estate market is quite sensitive to external events and unpredictable. So, the process of the house price prediction is really complex, and classical methods do not give the desired result in this case [1]. It's expedient to use machine learning methods for such situations of fuzzy variables prediction [2, 3].

To choose a rational method for house price prediction we propose to analyze the following models: *LASSO Regression*, *Elastic Net Regression*, *Ridge Regression*, *Gradient Boosting Regression*, *XGBoost*.

The study was conducted on a dataset of 1300 houses [4]. The "sales price" was set as the target variable.

We identified 79 independent variables to model the target variable. All data are equally divided into training and test parts.

The result of logarithmic normalization of the target variable for its qualitative modeling is further presented in Fig. 1.

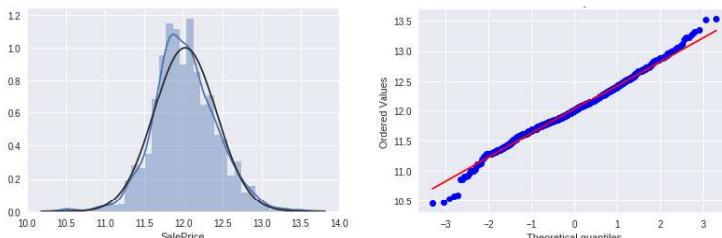


Figure 1. Comparison of distribution graphs of the logarithmically normalized "Sales Price" series with the normal distribution.

The selected models were evaluated for the quality of the model prediction – RMSE error and standard deviation (Table 1).

Table 1  
Models prediction results

Model's score	RMSE	Standard deviation
Lasso	0.1115	0.0074
ElasticNet	0.1116	0.0074
Ridge	0.1153	0.0075
Gradient Boosting	0.1177	0.0080
Xgboost	0.1161	0.0079
<i>Averaged base models</i>	<i>0.1091</i>	<i>0.0075</i>

The gradient boosting model was found to be the least accurate. So, we tried to build an aggregate real estate price forecasting model without it. Obtained result: RMSE = 0.1091, and the standard deviation is 0.0075. This approach improves prediction accuracy by 5%.

**Conclusions.** We established that better to use a combination of models to obtain the most accurate house price prediction in the real estate market. We determined that such an approach will be the basis for the creating of information technology, and this will be a prospect for further research.

#### References:

1. YanY. Hui B. (2007). Method for Housing Price Forecasting based on TEII. Methodology Systems Engineering - Theory & Practice, Volume 27, Issue 7, 1-9.
2. Park B., Baeb J. (2015). Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data, Expert Systems with Applications, Volume 42, Issue 6, 2928-2934
3. Segnon, M., Gupta, R., Lesame, K. et al. (2020). High-Frequency Volatility Forecasting of US Housing Markets. J Real Estate Finan Econ. <https://doi.org/10.1007/s11146-020-09745-w>
4. Dean C. (2011). Journal of Statistics Education, Volume 19, Number 3
5. Bailey MJ, Muth RF, Nourse HO. 1963. A regression method for real estate price index construction. Journal of the American Statistical Association 58: 933–942.

**<sup>1</sup> Oleksandr Burmistenko**

Doctor of Technical Sciences, Full Professor

**<sup>2</sup> Tetyana Bila**

Candidate of Technical Sciences, Associate Professor

**<sup>3</sup> Volodymyr Statsenko**

Candidate of Technical Sciences, Associate Professor

**<sup>4</sup> Dmytro Statsenko**

Candidate of Technical Sciences, Associate Professor

<sup>1-4</sup> Kyiv National University of Technologies and Design

## INFORMATION ANALYSIS OF THE BULK MATERIALS CONTINUOUS DOSING PROCESS

**Abstract.** This work presents the principles of bulk material flows information processing at the plate feeders outlet. The experimental stand structure and the results of data processing are shown.

**Keywords:** Data processing, data analysis, dosing, bulk materials, plate feeder

**Introduction.** Bulk materials continuous dosing processes are used in the manufacture of products from polymer materials. The basis of such products is a solid polymer granules mixture with various additives. Mixing complexes are used to obtain them, which include equipment for storing bulk materials (hoppers), dosing devices (feeders) and a mixer [1]. The quality of the mixture is largely determined by the compliance of its percentage composition with the given recipe. Therefore, the accuracy of dosing devices is an important parameter that significantly affects the quality of products.

The movement of bulk materials is a complex discrete process during which flow ruptures, formation of arches, formation of lumps and other phenomena can occur that reduce the accuracy of dosing. The problem gets worse with the continuous operation equipment [2]. Such equipment makes high demands on the accuracy dosing, because bulk material moves in a continuous flow and any deviations lead to the appearance of local changes in the percentage composition of the mixture, which are almost impossible to compensate. The output flow pulsations magnitude allows you to indirectly determine the possible equipment deviations from the specified mode. This information can be used to control both the feeder operating modes and the mixing complex as a whole. These problems determine the relevance of the analysis of the flow bulk materials parameters at the continuous feeders outlet.

**The research results.** This work discusses plate feeders, which are widely used in industry, because they provide minimal mechanical impact on bulk material and the ability to accurately control its performance. The measurement of the bulk materials flow parameters was carried out using an experimental stand (Figure 1), which included a hopper (H), a plate feeder (F), a flow shaper (S), a strain gauge mass sensor (GS), an analog-to-digital converter (ADC), a microcontroller (MC), a personal computer (PC) and a mixer (M).

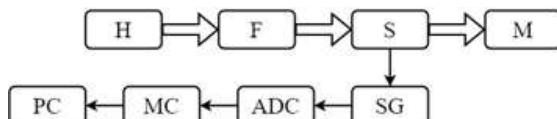


Figure 1 – Stand structure for research of the plate feeders work

A plate feeder transported bulk material to a flow former with a weight sensor installed under the surface. The mass of material determined the sensor signal. Using an ADC, the sensor signals were converted into digital form and read out by the microcontroller at specified time intervals. Then they were transferred to a personal computer for further analysis. An example of the obtained data is shown by dots in Figure 2, a.

Data analysis included two stages:

1. Linear regression equation coefficients calculation passing through the obtained points. The coefficients were determined by using the least squares method. The obtained values are shown in Figure 2, a. Ideally, in steady-state operation, the straight line slope tangent should be equal to zero, and the free term of the equation should equal the specified performance.

2. Spectral analysis of signals using fast Fourier transform (Figure 2, b). The data obtained make it possible to numerically estimate the magnitude of the pulsations and determine possible problems associated with the operating modes of the equipment. For example, the appearance of pulsations with a frequency equal to the plate feeder rotation frequency indicates its incorrect installation.

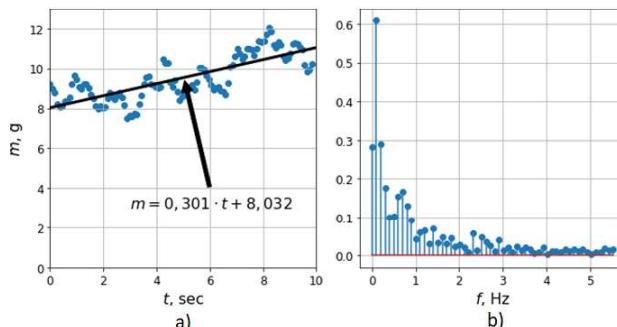


Figure 2 – Sensor Signal Analysis Results

The obtained results can be used as a data source for the dosing equipment control system.

#### References:

1 V. Statsenko, O. Burmistenkov, T. Bila, D. Statsenko. Determining the motion character of loose materials in the system of continuous action «hopper – reciprocating plate feeder». Eastern-European Journal of Enterprise Technologies. Volume 2/1 (98), 2019, P.21-28. doi: 10.15587/1729-4061.2019.163545.

2 R. Weinekötter, L. Reh. Continuous mixing of fine particles. Part. Syst. December 1994, Volume 12, P.46–53.

**<sup>1</sup> Serge Dolgikh**

M.Sc. Theoretical Physics, Telecommunications Engineering, Senior Project Engineer

**<sup>2</sup> Oksana Mulesa**

PhD, Associate Professor

*<sup>1</sup>Solana Networks, Canada*

*<sup>1</sup>National Aviation University, Ukraine*

*<sup>2</sup>Uzhgorod National University, Ukraine*

## **COVID-19 EPIDEMIOLOGICAL FACTOR ANALYSIS: IDENTIFYING PRINCIPAL FACTORS WITH MACHINE LEARNING**

**Abstract.** Based on a set of Covid-19 statistical data of national and subnational jurisdictions at the time point of approximately two months after the local onset of the pandemics, an analysis of the factors with strong influence on the reported local outcomes was performed with several different statistical methods. The consistent conclusion of the analysis confirmed epidemiological policy and management as the dominant factors in the outcome. The methods in the study can be used to evaluate principal factors at future time points to reach a confident conclusion.

**Keywords:** epidemiology, machine learning, regression, factor analysis

### **Problem Statement**

The task of modeling and forecasting time-series processes of different nature is essential and arises in different fields such as planning [1], the study of the dynamics of climate change [2] and importantly in the current situation, health science and epidemiology.

Known models and forecasting methods are based on using integrated information about the background of the predicted processes [1, 3]. Among the tasks of forecasting an important place is occupied by the methods of factor estimation and time-series analysis that includes a variety of methods and approaches including fuzzy sets [4], expert models and methods [5], genetic and neural network methods [6, 7] and others.

A common challenge in the analysis of statistical data related to a developing situation, such as in this work, the developing epidemiological scenario related to a dangerous infection with potentially high impact on health and safety of population, economy and the society as a whole is evaluation of methods and models with the objective of identifying the approaches that could be most effective in describing the process that is being studied.

To avoid or reduce the possible ambiguity related to the selection of the method of analysis of statistical data, in this work we used several common methods of statistical analysis specifically, evaluation and ranking of factor influence with an expectation that if consistency between the results of different methods can be achieved, it would enhance the confidence in the result that can be essential for the development of reliable and effective policies based on the conclusions of factor analysis.

## Methods

With variety of statistical methods and techniques used to evaluate the correlation hypothesis as discussed above, we set out to provide an analysis of principal factors influencing the development of the epidemics in the national and subnational jurisdictions based on the available data for the first group of countries that were exposed to Covid-19 pandemics in late January – beginning of February, 2020. This objective is approached by applying several commonly used methods of factor analysis and ranking, looking for consistency of results between different methods. A consistency between the results of different methods would improve confidence in the findings, providing a grounded and reliable statement of their influence on the outcome.

The analysis of scientific publications showed that the following factors have a strong influence on the development of the epidemic including but not limited to the following: the time of the local development of the epidemics; traditions, social and lifestyle factors; demographics including gender and age; the level of the economic and social development; quality standard and epidemiological efficiency of the public healthcare system, and not in the least, the quality of public health policy making and execution. Based on the identified selection criteria and publicly available epidemiological information from a number of trusted sources as indicated below, the dataset of 18 cases of national and subnational public health jurisdictions was constructed. The data included one provincial jurisdiction in Canada (Ontario), one state (California) and one municipal jurisdiction in the USA (New York City); given the high geographical variation of the impacts, data with more detailed geographical breakdown is expected in the future studies. The time point at which the data was collected was *TZ + 3 months*, i.e. approximately two months of the local development of the epidemics in the selected group of jurisdictions.

To evaluate the consistency of results produced by different methods of statistical analysis, several methods were used to evaluate the influence of the selected factors on the overall impact of the developing epidemics:

1. Calculation of correlation between the resulting effect and a specific factor;
2. Linear regression analysis by single factor and a combination of factors;
3. Evaluation of factor importance with Random Forest regression;
4. Evaluation of factor influence or rank with a feature ranking method.

## Results

The methods applied to the dataset of early epidemiological data of selected jurisdictions demonstrated consistent results with good agreement between methods. The findings confirmed the importance of clear, timely and evidence-based epidemiological policy as the factor with the highest influence on the development of the epidemiological scenario. This finding is consistently produced by all methods of analysis used in the study.

The analysis offered additional arguments in support of the hypothesis of some form of general population-wide protection effect against Covid-19 as an effect of previous universal immunization program with Bacillus Calmette–Guérin vaccine

(BCG), that has been reported in a number of earlier results adding arguments to the rationale for further studies of the possible correlation and the mechanisms of such general protection with potential benefits that may extend beyond Covid-19 pandemics. Additionally, it established potential significance of secondary factors such as smoking prevalence as in the epidemiological impact, consistently confirmed by several independent methods used in the analysis.

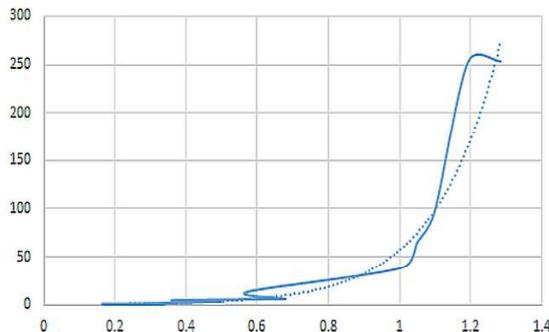


Figure 2 – Epidemiological outcome vs. combined dominant factors

Overall, the authors expect that the methods of factor influence analysis demonstrated in this work can be applied repeatedly over a time series of epidemiological data, allowing to reach confident conclusions by establishing and analyzing the trend over an extended period of time.

#### References:

1. V. N. Kuharev, V. N. Sally, A. M. Erpert, Economic-mathematical methods and models in the planning and management, Kiev, Vishcha School, 1991.
2. A. S. Kozadaev, A. A. Arzamasians, Prediction of time series with the apparatus of artificial neural networks. The short-term forecast of air temperature. Bulletin of the University of Tambov, Series: Natural and Technical Sciences, №3, is 11, 2006, pp. 299-304.
3. V. Ye. Snytyuk, Forecasting. Models. Methods. Algorithms: Tutorial, Kyiv, 2008.
4. O. Mulesa, F. Geche, V. Voloshchuk, V. Buchok and A. Batyuk, Information technology for time series forecasting with considering fuzzy expert evaluations, 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Lviv, 2017, pp. 105-108.
5. A. S. Mendel, Method counterparts in predicting short time series: expert-statistical approach, Machine Telemechanics, № 4, 2004, pp. 143-152.

**<sup>1</sup> Mykhailo Dvoretskyi**

A senior lecturer at the Department of Software Engineering

**<sup>2</sup> Svitlana Dvoretska**

A senior lecturer at the Department of Software Engineering

**<sup>3</sup> Hlib Horban**

PhD., an associate professor at the Department of Software Engineering

**<sup>4</sup> Yuriy Nezdoliy**

A senior lecturer at the Department of Software Engineering

*<sup>1,2,3,4</sup> Petro Mohyla Black Sea National University*

## **USING THE ANALYTIC HIERARCHY PROCESS FOR OPTIMIZATION THE DATABASE STRUCTURE OF A DISTRIBUTED CORPORATE INFORMATION SYSTEM NODE**

**Abstract.** One of the research aims is to determine and build a mathematical model of the optimality criteria for the structure of a remote node of the distributed corporate information system database. The statistics of user SQL-queries activity is taken into account and presented in the form of a multidimensional database. Criteria of the model effectiveness are formulated and the problem of multicriteria optimization is solved. Choosing the best alternative makes possible to determine the optimal level of the data representation marker.

**Keywords:** corporate information system, distributed database, SQL-query, multicriteria problem, analytic hierarchy process.

In information systems development, there is a trend of transition from local to distributed databases (DDB). Within one company there is a need to automate different types of accounting. The attempt to automate all types of accounting leads to so-called "universal" corporate information systems. This approach has many disadvantages [1], which can be eliminated by using separate specialized solutions. But this path leads to use of several databases (and perhaps DBMS) that require their synchronization [2-3].

A key factor influencing the reliability and accessibility of such databases is the so-called localization of links. If the database is distributed so that the data hosted in a node is called exclusively by its user, it indicates a high level of link localization [4].

A combined data distribution strategy is the best in terms of combining the benefits of strategies with and without duplication. But when using it, in addition to the task of synchronizing duplicate information, the task of designing the structure of the database is actual, depending which node data belonging to [4-5]. The performance of the system will directly depend on the decision on the need for partial or complete duplication of data.

The purpose of the research is to create a mathematical optimization model and subsequent choosing the best alternative to the marker of data representation of the remote node of distributed CIS. The next obtained multicriteria problem need to be

solved to determine the optimal level of data representation marker. The solution of the problem is also complicated by the fact that the solution space is defined on a set of real numbers, and therefore the set of solutions contains a large number of alternatives. The research is related only to the relational databases.

To avoid the need for further replication some data that required on the DDB node can only be presented on the central node of the database and participate the query through the use of distributed queries. Due to the fact that to represent the data on the remote node it is necessary to use elements of both vertical and horizontal data fragmentation (both projection and selecting), the node relation is a subset or attributes and tuples of the base relation R.

The model of presenting user queries should support the possibility of their further classification according to belonging to a particular workplace, location, user role and other criteria that can be added to the model. That is, the user query is defined as

$$Q = \langle \text{DateTime}, \text{WorkplaceType}, \text{Location}, \text{UserRole}, \text{Application}, R, A, \text{tup} \rangle$$

For the dimensions elements the term of data representation marker is proposed. It reflects the level of data representation necessity at the node of distributed CIS.

So, we have a model where each dimension attribute has a value, a marker and a weight  $A_{\text{dim}} = \{\text{Val}, \text{Mrk}, \text{vol}\}$ , where  $\text{Mrk} = \{"\text{obligatorily}", "\text{necessary}", "\text{neutral}", "\text{not required}", "\text{forbidden}"\}$ , and  $\text{vol}$  – weight (ignored for the values of the marker "obligatorily" and "forbidden"). By converting a non-numeric linguistic variable of markers into a numeric value ("obligatorily" – "2", "necessary" – "1", "neutral" – "0", "not required" – "-1", "forbidden" – "-2"), the aggregation function was defined:

$$\text{Aggregate}_{i=1}^n \text{Mrk}_i = \begin{cases} 2, & \text{if } \exists \text{Mrk}_i = 2 \\ -2, & \text{if } \exists \text{Mrk}_i = -2 \wedge \nexists \text{Mrk}_i = 2 \\ \sum_{i=1}^n (\text{Mrk}_i * \frac{\text{Vol}_i}{\sum_{i=1}^n \text{Vol}_i}) \end{cases}$$

When deciding on the data representation on a remote node, we consolidate the rows of the fact table by the tuple  $\langle R, A, \text{tup} \rangle$  and calculate the value of the marker for each of its elements by formula (5). And based on following the decision about data representation is made:

$$\text{Repr}(\text{Node}, R, A, \text{tup}) = (\text{Aggregate}(R, A, \text{tup})_{i=1}^n \text{Mrk}_i > \text{koef}_{\text{repr}}^{\text{node}})$$

When deciding on the data representation on a remote node, we consolidate the rows of the fact table by the tuple  $\langle R, A, \text{tup} \rangle$  and calculate the value of the marker for each of its elements by previous formula. And based on following the decision about data representation is made:

$$\text{Repr}(\text{Node}, R, A, \text{tup}) = (\text{Aggregate}(R, A, \text{tup})_{i=1}^n \text{Mrk}_i > \text{koef}_{\text{repr}}^{\text{node}})$$

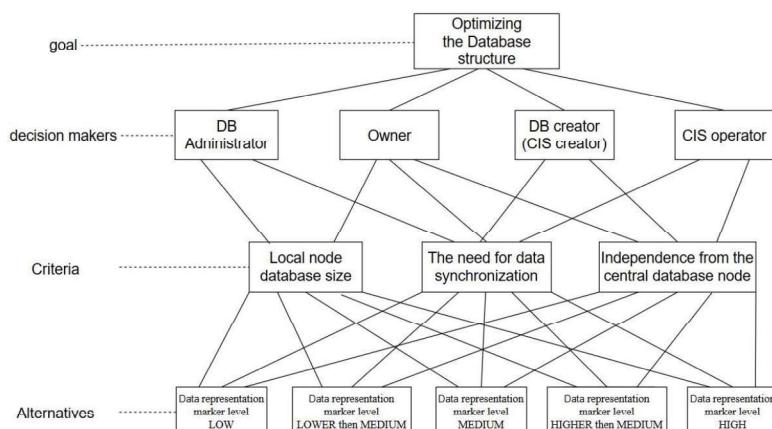
where  $\text{koef}_{\text{repr}}^{\text{node}}$  – the threshold coefficient of data representation in a certain node Node, that is defined at the range of [-1, 1].

When planning the structure of the database of the remote node of distributed CIS, several factors will be involved - availability and speed of data obtaining,

independence from the central DB node, the DB size, the level of data reliability, the need for further synchronization [6-7]. Criterion of independence from the central database node, and, accordingly, the availability and access speed directly depend on the representation of user SQL-query data on the node of distributed CIS. The criterion of the local database size affects both the performance of queries to the local database and the power of computing resources required to perform database and CIS administration operations. The criteria of the need for data synchronization is the ratio of the number of modified data queries to the total number of queries. In the research the mathematical models of mentioned criteria were obtained.

A multicriteria problem, that was obtained, need be solved to determine the optimal level of data representation marker. The analytic hierarchy process (AHP), which is a general methodology for solving a wide class of decision-making problems, allows to combine a relatively simple mathematical apparatus with knowledge and experience of the decision maker.

When compiling the hierarchy, following relationship between the levels elements was used: goal - stakeholders - criteria - alternatives. The value of the data representation marker (alternative) is a real number in the interval [-1, 1]. It leads to potential large number of alternatives at the 4th level of the hierarchy and therefore the matrices of pairwise comparisons by criteria can become very big. This complicates estimation process for the decision makers. It is proposed to simplify the task by reducing the number of alternatives to 5: "low" (L) – "-1", "lower them medium" (LM) – "-0.5", "medium" (M) – "0", "higher then medium" (HM) – "0.5", and "high" (H) – "1". The level of "decision makers" is represented by the elements "Owner", "Database Administrator", "Database Developer" and "CIS Operator". The obtained hierarchical model is shown in Fig. 1.



**Figure 3:** Hierarchical model of the distributed CIS node structure optimization problem

Solving a multi-criteria problem and finding the optimal level of data representation at a remote node increases the level of data availability and efficiency of distributed CIS. Efficiency is defined as the ratio of result and resources, so taking into account the vector of relative weight of the optimality criteria of the model, we calculate the efficiency as

$$Eff = \frac{F_{availab} \times W_1^{criteria}}{F_{size} \times W_0^{criteria} + F_{synchro} \times W_2^{criteria}}$$

Thus, the results of the research allow to increase the efficiency of using the distributed CIS node of the subject area by 25% compared to the presentation of only critical data, and by 11% compared to the presentation of all necessary data of the central database, respectively. The research can be followed by presenting the obtained vector of global priorities in the form of fuzzy sets of one variable. Dephasing the obtained results can make numerical value of the optimal level of data representation at the DCIS node more accurate.

## References

1. M. Dvoretskyi, S. Borovlova, Web-application of warehouse accounting in non-automated points of sale, Science works "Petro Mohyla Black Sea National University", Rel. 308. T. 320, Series: Computer technologies, 2018, pp. 45–52 (in Ukrainian).
2. M. Tamer Özsü, Patrick Valduriez. Principles of Distributed Database Systems 3rd ed. Springer, 2011.
3. Automatic synchronization of distributed databases in split mode, [Online]. Available: [http://stimul.kiev.ua/materialy.htm?a=avtomaticheskaya\\_sinkronizatsiya\\_raspredelennykh\\_baz\\_dann\\_ykh\\_v\\_razdelennom\\_rezh](http://stimul.kiev.ua/materialy.htm?a=avtomaticheskaya_sinkronizatsiya_raspredelennykh_baz_dann_ykh_v_razdelennom_rezh) (in Russian)
4. H. Garcia-Molina, J. D. Ullman, and J. Widom, Database Systems: The Complete Book 2nd Edition, Pearson, 2008.
5. Dusan Petkovich. Microsoft SQL Server 2019: A Beginner's Guide, Seventh Edition 7th Edition, Kindle Edition, Mc-Graw-Hill Education, 2020.
6. M. Dvoretskyi, S. Dvoretska, Y. Nezdoliy, S. Borovlova, Data Utility Assessment while Optimizing the Structure and Minimizing the Volume of a Distributed Database Node, in: Proceedings of the 1st International Workshop on Information-Communication Technologies & Embedded Systems (ICTES 2019), Mykolaiv, 2019. pp. 128–137
7. M. Fisun, M. Dvoretskyi, A. Shved and Y. Davydenko, Query parsing in order to optimize distributed DB structure, in: Proceedings of the 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Bucharest, 2017, pp. 172–178. doi: 10.1109/IDAACS.2017.8095071.

**<sup>1</sup> Ievgen Fedorchenco**

Senior Lecturer

**<sup>2</sup> Andrii Oliynyk**

PhD, Associate Professor

**<sup>3</sup> Alexander Stepanenko**

PhD, Associate Professor

**<sup>4</sup> Anastasiia Kharchenko**

Student

**<sup>5</sup> Marharyta Saman**

Student

<sup>1-5</sup>Department of Software Tools, National University "Zaporizhzhia Polytechnic"

## **RESEARCH AND DEVELOPMENT OF A GENETIC ALGORITHM FOR DIAGNOSING THE STRENGTH OF THE BLADE STRUCTURE IN GAS TURBINE ENGINES**

**Abstract.** The problem of developing a method of technical diagnostics on the basis of data for the study of the structural strength of the blades is considered. It has been determined that evolutionary methods, including genetic algorithms, are effective methods of computational intelligence that can be used to build diagnostic models. The proposed methods and tools can be used to predict the state of critical load points in the diagnosis of gas turbine blades of aircraft engines during operation.

**Keywords:** evolutionary method, genetic algorithm, optimization, technical diagnostics, software, forecasting

### **Introduction**

The relevance of the research topic lies in the presence of the need for accurate and fast methods of searching for critical points of functions of different complexity, especially multimodal and multifunctional. Even in such cases, the application of standard search methods complicates the search, makes it more costly and increases the time of information processing. Such a problem is especially acute in areas where diagnostic processes are critical to time and high accuracy, such as aircraft construction [1].

### **Analysis of literature data and problem statement**

Despite the constant development of methods and algorithms for solving technical diagnostics, this task is relevant. This is due to the fact that it is impossible to create a universal method, because the task of technical diagnosis is divided into subtasks. Such tasks include preventive technical diagnostics, technical diagnostics of failures, diagnostics of maintenance of the current state of the object, etc. These subtasks can also be divided by the nature of the diagnosis - it can be a diagnosis by one or more parameters, and so on. The described reasons are the basis for the creation of new methods of technical diagnosis [1].

### Development of an algorithm for technical diagnosis

Algorithm of the modernized GA with leader method based on GA with decreasing population size. It was decided to develop a modified method "GA with leader" based on GA with population reduction, which proposes to change the population size depending on the number of most adapted individuals in the population, thus reducing the amount of computation to obtain the optimal solution.

In the proposed method for generating a new set of solutions from the input data, the number of individuals corresponding to the size of the population, the most adapted to the analysis, is selected. Then, when selecting the parent pair from the current generation, a certain number of pairs is selected by the probable rank. In each iteration (in each generation) this number is different. Each pair is selected taking into account the values of the objective functions of the most adapted and least adapted individual in the population:

$$e^{\frac{f_j - f_{\text{worst}}^t}{f_{\text{best}}^t - f_{\text{worst}}^t}} < \text{rand}(e^{-1}; 1), j = \overline{1; N} \quad (1)$$

where,  $N$  is the size of the current population;  $f_i$  - the value of the fitness function of the  $j$ -th individual;  $f_{\text{best}}^t$  - the best value of the fitness function of the current population  $t$ ;  $f_{\text{worst}}^t$  - the worst value of the fitness function of the current population  $t$ ;  $\text{rand}(e^{-1}; 1)$  - random number from  $e^{-1}$  to 1 [2].

During crossing, a descendant  $G_i$  is created, which is located at some distance from the ancestor with the best values of the fitness function  $G_1$  in the direction from the ancestor with the worst value of the fitness function  $G_2$ . Determining the value of the  $i$ -th gene of the  $g_i$  chromosome-offspring is determined by the formula:

$$g_i = k(g_{1i} - g_{2i}) + g_{1i} \quad (2)$$

where  $k \in [0; 1]$  - the actual coefficient specified by the user at the stage of initialization of genetic search.

At the stage of mutation, starting from the first chromosome, the whole population is reviewed, and for each chromosome  $H_j$ , drop numbers  $x_i$  from the interval  $[0; 1]$  are assigned. If this number is less than the probability of mutation, then the current chromosome  $H_i$  is mutated. In the selected chromosome there is a mutation of genes by some value:

$$g_{ij}^* = g_{ij} + \Delta g_{ij}, \quad (3)$$

where  $i$  is the gene number in the chromosome;  $j$  - chromosome number;  $g_{ij}$  - gene for mutation;  $g_{ij}^*$  - gene after mutation [3].

The values of the  $i$ -th gene  $g_{ij}$  of the chromosome  $G_j$  after mutation can be calculated by formula:

$$g_{ij}^* = \begin{cases} g_{ij} + \Delta(p, \max_i - g_{ij}), & 1 \leq i \leq w \\ g_{ij} + \Delta(p, g_{ij} - \min_i), & w < i \leq K \end{cases}, \quad (4)$$

where

$$\Delta(p, y) = y \left( 1 - e^{(1-p/P)^v} \right), \quad (5)$$

where  $c$  is a randomly generated number in the interval  $[0; 1]$ ;  $p$  is the number of the current iteration;  $P$  is the maximum number of iterations;  $v$  is a parameter that

determines the degree of homogeneity (uniformity);  $\min_i$  and  $\max_i$  - the minimum and maximum value of the  $i$ -th parameter in the solution with the help of the genetic method of the problem;  $w$  is a number equal to  $|K / 2|$ ;  $K$  is the number of genes in the chromosome [4].

The new generation is formed from the existing set of solutions obtained as a result of the application of crossing, mutation and inversion operators. The probability of an individual to be selected for a new generation is calculated by formula (6) [5]:

$$P(X^i) = \frac{-\text{fitness}(X^i) + D}{N \cdot D - \sum_{j=1}^N \text{fitness}(X^j)} \quad (6)$$

After that, the criteria for stopping the evolutionary search are checked (achievement of an acceptable value of the objective function, absence of significant improvements of the values of the objective function during a certain number of iterations, exceeding the maximum possible search time, etc.). In case of non-satisfaction of the stopping criteria, the stages of crossing, mutation and inversion are repeated.

### The results of the algorithm

Table 2. Influence of population size on the accuracy and speed of algorithm operation during processing of unimodal functions (number of generations - 50, total discharge - total – 256, of them into a fractional part - 16)

Population size	5	10	20	40	80	160	320	640
<b>Canonical</b>								
Minimum	4	3,9867	4	4	3,7968	3	3	3
time, s	0,8548	0,8392	0,8299	0,8361	0,8392	0,8361	0,8486	0,8642
<b>Genitor</b>								
Minimum	4	4	4	4	3	3	3	3
time, s	1,0077	0,9360	1,1481	1,0826	1,3260	1,5412	2,1122	3,1512
<b>CHC</b>								
Minimum	4	4	4	4	3	3	3	3
time, s	0,9453	0,9703	1,0358	1,0670	1,2417	1,4976	2,0716	3,1512
<b>Island model</b>								
Minimum	-	1.631e +137	2.070e +120	1.198e +51	3	3	3,2000	3
time, c	-	1,7752	1,7971	1,8751	1,9812	2,1590	2,5833	3,4320
<b>Bidirectional GA (DAGA2)</b>								
Minimum	-	5,288e +122	6,549e +60	3,095e +53	1,475e +19	3	3	3
time, s	-	2,7268	2,6239	3,0544	3,7253	5,3071	8,6518	15,703
<b>GA with a decrease in population size</b>								
Minimum	-	3	3	3	3,2000	3	3	3,2000
time, s	-	1,2261	1,3010	1,4664	1,8782	2,7175	4,3118	7,7064

From the results obtained in Table 1, it was concluded that the accuracy of all methods manifests itself at a population size of 80 or more individuals, and becomes stable in the range from 160. The only exception is the method with a decrease in population size, which shows a stable result for all values of the population size.

### **Conclusions**

It has been determined that technical diagnostics is a field of knowledge, that it consists of theory, methods and means for identifying the state of objects. It is noted that, as a rule, the state of an object is determined based on the available observations of it (measured values of input parameters) and a mathematical model that describes the relationship between the input and output parameters of the objects under study, and which is built on the basis of the training sample data. It has been determined that evolutionary methods, including genetic algorithms, are effective methods of computational intelligence that can be used to build diagnostic models.

As a result of the research, the genetic algorithm was refined and adjusted to reduce the proportion by adjusting its parameters to increase the speed of evolutionary optimization. Recommendations are given for tuning the initial parameters of the evolutionary search when using the proposed modification. The parameters of the method, in particular, the size of the population, the number of generations, the bit width and the type of crossover are selected in such a way as to minimize the operating time of the module being developed and to obtain an accuracy within acceptable limits.

### **References:**

1. R. Thompson, Automotive Technology: A Systems Approach, 6th ed., Cengage Learning, Inc, Clifton Park, 2014.
2. M. Busch, Mike Bush on engines: what every aircraft owner need to know about design, operation, condition monitoring, maintenance and troubleshooting of piston aircraft engines, 1st ed. CreateSpace Independent Publishing Platform, 2018.
3. S. Russel, P. Norving, Artificial Intelligence: A Modern Approach, CreateSpace Independent Publishing Platform, 2016.
4. F. Shen, A fast multi-tasking solution: NMF-theoretic co-clustering for gear fault diagnosis under variable working conditions, Chinese journal of mechanical engineering 33, (2020), pp. 16. DOI: 10.1186/s10033-020-00437-3.
5. W. Zamboni, G. Petrone, G. Spaquuolo, D. Beretta. An evolutionary computation approach for the online/on-board identification of PEM fuel cell impedance parameters with a diagnosis perspective, Energies 12, (2019), pp. 4374. DOI: 10.3390/en12224374.

**<sup>1</sup> Hlib Horban**

PhD., an associate professor at the Department of Software Engineering

**<sup>2</sup> Ihor Kandyba**

A lecturer at the Department of Software Engineering

**<sup>3</sup> Mykhailo Dvoretskyi**

A senior lecturer at the Department of Software Engineering

**<sup>4</sup> Anzhela Boiko**

Ph.D., an associate professor at the Department of Computer Engineering

<sup>1-4</sup> Petro Mohyla Black Sea National University

## PRINCIPLES OF SEARCHING FOR A VARIETY OF TYPES OF ASSOCIATIVE RULES IN OLAP-CUBES

**Abstract.** The classification of association dependencies which can take place among multidimensional data is presented in the article. The representation of templates of inter-dimensional association rules is considered. Generation methods of inter-dimensional and intra-dimensional association rules are presented. Formulas for calculating objective and subjective characteristics of significance of these association rules types are presented.

**Keywords:** OLAP, Data Mining, multidimensional data, association rule, support, confidence, lift, leverage, template, dimension, measure, attribute, combination, set, difference.

Technologies of on-line Analytical process (OLAP) [1, 2] and data processing [3, 4] are typically employed in trendy data analysis systems and in decision support systems, that alter additional or less effective knowledge analysis. OLAP technology permits conducting user-defined operation like consolidation, detalization, data slice, cube rotation et al. At the identical time data processing investigates some cumulated hidden knowledge that was unknown before that and will be enough helpful within the data analytics process, upon that knowledge is taken from data sheets pre-spawned likewise by means that of database management systems (DBMS). One of the foremost common tasks of Data Mining is association, that represents detection of regularities between connected objects, an example of which can be the rule that event Y follows event X [5]. X is named a condition or an antecedent, and Y is named a consequent. Rules of that sort are known as association rules.

Data Mining strategies and algorithms [6], together with association rule mining likewise, are chiefly supported processing bestowed in tabular type, wherever sets of analyzed knowledge are settled either in one column or in one line, so that they add one dimension. But such knowledge regularities could happen even in three-dimensional data [7]. If to think about a three-dimensional cube rather than relational table data, then an item set for association rule mining may be bestowed as a collection of attribute values for every dimension, likewise as sets of values within the plurality of

dimensions.

The main elements of OLAP cubes are dimensions and measures. Dimension could be a values sequence some of the parameters to be analyzed. Samples of dimensions is time, geographic location, etc. Typically, dimensions contain extra data that permits users to investigate actual knowledge. Values that are obtained at the intersection of cube dimensions and represent quantifying facts are referred to as measures.

Mathematically the hypercube is suitable to represent by following sets:

D – a set of hypercube dimensions for a specific subject area:

$$D = \{D_1, D_2, \dots, D_i, \dots, D_n\},$$

where  $D_i$  –  $i^{th}$ -dimension,  $n$  – the quantity of dimensions;

A – a set of attributes (values of elements) of hypercube dimensions:

$$A = A_1 \cup A_2 \cup \dots \cup A_i \cup \dots \cup A_n,$$

where  $A_i$  – a set of attributes of dimension  $D_i$ , that successively are often diagrammatic as:

$$A_i = \{A_i^1, A_i^2, \dots, A_i^k, \dots, A_i^m\},$$

where –  $k$ -attribute of  $i^{th}$ -dimension,  $m$  – the quantity of attributes in  $i^{th}$ -dimension;

M – a set of values of hypercube measures:

$$M = \{M_{I_1, I_2, \dots, I_i, \dots, I_n}^1, M_{I_1, I_2, \dots, I_i, \dots, I_n}^2, \dots, M_{I_1, I_2, \dots, I_i, \dots, I_n}^z\},$$

where  $I_i$  – attribute index of  $i^{th}$ -dimension,  $n$  – the quantity of dimensions,  $M_{I_1, I_2, \dots, I_i, \dots, I_n}^l$  – l-measure for the cube cell with  $I_1, I_2, \dots, I_i, \dots, I_n$  index,  $z$  – the quantity of hypercube measures.

If to contemplate OLAP cube rather than relational data, then associate item set of association rules will represent a collection of values (attributes) of every dimension. Association rules that arise in multidimensional data will be classified by the subsequent sorts:

1. Inter-dimensional association rules – rules between attributes of different dimensions:

$$(A_i^x \in D_I) \wedge \dots \wedge (A_j^y \in D_J) \rightarrow A_k^z \in D_K;$$

2. Intra-dimensional association rules:

$$(A_i^x \in D_I) \wedge \dots \wedge (A_i^y \in D_I) \rightarrow (A_i^z \in D_I) \wedge \dots \wedge (A_i^w \in D_I);$$

3. Hybrid association rules – dependencies between dimensions, but some operands can be attributes of the same dimension:

$$(A_i^x \in D_I) \wedge \dots \wedge (A_j^y \in D_J) \rightarrow (A_j^y \in D_J) \wedge \dots \wedge (A_k^z \in D_K).$$

Hybrid association rules may be known as repetition association rules in distinction to different rules thought of, that essentially represent association rules while not repetitions.

As modern databases can be very large (up to gigabytes and terabytes), you need efficient algorithms to find reflection rules that can be scaled up and that will allow you to find a solution within a reasonable time.

One such algorithm is Apriori, first proposed by Sricant and Agraval [8]. Originally it was developed for relational databases and allowed the generation of frequent data sets from transaction tables.

Frequent subject set in multidimensional data means a set of attribute values for the relevant measurements, the value for which is below the threshold for the minimum support value, which is set by the end user based on his own experience.

This results in frequent subject sets from data first with one dimension, then with two, etc. Finally, frequent subject sets can be found with n dimensions, where n is the total number of measurements in a cube.

In general, let the set of all frequent sets of topics in the OLAP cube be a set of S:

$$S = \{S_1, S_2, \dots, S_i, \dots, S_n\},$$

where i is the number of elements in a subject set,  $S_i$  is a lot of frequent subject sets with the number of elements and, n is the total number of elements in a cube.

In turn, sets of  $S_1, \dots, S_n$  contain different subject sets for each of the measurements or sets of measurements if the number of elements in the set is greater than one.

In other words:

$$S_1 = \{s_1, s_2, \dots, s_n\},$$

where  $s_1$  is a set of frequent single element subject sets in the first dimension of the cube,  $s_2$  in the second dimension and  $s_n$  in the n dimension.

In turn, many two-element subject sets can be presented as follows:

$$S_2 = \{s_{12}, s_{13}, \dots, s_{mn}\},$$

where  $s_{12}$  is a set of frequent subject sets for the first and second dimensions,  $s_{13}$  for the first and third dimensions,  $m \neq n$ .

Let k be the number of elements in the subject set. So, in general:

$$S_k = \bigcup_{i=1}^{C_n^k} \underbrace{\{s_{i_1, i_2, \dots, i_k}\}}_k.$$

It is clear that when creating frequent OLAP cube subject sets, they will not include all the elements included in the corresponding cube measurements. To include an element or a collection of them in such sets, you must first calculate the support for that collection.

It is proposed to create a frequent subject set in the form of a list, where the first element is a sublist containing the sequence numbers of cube measurements according to which the set is generated [24]. in a single element set, such a list contains only one element. This sublist in the first element in the further generation of associative rules

based on subject sets is necessary to identify the measurements for which all sets have been created.

Among multidimensional data similar to tabular one, it is possible to find certain association dependencies represented in the form of rules that can be classified as inter-dimensional, within one dimension and hybrid. The approach to construction of templates of inter-dimensional association rules is proposed by generating all possible combinations of dimensions in OLAP-cube, which allows obtaining possible association rules, as well as the approach to construction of association rules within one dimension by generating all possible combinations of values of a certain dimension, among which search for dependencies is carried out. Appropriate methods have been developed for generating inter-dimensional association rules and association rules within one dimension. In the future, it is planned to study methods of hybrid association rule mining among multidimensional data.

### References:

1. E. Thomsen, [OLAP Solutions: Building Multidimensional Information Systems], John Wiley & Sons, New York, 2002, pp. 50–688.
2. R. Wrembel, C. Koncilia, [Data Warehouses and OLAP: Concepts, Architectures and Solutions], Idea Group Inc., 2007, pp. 12–237.
3. D. Hand, H. Mannila, P. Smyth, [Principles of Data Mining], Massachussets Institute of Technology, Cambridge, 2001, pp. 2–546.
4. N. Ye. (Ed.), [The Handbook Of Data Mining], Lawrence Erlbaum Associates Publishers, 2003, pp. 254–690.
5. C. Zhang, and S. Zhang, [Association Rule Mining: Models and Algorithms], Springer-Verlag, Berlin, 2002, pp. 12–238.
6. Kovalenko, Y. Davydenko and A. Shved, Formation of Consistent Groups of Expert Evidences Based on Dissimilarity Measures in Evidence Theory, in: Proceedings of the 14th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2019, pp. 113–116. doi: 10.1109/STC-CSIT.2019.8929858
7. H. Zhu, [Online analytical mining of association rules. Master's thesis], Simon Faster University, Burnaby, 1998, pp. 5–51.
8. Symeonidis and P. Mitkas, Agent intelligence through Data Mining, Springer Science+Business Media, Heidelberg, 2005, pp. 3–200.

<sup>1</sup>**Andrii Khlevnyi**

PhD, Associate professor Department of technology management

<sup>2</sup>**Bohdan Koval**

Master's degree student Department of technology management

<sup>3</sup>**Svetlana Shabatskaya**

Associate professor Department of Medical and Biological Physics and Informatics

<sup>1,2</sup>*Taras Shevchenko National University of Kyiv*

<sup>3</sup>*National Medical University*

## **DEVELOPMENT OF A FRAUD DETECTION SYSTEM IN PAYMENT SERVICES USING CRISP-DM METHODOLOGY**

**Abstract.** The paper presents the development of a system for detecting fraud in payment services using CRISP-DM methodology. It appoints that the methodology partially satisfies the requirements of financial institutions and proposes its adaptation.

**Keywords:** methodology, data analytics, payment services.

The demand for various payment services is constantly increasing nowadays. It is coupled with the development of technology and increasing need for such systems. High transaction speed, transparency, full control over payments, and ease of use are essential for clients. However, in addition to the benefits of modern payment services, there is a dynamic problem associated with maintaining the integrity of transactions. Therefore, one of the urgent issues for banks is to solve the problem of detecting and preventing illegal actions with their financial resources. The best way to deal with fraud is to prevent it. The warning is possible due to systematic data processing, with fairly well defined stages. To address this issue, it is appropriate to apply a methodological approach. Given the basics of data analysis [1,2] it has been determined, that the use of CRISP-DM is appropriate for the development of a system that will detect fraud at the transaction stage. Let's take a look at the basic principles and concepts underlying the methodology of developing a system for detecting fraud in payment services.

According to CRISP-DM [2] the lifecycle of a fraud detection system project consists of 6 stages. It has been confirmed that any methodology does not work without adjusting it to the relevant business needs [3]. Therefore, the sequence of steps is not considered from the point of their strict compliance and depends on the requirements of the applied field - financial system. According to the results of the study, all development stages of a system for detecting fraud in payment services have been taken. It has been discovered that at 3 - 5 stages data analysts play the key role, and at stages 1, 2, and 6 the key is the project team. Let's look through these stages and fill them with the main tasks from the point of developing a system for detecting fraud in payment services.

*1. Business understanding.* At this stage, we need to explore the business processes of the financial institution that owns the payment services. Identify the main types of fraud the institution deals with, and its competitors. We need to set a business goal - to save money and increase customer loyalty by minimizing fraudulent

transactions. Investigate competitors and identify the risks that occur in the development of such systems, prepare the project plan to develop a system for fraud detection in payment services and its expected results, establish responsible persons for each stage of the methodology.

2. *Data understanding.* To implement this stage, we need to determine data sources for each type of fraud, configure basic parameters of the data collection, set attributes, perform data cleaning and pre-processing, identify valuable subsets to form hypotheses of fraud in the system, so we can further analyze hidden patterns.

3. *Data preparation.* Data analyst is responsible to pre-process the data before modeling at this stage.

4. *Modeling.* Here we choose the methods using which the transactions' analysis and fraud detection will be resolved, and implementing the model. Also, we determine the methods of testing, training and evaluation of models.

5. *Evaluation.* The list of approved models and the stakeholders' feedback on the results is determined at this stage. Key persons prepare alternative solutions and strategies, that may be used to improve the models, the following steps are appointed.

6. *Deployment.* Develop a specific methodology for implementing a fraud detection system in payment services of financial institutions.

#### References:

1. Foster Provost and Tom Fawcett. Big Data. Mar 2015. 51-59. <http://doi.org/10.1089/big.2013.1508>
2. Shafique U. Qaiser H. A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA) International Journal of Innovation and Scientific Research ISSN 2351-8014 Vol. 12 No. 1 Nov. 2014, pp. 217-222 © 2014 Innovative Space of Scientific Research Journals <http://www.ijisr.issr-journals.org/>
3. Khlevna J.L., Khlevyi A.O. Osnovy` formuvannya meta-metodologiyi upravlinnyxa proektamy'. XIV- Mezhdunarodnaya prakty` cheskaya konferency`ya "Upravleny'e proektamy': sostoyany'e y' perspektiv'y", My'kolayiv, 11 - 15 September 2018p.
4. Fraud Detection Techniques: Data and Technique Oriented Perspective / S. Sorournejad, Z. Zojaji, R.E. Atani, Amir Hassan Monadjemi / Cornel University Library, 2016. Mode of access: <https://arxiv.org/ftp/arxiv/papers/1611/1611.06439.pdf>.

<sup>1</sup> Nikolay Kiktev

<sup>2</sup> Taras Lendiel

<sup>3</sup> Volodymyr Osypenko

<sup>1,2</sup> National University of Life and Environmental Sciences of Ukraine

<sup>1</sup> Taras Shevchenko National University of Kyiv

<sup>3</sup> Kyiv National University of Technologies and Design

## APPLICATION OF THE INTERNET OF THINGS TECHNOLOGY IN THE AUTOMATION OF THE PRODUCTION OF COMPOUND FEED AND PREMIXES

### 1. Introduction

An increase in the efficiency of using the technological opportunities of agricultural units is achieved through extensive automation of production processes and the development of a computer control system. In recent years, there is actively developing a new direction - cloud technologies and the Internet of Things (IoT), which can also be used in the automation of agricultural production, including in the production of poultry feed and premixes. The IoT is the most advanced tool in industrial automation.

### 2. Problem Statement and Literature Survey

Given the above, to increase the efficiency of feed production and product quality it is necessary: to improve the control system of technological processes in the production of poultry feed by developing a two-tier computer-integrated control system; to apply modeling of individual technological processes; at the second level of the integrated control system to apply software to calculate the best feed recipe for nutritional qualities; using of modern cloud technologies and the Internet of Things for remote monitoring of the process via the Internet. Purpose of the research. This work solves the problem of developing a hardware and software complex for managing the production of poultry feed using cloud technologies and the Internet of things. With the advent of production automation, this area also applied to poultry feed production. Thus, in [3] researchers investigate how to obtain the optimal composition of feed. It is usually performed using the simplex method [4]. This technology will be included in AquaSim, a set of custom IoT productivity tools for Skretting [2].

### 3. Formulation and solution of the problem of optimizing the feeding ration of animals. Control system operation algorithm

The objective function of the feeding optimization problem can be written as follows:

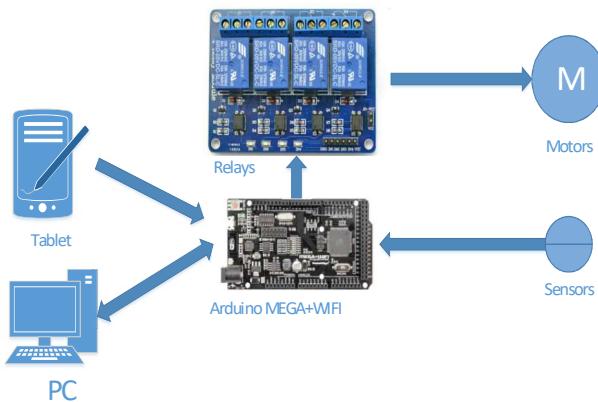
$$Z = \sum_{j=1}^m C_j X_j \rightarrow \min,$$

where  $C_j$  - cost or purchase price of the  $j$ -th type of feed;  $X_j$  - the required amount of the  $j$ -th type of feed in the daily diet, under restrictions (conditions) - nutrients in the diet contain at least the required amount:

$$\sum_{j=1}^m A_{ij} X_j \geq B_i,$$

where  $A_{ij}$  - the content of the  $i$ -th nutrient per unit of the  $j$ -th type of feed;  $B_i$  - daily requirement of the animal in the  $i$ -th nutrient.

The operation of the feed production control system is as follows. The grain comes from the loading hoppers through the pipeline to the ripper passing the flow sensor, which removes the flow rate. The signal from the flow sensor is fed to the flow meter which control signal through the actuator closes or opens the valves of the grain loading hoppers. After the baking powder, the raw material enters the hammer crusher. It is proposed to create a control system for the technological process of feed production using the technology of the Internet of Things, namely the principle of remote control and monitoring of technological processes. In project we use an Arduino MEGA2560 + WiFi R3 controller from RobotDyn [1]. In addition to Wi-Fi, the microcontroller is able to run programs from external flash memory with SPI interface. Schematically, the control system is shown in Fig. 1. The system works as follows: after power supply, the entire system is initialized; the current value of time  $t$  is compared with the critical value of time  $tk$  (time for which the whole technological process); measurement of technological parameters; check of wireless communication; under the condition of wireless connection, the measured value is displayed; waiting for the command; go to step 2; in the absence of wireless communication, the system goes into automatic mode; after the transfer of control action to the actuators, the data is sent to the personal computer of the general control system; go to step 2; when the critical time value is reached, the program is terminated. The control system is programmed in the "Arduino IDE" environment.



**Figure 4:** Symbolic circuits of the control system

#### 4. Processing Experimental Results and discussion

Conducted experimental researches have shown that there is developed the control system of technological process of production of compound feed with use of technology of the Internet of things. There is implemented an algorithm of control of technological process with use of technology of the Internet of things and is considered at the same time the objective function of a problem of optimization of feeding. In

particular, there is created a prototype of the installation with the proposed module for monitoring and maintaining technological parameters. The layout included an Arduino MEGA2560 + WiFi R3 controller from RobotDyn, level sensors, actuators and a PC.

### 5. References:

1. V. Lysenko, T. Lendiel, and D. Komarchuk. "Phytomonitoring in a greenhouse based on arduino hardware." 2018 International Scientific-Practical Conference (PIC S&T) (2018): 365-368.
2. K.R. Raghunandan, L.J. Quadras, S. Gurunandan, S.S. Karthik. "Usage of Internet of Things in Agriculture Automation." International Journal of Computer Trends and Technology (IJCTT), Vol. 58, Issue 1, April (2018): 35-39.
3. A.O. Kashkariov. "Pro efektyvnist' skladannya receptiv kombikormu." [On the effectiveness of compound feed recipes]. Konferenciya molodyh vchenyh TDAU, (2010): 1-3. (In Ukrainian).
4. N.A. Kiktev. "Infomacionnye tekhnologii v reshenii zadach upravleniya proizvodstvom kombikormov." [Information technologies in solving problems of compound feed production management]. Innovacii v sel'skom hozyajstve, № 3 (13), Moscow (2015): 53-57. (In Russian).

**Natalia Kondruk**

Candidate of Technical Sciences, Associate Professor of the Department of Cybernetics and Applied Mathematics

*Uzhhorod National University*

**SEGMENTATION OF DATA SETS BY DIFFERENT TYPES OF CLUSTERS**

Clustering is a powerful tool in the field of Data mining, when there is no a priori information about the relationships between data. Currently, many cluster analysis algorithms are successfully used in various application areas, where there is a need to divide similar in certain features objects into subsets [1-3]. A crisp split into clusters is possible only with very different features of clustering objects. Therefore, fuzzy methods are increasingly used to solve real problems, in which the division of objects is carried out to determine the degree of belonging of objects to clusters.

All existing methods can be classified according to the similarity measures they use [3, 4]. On the other hand, it determines the different geometric shape of the formed clusters and allows obtaining qualitatively different applied interpretations of the obtained homogeneous segments of data sets. Therefore, it is the specifics of applied problems that make it impossible to automatically transfer methods to another application area without the risk of deliberately obtaining a bad solution. Therefore, it is advisable to develop an information system that would have a fairly wide range of tools for grouping objects by different similarity measures. This makes it possible to effectively solve a lot of applied problems in different subject areas. The main works in which the technology is presented, which allows to solve this problem are presented in [4-7].

The focus of the system is a single-level clustering method based on fuzzy binary relations described in [5]. The flexibility of this algorithm allows you to form different geometric shapes of clusters of datasets by simply changing the appearance of the degree of similarity of objects. In this case, the similarity of the objects  $O_i$  and  $O_j$  by some criterion is characterized by a fuzzy binary relation  $R$  on the set of vector features with the membership function  $\mu_R$ . The closer the value  $\mu_R$  is to 1, the more similar the objects will be to this criterion. Thus, in [4-7], three types of similarity measures of objects are proposed: length-based, angular and distance.

To form elliptically similar clusters, it is expedient to use the "distance" similarity measure, which is described by a fuzzy binary relation  $R^D$  [4]. The fuzzy binary relation  $R^K$  [7] characterizes the angle of deviation between the feature vectors. Its use makes it possible to carry out clustering with conical clusters. The length-based similarity measure  $R^L$  allows splitting the feature vectors of objects into clusters by concentric spheres [5].

Conical clustering can be effectively used to solve multi-criteria linear programming problems with a large-scale criterion space [7], which arise, in particular, in mathematical modeling of balanced nutrition problems. One of the steps in solving such problems is to cluster their criteria space. In this case, the relationships between the criteria are determined by their angular similarity  $R^K$ . Clustering by elliptical

clusters is most common in many application problems, as the similarity of objects is based on a "distance" similarity measure. Also in [6] two synthetic sets of two-dimensional data of Gaussian type are generated and efficiency of application of a clustering method based on fuzzy binary relations at various indices of an estimation of quality of partition is investigated. Clustering by concentric clusters (clusters in the form of concentric spheres) [5] made it possible to group objects by length-based similarity of their feature vectors and to obtain a qualitatively new applied meaningful interpretation of the formed homogeneous groups in practice. In addition, this approach allows for both crisp and fuzzy data clustering.

In perspective researches the combined index of an estimation of clustering quality which is adapted to use of various similarities measures of a fuzzy binary relations method will be created; development of a software system that will ensure the segmentation of data sets into different geometric shapes clusters without prior determination of the clustering threshold.

#### References:

1. T. Sajana, C. S. Rani, K. V Narayana, A survey on clustering techniques for big data mining. Indian journal of Science and Technology, 9(3), 2016, pp. 1-12. doi: 10.17485 / ijst / 2016 / v9i3 / 75971
2. A. Amelio, A. Tagarelli, Data mining: clustering. Encyclopedia of Bioinformatics and Computational Biology, 2018, pp. 437-48. doi: 10.1016 / B978-0-12-809633-8.20489-5
3. K. Chitra, D. Maheswari, A comparative study of various clustering algorithms in data mining. International Journal of Computer Science and Mobile Computing, 6(8), 2017, pp. 109-115.
4. N. Kondruk, Clustering method based on fuzzy binary relation, Eastern-European Journal of Enterprise Technologies, 2017, pp. 10–16. doi:10.15587/1729-4061.2017.94961
5. N. Kondruk, Use of length-based similarity measure in clustering problems, Radio Electronics, Computer Science, Control, 2018, pp. 98–105. doi:10.15588/1607-3274-2018-3-11.
6. N. E. Kondruk, A comparative study of cluster validity indices, Radio Electronics. Computer Science. Control, 4, 2019, pp. 59 – 67. doi: 10.15588/1607-3274-2019-4-6.
7. N. E. Kondruk, M. M. Malyar, Structuring of the criterional space by an angle similarity measure, Scientific Bulletin of Uzhhorod University. Series of Mathematics and Informatics, 2020, pp. 85 – 91. doi: 10.24144/2616-7700.2020.1(36).85-91

<sup>1</sup>**Bohdan Koval**

Master of the Department of Management Technology

<sup>2</sup>**Iulia Khlevna**

Doctor of Technical Sciences, Associate Professor of the Department of Technologies Management

<sup>1,2</sup>*Taras Shevchenko National University of Kyiv*

## FRAUD DETECTION TECHNOLOGY IN PAYMENT SYSTEMS

**Abstract.** The paper outlines the relevance of fraud prediction from the standpoint of the integrity of the study. It proposes the solution - the development of technology to detect fraud in payment systems and gives the definition of such technology. It has been established that in terms of technology it is important to develop an effective and optimized model for the classification of fraud in payment systems from the standpoint of all stages of the study.

**Keywords:** data science, machine learning, deep learning, data visualization, binary classification.

The fast expansion of the practice of financial institutions of user autonomy is a requirement of today. Financial institutions and customers face new challenges related to fraudulent malicious activities. The consequences of which are violations of the integrity and truthfulness of transactions, financial losses, reduced customer loyalty and their loss. To prevent this, we need to transform approaches and means to monitor, detect and control illegal actions. Of course, the best way to combat fraud is to prevent it. Such a signal is possible through the development of a system based on the prediction of fraudulent actions at the level of suspicious transactions and forecasting the probability of its occurrence.

According to the analysis [1, 2] it has been pointed, that the main essence of the presented works is the models of classification of transactions for fraud by different methods. The use of combined methods is presented in the works [3, 4]. It has been revealed that the research of the model of user behavior with the subsequent assignment of the transaction to fraudulent or non-fraudulent is the most relevant. Various methods and algorithms can be used to solve this problem. However, there is no powerful algorithm in the literature on credit card fraud that would be the standard for all financial institutions. [5]. Therefore, the study of fraud detection models, changes in their parameters, combination of algorithms to maintain each other's advantages and cover their weaknesses in detecting fraud with financial payment systems from the standpoint of systematization, namely in the form of consistent actions, is the technology of scientific and practical interest.

Technology of fraud detection in payment systems (TFDPC) has been proposed. TFDPC is a set of systematized ways to provide forecasting, detection and control of fraudulent transactions in financial systems. At the basis of such methods are models, methods and algorithms of machine learning.

In terms of the construction of TFDPC, it has been proposed to develop an

effective and optimized model for the classification of fraud in payment systems from the standpoint of all stages of the study. To implement the full cycle of TFDPC we chose the Python3 programming language for its simple syntax, broad support of the programming community and a huge amount of available documentation.

It has been established that to implement the solution it is advisable to choose the following libraries: pandas, numpy, matplotlib, scikit-learn, K-nearest neighbors, random forests, xgboost and others.

A dataset has been formed, which is based on a set of transaction data of an anonymous payment system, namely data of banking operations performed by individuals alone and consists of 6,362,620 records. 10 transaction attributes are selected.

Research analysis and data preparation consisted of: determining which type of transactions is most often fraudulent (and accordingly adapting the data frame), substantiation of anomalies in the funds transfer, analysis of anomalous transactions by initiator, quantitative detection of certain anomalies in transactions.

We apply visualization of the difference between fraudulent transactions and regular transactions and determine the correlations of attributes in regular and fraudulent transactions (Fig. 1).

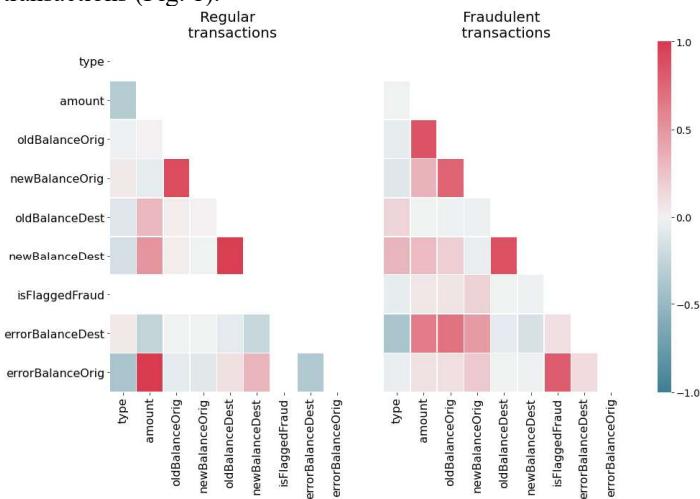


Figure 1. Heats maps of attributes correlation in regular and fraudulent transactions

The metrics used to evaluate the model are based on the area under the response accuracy curve (AUPRC), rather than the usual area under the recipient performance curve (AUROC).

Models were created and analyzed, focusing on the detection of anomalies and supervised training: logistic regression, K-nearest neighbors, support vectors machine (SVM), the Bayesian classifier. The best result is achieved by applying an algorithm based on ensembles of decision trees that works effectively on unbalanced data. Among

these algorithms (based on decision tree ensembles) there are 2 most effective - Random Forest and XGBoost, and the last, gradient boosting algorithm, still shows the best result. In addition, XGBoost allows you to weigh the positive class (fraud) more efficiently than the negative class (no fraud) - which allows you to more efficiently process unbalanced data.

The constructed algorithm has AURPC score of 0.9986, which indicates a very high efficiency of the classifier. The accuracy of the model implemented using the technology of extreme gradient boosting is 99.97%.

The obtained results are not just high, but the technology can be recommended for use in business and banking, because out of 554,082 test transactions, only 3 transactions that were classified as genuine (non-fraudulent) turned out to be fraudulent, 166 actually genuine transactions were identified as fraudulent. Accordingly, the construction of TFDPC embodies: research analysis, data visualization with subsequent adaptation of the data set, technology creation using existing classification algorithms, visualization of the obtained model and results.

#### References:

1. Fraud Detection Techniques: Data and Technique Oriented Perspective / S. Sorournejad, Z. Zojaji, R.E. Atani, Amir Hassan Monadjemi / Cornel University Library, 2016. Mode of access: <https://arxiv.org/ftp/arxiv/papers/1611/1611.06439.pdf>.
2. Lebichot, B., Le Borgne, Y.-A.: Deep-Learning Domain Adaptation Techniques for Credit Cards Fraud Detection. In: Oneto, L., Navarin, N., Sperduti, A., Anguita, D. (eds.) Recent Advances in Big Data and Deep Learning, pp. 78–88. Springer, New York (2019)
3. Kuznyecova N.V. Analiz ta prohnozuvannya ryzykiv shaxrajstva z kredytnymy kartkamy. Informatics and Mathematical Methods in Simulation Vol. 8 (2018), No. 1, pp. 16-25
4. Kuznietsova, N.V. Scoring Technology for Risk Assessment of Fraud in Banking / Selected Papers of the XVI International Scientific and Practical Conference "Information Technologies and Security" (ITS 2016). — 2016. — Pp. 54-61 .
5. MasoumehZareapoor, Seeja.K.R, M.Afshar.Alam, "Analysis of Credit Card Fraud Detection Techniques: based on Certain Design Criteria", International Journal of Computer Applications (0975 – 8887) Volume 52– No.3, 2012.
6. Bailey MJ, Muth RF, Nourse HO. 1963. A regression method for real estate price index construction. Journal of the American Statistical Association 58: 933–942.

**<sup>1</sup> Yaroslav Linder**

Ph.D. in Physics and Mathematics, associate professor

**<sup>2</sup> Maksym Veres**

Ph.D. in Physics and Mathematics, associate professor

**<sup>3</sup> Kateryna Kuzminova**

*<sup>1,2,3</sup> Taras Shevchenko National University of Kyiv*

## **MODELING AND PREDICTION OF COVID-19 USING HYBRID DYNAMIC MODEL BASED ON SEIRD WITH ARIMA CORRECTIONS**

**Abstract.** The stages of the proposed method are building a SEIRD compartment model with vital dynamics, estimating its parameters, calculating and predicting the difference between the SEIRD model solution and the observed data using the ARIMA model, and adjusting model prediction using this newly obtained data on the residuals. The proposed method was tested on the data on the epidemic's dynamic in Ukraine. The validation results indicate the method's aptitude to real-world usage.

**Keywords:** COVID-19, SEIRD, ARIMA, Hybrid Dynamic Model.

### **Introduction**

As the coronavirus pandemic continues to rattle the world, humanity craves for means to alleviate the situation if not overcome the crisis entirely. Quality estimations and predictions of future dynamics of the disease spread will ensure better prevention and thorough preparation for exacerbations of the problem (such as expected rises in infection cases after the holidays or lockdown lifts). Rational use of resources may help avoid future boiling points for the healthcare and other systems critical to the delivery of the COVID-19 response.

The susceptible exposed infectious recovered model (SEIR), which is based on differential equations, is one of the most widely adopted methods for modeling the epidemic of the COVID-19 outbreak [1]. The SEIR model replicates the “time-history” of any epidemic or pandemic outbreak, and it presents the model of dynamic interaction between people with four different health conditions or phases of the pandemic, namely the susceptible (S), exposed (E), infective (I), and recovered (R). SEIRD model, as a generalization of the SEIR model, has an additional variable – Deceased individuals. A “Formal Characterization and Model Comparison Validation” based on the SEIRD model, which uses the data from Korea and Spain, is proposed by Casas et al. [3]. The proposed model showed the predicted parameterization with empirical evidence and a decision support system (DSS) is implemented to study the nature of the pandemic in Catalonia [3]. A data-driven model to predict the spread of Covid-19 for an upcoming week using the SEIRD model is studied and tested for datasets obtained from Italy, India, and Russia [2].

### The hybrid dynamic model framework

Upon investigation, we introduce a novice model based on an enhanced SEIRD model and ARIMA model. As shown in Figure 1, the stages of the proposed method are building a SEIRD compartment model with vital dynamics, estimating its parameters, calculating and predicting the difference between the SEIRD model solution and the observed data using the ARIMA model, and finally adjusting model prediction using this newly obtained data on the residuals.

The model consists of such stages:

- At the first one, we estimate SEIRD model parameters using historical data, trying to lessen the difference between the model's output and observed data. This model is responsible for long-term prediction (i.e., 60 days or 100 days).

- Calculate residuals between observed infected, recovered, and deceased percentage of the population and corresponding solutions of the SEIRD model.

- Build three ARIMA models on the time-series of each of these residuals. Prediction of these ARIMA models will compensate residuals between the SEIRD model and historical data in order to make predictions mode accurate.

- Validate the prediction of the obtained model using the data on the number of infected, recovered, and deceased individuals as of the most recent days, data on which was not included while working with the model on previous stages.

The compartments of the model are as follows:

- $S(t)$ : Susceptible individuals - stock of healthy people who may be infected; population inflow due to births is taken into account.
- $E(t)$ : Exposed individuals - virus carriers in the latent stage, during which they are not virus spreaders. Usually corresponds to an asymptomatic phase of the disease.
- $I(t)$ : Infectious individuals - virus carriers able to spread the disease to individuals in contact with them.
- $R(t)$ : Recovered individuals - stock of healthy people who are immune to COVID-19.
- $D(t)$ : Deceased individuals - population loss due to the disease, natural deaths included.

The model itself is comprised of a system of differential equations:

$$\left\{ \begin{array}{l} \frac{ds}{dt} = \Lambda N - \mu S - \frac{\beta SI}{N} \\ \frac{dE}{dt} = \frac{\beta SI}{N} - (\mu + \sigma)E \\ \frac{dI}{dt} = \sigma E - (g + \mu)I \\ \frac{dR}{dt} = g(1 - \mu_{COVID}(t))I - \mu I \\ \frac{dD}{dt} = g \mu_{COVID}(t)I \end{array} \right. \quad (1)$$

with constraints at time  $t=0$   $S=S_0$ ,  $E=E_0$ ,  $I=I_0$ ,  $R=R_0$ ,  $D=D_0$  and parameters  $\Lambda$  – population's birth rate;  $\mu$  – population's mortality rate;  $\beta$  – rate of virus transmission,

which is the probability of transmitting disease between a susceptible and an infectious individual;  $\sigma$  – rate of latent individuals becoming infectious (average duration of incubation is  $1/\sigma$ );  $g$  – recovery rate, which can be initially estimated as  $g = 1/D$ , where  $D$  is the average duration of infection;  $\mu_{COVID}(t)$  – death rate due to COVID-19, which is estimated by an inverse exponential formula  $\mu_{COVID}(t) = \alpha e^{-\xi t}$ . The population size  $N(t) = S(t) + E(t) + I(t) + R(t)$  is not fixed due to its global birth and mortality rates taken into account at any given time  $t$ .

## Results

In this section, we will provide results of hybrid model approbation on data from the Ukrainian finance analytics website [4].

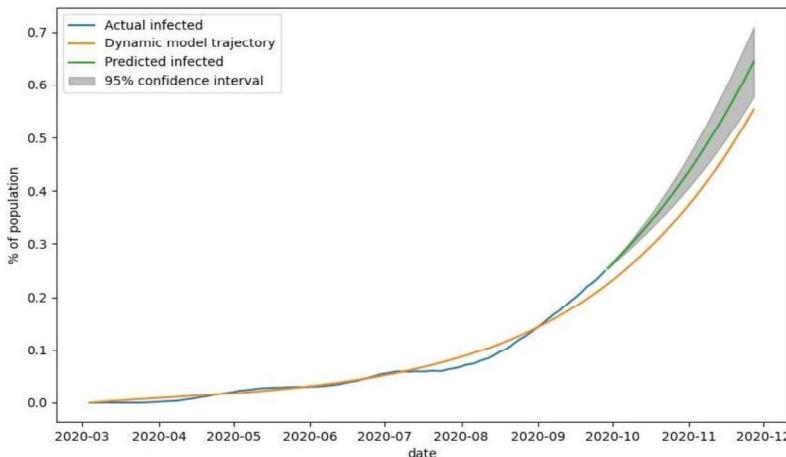


Figure 1 – The observed number of infected individuals (blue), number of infected individuals modeled with SEIRD model (yellow), and predicted number of infected individuals (green) by SEIRD model and corrected by ARIMA residual prediction with 95% confidence interval (grey)

The analysis of modeling and prediction of the number of infected individuals (Figure 5) shows that the number of observed cases of the disease grew steadily during the first half of the outbreak (mid-July) and is very accurately modeled with our method. The deviation of the predicted number of infected individuals from the observed data in the second half of July and August is most likely caused by the insufficient number of tests for COVID-19 performed during this period. The inconsistency in testing and changing levels of quarantine severity explain further deviations of observed data from the output of the SEIRD model. The prediction, corrected by ARIMA residual estimation, steadily increases, with optimistic and pessimistic scenarios (lower and upper bounds of the grey area, respectively) deviating by less than 0.1%.

Table 1

Quality measures of the fitted model for validations set

	MAE	MSE	MSLE	Normalized MAE	Normalized MSE	Max. deviation
Infected	1.13 $\cdot 10^{-4}$	$2.51 \cdot 10^{-8}$	2.50 $\cdot 10^{-8}$	3.59 $\cdot 10^{-2}$	2.62 $\cdot 10^{-3}$	8.6%
Recovered	2.76 $\cdot 10^{-4}$	$9.25 \cdot 10^{-8}$	9.21 $\cdot 10^{-8}$	1.1 $\cdot 10^{-1}$	1.66 $\cdot 10^{-2}$	15.4%
Deceased	9.28 $\cdot 10^{-6}$	1.24 $\cdot 10^{-10}$	1.24 $\cdot 10^{-10}$	8.41 $\cdot 10^{-2}$	1.11 $\cdot 10^{-2}$	15.5%

As shown in Table 1, all measures of the prediction quality for the infected, recovered, and deceased fractions of the population are very low.

## Discussion

The proposed hybrid model consists of a dynamic SEIRD model with vital dynamics and decaying COVID mortality rate and three ARIMA models that cancel out dynamic model residuals and enhance prediction quality. The model was tested on Ukrainian COVID statistic data. Obtained validation results allow us to draw conclusions that the proposed hybrid model has good prediction ability and decent performance. Obtained long-term predictions reflect the general dynamic of the outbreak and are especially useful for the healthcare system workers and government officials. Obtained short-term predictions allow us not only to forecast the future number of infected, recovered, and deceased patients but only estimate forecast error under adverse or optimistic circumstances. The proposed method can be used as an effective tool for prediction and analysis of the dynamics of the COVID-19 pandemic.

## References:

1. Hethcote, H.W., 1989. Three basic epidemiological models. In Applied mathematical ecology (pp. 119-144). Springer, Berlin, Heidelberg.
2. Rapolu, T., Nutakki, B., Rani, T.S., and Bhavani, S.D., 2020. A Time-Dependent SEIRD Model for Forecasting the COVID-19 Transmission Dynamics. medRxiv.
3. Fonseca i Casas, P., Garcia Carrasco, V. and Subirana, J., 2020. SEIRD COVID-19 Formal Characterization and Model Comparison Validation. Applied Sciences, 10(15), p.5162.
4. Ukrainian finance analytics website, 2013. URL: <https://index.mnfin.com.ua/ua/reference/coronavirus/ukraine/>.

<sup>1</sup> Volodymyr Mikhieiev

Student of the Department of Management Technology

<sup>2</sup> Olga Mezentseva

Candidate of Economic Sciences, Associated prof. of the Department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## ANALYSIS AND FORECASTING OF ENVIRONMENTAL POLLUTION BY CARBON DIOXIDE

**Abstract.** The main idea of this work is to analyze carbon dioxide pollution and apply several methods of mathematical modeling for this. In this case - the Gaussian model of the distribution of pollutants in the atmosphere and the matrix method of the effect of carbon dioxide on the state of the atmosphere.

**Keywords:** Gaussian model, forecasting, air pollution

The solution to the problem of managing the impact of industrial facilities on atmospheric pollution by emissions in the theoretical and methodological aspect requires the development of technologies for assessing the state of the environment and the dynamics of changes in the ecological and economic situation, as well as the formulation of problems related to environmental and mathematical modeling and the use of forecasting models. Modeling and forecasting, as the main tools of the system for managing the impact of industrial facilities on atmospheric pollution, allow not only to develop scenarios and options for ecological and economic development, but also to determine the maximum level of air pollution through the optimum of the objective function.

In the last decade, the expansion of the framework of the environment protection management system due to the inclusion of an environmental safety block in it required the solution of many methodological, regulatory, legal, and information problems. One of which consists in the methodological support of diagnostics of environmental safety and assessment of its state in order to rank disadvantaged areas and determine the priorities for their development[1].

One of the most effective methods is based on the Gaussian scattering model. When pollutants escape into the atmosphere, a cloud is formed, which is carried away along with the surrounding atmospheric air in the direction of the wind. In the process of movement, turbulent mixing with the surrounding atmospheric air occurs, which leads to the expansion of the cloud in space and a change in the concentration of pollutants in it. As a result, a spatial distribution of the concentration of pollutants is formed, which in the most general case is described by the normal (Gaussian) law[2].

In the case of using the matrix method, the values of the air pollution indicator, obtained in the form of relative values, characterize the growth rates of emissions (pollution) for individual industries and the average growth rate for industry. Forecasting is performed based on the dependence  $Y = f(X, t)$  and a specific analytical expression defined for each industry. In this case, the calculation of the forecast values

of the growth rates of the mass of emissions by each industry is carried out by a simple substitution of time periods for the subsequent forecast interval[3].

**Conclusions.** Especially important and urgent are the tasks of monitoring the concentration of pollutants and forecasting the state of the air basin in industrial regions. Solving these problems will allow us to rationally approach the issue of locating new enterprises in the region, make administrative decisions in the field of environmental safety, and develop effective measures to reduce the level of air pollution. Predicting the level of air pollution will warn the population about possible dangers and strengthen environmental control by society. Using monitoring systems, organizations will be able to adjust the work schedule of employees in the open air, and government agencies will warn about the dangers of various events on days when the norms are exceeded, and industrial enterprises can regulate the level of emissions of harmful substances into the atmosphere.

#### References:

1. Vyvarets A.D., Belik I.S., Stepanova N.V., Leontyev Y.V., Nikulina N.L., Atmospheric pollution assessment industrial emissions, 2017. URL: [https://elar.urfu.ru/bitstream/10995/40861/1/ozapv\\_2006.pdf](https://elar.urfu.ru/bitstream/10995/40861/1/ozapv_2006.pdf).
2. Antonova M.A., Vorobev A.V., Vorobev A.V., Dutova E.M., Pokrovskiy V.D., Modeling the distribution of pollutants in the atmosphere, 2019. URL: [http://archive.tpu.ru/bitstream/11683/55288/1/bulletin\\_tpu-2019-v330-i6-17.pdf](http://archive.tpu.ru/bitstream/11683/55288/1/bulletin_tpu-2019-v330-i6-17.pdf).
3. Matveev Y.N., Maslenikov B.I., Karelskaya K.A., Stukalova N.A., Mathematical modeling of processes the spread of a pollutant in soils and the atmosphere, 2016. URL: <https://naukovedenie.ru/PDF/65TVN516.pdf>.

**Julia Minaeva**

Ph.D., Associate Professor, Associate Professor at the Department of Intellectual Technologies.

*Taras Shevchenko National University of Kyiv*

## **PROCESSING OF MULTIDIMENSIONAL AND MULTI-ASPECT DATA IN CONDITIONS OF UNCERTAINTY**

Currently, the fuzzy set theory is a powerful mathematical apparatus that has a wide range of applications. However, some difficulties and inaccuracies in the fuzzy set theory (FST) application were found, in particular, in conditions of insufficient information, when the membership function (MF) assignment is impossible or associated with difficulties [1].

Fuzzy variables (FV), fuzzy numbers (FN) are the main objects of FST, their use reflects the uncertainty [2]. The limits of fuzziness are determined by the expert, using the available a priori knowledge of the system. Fuzzy numbers are in many aspects similar to probability distributions, and in [3] is brought an example of constructing a fuzzy number — an analog of a normal probability distribution — by using the convolution theorem.

Fuzzy set theory declared the universality of its models and the generality of their application, although several processes are not amenable to meaningful (formal) representation in the form of fuzzy sets (FS) due to their complexity lack of study. For example, the process of using FST in the study of biomedical processes is associated with some difficulties, the main of which is the omission or distortion of data, which does not allow the expert to assign a membership function. A new class of tasks has been identified that have a high semantic readiness for the use of the FST apparatus, but the incompleteness of the data prevents this [4].

FST was created as a means of solving problems (primarily - management) under conditions of uncertainty in the 2D data space. Further expansion of FST to 3D space is caused by new types of tasks that are difficult to transform into 2D space without losing the task representation adequacy. This emphasizes the relevance of finding new methods, models, and tools of solving problems under conditions of uncertainty in a multidimensional space, taking into account the multifaceted nature of the data.

Fuzzy sets (FS), and not only FN, are a product of the human mental activity, the universal set must be calculated based on the initial data set, in particular, the definition of the universal set (US) interval [min/max] requires data, but the vast majority of works almost "leave behind" this fact. This is especially noticeable in the case of the influence of BIG DATA where, on the one hand, the use of FST can give a certain effect, on the other hand, the fundamental difficulties in determining the US limit the possibilities of using an effective mathematical apparatus.

Under uncertainty, a multidimensional (multi-aspect) object can be represented by a subset of ordered sequences (multi FS analog) — a subset of OS or a subset of ordered pairs — SOP, (analog of FS) [1], which are equivalent in terms of proximity of F-norms. In our case, the application of a subset of ordered pairs and subset of ordered

sequences for the analysis and modeling of uncertainty is based on the condition of invariance of norms).

A multi-aspect object is considered as:

$$\left\{ \begin{array}{c} x_1^{(1)} \quad x_1^{(2)} \dots x_1^{(n)} \\ \vdots \\ x_m^{(1)} \quad x_m^{(2)} \dots x_m^{(n)} \end{array} \right\} \quad (1)$$

$\underbrace{\qquad\qquad\qquad}_{\text{Aspects}}$

Fuzzification (MF calculation) is offered in 2 cases:

- the expert can assign heuristic FNs for each aspect;
- the choice or assigning of MF is limited.

A 3D initial data set, as a rule, is presented in the form of slices or fibers, which significantly complicates the process of forming FS, requires, in turn, new assumptions, in particular, the decision that can be obtained in one direction not always can be taken as a general, also, it is difficult to take into account distorted or omitted data, pre-processing is required, especially in terms of restoring some subset of the original data. In 3D space, there is, on the one hand, the need to take into account the uniqueness of the problem, on the other hand - the ability to bring the initial problem to a level where it is possible to apply standard methods for one- or two-dimensional spaces. Let's note, that 2 type-FS are objects of 3D space, fuzzy sets of such type are extensions of 1 type - FS, which belong to objects from 2D space

In conditions of uncertainty, the object under study has a number of hidden properties that can be detected by structuring the object (data set) in the form of 2D or 3D tensors, performing the next step of tensor decomposition, and obtaining a fuzzy-like subset of ordered pairs similar to FS. In addition, a powerful blurring tool that allows obtaining SOP is using of special matrices. This gives the ability for blurring not only a subset of values, such as a universal set but also a more complex object, such as a block matrix.

### References:

1. Minaeva, Julia & Filimonova, Oksana. Alternative subsets of ordered pairs and their application in decision-making problems under conditions of uncertainty. Management of Development of Complex Systems, 41, 68 – 82 (2020). dx.doi.org/10.32347/2412-9933.2020.41.68-82.
2. Hanss. Applied Fuzzy Arithmetic. An Introduction with Engineering Applications. Springer-Verlag Berlin Heidelberg 2005- 260 p.
3. Kofman A. Introduction to the theory of fuzzy sets: Per. with French. M.: Radio and communication. 1982.
4. E. Acara, D. M. Dunlavy, T. G. Kolda, M. Morup. Scalable Tensor Factorizations for Incomplete Data. arXiv:1005.2197v1 [math.NA] 12 May 2010.- 34p

<sup>1</sup> **Anastasiia Mudra**

Student of the Department of Management Technology

<sup>2</sup> **Olga Mezentseva**

Candidate of Economic Sciences, Associated prof. of the Department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **EXAMINATION OF THE DEPENDENCE BETWEEN CRIMINAL'S APPEARANCE AND HIS OFFENSE USING MACHINE LEARNING**

**Abstract.** The main idea of this paper is to present the relevance of researching of dependency between tendency to some actions, that in our case is type of offense, and appearance, in particular, face. Also describes the solution method – using convolutional neural network.

**Keywords:** convolutional neural network, Criminal's facial features, Offense

Not so far after the discovering of photography, some scientists began to notice similarities in the photographs of criminals taken after their arrest. If you believe their words, criminals are united by common facial features.

Modern scientists have already tried to prove this theory using the capabilities of artificial intelligence. Such experiments were held in America and China, but the results were not enough. This situation only proves the relevance of the topic and the fact that there is still a lot of work to be done in this direction.

A new view on the problem, that is describing in this article is trying to connect criminal's appearance with a type of his crime. Why exactly appearance? A growing number of studies have linked facial images to personality. It has been established that humans are able to perceive certain personality traits from each other's faces with some degree of accuracy. Studies focusing on the objective characteristics of human faces have found some associations between facial morphology and personality features. For instance, facial symmetry predicts extraversion. And, actually, we try to establish the connection between personality and the offence's type, because there is a wide range of different types of crime, that requires different traits from assailant and have different impacts on victims [1].

The databases of criminals are in every country, but they are confident. That is why the first information for model are put from Interpol from "Red Notices" category. Red Notices are issued for fugitives wanted either for prosecution or to serve a sentence. A Red Notice is a request to law enforcement worldwide to locate and provisionally arrest a person pending extradition, surrender, or similar legal action. Naturally, the data set needs more additions [2]. It is necessary to repeat the experiment with more people of different ages, gender, ethnic groups and with more information about them, like motives.

The most logical solution for solving this problem is to use the capabilities of a neural network. There is a need to process a lot of information, that, first of all, contains photos of criminals. It is the hardest part for neural network. When processing images,

there is a need to scan photos from different angles. That is why we decided to use convolutional neural network, which successful work with images is confirmed. Its architecture includes 2 main paradigms: local perception and shared weights.

Further, it technology can be used in systems of face recognition. Coupled with an automated biometric software application, this system is capable of identifying or verifying a person by comparing and analyzing patterns, shapes and proportions of their facial features and contours. Unlike a person [3], the computer vision algorithm has no subjective "baggage", emotions, prejudices regarding experience, race, religion, political beliefs, experience. It doesn't get tired, it doesn't need sleep or food. Thus, sometimes, it can help different specialists to improve their work.

**Conclusions.** Identifying patterns between a person and a crime can provide an opportunity to make a set of rules by which people and their tendencies can be classified and described. The results from the study will not try to call someone a criminal, rather it is an attempt to find out if there is an objective dependence of appearance and inclination to certain behavior. If the correlation will be confirmed, then further these data can be used in order to prevent the development of undesirable behavior. For example, it can be a classification of young people and further conversations with a psychotherapist, who, as a more educated specialist, can help solve some internal conflicts that a person does not admit, ignores or when a person does not have someone, who can help in overcoming this issues.

#### References:

1. Xiaolin Wu, Xi Zhang, Automated Inference on Criminality using Face Images, 2017. URL: <https://arxiv.org/pdf/1611.04135.pdf>.
2. Osin E., Novokshonov A., Shutilov K., Davydov D., Kachur A., Assessing the Big Five personality traits using real-life static facial images, 2020. URL: [https://www.researchgate.net/publication/341568689\\_Assessing\\_the\\_Big\\_Five\\_personality\\_traits\\_using\\_real-life\\_static\\_facial\\_images](https://www.researchgate.net/publication/341568689_Assessing_the_Big_Five_personality_traits_using_real-life_static_facial_images).
3. Nurul Azma Abdullah, Md. Jamri Saidi, Nurul Hidayah Ab Rahman, Chuah Chai Wen, Isredza Rahmi A. Hamid, Face recognition for criminal identification: An implementation of principal component analysis for face recognition, 2017. URL: <https://aip.scitation.org/doi/pdf/10.1063/1.5005335#:~:text=Face%20Recognition%20for%20Criminal%20Identification%20is%20a%20face%20recognition%20system,be%20removed%20from%20the%20image>.

<sup>1</sup> **Dmytro Orlovskyi**

PhD, Associate Professor

<sup>2</sup> **Andrii Kopp**

Senior Lecturer

<sup>1,2</sup> National Technical University "Kharkiv Polytechnic Institute"

## A BUSINESS INTELLIGENCE DASHBOARD DESIGN APPROACH TO IMPROVE DATA ANALYTICS AND DECISION MAKING

**Abstract.** This paper considers a problem of dashboard design, since usage of inappropriate visuals may mislead users and shift their focus to wrong things. Bar charts, line charts, and pie charts are considered as the most common visualization graphs. Proposed approach includes dataset preparation and analysis phases. While dataset preparation phase is mostly focused on star schema transformation into flat structures, dataset analysis phase proposes recommendations on which visualizations may be placed on a designed dashboard.

**Keywords:** Data Analytics, Business Intelligence, Dashboard, Star Schema.

In today's competitive market situation, it is extremely important for small business and large corporations to have permanent access to analytical reports regarding their business activities. Such access may be granted by modern data analytics and data visualization techniques covered in this section. However, it is a challenging problem to design information technologies in this field.

The dashboard design problem includes selection of visualizations, such as graphs and charts, which should be placed in a limited space. If choose inappropriate visualization charts that do not fit nature of data presented in datasets prepared for visualization, developed Business Intelligence (BI) dashboard applications may mislead business users and shift their focus and attention to unimportant or wrong things. Thus, dashboard design problem is extremely relevant nowadays when big data volumes processing and analysis is vital for making business decisions. Since the problem is not trivial and complex enough, it is required to propose a BI dashboard design approach.

Proposed approach is based on the relational algebra methods used to process Data Mart (DM) and Data Warehouse (DW) data structures in order to prepare datasets for analysis and suggest appropriate visualizations that should be placed on dashboards. The dashboard design process, which is underlying for a proposed approach, includes steps related to dataset preparation and dataset analysis that lead to recommendations on dashboard visualizations.

Proposed dashboard design process generalizes all the tasks required to prepare DM or DW measures and dimensions for visualization. At first, it is necessary to transform a star schema structure into a flat dataset. Since the star schema data warehouse model is one of the simplest but, on the other hand, is the most widely used data structure in BI domain, in this paper will be considered exactly this kind of storages [1]. This may be done using the SQL language join operators [2]. Then, it is required to prepare subsets of the general flat dataset. Prepared subsets will be used as

data sources for future visualizations (graphs and charts) that should be then placed on a dashboard. As well as the generic dataset, these subsets may be also prepared using the SQL language and its powerful selection and projection capabilities, and analytical functions [2]. After that, thresholds for pie charts and bar charts should be selected. When all the previous steps are completed, recommendations regarding the visualization charts and graphs, which should be used to display on a dashboard prepared data subsets, may be obtained.

With respect to the generated recommendations, the content of a designed dashboard may be created by data analysts or other stakeholders. The dashboard design process diagram is shown in Figure 1.

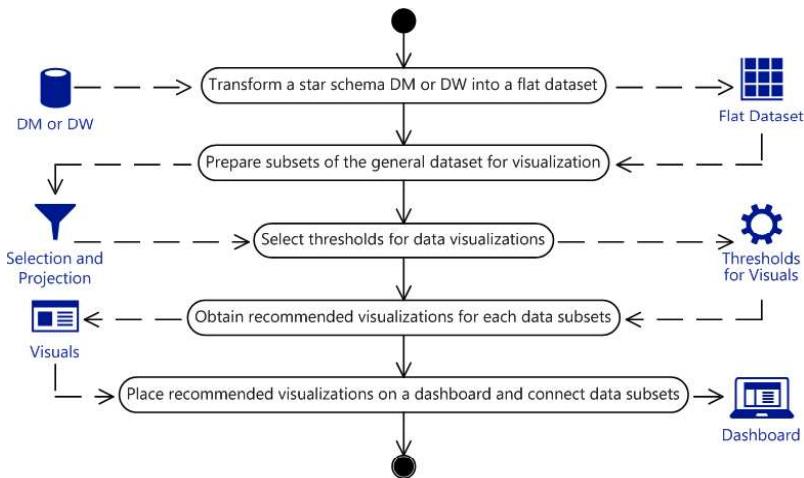


Figure 1 – Dashboard design process diagram

Described process allows users to choose appropriate visualization charts and graphs. Outlined approach is based on data mart or data warehouse transformation from the star schema, extremely popular and quite simple data structure, into the flat dataset that should be used to produce data subsets, which then may be used as data sources for visualizations, which are usually bar charts for comparisons, line charts for trends over time, pie charts for parts of a whole, and cards for scalar values [3]. Future research includes software implementation of the proposed approach, as well as the research of placement of visualizations on a dashboard's space.

#### References:

1. P. Bhatia, *Data Mining and Data Warehousing: Principles and Practical Techniques*, Cambridge University Press, 2019.
2. R. Ghilala, *Analytic SQL in SQL Server 2014/2016*, John Wiley & Sons, 2019.
3. R. Telg, T. A. Irani, *Agricultural Communications in Action: A Hands-On Approach*, Cengage Learning, 2011.

<sup>1</sup> **Vladyslava Rudenko**

Student of the Department of Management Technology

<sup>2</sup> **Olga Mezentseva**

Candidate of Economic Sciences, Associated prof. of the Department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## INFLUENCE ANALYSIS OF DIFFERENT MANAGEMENT METHODOLOGIES ON THE RESULT OF BIG DATA PROJECTS

**Abstract.** The paper considers various methodologies of project management and their interaction on BIG DATA projects. In essence, big data project management relies on general project management methodologies, but not all of them can be successfully applied to the full. Because this is a relatively new industry, big data may require something new or at least a combination of standard approaches.

**Keywords:** analysis of project management methodologies, BIG DATA, Data Science

Since Data Science, and Big Data in particular, is still evolving and lacks direct form, there can be no single answer to the question of which methodology works best in such projects.

In general, the following methodologies can be distinguished:

• CRISP-DM as a traditional approach to project management in the field of Data Science.

- Waterfall as a traditional approach.

- Scrum as an Agile approach.

- Kanban as an Agile approach.

This is not a complete list of methodologies used in big data projects. However, tools based on the above approaches will help identify key points where each approach can be useful to develop a model that will work for a specific Big Data project.

Cross Industry Standard Process for Data Mining (CRISP-DM) is a standard that describes general processes and approaches to data analytics used in industrial data-mining projects, regardless of the specific task and industry [1]. The CRISP-DM standard includes six iterative phases in data processing project management: understanding a business problem, understanding and retrieving data from different sources, preparing data, modeling data when building and evaluating a model, and actually evaluating that includes visualization and communication, and deployment and maintenance with final reports and project overview [2]. This methodology has a flexible cyclical nature and a focused approach, but it does not work for teams and does not cover communication issues at all. CRISP-DM can be called a sequence of works required to perform in Data Science projects [3]. This list of jobs can be kept in mind when managing Big Data projects, but it is too general, so other tools should be considered.

An approach such as Waterfall gives a clear and consistent picture of all the tasks that have been identified since the beginning of the project. The project or its individual phases are divided into smaller parts that depend on each other. In working with data, it is advisable to try the project in action as early as possible in order to quickly check and test hypotheses. If we stick to the waterfall style, we first need to make a complete and polished model, and then apply it in a work. Changes in this approach are not expected, however, in Big Data projects they may occur. This methodology does not work for the data processing. However, it may be useful to plan the use of some methodology tools, such as a Gantt chart. Thus, the main disadvantage of Waterfall is its lack of coverage of change management, which is not compatible with big data processing projects.

The main problem of Data Science projects in general is the misunderstanding between them and business goals [4]. Big Data projects add specific issues such as huge amounts of data and continuous changes. Agile methodologies can handle these problems. Scrum is one of the world's most common Agile approaches, in which large projects are divided into smaller phases, called sprints, and last from 1-2 weeks to 1-3 months. Each sprint has a fixed time frame and should achieve the results that have been set at the meetings. Scrum is largely focused on customer feedback. It is adaptive and flexible, with a high degree of autonomy, which in terms of Data Science allows you to optimize predictability. However, unlike Scrum in Software Engineering, where there is a constant increment to demonstrate results, Big Data has no feedback material. Sprint requirements are changing. In working with data, there is an iterative between the phases of data preparation and modeling, where everything can change. For example, a new hypothesis appeared, data were prepared, the hypothesis was tested, and everything planned for the end of the sprint is no longer of value. Everything needs to be reworked because the concept has changed. If you take into account this feature and slightly change the classic approach of Scrum, it can work well on Big Data projects. The continuous flow of a huge amount of information from the Internet, corporate systems or devices falls under the definition of Big Data at high speeds of download or accumulation [5]. Response speed and high frequency of result presentation is one of the 12 basic principles of Agile Manifesto.

Consider another Agile approach - Kanban. This methodology uses the board as a project and the cards as a task. The traditional kanban board includes three columns – To Do, In progress and Done [6]. For data processing projects, you can add additional columns: data preparation, development, coding, testing, and so on. It is necessary to improve the standard approach until it is suitable for working directly with the Big Data project. To use kanban in data analysis, some of the separate phases can also be combined - for example, modeling and data preparation can become one phase, because working with data and experiments occur at the same time. The approach places more emphasis on work in progress without reference to dates and roles, the work in it is highly visualized [7]. Disadvantages of the methodology - the lack of emphasis on dates and deadlines - can, on the contrary, be called advantages for big data processing projects.

It is difficult to find one methodology that would work well during the project life cycle in the Big Data field. Big data processing projects can combine several tools

of different methodologies in their management, and this is completely natural. It is worth exploring which tools influence the success of Big Data projects and synthesize them into a single approach that can have a positive impact on project results.

**References:**

1. Shearer C, The CRISP-DM model: the new blueprint for data mining, J Data Warehousing (2000); 5:13—22.
2. V. Gogunskii, O. Kolesnikov, G. Oborska, , ... S. Harelkik, , D. Lukianov. Representation of project systems using the Markov chain, in: Eastern-European Journal of Enterprise Technologies, volume: 2(3-86), 2017, pp. 60-65. doi:10.15587/1729-4061.2017.97883
3. Óscar Marbán, Gonzalo Mariscal and Javier Segovia (2009); A Data Mining & Knowledge Discovery Process Model. In Data Mining and Knowledge Discovery in Real Life Applications, Book edited by: Julio Ponce and Adem Karahoca, ISBN 978-3-902613-53-0, pp. 438-453, February 2009, I-Tech, Vienna, Austria.
4. V. Morozov, O. Mezentseva, M. Proskurin. Application of game theory for decisions making on the development of it products, in: Advances in Intelligent Systems and Computing, volume: 1, 2021. doi: 10.1007/978-3-030-54215-3\_24
5. K. Senthuran, R. Sithamparanathan, S. Evans. Markov Decision Process-Based Opportunistic Spectral Access, in: IEEE Wireless Communications Letters, volume: 5 , issue: 5, 2019, pp. 544-547. doi: 10.1109/LWC.2016.2600576
6. Georg N. Krieg. Kanban-Controlled Manufacturing Systems. — Germany: Springer-Verlag Berlin Heidelberg, 2005. — 236 c. — ISBN 3-540-22999-X.
7. D. Anderson: Agile Management for Software Engineering - Applying the Theory of Constraints for Business Results. Prentice Hall, 2004, ISBN 0-13-142460-2

**<sup>1</sup> Tetiana Shelest**

Assistant Professor, Technology Management Department, Faculty of information Technology

**<sup>2</sup> Veronika Yeremieieva**

Master of degree, specialty Project Management, Faculty of information Technology

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **ANALYSIS OF THE POSSIBILITY OF USING VR TECHNOLOGIES IN ENVIRONMENTAL AWARENESS PROJECTS**

Reorganizing the projects connected with eco awareness is important in Ukraine because the waste management system is imperfect and has potential for development.

The main goal of this thesis is to collect data that will confirm the reasons that make the system imperfect, to analyze the ways of solving this problem and practices that will allow Ukrainians preserve the environment.

According to USA Today research, Ukraine ranks 9th out of 10 countries in the world that produce the largest amount of waste in relation to the population. According to the statistics of 2019, Ukraine produces 10.6 metric tons of waste per capita per year, and the total amount is approximately 474 106 065 metric tons. From this waste, only 3.2% are recycling [1].

Despite the fact that Ukraine produces a lot of waste and do recycle only small part of it, we do also import sorted waste from Europe. In 2018 we have imported around 100 thousand tons of plastic for 40 million dollars, glass waste for 11 million dollars and paper for 80 millions [2].

As a conclusion, Ukraine imports sorted waste from Europe because can not provide good quality of sorting in our country. The key problem in this case is that people do not know how to do it in a right way.

Analyzing the data of 2018 year from questionnaire "Environmentally conscious citizen - the key to successful implementation of the Association Agreement" was informative for our study. Let's pay attention on two indicators: ways to solve environmental problems and practices that preserve the environment [3].

Table 1 – Ways to solve eco problems [3]

<b>Ways to solve problems</b>	<b>Positive answers, %</b>
Increase information on environmental issues	32.4%
Ensuring better application of current environmental legislation	27.3%
High penalties for violations of environmental legislation	44.7%
Implementation or dissemination of training programs	26.8%

Basing on this statistics (tab. 1), we can ensure that Ukrainians do need more information about ways of sorting. Moreover, introducing trainings, workshops and lectures to increase eco awareness can build new eco habits.

Here (tab. 2) we can see that more than 60% think that sorting waste is the key to build ecofriendly environment and around 40% are willing not to buy plastic for one use.

Table 2 – Practices that preserve the environment [3]

<b>Practices that preserve the environment</b>	<b>Positive answers, %</b>
Choosing more eco friendly transport	34.7%
Avoid buying disposable plastic products	39.3%
Sort the most part of waste	60.2%
Reduction of energy consumption	31.3%

This analysis showed that people are ready to change their environment, learn the sorting rules and rebuild their habits. But in Ukraine we do not have any learning content about eco awareness that would be interesting for kids as well as for adults.

Educational systems are now being radically restructured and raising the awareness of citizens can be regulated through multi-touch interactive VR technologies.

It is important that we can improve the experience of participants through manipulating and controlling VR environment. Through thematic educational modules it is possible to improve the experience of learners and increase their sensitivity to current environmental challenges [4].

We can install stationary interactive racks near supermarkets and sorting tanks. Each rack will have several VR modules, between which the user can switch and find out the most interesting information. Modules can include: an overview of things that can be made from recycled materials, a module with information about waste cycle in Ukraine, sorting rules and tips on how to minimize the waste amount.

With the help of interactive communication, the information is perceived much better. Therefore, the use of VR environment in various projects to increase environmental awareness, will introduce a completely new and experimental learning. Taken together, this will give the people an impetus to increase their eco-consciousness and learn the rules of waste sorting.

#### **References:**

1. H. Byrnes, T. Frohich, USA Today: Canada produces the most waste in the world. The US rank. URL: <https://www.usatoday.com/story/money/2019/07/12/canada-united-states-worlds-biggest-producers-of-waste/39534923/>
2. Will the planet turn into Plastic Porridge? URL: <https://re-solutions.com.ua/ru/rus-plastyk-nastupaet-zachem-ukrayna-pokupaet-chuzhye-otody/>
3. Environmental protection and citizens of Ukraine. Research of practices , values and judgements. May 2018, URL: <https://www.rac.org.ua/uploads/content/481/files/envportraitpollreport2018.pdf>
4. T. Mikropoulos, Virtual realities in environmental education., URL: [https://www.researchgate.net/publication/226299007\\_Virtual\\_realities\\_in\\_environmental\\_education\\_The\\_project\\_LAKE](https://www.researchgate.net/publication/226299007_Virtual_realities_in_environmental_education_The_project_LAKE)

<sup>1</sup> Serhiy Shtovba

Doctor of science, professor

<sup>2</sup> Mykola Petrychko

Master, PhD student

<sup>1</sup> Vasyl' Stus Donetsk National University

<sup>2</sup> Vinnytsia National Technical University

## AN INFORMETRIC ASSESSMENT OF VARIOUS RESEARCH FIELDS INTERACTIONS ON BASE OF CATEGORIZED PAPERS IN DIMENSIONS

**Abstract.** We calculated the level of interactions between all the pairs of the research groups and between all the pairs of the research divisions for 4 five-year periods. Paired interaction for research divisions shows that every consecutive five-year period has decreased irregularity of its distribution of interaction. Cumulative interaction shows that all research divisions tend to have their stickiness decreased over time. Paired interaction for research groups showed the same results but difference between two consecutive periods is greater by a larger factor.

**Keywords:** informetrics, research group, research division, interactions, interdisciplinary, Dimensions, ANZSRC, Jaccard index, stickiness index, distribution, categorization.

Interdisciplinarity is a fashion direction in modern education and science. Interdisciplinarity is impossible without interactions between research from various fields. For quantitative assessment of level of interdisciplinarity, of level different research fields interactions many approaches are proposed in recent years. Currently the most widely used approaches to quantitative measuring of interdisciplinarity use bibliometric. They take into account co-authorships, collaborations, references, citations and co-citations. There are a few researches related to measuring interdisciplinarity using bibliometric [1-5]. In this research we used the same approach to assessing interdisciplinarity as in [5] - Jaccard index. We measured interdisciplinarity for Australian and New Zealand Standard Research Classification (ANZSRC) system. Our assessing is based on Dimensions' categorized papers where classes are research divisions and groups from ANZSRC. There are 22 research divisions with 157 groups. Now, Dimensions indexes over 110M research papers. Each paper is assigned to one or several research divisions and groups. Such categorization is carried out by Dimension itself by special software based on machine learning guided by topic experts. The title and abstract of the paper are source data for this categorization.

We analyzed the interaction of research divisions and research groups for 4 five-year periods. We analyzed paired and cumulative interaction for both research divisions and research groups. Only research groups of different research divisions were considered. Paired interaction allowed finding pairs that interact the most and analyze the behavior of different interactions over time.

Using paired interaction, we analyzed leaders and the most changeable research

divisions and groups in terms of interaction. Paired interaction for research divisions shows that every consecutive five-year period has decreased irregularity of its distribution of interaction. This also confirms the Gini index and the number of pairs with non-zero interaction. Cumulative interaction shows that all research divisions tend to have their stickiness decreased over time. The same effect shows Gini index for the stickiness index that decreases from 0.197 in the first period to 0.188 in the last period.

Paired interaction for research groups showed the same results but difference between two consecutive periods is greater by a larger factor. Cumulative interaction for research groups showed that with time research groups tend to have a more equally distributed stickiness index. It is also observed from the number of research groups that increased their stickiness index from zero to greater than zero and Gini index.

On base of interacting assessments, we identified five research groups that have been assigned suspiciously to divisions in ANZSRC. They are as follows: 0105 Mathematical Physics, 0602 Ecology, 0909 Geomatic Engineering, 1502 Banking, Finance and Investment and 2001 Communication and Media Studies. All the mentioned research groups have more similar research groups in concurrent divisions with statistically significant level. Those five research groups also have strong semantic ties with concurrent divisions, this can be an argument for adapting of ANZSRC.

#### **References:**

1. Wagner, S., Roessner, D., Kamau, B., Klein, J., Boyack, K., Keyton, J., Rafols, I., Borner, K.: Approaches to Understanding and Measuring Interdisciplinary Scientific Research (IDR): A Review of the Literature. *Journal of Informetrics*. 5, 14-26 (2011). doi: 10.1016/j.joi.2010.06.004.
2. Porter, A.L., Cohen, A.S., Roessner, J.D., Perreault M.: Measuring researcher interdisciplinarity. *Scientometrics*. 72, 117–147 (2007). doi: 10.1007/s11192-007-1700-5.
3. Noorden, V.R.: Interdisciplinary research by the numbers. *Nature*. 525, 306-307 (2015). doi: 10.1038/525306a.
4. Karlovcec, M., Mladenic, D.: Interdisciplinarity of scientific fields and its evolution based on graph of project collaboration and co-authoring. *Scientometrics*. 102, 433-454 (2015). doi: 10.1007/s11192-014-1355-y.
5. Shtovba S., Petrychko M. Jaccard Index-Based Assessing the Similarity of Research Fields in Dimensions // CEUR Workshop Proceedings, Vol. 2533 “Proc. of the First International Workshop on Digital Content & Smart Multimedia”. – 2019. – P. 117-128.

**<sup>1</sup>Taborovskyi Andrii**

Master's degree student Department of technology management

**<sup>2</sup>Kolesnikova Kateryna**

Doctor of technical sciences, Professor of the Department of Technologies Management

**<sup>3</sup>Khlevnyi Andrii**

PhD in Engineering Science, Assistant Professor of Department of Technologies Management.

*<sup>1-3</sup>Taras Shevchenko National University of Kyiv*

## **THE IMPACT OF AUTOMATED AND ROBOTIC WAREHOUSES ON THE SCOPE OF SUPPLY CHAIN PROCESS**

**Abstract.** The essence and purpose of systems of automation and robotics of warehouses are considered, their advantages are defined, the offered software-analytical decision is described in the form of expert system.

**Keywords:** automation, robotics, supply chain, warehousing, warehouse capacity.

Warehousing is an important and integral part of every business. Its task is to save stocks of raw materials and finished products. It plays an important role in the movement of tangible assets, raw materials, materials, fuel, tools, equipment, spare parts, clothing, and other products.

Warehouse - a room or a set of rooms intended for storage of materials. Supply Chain Management combines the management of supply and demand within and between companies. A recent network of companies that cooperate with the aim of offering products and services is called an expanded enterprise. [2]

Automation and robotization of warehousing allows to introduce the technical, logistical and analytical measures that increase the level of productivity of facilities in the warehouse and reduce the impact of the human factor on the activities of the warehouse. The number of resources spent on the technical implementation of automated and robotic components of the warehouse is significantly less than the number of resources required for the maintaining of non-automated and non-robotic warehouse.

An urgent task is the choice of bots-carriers for their use in the department of planning the placement of boxes with goods for the automated warehouse. A solution in the form of an expert system is proposed. Its task is to recommend to the user on decision-making when choosing a manufacturer of carrier bots for automated warehouse.

During the development of the expert system, knowledge was extracted by analyzing the characteristics of carrier bots by the method of main components, cluster analysis and building a decision tree.

According to the results of the principal components method, three main components Comp1, Comp2 and Comp3 have been identified.

The first main component is determined by more than 66.3% of the following indicators: power, charge duration, charging speed.

The second main component is determined by more than 78.6% of the following indicators: the number of charging cycles, charging duration, price. The third main component is determined by more than 79% of the following indicators: power, charge duration, price.

The results of component and cluster analyzes were analyzed when constructing decision trees using the See5 / C 5.0 system. The result of the See5 system was expressed in the form of decision trees and 42 if-then rules that provide answers to the manufacturer's belonging to the class according to its characteristics.

The results obtained during the construction of decision trees completely coincided with the results of component and cluster analyzes.

The created expert system can become a prototype for a real decision-making system for an automated warehouse.

#### **References:**

1. Belinskyi P. I. Management of manufacturing and operations : P.I. Belinskyi, Y. Fedcovich national universist of Chernivtsi. Kyiv, 2005 – 380 p.
2. Donald Bowersox Integrated Supply Chain : Bowersox D, Olymp-business, 2017 – 500 p.
3. Dibsko V.V. Warehouse management in Supply Chain: V.V. Dibsko, Alfa-press, 2014 – 466 p.

**<sup>1</sup>Oksana Tereshchenkova**

Ph.D., associate professor

**<sup>2</sup>Kostyantyn Kondrashov**

graduate student

*<sup>1,2</sup>Kherson State Maritime Academy*

## **INFORMATIONAL EXPERT SYSTEM FOR MINIMIZING THE TIME FOR SEARCHING OF FAILURES OF SHIP ELECTRICAL EQUIPMENT**

**Abstract.** This article is devoted to obtaining and comprehensively studying methods and models for fault trees and decision trees construction for identifying a defect in specific diagnostic objects and their structural units, as well as methods used to defects finding. The basic subjective and objective conditions that affect the time spent by maintenance personnel on the readapting to work of the failed ship system are systematized. The article substantiates the need of the transition from existing paper documentation to electronic maintenance documentation using the expert system.

**Keywords:** object of diagnostics, structural units, complex technical system, decision maker, expert system, alarm monitoring system, decision support system.

Each ship system can be represented as a complex of structural units interconnected and interacting with each other. In turn, each structural unit can be decomposed into many simple elements interconnected and interacting with each other.

Accordingly, the more structural units create a system, the more complex the system becomes and the more difficult it is to identify a malfunction in this system.

Thus, conventionally, all ship systems can be divided into 5 levels, according to the criterion of complexity when troubleshooting in this system.

Analysis of the failure diagnostic tools used by the operator in real navigation conditions to find and eliminate the causes of malfunction of shipboard automated systems and mechanisms is an actual problem. Quick search and elimination of a defect affect the level of ship's safety.

The troubleshooting process is the most difficult at electrical equipment repairing, as modern automated systems are a complex interconnected network of electrical and electronic circuits. The task of faulty element finding is finding of the sequence of checks when a minimum of time is spent on defect searching [1-3].

To be able to show the possible amount of time spent searching and repairing, an experiment was conducted on one of the container ships of the shipping company Mediterranean Shipping Company (MSC) m/v MSC "Brunella" [4, 5].

We used the archive logbook of the Kongsberg K-Chief 600 alarm monitoring system, the container ship MSC "Brunella", it was built in 2015. The total number of parameters controlled by the AMS is 3410 units [6].

Data of ship failures recorded in the ship's logbook for six months were conditionally divided by the level of complexity of the systems in which they occurred and are summarized in the table shown in Tab.1.

The purpose of the experiment was to calculate the average number of possible

causes of these failures, as well as the number of possible ways to eliminate them and the time taken to eliminate them.

Based on the obtained data, we construct a variational series of observations for the number of malfunctions, occurred within six months on the ship, and the number of their possible causes.

Let's find the relative frequency of events for 6 months  $W_i$ , for each level of complexity of the systems.

**Table 3**

The number of malfunctions recorded by the AMS system for six months				
Simple elements	Simple systems	Medium difficulty systems	Complex systems	Very complex systems
<b>Total – 2</b>	<b>Total – 126</b>	<b>Total – 198</b>	<b>Total – 168</b>	<b>Total – 12</b>

$$W_i = \frac{N_i}{n} \quad (1)$$

where  $N_i$  is the number of failures at a given interval;  $n$  is the total number of malfunctions within six months.

Let's substitute the numerical data, we obtain:  $W_1 = 0.046$ ;  $W_2 = 0.241$ ;  $W_3 = 0.379$ ;  $W_4 = 0.311$ ;  $W_5 = 0.023$ .

Find the numerical parameters: average value and variance.

Sample average:

$$\bar{X}_B = \frac{1}{n} \sum_{i=1}^m * N_i * X_i \approx 15 \quad (2)$$

Dispersion of discrete random variance:

$$D_B = \frac{1}{n} \sum_{i=1}^n * (X_i - \bar{X}_B)^2 = 28,88 \quad (3)$$

Then, the standard deviation (standard error):

$$\sigma_s = \sqrt{28,88} \approx 5 \quad (4)$$

Thus, the average number of possible causes of an accidental failure recorded by the AMS system  $\bar{X}_B = 15$  with the standard deviation of  $\sigma_B = 5$ .

One-sigma interval (confidence probability is 67%) for the given random variable is from 10 to 20 possible reasons.

This means that very often even experienced electricians will spend quite a lot of time guessing about the causes of the breakdowns and how to fix it.

The increasing of SAS effectiveness can be achieved in two methods.

The first method is the highly qualified personnel training. In order to quickly search and eliminate the OOD defect, the decision-maker must have the extensive

knowledge, experience and a wide range of personal qualities. In addition, he should be able to adapt to objective reasons that make troubleshooting difficult.

The problem is that availability of these qualities in one decision-maker (it is extremely unlikely), the process of defect searching can be taken place rather long. It is due to the information content received by the operator, in each case, is often excessive. The same OOD is represented by different models, and the information content about its elements and connections, as well as various features significantly exceeds the level necessary for defect searching.

So, it is impossible to draw up quickly a clear pattern of action at defect searching. The decision maker is always forced to keep in mind all the methods and algorithms of checks, to understand when to replace one method by another. In the process of searching of the same defect, he should constantly think about what to use at a given time. In this case, the factor of the human psyche works as limitedness to process a large amount of information (from 5 to 9) per unit of time.

As a result, even a competent decision maker falls into the mandatory time frame; it increases the troubleshooting process.

The second method is increasing the reliability of OOD by strengthening of the control over the operability of the main OOD nodes and the connections between them.

The problem here is that the structural, circuit and technological capabilities for improving the reliability of ship systems are limited, and, in practice, exhausted. Moreover, increasing of the OOD reliability due to the structural complication of diagnostic systems involves growing the number of measurements with dimension enhancement of the diagnosed circuit. It requires an increase of the control points in OOD; it inevitably raises a new problem related to the diagnostic systems reliability. In addition, their false positives can trigger a chain of incorrect operator actions leading to an accident or disaster.

As a result, even complex diagnostic systems help to reduce the number of failures of electrical equipment by timely informing the operator about violations in the operation of a particular mechanism, but, unfortunately, it doesn't contribute to the quick searching and elimination of a defect, in the case of ship system failure. And it requires the high qualification of the service personnel and a longer duration of the checks. In the conditions of autonomous navigation and with low qualification of the staff it can lead to undesirable consequences.

All described above clearly demonstrates the urgent need for the implementation of special information expert systems, which allow, even with low qualifications of the service personnel and low efficiency of control of OOD, to quickly search for defects in a failed ship system.

The proposed system will be built on the basis of knowledge, which includes the experience of experts in repair and troubleshooting. The knowledge base is formed on the basis of expert evaluation (experts are electricians with experience of at least 5 years, as well as superintendents of crewing firms with the same experience).

The system uses the approach that implements the task of separating of information stored in a common database and directly in the knowledge base (a set of decision tables).

To implement this approach, linking variables (link tables) are used.

By the use of these communication tables, a variable from the knowledge base is connected with the data stored in a common database of equipment and ready-made troubleshooting algorithms.

The knowledge base includes the knowledge and assessments of experts in failures, as well as databases with structural diagrams, principled schemes of elements and components, as well as troubleshooting algorithms.

Filling occurs from ship's logbooks. The number of malfunctions detected by the AMS system for vessels of the type container ship is recorded. For entry into the database, faults are ranked by their level of complexity. All entries are transmitted to the crewing company by the superintendent. The database is filled on the basis of the data of the logs collected from all ships of the crewing during the entire period of ship running.

The final product is software that provides the operator with complete, but not redundant information on the necessary malfunction, as well as a clear sequence of actions for its quick elimination.

The decision-making operation in the ES of a ship electrical engineer is: the registered error of the AMS is entered into the system window. The user receives all the necessary documentation of the unit that gave the error signal, as well as a set of strategies for troubleshooting. The variability of possible problems increases with the complexity of the mechanism. It becomes necessary to choose the most effective strategy to reduce the time of elimination. There is a table of opinions of experts who had the similar problems.

#### References:

1. B. Palyukh, T. Kakatunova, O. Baguzova, Intelligent decision support system for managing complex objects using dynamic fuzzy cognitive maps, Software Products and Systems, 2013.
2. C. Moreno, E. Espejo, A performance evaluation of three inference engines as expert systems for failure mode identification in shafts, Engineering Failure Analysis, 2015.
3. E. Liberado, Novel expert system for defining power quality compensators, Expert Systems with Applications, 2015.
4. KONSBERG. Standard K-Chief 600 Alarm and Monitoring System / 354760 / Rev.D March 2013 © Kongsberg Maritime AS
5. KONSBERG. Kongsberg K-Chief 500/600 Marine Automation System Installation Manual / 311956 / F March 2013 © Kongsberg Maritime AS.
6. Y. Krainyk, Y. Davydenko and V. Tomas, Configurable Control Node for Wireless Sensor Network, in: Proceedings of the 3rd International Conference on Advanced Information and Communications Technologies (AICT), Lviv, Ukraine, 2019, pp. 258–262. doi: 10.1109/AICT.2019.8847732

**Anastasiia Vavilenkova**

Doctor of technical science, professor, docent

*National Aviation University I*

## RAGULARITY OF CONTEXT UNITS IDENTIFICATION IN ELECTRONIC TEXT DOCUMENTS

In this article had been analyzed actual software services, that can build relation's tree and make syntactical analysis. Each of them transforms primary text into the data structure with special features. The author proposed to use components of logic and linguistic models for automatic generation of grammar collocations. Also author suggested the rules for context units identification for complex sentences of natural language.

All early efforts to extract knowledge from the textual information by difference scientists leads in grammars by Homskiy and transformation grammars [1–2]. Grammars of regularity do not come up with collocations like analysis ones. However, almost all linguistic theories describe linear sequence of the sentence units by mean of hierarchic structure of gramma components [3]. Traditional ways of analysis by key parameters and standard answers are not possible to analyze natural language text at all and are not helpful for context analysis [4–5].

Today the main aim of natural language researches is automatic creation of context data structures for formalization of logical links by mean of particular algebraic construction. From the other hand, these researches give practical value for automatic analysis and synthesis of natural language texts by computer technologies.

Despite almost a century of research in artificial intelligence, context units identification still can not be realized in correct form for complex sentences.

One of the most essential thing for all systems that works with natural languages must be using the process of grammatical analysis [6].

There is a full correspondence between grammar structure and logic form of natural language sentence of natural language sentence [7]. Considering grammar organization of the sentences, we have such graduation of sentences' members [8]:

- subject of the sentence – subject  $x$ ;
- predicate of the sentence – relation  $p$ ;
- object of the sentence – object  $y$  or subject-matter of relation  $z$ ;
- definition – characteristic of subject  $g$ , characteristic of object  $q$  or characteristic of subject-matter of relation  $r$ ;
- circumstance – characteristic of relation  $h$ .

A set of words connected between each other by logic links, will be lettered  $sp_j, j = \overline{1, m}$ , where  $m$  – amount of the collocations in the sentence.

According to the Ukrainian and English language rules, collocations can be formed between those members of sentence [9–10]:

- “definition – subject” –  $sp_j = g \cup x$ ;

- “predicate – object” —  $sp_j = p \cup y$  ;
- “definition – object” —  $sp_j = q \cup y$  ;
- “object – object” —  $sp_j = y \cup z$  ;
- “object – object” —  $sp_j = r \cup z$  ;
- “circumstance – predicate” —  $sp_j = h \cup p$  .

The author formulated special rules for identification context units according to the rules for creating different collocations in flexional natural languages, examples of what were depicted above. It was developed 32 rules with additions for punctuation symbols in complex sentences and for considering homogeneous parts of the sentence. Some of these rules are represented below.

1. If the first word is adjective, numeral, pronoun or participle and the part of speech for second word is noun, their characteristics of case, number and genus are similar, the words are made collocation. For example, “*mathematical modelling*”, “*computer modelling*”, “*three pets*”, “*her name*”, “*designed room*”.

$$\begin{aligned} &\text{if } ((cm(S_i) = 2) \text{ and } (cm(S_{i+1}) = 1)) \text{ and } (g(S_i) = g(S_{i+1})) \\ &\text{and } (n(S_i) = n(S_{i+1})) \text{ and } (k2(S_i) = k2(S_{i+1})) \\ &\text{then } (S_j = S_i \cup S_{i+1}) \end{aligned} .$$

2. If the first word is noun and second word is noun of personal name too, their characteristics of case and number are similar, the words are made collocation. For example, “*Dnipro river*”.

$$\begin{aligned} &\text{if } ((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 1)) \text{ and } (g(S_i) = g(S_{i+1})) \\ &\text{and } (n(S_i) = n(S_{i+1})) \text{ then } (S_j = S_i \cup S_{i+1}) \end{aligned} .$$

3. If the first word is verb and the second word is noun in genitive case, the words are made collocation. For instance, “*read book*”.

$$\begin{aligned} &\text{if } ((cm(S_i) = 5) \text{ and } (cm(S_{i+1}) = 1)) \text{ and } \\ &((g(S_{i+1}) = 2) \vee (g(S_{i+1}) = 4)) \vee (g(S_{i+1}) \neq 1) \\ &\text{then } (S_j = S_i \cup S_{i+1}) \end{aligned} .$$

4. If the first word is verb, second word is preposition and third word is noun in subjective case the words first and third are made collocation. For example, “*created for children*”.

$$\begin{aligned} &\text{if } ((cm(S_i) = 5) \text{ and } (cm(S_{i+1}) = 9) \text{ and } (cm(S_{i+2}) = 1)) \text{ and } \\ &(g(S_{i+2}) \neq 1) \text{ then } (S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned} .$$

5. If the first word is verb and second word is pronoun not in subjective case, the words are made collocation. For instance, “*integrated scheme*”.

$$\begin{aligned} &\text{if } ((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 1)) \text{ and } (g(S_i) = g(S_{i+1})) \\ &\text{and } (n(S_i) = n(S_{i+1})) \text{ then } (S_j = S_i \cup S_{i+1}) \end{aligned} .$$

Using developed rules and according finding regularity it had been possible to create a system for context units identification. For the complex Ukrainian language sentence the system creates such context units (Figure 1).

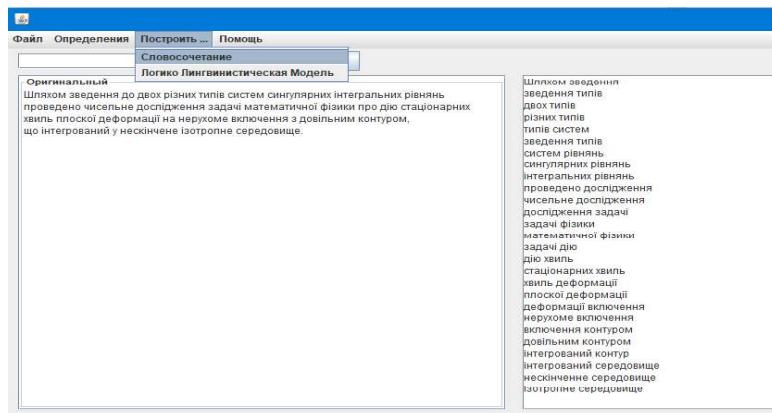


Figure 1 – Results of automatic identification of context units

It had been demonstrated the outcome of using these rules. It is a program, that extract all collocations from different types of natural language sentences.

#### References:

1. V. Evans, Lexical concepts, cognitive models and meaning-construction. in: Cognitive Linguistics, Edinburg university press Publ. Vol. 17, 2006, pp. 73-107.
2. F. H. George The foundations of cybernetics, Gordon and breach science publishers Ltd., U.K., 1977.
3. W. Che, Y. Zhang, Deep learning in lexical analysis and parsing, Springer Nature Singapure Pte Ltd., ch.4 in: Deep learning in Natural Language Processing, 2018, [http://doi.org/10.1007/978-981-10-5209-5\\_4](http://doi.org/10.1007/978-981-10-5209-5_4).
4. Y. Zhang Discriminative syntax-based word ordering for text generation, in: Computational linguistics, Vol.41, 2015, pp. 503-538.
5. D. Chen, C. D. Manning, A fast and accurate dependency parser using neural networks, in: proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp.740–750.
6. A. Bashmakov, I. Bashmakov Intellectual information technologies, M., MGTU Publish, 2005.
7. A. Vavilenkova, Basic principles of the synthesis of logical-linguistic models, in: Cybernetics and systems analysis, Vol. 51(5), 2015, pp. 826-834, <http://doi.org/10.1007/s10559-015-9776-z>.
8. V. Shyrokov, System semantics of explanatory dictionaries, in: Cognitive Studies, Vol. 12, 2015, pp. 95-106.
9. A. Vavilenkova, Analysis and synthesis of logic and linguistic models, TOV “SIK GROUP Ukraine”, 2017.
10. A. Broder, Identifying and Filtering Near-Duplicate Documents, in: proceedings of the 11<sup>th</sup> annual symposium, Canada, Montreal: Springer, 2000. pp. 1–10.

<sup>1</sup> **Hlib Yefremov**

Student of the Department of Technology Management

<sup>2</sup> **Kateryna Kolesnikova**

Doctor of Technical Sciences, Professor of the Department of Technologies  
Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## OPINION MINING METHODOLOGY IN MARKET RESEARCH

**Abstract.** The objective of this paper is to present the relevance of using opinion mining methods and tools in market research. During the research, we've established that opinion mining methods proved to be of significant capability to provide insights and conclusions needed for developing market strategy.

**Keywords:** Opinion Mining, Data Mining, Sentiment Analysis, Marketing.

Marketology is moving at a great pace to becoming a lot more significant and valuable than it was a couple of years ago. With the number of gadgets and digital services common people are using gradually growing, the amount of data trail they leave in cyberspace is growing almost exponentially, left basically as dead weight, unused. Analyzing these data trails is, most certainly, crucial for obtaining and/or maintaining a good grasp of your clients' preferences and opinions. Opinion mining methods is a great way to find out the latter.

Textual information in the world can be broadly classified into two main categories, facts and opinions. Facts are objective statements about entities and events in the world. Opinions are subjective statements that reflect people's sentiments or perceptions about the entities and events. Much of the existing research on text information processing has been (almost exclusively) focused on mining and retrieval of factual information, e.g., information retrieval, Web search, and many other text mining and natural language processing tasks. Little work has been done on the processing of opinions until only recently. Yet, opinions are so important that whenever one needs to make a decision one wants to hear others' opinions. This is not only true for individuals but also true for organizations[1].

One of the main reasons for the lack of study on opinions is that there was little opinionated text before the World Wide Web. Before the Web, when an individual needs to make a decision, he/she typically asks for opinions from friends and families. When an organization needs to find opinions of the general public about its products and services, it conducts surveys and focused groups. With the Web, especially with the explosive growth of the user generated content on the Web, the world has changed. One can post reviews of products at merchant sites and express views on almost anything in

Internet forums, discussion groups, and blogs, which are collectively called the user generated content. Now if one wants to buy a product, it is no longer necessary to ask one's friends and families because there are plentiful of product reviews on the Web which give the opinions of the existing users of the product. For a company, it may no longer need to conduct surveys, to organize focused groups or to employ external consultants in order to find consumer opinions or sentiments about its products and those of its competitors. Finding opinion sources and monitoring them on the Web, however, can still be a formidable task because a large number of diverse sources exist on the Web and each source also contains a huge volume of information. In many cases, opinions are hidden in long forum posts and blogs. It is very difficult for a human reader to find relevant sources, extract pertinent sentences, read them, summarize them and organize them into usable forms. An automated opinion mining and summarization system is thus needed[2]. Opinion mining, also known as sentiment analysis, grows out of this need.

Given a set of evaluative text documents  $\mathbf{D}$  that contain opinions (or sentiments) about an object, opinion mining aims to extract attributes and components of the object that have been commented on in each document  $d \in \mathbf{D}$  and to determine whether the comments are positive, negative or neutral[3,4].

Opinion mining and summarization process contain three main steps, first is Opinion Retrieval, Opinion Classification and Opinion Summarization(Figure 1).

Opinion Retrieval is the process of gathering review text from review websites. Different review websites involve reviews for products, movies, hotels and news. Information retrieval techniques like web crawler can be applied to accumulate the review text data from many sources and store them in database. This step includes retrieval of reviews, micro blogs, and user's comments.

Next basic step in opinion mining is classification of review text. Given a review document  $\mathbf{D} = \{d_1, \dots, d_l\}$  and a categories set  $\mathbf{C} = \{\text{positive}, \text{negative}\}$ , sentiment classification is to classify each  $d_i$  in  $\mathbf{D}$ , with a tag expressed in  $\mathbf{C}$ . The method involves classifying review text into two forms namely positive and negative.

Summarization of opinion is a main part in opinion mining process. Summary of reviews provided should be established on lineaments or subtopics that are mentioned in reviews/blogs/comments etc.

**Conclusions.** We established that opinion mining is an emerging methodology of data mining applied to summary the knowledge from large volume of data, left by people in the Web, and it's a promising new domain which can greatly increase the quality of the potential market research for your needs. For example, it is critical for a product manufacturer to know how consumers perceive its products and those of its competitors. This information is not only useful for marketing and product benchmarking but also useful for product design and product developments.

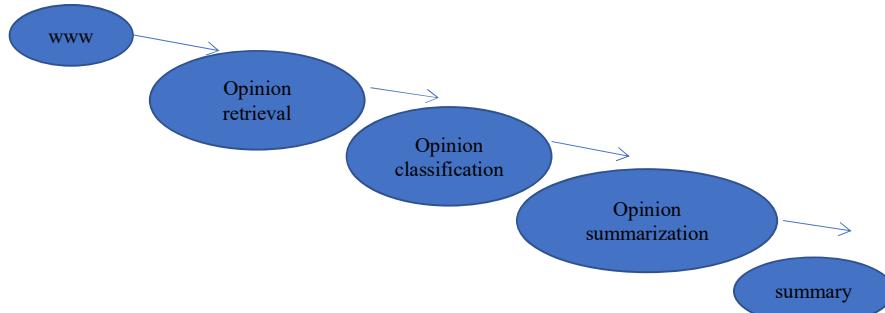


Figure 1. Architecture of Opinion mining.

#### References:

1. Carenini, G., Ng, R. and Zwart, E. "Extracting Knowledge from Evaluative Text". Proceedings of the Third International Conference on Knowledge Capture (K-CAP'05), 2005.
2. Ayesha Rashid, Naveed Anwer, Dr. Muddaser Iqbal, Dr. Muhammad Sher "A Survey Paper: Areas, Techniques and Challenges of Opinion Mining", International Journal of Computer Science, November 2013
3. McKinsey&Company. "Marketing & Sales Big Data, Analytics, and the Future of Marketing & Sales" (March 2015).
4. Michael Svilars, Arnab Chakraborty and Athina Kanioura "From hype to real help: Finding valuable consumer insight in a stream of data".
5. Bailey MJ, Muth RF, Nourse HO. 1963. A regression method for real estate price index construction. *Journal of the American Statistical Association* 58: 933–942.
6. Bailey MJ, Muth RF, Nourse HO. 1963. A regression method for real estate price index construction. *Journal of the American Statistical Association* 58: 933–942

<sup>1</sup> **Vladyslav Yeshchenkov**

Student of the Department of Management Technology

<sup>2</sup> **Olga Mezentseva**

Candidate of Economic Sciences, Associated prof. of the Department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **IDENTIFICATION OF THE MAIN PROBLEMS OF COLLECTION AND ANALYSIS OF SPEECH DATA USING MACHINE LEARNING**

**Abstract.** The main idea of this paper is to identify and present the main problems that data analysts face when collecting and analyzing speech data. The development of human-machine interfaces has been developing very rapidly nowadays. Invention of voice assistants, smart house technologies is a confirmation of that. Speech technologies are the latest technologies of the 21st century that are used to control computers, cars, and household appliances using voice.

**Keywords:** machine learning, data analysis, automatic speech recognition.

Despite the large number of researches, modern speech recognition systems remain insufficiently perfect, many problems associated with the process of automatic speech recognition remain unsolved. This indicates the urgency of studying the process of recognizing speech signals and the development of algorithms and methods for implementing this process.

For a machine learning system being able to learn, a huge amount of data is required. This becomes a problem when we speak about audio data. We face into privacy and data security issue. Biometrics like voice, face, fingerprints, and other personal traits are widely used as robust features to identify individuals in authentication systems. It is important to keep the biometric data secure to protect the privacy of users, and we require privacy-preserving machine learning algorithms that can perform the authentication using the secure data. That's why the speech data that collected by personal devices is confidential and barely can be used in data analysis. The actual workaround is using of data generated through paid research and studies, but it is very limited approach.

Speech recognition systems can be sufficiently accurate when trained with enough data having similar characteristics to the test conditions. However, there still remain many circumstances in which recognition accuracy is quite poor. These include moderately to seriously noisy or reverberant noise conditions, and any variability between training and recognition conditions with respect to channel and speaker characteristics (such as style, emotion, topic, accent, and language). While systems are getting better there's still a big difference in their ability to understand different accents of Ukrainian for example. And even a simple cold can be a reason for voice commands not to work as well as usual. In other case, when there is too much background noise speech recognition will be challenging. Making it especially hard to use them effectively in the urban outdoors or large public spaces/offices. For the purpose of

cleaning speech signal from noise is currently successfully applied method of wavelet transformation of an audio signal. Using this method, it seems possible to isolate voice on audio recording, even if present strong background noise [1].

Speech has no natural pauses between the word boundaries, the pauses mainly appear on a syntactic level, such as after a phrase or a sentence. This introduces a difficult problem for speech recognition — how should we translate a waveform into a sequence of words? One way to simplify this process is to give clear pauses between the words. This works for short command-like communication, but as the possible length of utterances increases, clear pauses get cumbersome and inefficient. Additional complexity appears in languages which are phoneme-based, which is typical for Slavic languages. So, the words get extra parts and become different to those from dictionaries. This complicates the process of speech analysis.

Natural language has an inherent ambiguity, i.e. we can not always decide which of a set of words is actually intended. The main ambiguity concept is homophones. The concept “homophones” refers to words that sound the same, but have different orthography or meaning. How can we distinguish between homophones? It’s impossible on the word level in ASR, we need a larger context to decide which is intended [2].

**Conclusions.** In this paper, we identified the main problems of automatic speech recognition, but not all of them. But one thing we can say certainly, audio data analysis is a challenging process. And as we can see, the most problematic issues are connected to the large search space and the input data variability. Thus, the speaking style, speed and even gender of person can cause errors in speech recognition or learning the system. This is why the search for ideal methods and algorithms for speech processing is still ongoing. And with the growing demand for human-machine interfaces, it becomes even more relevant.

#### **References:**

1. M.Z.Hein, The current state of the problem of processing, analysis and synthesis speech signals, 2018. URL:  
<http://www.mathnet.ru/links/6fc916ae022b05118d03dd6b04428d3c/cn182.pdf>.
2. Markus Forsberg, Why is speech recognition difficult, 2003. URL:  
[https://www.researchgate.net/publication/228763868\\_Why\\_is\\_speech\\_recognition\\_difficult](https://www.researchgate.net/publication/228763868_Why_is_speech_recognition_difficult).

<sup>1</sup> **Dmytro Zhovtukhin**

Master student of the Department of Management Technology

<sup>2</sup> **Oleksii Yehorchenkov**

PhD, Associate Professor of the Department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## CLASSIFICATION OF BOTTLES IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

**Abstract.** The purpose of this paper is to present the relevance of used bottles classification, as well as to developed efficient deep learning models. During the research, we've trained several convolutional neural networks to choose optimal one. As a result, we've obtained separate models with height accuracy score and small time of performance on test dataset.

**Keywords:** Deep Learning, Neural Network, Image Classification.

Annually in Ukraine about 12-15 million tons of household rubbish are generated [1], 40% of which is used for packing. Almost 96% of all trash, including plastic, is sent to landfills, where it poisons the earth for years. For example, a plastic bag decomposes for 500 years, an ordinary bottle of water - a millennium. Ukrainians started sorting garbage only in 2018, but despite this, the problem of recycling is still acute.

To facilitate the process of sorting garbage, we offer to use a system of bottles classification based on the processing of their images by convolutional neural networks. Automating the sorting process will help to speed up the recycling and, in the long run, save money that is now spent on manual labor. The reason of choosing this method is its similarity to human analysis. The machine, as well as person, finds some patterns with differ one type of bottles from others. Same approach very popular in machine learning world to classify objects that can be represented by their images.

For training was chosen 2 popular architectures: ResNet-18 [2] and Mobilenet\_v2 [3]. First one famous for its accuracy, second for small computational needs. In ResNet was added one block of Linear, ReLu, Dropout layers, Mobilenet has two additional blocks.

Due to the fact that the available dataset was not found, for some time was collected personal dataset with such classes: plastic bottles with 1206 images, glass bottles with 561 images and 902 images of aluminum cans. Each class was split on 3 sets for train, validation and test. For test - 100 images by class and others images divided into test and validation by ratio 8:2.

The dataset is small for independent training. Two approaches have been used to solve this problem. First, a pre-trained model on PyTorch was used. So, all layers except few last one was frozen to save previous coefficients. Secondly, augmentation was added to the training pipeline. Sometimes one of these changes could occur: horizontal flip, vertical flip, random affine transformation, color changes.

As metric was used accuracy, as loss function – weighted cross entropy, optimizer - Adam. Epoch number – 75, batch size – 25, initial learning rate – 0.001 with decay rate optimization. GPU provided by Google Colaboratory was used for training acceleration.

As the result we have 4 models: ResNet and Mobilenet\_v2 without augmentation and with it. Their performance can be seen on Table 1.

Most important features are accuracy, size and speed. Most errors occur plastic bottles class that was false predicted by developed models. Size of ResNet-18 are bigger and speed are slow as was expected. But its accuracies are better. In the other case, Mobilenet\_v2 fast and small and not so bad in accuracy score.

It is interesting that the best loss score is 0. It may cause by overfitting because model perfectly learned training set. So, it is better to avoid of using that model.

Table 1  
Models performance results

	Test Accuracy	Best Loss	Size, MB	Training time, sec.	Mean speed on CPU/GPU, sec.
ResNet-18	0.9967 (best)	0.000 (best)	42.94259 (worst)	1325.55	0.188/0.0138
ResNet-18 with augmentation	0.9967 (best)	0.005	42.94212	4080.65 (worst)	0.198/0.0139 (worst)
Mobilenet_v2	0.9834	0.004	11.32676 (best)	1269.44 (best)	0.102/0.0157 (best)
Mobilenet_v2 with augmentation	0.98 (worst)	0.017 (worst)	11.32725	3821.54	0.148/0.0157

**Conclusions.** We established that there isn't model that can provide both high accuracy and quick performance at the same time. If the bottles classification system must run in real time on a device with low computational potential, is better to choose Mobilenet\_v2 architecture with no augmentation during training process. For the best accuracy score, ResNet-18 architecture is more wisely to choose. Despite slightly worse results of speed and loss score, we recommend using augmentation in order to avoid overfitting. Also, this work can be the basis for classification system of used bottles and shows a prospect for further research.

#### References:

1. State Statistics Service of Ukraine: <http://www.ukrstat.gov.ua/>.
2. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun (2015). Deep Residual Learning for Image Recognition.
3. Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, Liang-Chieh Chen. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks.



# **DIGITAL PROJECT MANAGEMENT TECHNOLOGIES**



<sup>1</sup> Yuliia Dehtiarova

Master student of the Faculty of Information Technology

<sup>2</sup> Viktor. Morozov

PhD in Engineering Science, Full Professor, Head of the Department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## PRACTICAL IMPLICATION OF DIGITAL PROJECT MANAGEMENT TECHNOLOGIES

**Abstract.** The report describes how digital project management technologies are used on the example of a platform for buying and selling currency. It presents several vital processes in this project: project and recourse planning, backlog and task management, meetings and communication, data management.

**Keywords:** management processes, project planning, recourse planning, backlog & task management

COVID-19 pandemic forced firms to adopt changes in their working processes in favor of digital technologies. It's now become a routine to conduct online meetings, put tasks on the online boards, and use more digital instruments for project management.

The use of a modern project approach to solving socio-economic problems and tasks, developing management processes, and focusing on achieving results with minimal time and money will form a new generation of leaders and improve socio-economic processes in the country.

Project management requires constant guidance and coordination of the project's human, material, and financial resources. During its life cycle, project management should be carried out through the use of modern management methods and techniques to achieve project results. As stated in [1], "an effective project manager must ensure that all necessary work is carried out promptly, thoroughly, and meet specific requirements. To this end, monitoring is carried out throughout the project implementation phase." As mentioned in [2, p.21], IT project managers must develop and use new methods and approaches, one of which is based on the use of proactive management methods.

Managing the project of developing a web platform for buying and selling currency is becoming especially important because, from January 1, 2019, Ukrainians have the right to sell and purchase foreign currency through payment devices - terminals and ATMs. Also, the provisions of the resolution of the National Bank №63 [3] came into force, which introduced new rules of currency exchange operations. In particular, it abolished the requirement for a paper document's mandatory issuance confirming the fact of currency exchange. As the innovations will be implemented gradually, today, there is a need to develop web platforms for placing ads to buy and sell currency to meet the urgent need for interactive services. According to a survey

conducted in the summer of 2019, 56 respondents (individuals) out of 80 consider it necessary to create this web platform.

The example of a platform for buying and selling currency shows how digital project management technologies are used. There are several vital processes in this project: project & recourse planning, backlog & task management, meetings & communication, data management. Digital tools play a crucial role in the successful implementation of all the processes.

Project & recourse planning. Pountney [4] highlights the critical role of recourse planning for IT projects. He pays particular attention to contemporary Human Resource Management theory. Pinha, D.C., and Ahluwalia, R.S. explain how project duration and costs can be reduced "by empowering project managers to assess different scenarios" [5]. It is rational to plan resources following the hierarchy of the influence of these restrictions developed by the project manager, depending on a particular project's characteristics. 1. Planning from resource constraints 2. Planning from time constraints.

Backlog & task management. Todd Sedano, Paul Ralph, and Cécile Péraire examined product backlog and proposed the "theory of product backlogs" to be widely applicable by organizations with different software development cultures [6]. One of the essential parts of project management is online meetings with developers to align with project processes. What tools can be used for it? The most widely used and spread now is Zoom, Google Meet, Microsoft Teams. A platform for buying and selling currency project meetings will be conducted through Zoom, which allows all the team to stay safe and be on the same page.

Data management. The process of data management consists of data quality, data design, data lineage, and data cost management. Many companies do not have a clear understanding of what needs to be considered in strategic data management. Meanwhile, there are several clear principles on which the Data Governance strategy is based: make your data accessible; users must work with consistent data; understand what to collect and what to throw away; ensure the safety of your data; take care of security. It is equally essential to ensure the safety of funds. Data is the same asset. In the modern world, it is sometimes even more valuable. So the same rules apply to it: like competent financial management, Data Governance helps to reduce the number of errors, increase business efficiency, and clearly understand what results can be achieved further. Data security is in the first place in the list of priorities of digital project management. It is essential to improve safety by classifying data according to different levels of risk.

All in all, risks in a market economy are an integral part of management. Uncertainty makes it impossible to avoid danger. It is necessary to learn to anticipate risk, assess its magnitude, plan measures to prevent it, and not exceed acceptable limits. Project planning and implementation occur in conditions of uncertainty caused by changes in the internal and external environment.

Risks arising in the process of project preparation and implementation require immediate response and decision-making to reduce them. Since making a decision is a

choice of possible options for developing events (scenarios), this scheme is based on risk analysis for each scenario. The main idea is to select the plan that will be optimal in terms of the "win-risk" ratio, taking into account the actions that need to be taken to mitigate the risks associated with this scenario.

Risk analysis requires, above all, fast and reliable information. In conditions of tougher competition caused by the globalization of markets, it will not be large enterprises that will win over small ones. Still, dynamic ones over those are slowly reacting to a change in the environment. Risk analysis is divided into two mutually complementary types: qualitative, the main task of which is to determine the risk factors and circumstances leading to risk situations, and quantitative, which allows you to calculate the value of individual risks and the risk of the project as a whole

### References:

1. Н. П. Юрчук Система моніторингу в управлінні ІТ-проектами/ Н. П. Юрчук//Електронне наукове фахове видання «ефективна економіка». – 2018. – № 4: [Електронний ресурс] – Режим доступу до ресурсу: [http://www.economy.nayka.com.ua/pdf/4\\_2018/58.pdf](http://www.economy.nayka.com.ua/pdf/4_2018/58.pdf)
2. В. В. Морозов Дослідження моделей взаємодії процесів в іт-проектах на основі проактивного підходу дослідження моделей взаємодії процесів в іт-проектах на основі проактивного підходу / В. В. Морозов, О. В. Кальніченко // Вісник Національного технічного університету «ХПІ». Серія: Стратегічне управління, управління портфелями, програмами та проектами. 2019. № 2 (1327). – с. 20-27: [Електронний ресурс] – Режим доступу до ресурсу: [http://repository.kpi.kharkov.ua/bitstream/KhPI-Press/40870/1/vestnik\\_KhPI\\_2019\\_2\\_Morozov\\_Doslidzhennia\\_modelei.pdf](http://repository.kpi.kharkov.ua/bitstream/KhPI-Press/40870/1/vestnik_KhPI_2019_2_Morozov_Doslidzhennia_modelei.pdf)
3. Про затвердження Правил з організації захисту приміщень банків в Україні. Постанова правління Національного Банку України від 10.02.2016 № 63: [Електронний ресурс] – Режим доступу до ресурсу <https://zakon.rada.gov.ua/laws/show/v0063500-16>
4. Ph. Pountney (2013). Importance of resource planning. The importance of resource planning during the initiation stage of an IT Project. Retrieved from [https://www.researchgate.net/publication/340883012\\_IMPORTANCE\\_OF\\_RESOURCE\\_PLANNING\\_G\\_The\\_importance\\_of\\_resource\\_planning\\_during\\_the\\_initiation\\_stage\\_of\\_an\\_IT\\_Project](https://www.researchgate.net/publication/340883012_IMPORTANCE_OF_RESOURCE_PLANNING_G_The_importance_of_resource_planning_during_the_initiation_stage_of_an_IT_Project)
5. D.C. Pinha, R.S. Ahluwalia. Flexible resource management and its effect on project cost and duration. J Ind Eng Int 15, 119–133 (2019). <https://doi.org/10.1007/s40092-018-0277-3>
6. T. Sedano, P. Ralph, C. Péreira (2019). The Product Backlog. Conference: International Conference on Software EngineeringAt: Montreal, CanadaVolume: 41st. DOI: 10.1109/ICSE.2019.00036

**<sup>1</sup> Snezhana Gamotska**

Assistant Professor

**<sup>2</sup> Petro Soroka**

PhD in Physical and Mathematical Sciences, Associate Professor

<sup>1,2</sup>Taras Shevchenko National University of Kyiv

## **CHOICE OF METHOD OF QUANTITATIVE RISK ASSESSMENT IN RISK MANAGEMENT TASKS OF IT PROJECTS**

**Abstract.** The paper presents a diagram of the algorithm for determining the probability of obtaining a finished product of sufficient quality, which reflects the total effect of risks on the date of completion of the project, its cost and volume of work performed.

**Keywords:** risk, risk management, uncertainty, IT project, qualitative risk analysis, quantitative risk analysis, simulation.

The task of risk management of IT projects is to reduce the impact of undesirable factors on the project life cycle and obtain results very close to the planned ones. Many different analytical models and methods of risk analysis are known. Their disadvantage is that its application introduces many assumptions and limitations for simplicity of description of the process and possible solution. These methods allow to generalize the process, idealizing and simplifying its elements. An alternative to risk assessment in IT projects may be the use of simulation, which allows you to bring the model as close as possible to the real situation.

The task of risk management of IT projects is one of the most difficult in project management, due to the specifics of such projects. The fact is that IT projects are characterized by a significant degree of uncertainty about the future outcome. It should take into account the stage of the life cycle of the developed product, as for each of them the risk management parameters will be different. This, in turn, requires the use of different methodological apparatus for their analysis and evaluation.

Project risks affect all interrelated constraints of the project, such as content, quality, schedule, budget, resources, and so on. To achieve the planned results and obtain at the end of the project a product of sufficient quality, it is necessary to effectively manage changes in the project and respond in a timely manner to any deviations that occur (Fig. 1).

Known risks are those risks that have been identified and analyzed. Appropriate actions can be planned for such risks. It is not possible to plan appropriate actions for unknown risks. To take them into account, a reasonable solution would be to allocate a

general contingency reserve that includes these unknown risks, as well as all known risks for which the development of specific response measures is for some reason impossible or unprofitable.

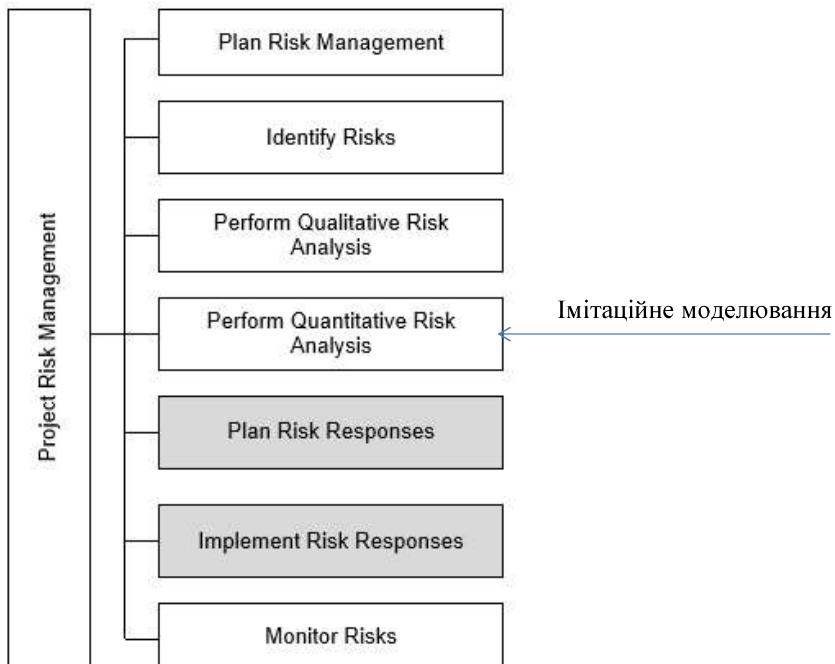


Fig. 1. Risk Management ANSI PMI PMBOK 6th Edition [1,2]

For planned risk avoidance, prevention, acceptance or compensation measures to be effective, their quantitative characteristics need to be assessed after risk identification and qualitative analysis. Approaches such as the method of sensitivity analysis of efficiency criteria, analysis of expected monetary value, analysis of decision tree, simulation modeling can be used for quantitative risk assessment [3].

Today, simulation is a priority in risk assessment in IT projects. It should also be borne in mind that the use of a simulation model does not exclude the need to use analytical models. They can be part of a directed experiment with a simulation model. Analytical modeling is carried out at the initial stage of research. Its results serve as a guide in the subsequent construction of the simulation model and can be used in its validation.

The simulation experiment is in many ways similar to the real thing. However, it has a number of advantages. A significant advantage is its cost, as it is much cheaper than the real thing. In addition, the simulation experiment allows you to calculate the behavior of systems that no longer exist or which do not yet exist. The efficiency and simplicity of this approach allows us to consider many situations with a very large number of combinations of initial conditions and possible effects on the process. The constant process of accumulation of simulation results leads to the formation, on the one hand, a fairly large, and on the other - a finite set of typical situations with a finite number of standard behaviors.

Quantitative risk assessment based on a simulation model. The task of determining the duration of the project can be divided into two: the task of project planning (determining the scope of the project, planning work and resources, etc.) and the task of quantifying the additional reserve of time, money, resources needed to prevent the consequences of risk situations. Next, consider the following problem, let the project be characterized by the presence of n important risks, each of which has a function that characterizes the probability of completion of the project by a certain date, using a certain amount, when performing a certain amount of work. You need to define a probability function that reflects the total effect of risks on the project completion date, its cost, and the amount of work done.

**References:**

1. A. Guide to the Project Management Body of Knowledge (PMBOK ® Guide) 6th Edition. PMI, 2017. 537p.
2. Features of the evolution risk management process in IT-projects / S. Gamotskaya, A. Vasylevskaya // Information Models and Analyses. V.6, N.2. 2017 – c. 183-194
3. Risk management in high-technology projects: overview and approaches / M.D. Godlevsky, A.A. Polyakov // Problems of the program. - 2006. - N 2-3. - S. 311-319.

<sup>1</sup> **Mariia Kambur**

Master student, department of Technology Management

<sup>2</sup> **Oleksii Yehorchenkov**

Associate professor, department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## SMART KITCHEN DEVELOPMENT PROJECT MANAGEMENT

**Abstract.** This thesis contains a consideration of public catering establishments automation problems and an exploration of smart kitchen system management and development process.

**Keywords:** project management, catering, IoT, web, embedded systems.

Currently, the field of catering services is flourishing in Ukraine. In 2018 alone, the number of catering establishments in Ukraine increased by 2.7 thousand units as seen on Figure 1. [1]

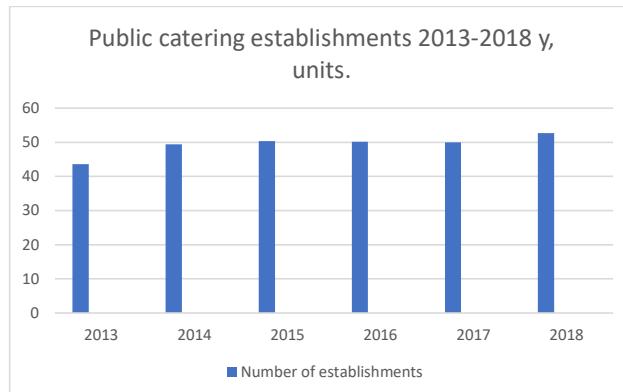


Figure 1 – Number of public catering establishments 2013-2018 years

The restaurant market is one of the most dynamic markets in Ukraine. It is extremely sensitive to the economic situation, exchange rate fluctuations, and social factors. For the last four years (2016-2020), it has been actively trying to recover from the political and economic crisis of 2014-2015. In total, about 5,600 catering establishments have disappeared during this period. According to Rest Consulting, about 1,500 of them closed without surviving the crisis, and a little more than 4,000 remained [2].

In 2020, the world experienced a shock from which it has not yet recovered. The COVID-2019 pandemic has dealt a significant blow to all areas of industry and service delivery, and especially to small and medium-sized businesses. We were able to watch cafes and restaurants close one after the other, as quarantined business owners simply could not afford to pay for food staff.

This problem could be partially solved by creating a comprehensive hardware and software system for a smart kitchen, which would automate most of the cooking process. Such a system still has no analogues on the market and is most often represented by separate devices that do not communicate with each other and still require close human supervision.

The introduction of such a system in catering establishments will allow business owners to reduce labor costs and optimize internal processes of processing and cooking.

The smart kitchen system is an automated control system for kitchen utensils via a web interface.

Thanks to the use of this product, the user can download a map of his kitchen in the web application and place the necessary equipment. Each of the items will have "On", "Off" buttons and a "Mode selection" panel.

All kitchen equipment will be upgraded and will include IoT modules to connect to the network and receive remote commands.

The product allows you to control an automated smart kitchen system (slicing, grinding, microwave, stove, kettle, dishwasher, etc.) via a web interface. The status of the system is displayed on the web interface and when you press the appropriate buttons, commands are sent to a specific device as seen on Figure 2.

Technology stack: embedded systems, microcontrollers, IoT, web programming, web design.

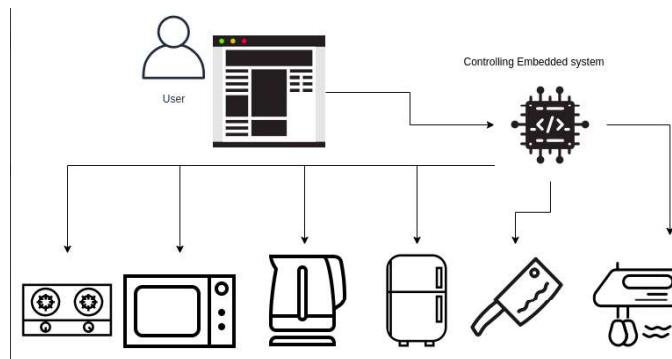


Figure 2 – Smart kitchen system architecture schematics

The system can solve a number of problems:

- 1) Excessive spending and wasteful use of products which is irrational from an ecological and economic point of view.
- 2) The incomprehensible interface of kitchen utensils which can lead to accidentally spoiling a dish or burning the appliance.
- 3) Mandatory presence in the kitchen which is uncomfortable does not allow you to do something else.
- 4) Introduction of IT in the food industry.

The specifics of project data management is the presence of three teams ranging in size from 2 to 8 people and therefore there is a strong need for careful communication planning between the three teams, as the products developed by these teams will be integrated with each other. The project manager must understand modern flexible project management methodologies, communications, and personnel management, as well as have a strong technical background. Also, there is a need to keep in touch with hardware suppliers which means neat procurement management.

Scrum is the Agile methodology which is preferred to manage the Smart kitchen system project. This methodology is best suited for this project, because we have three small teams (web development team, IoT development team, embedded systems development team), clear product requirements, the ability to increase functionality during development at the request of the customer, or when business requirements are changing.

The development process is divided into sprints (2 weeks).

Retrospectives, plannings and daily meetings with the project manager and the team are also involved, where they will discuss what was done the day before, what each team is going to do today and what were the difficulties in completing the tasks.

This synchronizes the work of the each team and adjacent teams, so everyone can be aware of project progress.

*Conclusions:* catering business is currently promising in Ukraine. However, it may be prone to destabilization in a pandemic or economic crisis times. The smart kitchen system is an innovative project, the product of which can partially solve the problems for the catering business. Such a project can be managed using flexible management methodologies such as scrum with an emphasis on communications management and product integration of different teams.

#### **References:**

1. Catering market analysis (HoReCa, catering, and food at gas stations). URL: <https://pro-consulting.ua/ua/pressroom/analiz-ryntka-obshestvennogo-pitanija-horeca-kejtering-takzhe-pitanie-na-azs>
2. L. Hirniak, V. Hlahola, CURRENT STATE, PERSPECTIVES AND TRENDS OF RESTAURANT BUSINESS DEVELOPMENT IN UKRAINE, Economics and Enterprise Management '16, 2018, pp. 71-78

**<sup>1</sup> Anastasiya Kovalenko**

Student of the Master's Project Management Program

**<sup>2</sup> Ivan Ivanov**

Student of the Master's Project Management Program

**<sup>3</sup> Viktor Morozov**

PhD, Professor, Head of The Department of Technology Management

*<sup>1,2,3</sup> Taras Shevchenko National University of Kyiv, Kyiv, Ukraine*

## **RESEARCH OF METHODS OF FORMATION OF THE INITIAL DESCRIPTION OF THE PROJECT OF CREATION AND START-UP OF THE ENTERPRISE ON PRODUCTION OF STREET FURNITURE MADE OF RECYCLED MATERIALS**

A century ago, humanity made a significant breakthrough in its development – a synthetic revolution took place. One of its main achievements was the invention of plastic. It was a durable and reliable material that became an alternative to expensive natural materials and therefore quickly gained popularity. In our time, the main advantage of plastic – reliability – became its main disadvantage [1]. Plastic products are usually used once and then thrown into trash. 9% of used plastic is recycled, 12% is incinerated, 79% ends up in landfills and the environment, polluting it and killing its inhabitants. According to researches, every year about 300 species of animals suffer from plastic, a significant number of which die. According to a report by the Ellen MacArthur Foundation, by 2025 the amount of unprocessed plastic will be 100-250 million tons per year, and by the middle of the century it will have reached 33 billion tons per year. Thus, there will be more plastic in the world's oceans than marine life.

The situation in Ukraine is also not comforting [2]. The number of official landfills alone is 6,000, and they occupy an area larger than allocated for natural reserves, 10% vs. 8%, respectively. There are about 35,000 unofficial ones. In total, landfills cover an area of 12,000 km<sup>2</sup>, and there are around 6 million tons of plastic there. At the same time, only 4% of all waste is recycled. In Europe, this number is an order of magnitude higher: 70% on average, in Sweden - 96%. The Swedes even buy garbage from neighboring countries, so that their recycling plants are not idle [3].

The problem of plastic waste worries many people. To attract public attention, many well-known companies collaborate with designers to create interesting and useful things from recycled plastic, which are also easy to recycle. For example, IKEA stores will soon have a collection of home textiles from shopper bags, pillowcases and tablecloths made from recycled plastic caught in the Mediterranean [4]. A Columbian company called Diseclar has gone even further, starting to make garden furniture from recycled plastic, which has an excellent aesthetic appearance, very similar to real wood products, but also strong and insensitive to weather conditions [5].

The global problem of accumulating used plastic in the environment and successful ideas for the use of recycled plastic have inspired the creation of a project to build and launch a factory for the production of outdoor furniture from recycled materials. Given, in particular, the current situation with inorganic waste in Ukraine,

this project is particularly relevant. The organization of enterprise construction and production start-up is quite a difficult task, as it requires careful planning of all processes and allocation of resources, identification of requirements of all stakeholders and coordination of many solutions with them, effective management of time, resources, budget and other important components. All this would be extremely difficult without the use of digital tools, so Airtable - one of the most highly valued services for creating workflows - comes to the rescue.

Compared to Microsoft Project, which is a more known alternative, Airtable is more user-friendly and has many advantages [6]. First of all, it can be used on iOS and Android platforms, which is important for the current fast pace of life. In addition, it has tools for collaboration, such as chat, commenting, communications calendar, email integration, task management, task planning, task scheduling, task tracking, workflow management. It can also integrate with various tools such as GitHub, Gmail, Google Calendar, Google Drive, Slack. Other useful benefits include automatic backup, automatic notifications, budgeting, bug tracking, built-in database, data filtering, data synchronization, social media integration, to-do list.

#### **References:**

1. Dmitrienko, Pollution of the planet with plastic is fraught with disaster, the scale of which is difficult to calculate, 2019. URL: <https://profile.ru/society/ecology/zagryaznenie-planety-plastikom-chrevato-katastrofai-masshtaby-kotoroj-slozhno-proschitat-133578/>
2. Mikhedov, Waste and society: how Ukraine is bent from garbage, 2018. URL: <https://platfor.ma/topic/otbrosy-y-obshhestvo-kak-ukrayna-zagyaetsya-ot-musora/>.
3. Society Ukraine is drowning in garbage - is there a way out, 2019. URL: <https://dengi.informator.ua/2019/03/22/ukraina-utopat-v-musore-est-li-vyhod/>
4. IKEA to release a collection made from recycled plastic, 2020. URL: <https://www.retail.ru/news/ikea-vypustit-kollektsiyu-sozdannuyu-iz-pererabotannogo-okeanicheskogo-plastika-30-sentyabrya-2020-198197/>.
5. Diseclar - recycled plastic furniture, 2019. URL: <https://gtracer.com.ua/ru/diseclar--recycled-plastic-furniture>.
6. Microsoft Project vs Airtable Comparison, 2020. URL: <https://www.getapp.com/project-management-planning-software/a/microsoft-project/compare/airtable/>.

**<sup>1</sup>Tatiana Latysheva**

PhD of Technical Science, Assistant Professor of the Department of Technologies Management

**<sup>2</sup>Dmytro Smishchenko**

Master student of the Intellectual Technologies Department

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **PROCESS OF EFFECTIVE PROJECT MANAGEMENT OF DEVELOPING MOBILE APPLICATION FOR CARSHARING**

**Abstract.** Covers process of effective project management of developing mobile application with help of modern management frameworks and development process organization.

**Keywords:** Application, mobile, SCRUM, shared economy, car sharing

Practice shows that recently the sharing economy is gaining popularity. Traditional consumption of goods and services is not the optimal use of resources. A person does not need to buy or maintain something to use it. In our example, we will consider car sharing. Economic and social research has shown that the idea of carsharing can meet the requirements for comfort and customer service price of a carsharing product (mobile application).

The current results of the project management study do not take into account or do not fully take into account the above problem. The purpose of the work is to study the processes of effective project management to create a mobile application for carsharing, which will be able to solve all the tasks posed by developers and meet user expectations. Therefore, the study of such processes seems relevant, and the problem requires a solution.

The sharing economy is a socio-economic system based on the sharing of human and physical resources [1].

Carsharing is an alternative model of car ownership and use. The idea of car sharing is not new in the world [2]. Thus, in Europe, the list of the first car-sharing companies includes the Swiss club "Mobility", which was founded in 1987. At the beginning of 2015, the members of the Mobility car-sharing club are about 120,000 Swiss citizens, who have the opportunity to use 2,700 cars of various categories. In the classic version, car sharing is the possession of a car with another person or people, i.e. the ownership of a share of the car with the right to use. The car-sharing model is designed to quickly and conveniently meet the traditional human need to move from one place to another [3], and also saves your own money by reducing the cost of sole proprietorship.

The project involves the creation of a mobile application for car sharing. Carsharing is a type of car use when one of the parties is not the owner. This is an option to rent a car from specialized companies (usually for long-distance and / or short trips) or private individuals (for any period and distance of the trip - by arrangement). This car rental model is convenient for periodic use of the vehicle or when you need a

car other than the brand, body type and load capacity from the usual one. Carsharing is one of the global directions of Sharing Economy development, when the population refuses to acquire goods in order to avoid responsibility and costs, but continues to have access to all the achievements of scientific progress, using their joint consumption.

The use of flexible SCRUM methodology is provided for the process of organizing the application development process. Scrum is the backbone of a process that includes a set of methods and predefined roles. The main actors are ScrumMaster, who oversees, leads and works as a project manager, product owner, a person representing the interests of end users and other stakeholders in the product, and a team that includes developers [4].

Scrum is a framework in which people can solve complex adaptation problems while producing products with the highest possible value productively and creatively.

Scrum itself is a simple basis for effective team collaboration on complex products.

Thus, the sharing economy in terms of the use of personal transport gives citizens an advantage in terms of price and convenience of service. The customer of the product can receive all the benefits of having personal transport, without spending time and resources on service of the transport unit, as well as on its purchase. In the form of a simple mobile application, the project solves user problems. In a convenient way, the user can receive car sharing services without spending on the purchase of the car itself.

The mobile application development project involves a long development period and the presence of a development team. For a planned iterative development process, the flexible SCRUM methodology has been identified as one of the leading ones at the moment. Using this methodology in the process of developing a mobile application, the project guarantees the timely execution of work on application development, testing of the application and its release schedule for subsequent updates.

In the future, it is possible to consider additional ideas for the development team depending on the scale and profits of the project. If the application is successful and it is decided to expand it, new development teams may appear, which will work in parallel. In this case, to continue to monitor the progress of development work, product owners can implement the use of the SAFE methodology, which provides for quarterly planning and goal setting for development teams.

Thus, car sharing as an idea can help users of the product in their daily lives. Due to comfortable trips, customers of the application can get more comfort in daily trips than users of public transport or taxis for less money.

## References

1. Wang, C., & Zhang, P. (2012). The evolution of social commerce: The people, management, technology, and information dimensions. Communications of the Association for Information Systems, 31(1), 105–127.
2. The CarSharing Handbook (Part 1)"(2016). Rain Magazine.
3. Ноздріна Л.В. Управління проектами: підручник / Ноздріна Л.В., Ящук В.І., Пологай О.І./ За заг.ред.Л.В.Ноздріної. – К.: Центр учебової літератури, 2010. – 432с.
4. Web-resource, access: <https://www.scrum.org>

**<sup>1</sup> Oleg Loik**

Masters of the Mechanics and Energy Faculty

**<sup>2</sup> Mykhailo Triska**

Masters of the Mechanics and Energy Faculty

**<sup>3</sup> Pavlo Lub**

Ph.D., Associate Professor of Department of the Information Systems and Technologies

**<sup>4</sup> Andriy Sharybura**

Ph.D., Associate Professor of Department of the Machines Operation and Technical Service of prof. O.D. Semkovych

*<sup>1-4</sup> Lviv National Agrarian University*

## **INFORMATION TECHNOLOGY IN PROJECT MANAGEMENT OF THE AGRICULTURE TECHNOLOGICAL SYSTEMS DEVELOPMENT**

**Abstract.** The prospects of the IT usage for management decisions supporting in agricultural enterprises projects are analyzed. Methods of virtual projects statistical simulation are used. It is shown that the use of IT in the projects management of enterprises technical re-equipment allows to support management decisions and ensure the effectiveness of these projects.

**Keywords:** IT, projects, production development, information-analytical system, modelling.

The use of IT to manage production projects requires the development and use of specialized simulation models. In particular, it allows to accompany management decisions during the projects implementation for the development of agriculture technological systems. Due to this, at the initial stages of project implementation, goals are defined, components of the external and internal environment are analyzed, and the peculiarities of their interaction and cumulative impact on project performance indicators are revealed.

The goals of such project management information systems (PMIS) also include tasks for the projects development of agricultural enterprises material and technical re-equipment and the formation of production resources. This makes it possible to carry out the tasks of agricultural production projects while ensuring the efficient use of limited resources. In fact, such projects are aimed at taking into account material and information links at the level of the subject area – the fields crops area, a set of specialized machines attached to them, performers, and so on. Coordination of the given parameters by technological systems among themselves allows to provide an extremum of efficiency indicators of both separate projects of technological systems and their efficiency [1,2].

The use of PMIS to match the start-up time of crop harvesting projects (for example sugar beets) and the production area of the crop with the parameters of technical equipment of these projects plays an important role in ensuring the minimum specific total cost. The definition of these cost estimates is carried out on the basis of

functional indicators of production and technological processes, which we obtained on the basis of computer experiments with the developed statistical simulation model (MS Visual Studio C #) [1].

This model is based on the system-event reflection of daily projects work stages, which allowed to take into account: 1) stochastic influence of natural (agrometeorological and biological-subject) component on calendar dates of crop harvesting and naturally allowed time for technical equipment beet harvester); 2) daily weight gain of root crops, as well as the impact of this indicator on the daily rate of harvest; 3) the impact of the production area of the crop and the combine productivity on the duration of these technological process, and hence on the functional indicators of their efficiency.

Execution of the main stages of works modelling in projects and calculations for technical equipment (beet harvester and tractor trailers-loaders of root crops) of different capacity allowed to optimize the production area of  $S^{opt}$  culture (Fig.).

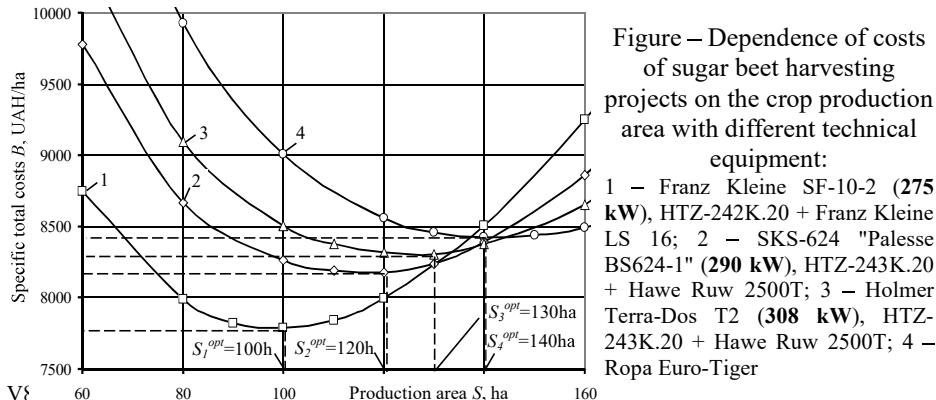


Figure – Dependence of costs of sugar beet harvesting projects on the crop production area with different technical equipment:

1 – Franz Kleine SF-10-2 (**275 kW**), HTZ-242K.20 + Franz Kleine LS 16; 2 – SKS-624 "Palesse BS624-1" (**290 kW**), HTZ-243K.20 + Hawe Ruw 2500T; 3 – Holmer Terra-Dos T2 (**308 kW**), HTZ-243K.20 + Hawe Ruw 2500T; 4 – Ropa Euro-Tiger

The development of PMIS based on statistical simulation models allows us to perform research on these projects, assess the content and timeliness of work and justify management decisions under probabilistic conditions of the project environment. The choice of one or another coordination of component projects (time of their start up, production area of culture with the parameters of technical equipment) must be considered in the context of a particular technological system of a some production environment.

#### References:

1. Lub, P., Sharybura, A., Pukas, V.: Modelling of the technological systems projects of harvesting agricultural crops. 14th International Conference on Computer Sciences and Information Technologies (CSIT). Vol. 3, 19-22 (2019).
2. Tryhuba, A., Ftoma, O., Tryhuba, I., Boyarchuk, O.: Method of quantitative evaluation of the risk of benefits for investors of fodder-producing cooperatives. 14th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Vol. 3, 55-58 (2019).

**Viktor Morozov**

PhD, Professor, Head of The Department of Technology Management

*Taras Shevchenko National University of Kyiv, Kyiv, Ukraine*

## USE OF MACHINE LEARNING METHODS IN DATA ANALYSIS FOR DIGITAL PROJECT MANAGEMENT

Modern conditions for the development of project management technologies are characterized by a significant impact of the mass of restrictions associated with the COVID-19 pandemic. This situation contributes to the search for new technologies, automation of solutions and the introduction of new methods of intelligent processing of project information. Among such methods, machine learning methods occupy a special place. Among them are methods of linear regression, methods using nonlinear models, methods of teaching representation, methods using specialized models [1]. At the same time, the latter are based on consumer choice theory, using polynomial models with logistic distribution, survival analysis models, and auction theory models.

The use of machine learning auction theory models can be used to solve the problems of automated supply of resources for the execution of projects on a competitive basis.

At the same time, the algorithmic approach determines the development of marketing services that can be offered by resource exchanges. An exchange or broker of any other type adds an extra layer of complexity between the provider and the client. This is determined by the fact that in addition to achieving the main marketing goals, both the supplier and the customer must optimize their selling and buying strategies accordingly [2].

The main purpose of a resource exchange is to maintain competition between buyers for a limited resource, such as advertising space. The standard way to solve this problem is to organize an auction, where each buyer places a bid, and the resource put up for auction goes to the participant with the highest bid. However, the auction rules can be configured in different ways [3].

In addition, it is important to understand that potential buyers take part in the auction, because for each of them the auction resource has a certain value. They try to make a profit, trying to get what they want at the lowest possible price. Consequently, it is critically important for bidders to correctly estimate the cost of a resource, and, as a result, we can classify all auctions by some types of cost, which will be reported in a report at the conference [4].

The optimization problem for open auctions may seem dynamic, but in reality it is static and similar to the problem of closed auctions.

The report will consider examples of different types of auctions, in particular, the Dutch auction and the English auction. As a result of such auctions, the participant can

learn from the observed rates.

Also, the Vickrey auction will be discussed in more detail in order to obtain tools for building optimization models that include auctions. The model of such an auction is more convenient for analysis and is widely used in practical applications. There are also other types of auctions that use more advanced analysis methods.

As a result of the research carried out, the author was able to formulate some conclusions proposed for discussion.

1. Many market problems, especially for the supply of resources on a competitive basis, can be expressed in the form of an optimization problem, in which the subject of optimization is the business result, and the variables of the business action.

2. The relationship between activities and business results can often be obtained from historical data. This can be accomplished using supervised learning methods.

3. The main goal of supervised learning is to assess the conditional distribution of the response based on changing input data. In many practical applications, this task can be reduced to finding the most likely outcomes. The two main types of supervised learning problems are classification and regression.

4. Training the model can be considered as an optimization problem, where it is required to select the model parameters that maximize the probability of the observed data following the model distribution.

5. Many supervised learning problems can be solved using linear models that determine either the relationship between input and output as a linear function. The simplest examples of linear models are linear and logistic regression.

### **References:**

1. Pountney, Philip. "Importance of resource planning. The importance of resource planning during the initiation stage of an IT Project." (2013). URL: [https://www.researchgate.net/publication/340883012\\_Importance\\_of\\_resource\\_planning\\_The\\_importance\\_of\\_resource\\_planning\\_during\\_the\\_initiation\\_stage\\_of\\_an\\_IT\\_Project](https://www.researchgate.net/publication/340883012_Importance_of_resource_planning_The_importance_of_resource_planning_during_the_initiation_stage_of_an_IT_Project)
2. V.V. Morozov, O.V. Kalnichenko, Research of models of interaction of processes in IT-projects on the basis of a proactive approach, Bulletin of the National Technical University "KhPI". Series: Strategic management, portfolio management, programs and projects, No. 2 (1327), pp. 20-27, 2019. URL:[http://repository.kpi.kharkov.ua/bitstream/KhPI-Press/40870/1/vestnik\\_KhPI\\_2019\\_2\\_Morozov\\_Doslidzhennia\\_modelei.pdf](http://repository.kpi.kharkov.ua/bitstream/KhPI-Press/40870/1/vestnik_KhPI_2019_2_Morozov_Doslidzhennia_modelei.pdf)
3. N.P. Yurchuk, Monitoring system in management of IT projects, Electronic Science Fakhove Vidannya "effective economy", No. 4, 2018. URL: [http://www.economy.nayka.com.ua/pdf/4\\_2018/58.pdf](http://www.economy.nayka.com.ua/pdf/4_2018/58.pdf)
4. I. Kravtsov, Machine learning for business and marketing, SPb.: Peter, 2019.

<sup>1</sup> **Viktor Morozov**

PhD, Professor, Head of The Department of Technology Management

<sup>2</sup> **Maksym Proskurin**

PhD student at the Department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

## ANALYSIS OF THE PROSPECTS FOR APPLYING METHODS FOR CUSTOMER CHURN PREDICTION USING MACHINE LEARNING IN INNOVATIVE STARTUP PROJECTS

Countries that have managed to establish a continuous process of generating new knowledge and innovative ideas and transforming them into innovative products are now the most efficient and have a leading role in the global economy. The experience of the USA, which brings 85% of innovative products to the market, Japan - 75%, Germany - 55%, Israel - more than 50% is indicative. Unfortunately, the share and innovations in the total volume of manufactured products in Ukraine do not exceed 2%.

Small companies such as startups create favourable conditions for innovation. Start-up is an innovative project for developing new products or services, formed to find a repeatable and scalable business model in conditions of extreme uncertainty. Based on an analysis of the number of startups in 137 countries, Startup Ranking has developed a ranking in which Ukraine took 42nd in 2018 (215 startups). First place went to the USA - 45 004 startups, second place in India - 5203 startups and third place in the UK with 4702 startups. [1].

In recent years, due to the development of cloud computing infrastructure, the SaaS (Software as a Service) software distribution model has become popular among IT startups, as it is more accessible both for creation and use by the end-user.

In 2018, the global SaaS market was estimated at \$134.4 billion and projected to grow to \$220.21 billion, with an annual growth rate of 13.1% by 2022. [2]. But startups working in B2B (Business-to-Business) have problems with long sales cycles because they include a wide range of contact persons, and the client's decision is collective, depends on many factors and is characterized by numerous interactions between companies.

Most SaaS startup founders put much effort into attracting new customers and MRR (recurring monthly income), but often forget about some of the most important parameters as Retention and Churn rates.

In essence, churn is the number of users who stop any interaction with the company. Depending on the area, this may mean that customers have deleted their account, cancelled their subscription, did not renew the contract, did not re-purchase or decided to leave for a competitor.

Churn rate is a metric that calculates the number of customers who leave a product over a given period of time, divided by total remaining customers. If the startup had 100 customers at the beginning of the month and 5 of them cancelled their subscription during the month, this means 5% customer churn (5/100). Here is an advanced formula for calculating customer churn:

$$\text{Churn Rate} = a = \frac{\Delta C_{\text{Cancels}}}{C * \Delta t} \quad (1)$$

where "a" is the customer churn rate, "C" = number of customers, "t" = term (in months), "C<sub>Cancels</sub>" = number of cancelled subscriptions for a period.

Churn is a direct reflection of the value of the product and the functions that the startup offers its customers. SaaS business should continuously optimize its development to reduce the churn rate.

At first, most startups experience difficulties with outflows. The churn rate can be 15% or more, although, for SaaS services, less than 3% is considered acceptable. It is essential to analyze customer churn in a startup because this will not only help to retain existing customers but also to adapt the product to the market needs. If everything is done correctly, then the outflow rate should decrease and stabilize.

The SaaS model allows software suppliers to collect data on the use of customers that are not available to traditional software suppliers. Although the SaaS and cloud computing market is overgrowing, as far as we know, from a scientific and practical point of view, the subject of churn in B2B SaaS startups has not been sufficiently researched.

**The goal of the abstract is to** analyze existing and future methods for predicting customer churn using machine learning, which can be applied to B2B SaaS startup companies.

Customer Churn Prediction is one of the classic problems in Data mining. While companies operating in the B2C (Business-to-Consumer) sector, for example, in the telecommunications or banking sector, regularly analyze customer behaviour and have been using customer churn forecasting for many years, B2B pays much less attention to the scientific literature.

There are many similarities between the telecommunications industry and the SaaS industry. As can be seen from Table 1, both areas can control the use of products by the number of times they are used and the duration of their service. Telecom providers record the details of each call, while SaaS providers record the details of each session. The session starts when the user logs in (logging in), and the session ends when the customer logs out, or the session ends due to inactivity. Thus, while most subscribers are active during the call, except the retention time, software users may not be active during the entire session.

Table 1

**Comparison of telecommunications industry data with SaaS industry data**

Telecom	SaaS
Number of calls	Number of logins (sessions)
Length of call	Length of session
Call targets	Modules used

Most studies on customer churn in telecommunications have studied the business to consumer relationship (B2C), while the SaaS supplier in our research only interacts with business (B2B). Thus, although customer attributes may be similar, further studies should use B2B SaaS data aggregated for all customer employees.

The scientific paper [3] discuss models and methods of proactive product management for complex IT projects. However, the interaction of clients with IT products is not tracked, and the direction of product development to increase client loyalty is not indicated.

Euler [4] has developed a solution tree to identify the types of telecommunications customers that are most likely to be addressed. Euler used the pre-processing capabilities of the KDD MiningMart system to obtain assumed characteristics that were not present in the original data. Coussement and Van den Pol used vector support machines to increase productivity in forecasting churn for the newspaper subscription service. The results of this work show that the interaction between customers and the provider is vital for analyzing churn. Coussement and Van den Pol continued to study the interaction between customer and provider, adding to their model the emotions of customers via email [5]. Hadd et al. [4] determined the prognostic characteristics of customer churn and found that decision trees outperformed neural networks and regression in terms of overall accuracy.

Numerous studies have investigated various machine learning algorithms and their potential for modelling outflows. Since predicting whether a customer will be lost is not a binary classification problem, several models such as Logistic regression have been tested [6], Decision trees [7] Random forest, supporting vector machines and neural networks [8]. Although our review points to numerous studies on SaaS, most work tends to focus on a few programmes. In particular, preliminary work on subscriber subscriptions tends to be based on subscriber data in mobile phone areas [9], credit cards [10] and the provision of Internet services [11].

After an analysis of information sources [1-11], we concluded that there are indeed no integrated models and methods for managing the development of commercial B2B SaaS products, which would allow us to react to various interactions of different classes of customers at the lowest cost. And also to determine, based on the analysis of such data, areas for further improvement and development of such products to increase economic efficiency.

This method significantly expands the opportunities for effective management of startup project development. In turn, the ability to proactively manage the development of innovative IT products based on the methods of predictive analytics to predict customer churn with the help of machine learning allows for taking into account the dynamic interaction on the processes of product development and subsequent management of the startup project.

### References:

1. Wang, Xin, Tapani Ahonen, and Jari Nurmi. "Applying CDMA technique to network-on-chip." IEEE transactions on very large scale integration (VLSI) systems 15.10 (2007): 1091-1100.
2. "StartupRanking," [Online]. Available: <https://www.startupranking.com>.
3. "Global Software as a Service (SaaS) Market Report 2020," 2020. [Online]. Available: <https://www.prnewswire.com/>.
4. V. Morozov, O. Kalnichenko, M. Proskurin and O. Mezentseva, "Investigation of forecasting methods the state of complex it-projects with using deep learning neural networks.," *Springer Series Aadvances in intelligent systems and computing.*, vol. 1020, pp. 261-280, 2020.
5. T. Euler, "Churn prediction in telecommunications using miningmart," *Proceedings of the Workshop on Data Mining and Business (DMBiz) at the PKDD*, 2005.
6. K. Coussement and D. VandenPoel, " Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers.," *Expert Systems with Applications*, p. 6127 – 6134, 2009.
7. S. Neslin, "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research American Marketing Association*, no. 43, p. 204–211, 2006.
8. A. Hiziroglu and O. Seymen, "Modelling customer churn using segmentation and data mining," *Frontiers in Artificial Intelligence and Applications*, p. 259–271, 2014.
9. J. Qi, L. Zhang and Y. Liu, "ADTreesLogit Model for Customer Churn Prediction," *Annals of Operations Research*, vol. 168, no. 1, pp. 247-265, 2009.
10. G. Nie, W. Rowe, L. Zhang, Y. Tian and Y. Shi, "Credit Card Churn Forecasting by Logistic Regression and Decision Trees," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15273-15285, 2011.
11. G. Madden, S. Savage and G. Coble-Neal, "Subscriber Churn in the Australian ISP Market," *Information Economics and Policy*, 1999.

<sup>1</sup> **Anhelina Naumenko**

Magister

<sup>2</sup> **Anna Kolomiets**

PhD in Economics, Assistant Professor

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **SPECIFIC CHARACTERISTICS OF PROJECT MANAGEMENT IN THE BANKING SECTOR**

**Abstract.** Defined the main differences of project management in the banking sector, researched essential bank projects and the role of the national regulator in them.

**Keywords:** project management, banking sector, commercial bank, tools, techniques, practices, activity sector, toolsets.

Banking institutions should permanently develop new products and services in the era of total digitalization. One of the ways to stay competitive for banks is to meet customer needs quickly and on time. According to Kerzner, the use of the best project management practices leads to added business value, greater benefit realization, and better benefit management activities [1].

The business function of the banking institution features some differences compared to the business function of commercial companies. Project management in the banking sector should consider specific activities of banking institutions [2]. The general regulator for banking institutions in Ukraine is the National Bank of Ukraine, which also sets certain milestones for the development and areas of activities for banks.

At first glance, it seems that the essential project of banks is to develop new products and services for customers. But the National bank of Ukraine approves regulations to maintain the stability of the banking sector. These regulations require banks to implement projects to improve effectiveness and decrease risks (Table 1).

Table 1  
Characteristics of essential projects of banks

	Product	Service	Software
Who requires the development?	Market, competitors	Market, competitors	National bank of Ukraine
Project results users	Customers	Customers, bank employees	Bank employees
Projected results of implementation	Increase in the number of customers	Increase in customer loyalty	Increase in effectiveness and reduction of risks
Examples	Deposit and credit products	Mobile applications for internet banking	Early warning systems

Usually, the banking institution works on the development and implementation of several projects at the same time. Employees from different departments are involved in these projects. This is the main reason why banks need to implement matrix organizational structures. Every department should have project managers to implement projects effectively.

The methodology of project management varies depending on the needs and results of a project. In case of development or implementation of services or products for customers, project managers and teams use Agile methodology for project management, which can be explained by the need to make changes during the development of projects. Competitors may implement options that are more profitable or convenient for customers, or some changes may take place in the market of banking services. Those are the main reasons why banks prefer Agile methodology for these needs [3].

With regard to the development and implementation of software for bank's employees, project managers prefer to use the Waterfall methodology. For these projects, the National bank of Ukraine clearly defines requirements that match regulations for results of projects: necessary functional, the roles in the system, project deadlines, etc. Banks can only make insignificant changes to these requirements, which are usually made in the initial phases of projects.

Also, the main feature in the project development for products, services, and software in the banking sector is that, in the case of delay or refusal of implementation of results of the project for products and services, the consequences for the banking institution will be delayed in time. That is, customers will gradually move to another bank where more convenient service or more profitable products. Accordingly, the bank will feel the financial consequences after a certain period.

At the same time, ignoring or having problems meeting the deadline requirements of the National bank of Ukraine will have immediate implications. Depending on the importance of the project this may result in a penalty charge, a decrease in a rating of the bank during the next audit by the National bank of Ukraine, or license revocation.

At project management in the banking sector, the manager must bear in mind the type of project result: whether it is a product, service or software; and who sets requirements: market or national regulator? It may influence not only a choice of a project management methodology but a functioning of a bank as a whole.

#### **References:**

1. Kerzner H. Project Management Best Practices: Achieving Global Excellence / Harold Kerzner. – New York: Wiley, 2015.
2. A Guide to the Project Management Body of Knowledge (PMBOK ® Guide) 6th Edition. PMI, 2017, 586p.
3. Badewi A. The impact of project management (PM) and benefits management (BM) practices on project success: Towards developing a project benefits governance framework / Amgad Badewi. // International Journal of Project Management. – 2016. – №4. – C. 761–778.

**<sup>1</sup>Ivan Oberemok**

PhD, associate professor

**<sup>2</sup>Nataliia Oberemok**

PhD, associate professor

*<sup>1</sup>Kiev National University of Construction and Architecture*

*<sup>2</sup>Taras Shevchenko National University of Kyiv*

## PRIORITY OF VALUES OF PROJECT STAKEHOLDERS

Changes in the market and increased competition lead to the fact that companies that are unable to adapt to these changes lose customers and, as a result, cease their activities. To prevent this from happening, many company executives think about the reasons for customer refusals and loss of competitiveness of their products [1-3].

There can be several reasons for the increased competition in the market, and first, it is a freer transfer of information between countries and territories, as well as a freer exchange of goods and resources. Attempts to restrict access to goods or services of foreign companies do not solve the problem, since the penetration of information also leads to the conquest of the market. Clients are acquainted with new, more advanced products or services of competitors and, as a result, the requirements for local manufacturers increase significantly. This situation is observed in all areas, even in those where until recently local producers dominated. The key differences between competitive products include their quality and complexity [8-10].

It is quite difficult to create a product that would meet the customers' requirements for quality parameters and new technical capabilities and at the same time have an affordable cost price. This is due to the growing differentiation of customer groups [6]. Each of the emerging social groups of customers has its own specific requirements for the company's products. When developing a product, companies are forced to focus on the requirements of customer groups for the quality and technical features of the product. Every year, the differentiation of social groups of clients only increases. Within the framework of already existing social groups of customers, separate subgroups are distinguished, the relation of which to the company's products is insignificant, but differ. Attempting to satisfy both subgroups of customers with the same type of product results in a drop in sales in both subgroups.

In addition to the emergence of new social groups of customers, the attitude towards the company's products in already existing subgroups is changing. This is due both to the directed work of competitors to conquer the market, and to the work of independent media. Information about new developments and product capabilities of

other companies forces customers to think about changing the supplier company and switching to another product.

To meet the changing requirements for the quality and technical complexity of products, companies conduct market research. And if earlier these marketing studies were carried out only when developing a new product, now companies are forced to conduct them on an ongoing basis. By introducing business processes of constant monitoring of the interest of social groups of customers in the products of the company, they try to maintain their leading positions in the market. Based on the results of marketing research, projects are initiated to amend the line of goods and services of the company. Marketing reports are also a document on the basis of which new products and services of the company are developed.

Analysis of current trends shows the orientation in conducting marketing research on customer values. Analysis of the values of target social groups allows you to formulate requirements for the product. By analyzing the values of social groups, the company's marketers assess the degree of influence of products on the needs of social groups of customers.

Implementing a competitive approach in a company has its own complexities. Changing requirements from customers demand the creation of unique products. The implementation of unique products is possible only if the project approach is used. The implementation of new requirements for quality and technological complexity of the product is possible only by using innovative technologies. The choice of a design approach for creating new products is due to the orientation of the approaches, methods and tools of this approach towards creating unique products using unique technologies. In this way, it is possible to ensure competitive production by considering the processes of creation and production of each batch of products as projects [4,5].

In the framework of project management of the consumers of the project products, not only the end users of the main project products are considered, but all the stakeholders of the project. Various stakeholder groups are involved in the project. Each of the stakeholder groups has its own set of values and has different attitudes towards the project product. The project team must create a project product that ensures that the priority values of all project stakeholders are met [7].

It is possible to provide a competitive advantage for a company by creating products that meet the values of stakeholders. The purpose of the research is to describe the approaches and methods for determining the priority values of stakeholders.

#### Research objectives:

- Principles for prioritizing values, project stakeholders.

- Development of a method for assessing the prioritization of the values of project stakeholders.
- Development of approaches to using the method within the project management processes.

**References:**

1. Kerzner, H. Project Management: A Systems Approach to Planning, Scheduling, and Controlling [Text] / H. Kerzner. – Ed. 10. – USA, New Jersey: Wiley, 2009. – 1120 p.
2. P2M «Program & Project Management for Enterprise Innovation» [Electronic resource]. – Project Management Association of Japan, 2016.
3. The Standard for Portfolio Management [Text]. – Ed. 3. – Project Management Institute, 2013. – 189 p.
4. Oberemok, I., Oberemok, N., “The stakeholder wide assessment of the impacts of accepted values of stochastic functions of project implementation on their success” International Journal “Information Theories and Applications” 25(4) (2018) 389-396.
5. Oberemok I., Oberemok N. “Development of a schedule of the project on the property board of values stakeholders” Journal of Technology and Exploitation ISSN 2451-148X in Mechanical Engineering Available online at: Vol. 4, no. 1 (2018) 36–42.
6. Morozov V.V., Khrutba Y.S., Khrutba A.S. “Peculiarities of stakeholder interaction in environmental projects” Management of complex systems development, Vol. 37, pp. 32-39, 2019.
7. Oberemok I.I. "Analysis of modern methods of project management of software development." International scientific-practical conference "Innovative development based on technological maturity in project management". Abstracts. K.: KNUBA: (2004) 56-57.
8. Chimshire V.I. "Quantitative assessment of the value of the project activity product." Management of complex systems development. 25 (2016) 80-85.
9. Bushuyev S., Kozyr B., Zapryvoda A. "Agile methods of product formation and the results of an innovative project" // XIV International Scientific and Technical Conference «Computer Sciences and Information Technologies», September 17 – 20, (2019): 136-141.
10. Yehorchenkova N.I. Awareness Management of Stakeholders During Project Implementation on the Base of the Markov Chain / Yehorchenkova N.I., Verenych O., Sharovara O., Dorosh M., Voitsekholvska M., Golyash I.// Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), September 18-21, 2019. Metz, France. P.259-263 (Scopus)

**Isus Raichuk**

Post graduate student, Department of management technologies of Information technology faculty

*Taras Shevchenko National University of Kyiv*

## **MODELS OF DIGITALIZATION OF BUSINESS PROCESSES OF PROJECT-ORIENTED ORGANIZATIONS BASED ON ARTIFICIAL NEURAL NETWORKS**

**Abstract.** The approaches to the integration of artificial neural networks into business processes chains of project-oriented organizations are analyzed. Three kinds of business processes are described. An application of the value-oriented methodology for the adjustment of feedforward neural network are proposed. Key practices of Kanban are analyzed. Conclusions regarding the conducted research are made.

**Keywords:** project management, artificial neural networks, business processes.

The field of application of artificial neural networks is expanding every year. To date, they can be used in almost all the areas of scientific and applied researches.

The purpose of this research is to analyze the different approaches to the integration of artificial neural networks into business processes chains of project-oriented organizations. Let outline the general characteristics of such an organization:

- the project team participants often work together;
- more organization resources are involved into the working projects;
- project managers are more independent and have considerable powers;
- organization items that report directly to the project manager or provide different additional services are often present [1].

Business process, as a set of interconnected resources and activities that has clearly defined get in and out, and as a result create a value, will be understood. There are three kinds of business processes:

- 1) management processes – business processes that operate the functions of systems. Cooperative and strategic management can be the example of management process;
- 2) the main processes – business processes which comprise the prime businesses of the company and create the major revenue streams. For example, supply, production, marketing, and sale;
- 3) the processes of providing – serving the main business processes. For example, bookkeeping, personnel and information management.

As is well known, artificial neural networks are used for solving the same type tasks, so their integration in the business processes is acquitted; such business processes solve the same type tasks (the searching tasks and decision making). Let's describe the most popular types of artificial neural networks. According to the structure, artificial neural networks are divided into three main types [2; 3]: feedforward neural network, recurrent neural networks (neural networks with feedback), completely linked networks.

In the feedforward neural network, neurons group into layers: an input layer, hidden layers, and an output layer. In such kinds of neural networks each neuron of a certain layer has the direct link to the next layer neurons. The links among neurons do not form the loops. The feedforward neural networks are mostly used for the classification of images, function approximation, prediction, management, etc.

In the recurrent networks the neuron state can influence its future state. In some neural recurrent networks, neurons can link by themselves, while the other recurrent networks – only indirect backward linkages: if an “A” neuron can send a signal to the input of a “B” neuron, the “B” neuron can send a signal to the “A” neuron the same way. The recurrent networks do not always have the accurate defined input and output neurons. In most cases, the recurrent neural networks are applied to the pattern identification, classification, associative memory, noise-tolerant signal transmission.

Completely linked neural networks allow the connection among all the neurons, except connection to links. Besides, the links have to be symmetric. A well-known example of such networks is a self-organized map [3]. Completely linked neural networks are mostly used for the formulation of associative memory, classification.

The neural networks are also classified by the types of studying [3]. The following three studying algorithms are: studying with a teacher, reinforcement studying, and a studying without a teacher.

The study with a teacher. For each input vector there is a target vector which represents a necessary output. These vectors are named a studying pair. The neural network usually studies on the several numbers of studying pairs. An input vector is represented; after an output of network is calculated, a result is compared to the proper target vector. The difference (a mistake) with a help of backward links brings into the network; weights are changed according to the algorithm which minimizes a mistake. The vectors of the training set are given gradually, the mistakes are calculated, and the weight is selected for each vector, until the mistake around the whole educational area reaches an acceptably low level which is defined as one of the training parameters.

While reinforcement studies, after finishing the training sequence, a network gets a reply, which defines whether the result of calculations is either correct or incorrect.

The study without a teacher does not need a target vector for outputs and does not require the comparison with proper answers. The training set consists of only input vectors. The training algorithm selects the weight for neural networks, the way the coordinated output vectors are got. It means a processing of similar input vectors produces the same outputs. As a result, the study defines the statistical properties of the training set and groups similar vectors into classes.

The neural networks can be classified according to the various types of problems that the neural network is able to solve [3]. There are regression and classification problems. For the regression problems the information at the output of the neural network is unbroken. When the classification problem is being preceded at the output of the neural network, discrete data is received. Feedforward neural networks are commonly used for solving the regression problems, whereas an application of any kinds of neural networks is possible while solving the classification problems [4; 5].

In fact, for solving the problems of forecasting implementation the progress of business processes which are regression problems, the application of feedforward neural networks is optimal. The study of such a feedforward network is easily implemented with the use of the study with a teacher algorithm. The application of value approach methodology (figure 1) which is described in the article [6] for the more accurate analysis of the situation, modeling the progress of business processes, and making the optimal decisions is proposed.

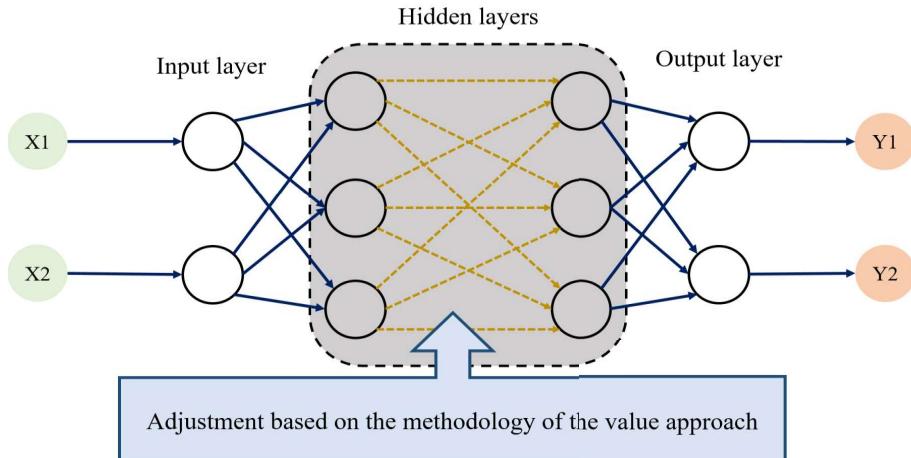


Figure 1 – An application of the value-oriented methodology for the adjustment of feedforward neural network

The modern project management world proposes a great amount of standards and approaches of project management (PMBOK, PRINCE2, P2M, Agile, etc). Since the digitalization projects are IT-projects which need a flexible methodology for their management, it is preferable to use Scrum [7]. For improving the understanding of the general overview of the project process, it would also be appropriate to use Kanban except for Scrum [8]. Scrum is widely used all over the world with the purpose to: discover and find the actual and profitable markets, technologies and product opportunities; develop new products and improve them; release the products and update them several times a day; develop and support cloud technologies (online, safely, on demand) and other environments for the product usage.

Therefore, due to the results of the analysis of Scrum framework, it is possible to make a conclusion that the framework is the one optimal decision for the implementation of digitalization projects. So, the usage of Scrum to manage the research project (the project of digitalization business processes based on the neural networks) is advisable.

In the project of digitalization of business processes based on the neural networks, Scrum is proposed to be used together with Kanban. Speaking of Kanban, it is necessary to remind its six key practices:

1. Visualize. Visualizing working processes help to understand the planning changes properly and implement them according to a schedule.

2. Limit the tasks during the working process. This concept foresees that in the project the system of “pulling” on some parts of the tasks or during the whole working processes of projects has to be used.

3. Operate the working stream. Each turn between the states in the stream has to be monitored, measured, and reported. Therefore, a constant monitoring is needed. Active stream management allows evaluating the positive and negative changing effects in the system.

4. Make the process clear. Until mechanisms or process become clear, it is often difficult or impossible to perform a discussion concerning its improvements.

5. Create feedback mechanisms. The organizations which did not create the second level of the feedback mechanisms (operation review) usually do not observe the improvement of the process beyond the local team level.

6. Make improvements in cooperation; make experiments involving models and the scientific method.

Thus, Kanban provides maximum clarity and realization of implementing project process, whilst Scrum provides the flexibility of project management.

To conclude, the basis for the digitalization of business processes is a selected feedforward neural network. The training of this network worth realizing based on a study with a teacher algorithm using the value approach methodology. The standard of management for digitalization business processes of project-oriented organization with the usage of neural networks project is proposed to base on the framework Scrumban. This project management approach allows providing a flexible, efficient and adaptive management of the project-oriented organization and also a constant improvement of the business processes by using of neural networks.

#### **References:**

1. Batenko L. P., The main forms of the projects organization structure, 2020. URL: <https://library.if.ua/books/96.html>.
2. Kriesel D. A., Brief Introduction to Neural Networks, 2007. URL: [http://www.dkriesel.com/en/science/neural\\_networks](http://www.dkriesel.com/en/science/neural_networks).
3. Rajesh Bordawekar, Bob Blainey, Ruchir Puri, Analyzing Analytics, 2015, Morgan & Claypool Publishers.
4. Osovsky S., The neural networks for the information processing. Translated from polish I. D. Rudynskii, 2002, Finances and statistic.
5. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An introduction to statistical learning, 2013, Springer Science+Business Media, New York.
6. Timinsky A., Raichuk I., Method of value-oriented stakeholder management of digitization project, 2019, Project management and development of production, Volodymyr Dahl East Ukrainian National University, 3(71).
7. Ken Schwaber, Jeff Sutherland, The Definitive Guide to Scrum: The Rules of the Game, 2017.
8. Max Rehkopf, Kanban, 2020. URL: <https://www.atlassian.com/agile/kanban>.

<sup>1</sup>**Anna Samonenko**

Master student, department of Technology Management

<sup>2</sup>**Oleksii Yehorchenkov**

Associate Professor, department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## PECULIARITY OF RPA PROJECTS

Robotic Process Automation (or RPA) is a form of business process automation technology based on metaphorical software robots (bots). According to the prediction of McKinsey & Company, Robotic Process Automation (RPA) will showcase an economic impact of \$6.7 trillion by 2025 [1].

In traditional workflow automation tools, a software developer produces a list of actions to automate a task and interface to the back-end system using internal application programming interfaces (APIs) or dedicated scripting language. In contrast, RPA systems develop the action list by watching the user perform that task in the application's graphical user interface (GUI), and then perform the automation by repeating those tasks directly in the GUI [2].

RPA tools have strong technical resemblance to GUI testing tools. These tools also automate user interactions with the GUI, and often do so by repeating steps performed by a user. RPA tools differ from testing tool in that they allow data to be processed between multiple applications. For example, receiving email containing an invoice, extracting the data, and then typing that into a bookkeeping system.

Here, are some reasons why Robotics Process Automation is advantageous

- A human can work average 8 hours a day whereas robots can work 24 hours without any tiredness.
- The average productivity of human is 60% with few errors as compared to Robot's productivity, which is 100% without any errors.
- Robots handle multiple tasks very well compared to a human being [3].

The next phase of RPA technology will combine artificial intelligence (AI) and machine learning to make it more powerful. AI gets more intelligent over time by assessing the data that RPA can provide. Instead of just completing a programmed action, RPA, with the help of AI is able to determine what action to take based on the data.

So according to this let us have a look at the peculiarity of PRA projects:

- RPA is similar to the traditional automation, because they both use software integration to automate business processes. However, RPA is more sophisticated than traditional automation. Traditional automation use APIs while RPA use software bots in order to understand the user behavior at the UI level. For example, in case of RPA, bots understand the user steps and repeat them. As the result of comparing these approaches it was defined that traditional automation has various limitations, such as restriction in application. Therefore, it is challenging to use traditional automation in

legacy systems. On the other hand, because RPA usually works on the UI layer, it is allowing avoiding these limitations. In RPA, no programming skills are required, user should know how to use RPA tools only, but in traditional automation, automation tests using programming languages should be written. Therefore, it makes RPA faster in implementation, than traditional automation. In addition, RPA offers high level of customization if compared to the traditional automation. RPA has integration with different applications like ERP, email, calendar and other.

- Typical lifecycle of RPA projects has 4 phases – Analysis, Bot Development, Testing and Support. On the phase of analysis team works together to analyze the business process for RPA development. After this team starts working on developing the automated workflow for requirements. On the phase of testing team runs testing cycles in order to understand the quality and find defects. On the last phase, team provides continuous support and helps in defect resolution.

As the result of comparing these approaches it was defined that traditional automation lifecycle to the RPA, it has more phases – analysis, design, development, testing, deployment and support. In general, both lifecycles are similar, but there is a difference on the stage of the development. In traditional automation, on this phase automation tests are implemented, but in the RPA is bot development.

It was also defined management of the RPA projects to the other IT projects, show no big difference. PM should decide what most suitable methodology for team is and implement it in the project. So there is no any specific methodology developed for the RPA project.

#### **References:**

1. The evolution of RPA: Past, Present and Future. URL: <https://www.uipath.com/blog/the-evolution-of-rpa-past-present-and-future>.
2. Robotic Process Automation. URL: [https://en.wikipedia.org/wiki/Robotic\\_process\\_automation](https://en.wikipedia.org/wiki/Robotic_process_automation).
3. Robotic Process Automation. URL: <https://www.businessdevelopment.blog/robotics-process-automation>.

<sup>1</sup> **Anton Sazonov**

Postgraduate

<sup>2</sup> **Natalia Yehorchenkova**

Dr. Tech. Sc. Associate Professor

<sup>1,2</sup> Kyiv National University of Construction and Architecture

## CONCEPT OF ORGANIZATION OF PORTFOLIO OF PROJECTS AND PROGRAMS OF FINANCIAL COMPANIES

**Abstract.** Addressed the problem of managing the portfolio of projects and programs in financial companies. Proposed a new method of portfolio management based on a process approach.

**Keywords:** portfolio of projects and programs, process approach, PDCA cycle, strategic management.

Today, Ukraine's financial sector is in a difficult macroeconomic climate under the rapidly developing stage, forcing financial companies to change traditional management methods. Competitive pressures for consumers and innovative discoveries in the information and technology environment force financial companies to act on new models and management methods. To increase customers' demand and proper quality of services, financial companies implement projects to improve the efficiency of production processes, create new products and services.

The world experience of improving quality has shown that it is possible to get certain outcomes only within the management process. However, the entry into the workforce of new products and services requires implementation of best global practices of project and process management, since these activities are limited at the time, resources, and coverage area. Taken into account the implementation of certain projects /programs requires a focused results-based management portfolio of projects/programs, thus there is a need for complex and systemic effectuation of project and process management methods to create an appropriate balance between them. It is also worthwhile to note that the existing issue lies in the fact that financial companies are always operating by process activity, that concentrates the relationship between many areas of activity by performing processes in which the outputs (results) of some are inputs (resources) for others, transferring the results of some to others creates added value, which ultimately affects the quality and cost of services. And the main models of project management are mainly focused on project organizations, in which the creation of a new unique result or product is the main source of profit.

Thus, there is a need to find an effective balance of methods and models that are combined to achieve maximum symbiosis of the two approaches to management and ensure the stable development of financial organizations in the current economy.

The authors are invited to suggest new methods and models of projects and programs portfolio in a financial institution through the use of the principles of process management, notably:

1) Sufficiency – the process should be regulated and clearly defined within the limits;

2) Appropriateness – the process should be carried out by the documents regulation;

3) Quality – the process should be completed with a result that meets the expectations of the process consumers (both internal and external).

4) Making the decisions based on facts - decisions on the effectiveness of activity should be based on systematic analysis of actual indicators and reliable certificates.

5) the principle of continuous process management by using the Shuhart-Deming cycle – PDCA (Plan – Do – Check – Act) [1].

6) According to the author's opinion, the PDCA principle can be used as the main method that needs improvement to build the cycle of operation of the project portfolio management process for its effective management. However, scrutiny of the latest research of the management process portfolio also should be taken into account [2].

According to the authors, the most important goal of managing the portfolio of projects for financial companies is the implementation of the strategy. Consequently, the project portfolio management process must meet the following conditions:

1) ensure the implementation of the Strategy;

2) ensure effective interaction of financial company units in the actualization of the portfolio of projects and programs;

3) to provide informing and reporting to stakeholders on the implementation of the portfolio of projects and programs.

The overall model could be used to achieve strategic goals and benefits is set out in Figure 1. Model for Achieving Strategic Goals and Initiatives.

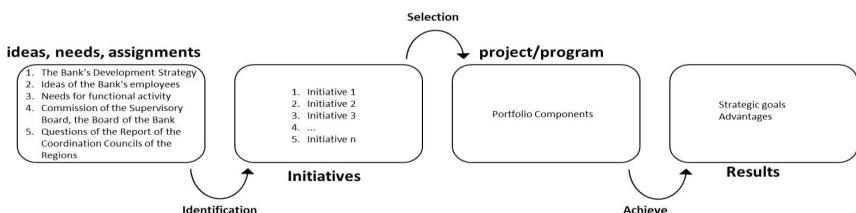


Figure 1 – Model of achieving strategic goals and initiatives

Thus, having examined the latest researches of process and project management there are positive conclusions on the implementation of one and the other approaches. However, there is a need to research new methods of project portfolio management, which will allow performing the process of managing the portfolio of projects in financial companies. In my view, such researches, can have successful results and be useful for use in the environment of financial companies.

#### References:

1. Deming, W. Edwards (1986). Out of the Crisis. MIT Center for Advanced Engineering Study – 419 pp.
2. Project Management Institute. The standard for portfolio management/14 Campus Boulevard Newtown Square, Pennsylvania 19073-3299 USA, 2017. – 127 pp.

<sup>1</sup> **Tetiana Shelest**

Assistant Professor, Technology Management Department, Faculty of information Technology

<sup>2</sup> **Anita Rudenko**

Bachelor of Applied Programming, Applied Information Department, Faculty of information Technology

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## ANALYSIS OF PREREQUISITES FOR THE APPLICATION OF IT PROJECTS IN CONSCIOUS CONSUMPTION MANAGEMENT

Our study aims to examine the prerequisites for the use of IT and their impact on the management of conscious consumption.

At this moment, the fast fashion business model is widespread in fashion technologies. It promotes the rapid production of cheap clothing following new fashion trends. Thanks to committed eco-activists, consumer preferences are changing, and the opposite model, ethical fashion, is gradually spreading. This business model ensures sustainable production, operation and disposal of clothing, footwear and accessories. Among consumers this model manifests itself in the form of conscious consumption. Conscious consumption is a thoughtful approach to shopping, responsibility for ethics and environmental friendliness, as well as reducing consumption [1]. On average, today the population buys 60% more clothes and accessories than in 2000, and 400% more than four decades ago. In the reporting documents for 2017, H&M stated unsold clothes worth \$4.3 billion [2]. And the luxury brand Burberry in 2018 burned its own products for \$37.8 million, confronting counterfeiting and reducing the level of exclusivity [3]. This consumption behavior is harmful to the environment. For example, in 2018, this industry sector emitted 2.1 billion tons of greenhouse gases (Fig. 1) - this is more than in the same period produced by all aircraft and ships in the world [4,5].

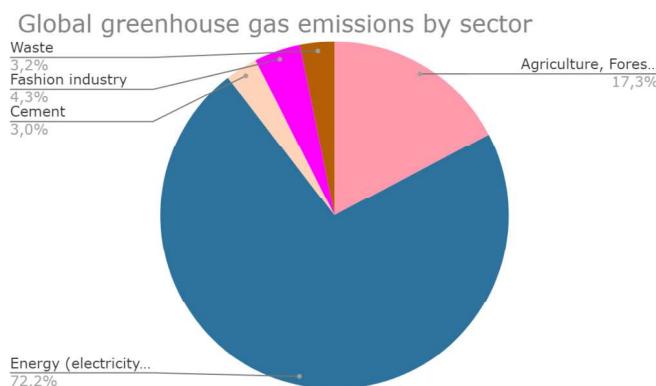


Figure 1 - Greenhouse gas emissions for 2018

Source: grouped by author by [4,5]

Nowadays conscious consumption is getting more popular, people are becoming more conscious and brands are gradually adapting to them. Buyers happen to be the main driver of change. At the end of 2018, the financial company HSBC surveyed more than 8500 different brands and it turned out that almost a third of them began to move to more ethical and environmentally friendly production [6]. A study of Ukrainian brands has shown that there are already brands that help improve the environmental and ethical situation in the country.

Continuing our research, we aim to find an innovative way to teach people to choose things wisely. The digitalization of society contributes to this, indeed, IT technologies can effectively help spread the idea of conscious consumption. We believe that it is necessary to initiate the development of a mobile application that will allow you to succeed in finding the matching clothes. The functionality of the mobile application enables selecting clothes and visualizing a combination of different pieces, which helps avoid impulsive and unnecessary purchases. The mobile application will also be a database on environmental-friendly and ethical companies, who will likely invest in the IT project on creating a mobile fashion application. This IT project will speed up the process of managing conscious clothing consumption.

Analysis of the Ukrainian brands shows that more and more companies are investing in the future. And companies can consciously use existing budgets, launch IT projects and attract human capital for positive transformations in the scope of conscious consumption.

#### **References:**

1. Sofia Pylypyuk It's easy to be sustainable in Berlin, 2018. URL: <https://www.the-village.com.ua/village/city/city-experience/276125-sustainable-fashion>
2. Daria Kosareva How luxury has become a responsible consumption. 2018. URL: <https://theblueprint.ru/fashion/industry/new-luxury-consumption>
3. The Guardian Burberry destroys £ 28m of stock to guard against counterfeits, 2018. URL: <https://retailers.ua/news/menedjment/7795-brend-burberry-za-posledniy-god-unichtojil-good>
4. McKinsey & company Fashion on climate, 2020. URL: <https://www.mckinsey.com/industries/retail/our-insights/fashion-on-climate>
5. Hannah Ritchie Sector by sector: where do global greenhouse gas emissions come from? Our world in data, 2020. URL: <https://ourworldindata.org/ghg-emissions-by-sector>
6. Varya Barkalova How does the fashion industry affect the environment? 2019. URL: <https://theblueprint.ru/fashion/industry/moda-i-ecologiya>

**<sup>1</sup> Grygorii Steshenko**

P.h.D, associate professor, Department of management technologies of Information technology faculty

**<sup>2</sup> Anatolii Buhrov**

Undergraduate student, Department of management technologies of Information technology faculty

**<sup>3</sup> Daria Horban**

Undergraduate student, Department of management technologies of Information technology faculty

**<sup>4</sup> Yuliia Timrova**

Undergraduate student, Department of management technologies of Information technology faculty

<sup>1,2,3,4</sup> Taras Shevchenko National University of Kyiv

## BASIC METRICS OF STARTUP EVALUATING

**Abstract.** The purpose of the article describing the key metrics in product development for startup projects.

**Keywords:** startup, key metrics, business value

### Why is that important?

Nowadays, lots of entrepreneurs present different sorts of metrics and indicators that illustrate the health and potential of their companies. Also, there are the same metrics that are being interpreted in different ways. As a result, it makes it difficult to understand whether the business is doing ok.[1]

Key metrics in product management:

**ARPU** (Average Revenue Per User) is the main indicator for predicting revenue. But it's not that straightforward, there are cases when companies projected revenue based on inflated ARPU, resulting in consistently falling behind their target.

ARPU analysis allows us to assess the correctness of pricing. It allows us to decide if we believe the numbers we see.

**CLV** (Customer lifetime value) - metric that shows us what profit the average client brings during his “life” (time client works with the company). One of the most famous venture capital investors, David Skok, said that many start-up companies fail just because they don't know about CLV. Or they know, but are too lazy to calculate. Thus, for these startups, the cost of acquiring a new customer outweighs the lifetime cost.[2]

**CAC** (Customer acquisition cost) - metric that shows how much it will cost to attract a new customer. This is essential, because if you know the value of a customer, then you know how much you can spend on attracting him. Well, if you spend on

acquisition less than one third of CLV (customer value), this is a generally accepted ratio.

Knowing the value of customers makes you understand how effective an advertising channel or a contextual campaign is and whether there are actions required to optimize it.

**MRR** (Monthly recurring revenue) - another indicator that speaks for itself, but the particular importance in this metric is that it is not an actual revenue, it just shows how fast a company is growing. Generally, it shows how much money we would receive from our entire customer base if they all paid every month and did not use discounts and promotions. It shows the real traction of a startup.[3]

But we should be aware of cases when attracted by the discount, customers made payments for a long period in advance leading to an increased revenue followed by promotion that was beneficial for clients. In this scenario companies simply drag customers' future months payments to the promotion period which entails a very noticeable drop in monthly payments in future.

### Cohort analysis

With this analysis, you can quickly notice changes. If the user database is large enough, its behavior changes slowly. If the flow of new interactions has halved in some months, you may not notice it right away due to the use of customers and their payments. Measuring the number of the conversion to paying customers gives an understanding of how successful the actions were in the field of contextual advertising, marketing, and also shows trends in customer behavior. Companies cannot influence such global events as rate races, but based on such information, they can adjust their plans and manage activities.

We'll conclude with a short description of why process performance and efficiency management is the focus of many companies today and how these five metrics are used by specialists around the world. The above metrics are of immense value to a business and should by no means be overlooked. Avoiding these critical metrics can lead to a premature failure. It is essential to understand that revenue is just one of many indicators.[2] If a company wants to have a complete vision of the health of the business it needs to master different approaches to examine it.

### References:

1. A Guide to the Project Management Body of Knowledge (PMBOK® Guide) Sixth Edition (2017). Project Management Institute, USA, 756 p.
2. Steve Blank. (2006) The Startup Owner's Manual: The Step-By-Step Guide for Building a Great Company, Microsoft Press, USA, 608 p.
3. The Seven Startup Metrics You Must Track // Young Entrepreneur Council. Available at: <https://www.forbes.com/sites/theyc/2014/06/20/the-seven-startup-metrics-you-must-track> (accessed 7 November 2020).

**<sup>1</sup> Oleh Suprun**

Assistant

**<sup>2</sup> Nina Klimenkova**

Bachelor student

*<sup>1,2</sup> Taras Shevchenko National University of Kyiv, intellectual technology department*

## **IT AUDIT AS A KEY COMPONENT OF INFORMATION SYSTEMS EFFECTIVENESS AND DATA SECURITY**

**Abstract.** In modern world, alongside with development of new information and calculation technologies, it is very important to ensure their effective work in combination, same as to provide high level of data security and protection on different scales. This paper provides the basic concept of IT audit and short description of modern methods used worldwide.

**Keywords:** information systems, IT audit, data protection.

Rapid development of the information society and the application of digital technologies in various spheres of human life and activities constantly pose new challenges and increase the requirements for the quality and efficiency of the methods and technologies used. For example, in recent decades the amount of information processed every hour in a variety of companies has increased dozens of times, requiring the development of new methods of big data computing [1] that use classic methods and at the same time the latest approaches to artificial intelligence, such as neural networks and evolutionary technologies, alongside with pictures identification and computing [2].

An information technology audit, or information systems audit, is an examination and evaluation of the management controls within an Information technology (IT) infrastructure [3]. The IT audit is rather different from other audits, like financial and logistic, since they evaluate the effectiveness and correctness of existing systems, but IT audit also calculates most of possible risks in future, considering the external circumstances, and overall system security at the same time.

The first country that mentioned the information security as part of government law was Great Britain. In 1984, the first Data Protection Act (DPA) was created which defined and formalized the first provisions and the importance of ensuring the information security of citizens and their personal data. In addition to setting a precedent, the UK first passed a law on the "improper" use of information and technology, even unintentionally, in 1990. It considers three types of criminal offenses like using data without authorization.

Shortly after, other European countries passed the laws about information security and illegal data usage, most of them are unified and global security programs are functioning these days.

However, modern IT-audit doesn't cover only the problem of security, it also investigates the problems and risks that the company may meet in future. Therefore, different approaches and methods of IT-audit were created, the most popular are:

- COBIT 5 is a set of standards and guidelines developed by the Association for the Audit and Control of Information Systems (ISACA) in collaboration with the Institute for IT Management (ITGI), which can be used by IT auditors due to its comprehensiveness and completeness;
- ITIL is the best description of international practices in the public and private sectors. It contains information about IT services to the company, documenting the processes, functions and roles of IT Service Management (ITSM);
- Prince 2 is a project management methodology that covers quality management, control and project organization consistently and in line with objectives. Prince 2 is used for IT projects, as well as for many other types of projects;
- ASL is a description of best practices used to standardize processes in Application Management, the basis for creating and maintaining information systems and applications;
- ISO / IEC 27001 and 27002. The international security standard ISO 27001 contains requirements for information security management and the information security management system (ISMS). ISO 27002 contains recommendations for information security management based on best practices.

At the same time, despite the importance and rapid growth of IT industry in Ukraine, the problem of IT audit is still at the initial level that may cause significant issues in future. The most obvious problem is the vulnerability of government infrastructure to external attacks, same as insecurity of national companies, that was seen in 2014 and 2016 years. On the other hand, all companies from Europe, same as Asia and America, require from each of their partners' presence of certificates that indicate the level of company security, such as ISO 27001 certificates. The main problem right now is the disregard of IT security problem by the governance. Even considering some steps, that were made after large attacks, these actions are rather local and do not solve the problem as a whole.

#### **References:**

1. N. I. Didenko; D. F. Skripnuk; O. V. Mirolyubova. "Big data and the global economy." 2017 Tenth International Conference Management of Large-Scale System Development (MLSD) 2-4 Oct. 2017, DOI: 10.1109/MLSD.2017.8109611
2. Popreshnyak, S., Suprun, O., Suprun, O., Wieckowski, T., "Personal documents identification system development using neural network.", 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT 2018, p. 129-134, 8526757
3. Gantz, Stephen D. (2014). The basics of IT audit: purposes, processes, and practical information. 2014: Syngress, an imprint of Elsevier.

<sup>1</sup> **Alexander Timinsky**

PhD, associated professor, Department of management technologies of Information technology faculty

<sup>2</sup> **Nikolay Kerdun**

Undergraduate student, Department of management technologies of Information technology faculty

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## MS PROJECT AS A DIGITALISATION TOOL OF PROJECT MANAGEMENT SYSTEM FOR PROJECT ORIENTED COMPANIES

The intensive development of information and communication technologies in recent years, together with modern socio-economic trends in the world economy, has led to the formation of new requirements for planning and project management in various areas of human activity.

One of the prerequisites for the effective implementation of projects in recent times has been the use of project management tools and tools based on the use of new information and communication technologies.

In 2020, the activities of almost any modern project, regardless of the duration, scope, budget, intended goals and objectives of the project, cannot be presented without the use of modern software.

Regardless of industry affiliation, a significant part of organizations structure all their activities in the form of projects (engineering, construction, IT, consulting, ship and aircraft enterprises), since each result of their activities is unique. But for organizations that have established service delivery processes, it may be necessary to allocate individual work in the form of projects.

The leading classic project management standard PMBOK of the American Institute of Project Management PMI in its latest edition pays a lot of attention to the digitisation of project management [1].

The development of special software for project planning and management is primarily due to the need for maximum integration of the IT sphere and the most effective methods, tools and tools of project management theory.

Today, according to Gartner Group, the Microsoft Project system is recognized as the leader in the market for project management systems. The company owns 75% of the project management systems market. The platform MsProject implemented corporate systems in such large organizations as Deutsche Bank, Merrill-Lynch, Delphi Automotive, etc. This is possible thanks to the combination of not only lower (compared to competitors) price and high quality in Microsoft solutions. In addition, the product is included in the MS Office package familiar to almost everyone, making it easier for staff to work in a familiar environment. All this significantly reduces the time for preparing the state, as well as the timing of the implementation of the system.

Microsoft Project has indeed become actually a standard on the market, as a means of individual work of project managers, and is expanding in the segment of heavy solutions. For corporate customers, MsProject Professional is the best solution to

cross-departmental issues where the key requirement is automatic scheduling, progress forecasting, and performance tracking.

Microsoft Project is the best choice for organizations using a matrix management structure, that is, project teams include the interaction of employees from different departments.

It is worth implementing Microsoft Project in the activities of project-oriented organizations based on the use of another Microsoft development to implement their own products - the MSF standard [2].

And desktop Microsoft Project and other solutions in the field of calendar and network planning of Microsoft (server version of Microsoft Project Server, cloud solution Project-Online) with a correct and fully functional implementation will allow project-oriented organizations

- Formalize and systematize project activities;
- Develop and monitor project plans;
- Minimize the cost of resources of all types (material labor, financial) when implementing projects;
- Complete the project on time and within the budget;
- Manage project risks, plan and implement effective risk response;
- Manage the project portfolio of the project-oriented company;
- Improve the culture of project management;
- Increase the technological maturity of the company in the field of scheduling and network planning.

Any time-tested scheduling solution can benefit a project-oriented company. Microsoft Project is not the only such tool, but one of the busiest. Scientific and practical research on the implementation of calendar and network planning solutions in the activities of project-oriented organizations may relate to the following: development of models, methods and algorithms for introducing Microsoft Project into the activities of a specific project-oriented organization, taking into account its specifics; the development of scientific foundations for the establishment and development of a corporate project management system; Create creative project team management models based on digitalized project management tools. These tools will ensure the modernity, flexibility and efficiency of the project management system for project-oriented organizations.

#### **References:**

1. A Guide to the Project Management Body of Knowledge (PMBOK® Guide) Sixth Edition (2017). Project Management Institute, USA, 756 p.
2. Turner, M. (2006) Microsoft® Solutions Framework Essentials: Building Successful Technology Solutions, Microsoft Press, USA, 342 p.

<sup>1</sup> Alexander Timinsky

PhD, associated professor, Department of management technologies of Information technology faculty

<sup>2</sup> Maksym Patsyuk

Undergraduate student, Department of management technologies of Information technology faculty

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## TEAM MANAGEMENT MODELS OF SEO-OPTIMIZATION START UP PROJECTS

**Abstract.** The approaches to the formation and management of start-up SEO optimization projects are described. Hierarchical development of the start-up team of the SEO-optimization project are proposed. The main advantages of this structure are considered. The Tier Meetings Framework for coordination and management of project teams are described. Conclusions regarding the conducted research are made.

**Keywords:** Digital project management, start up project, organizational structure, Tier Meetings Framework, SEO optimization.

Decades ago, the World Wide Web was used almost exclusively to find the information you needed. Today, the number of web portals exceeds one billion, and the number of users (according to the Facebook report for 2015) is more than three billion people, ie almost half of the world's population.

According to Forrester Research Inc., 81% of people find the right website through search engines. The data clearly shows that it is SEO that helps to attract more users to the site. Moreover, people who find a web resource through a search engine are the most interested target audience, because they enter keywords in the search field, on the basis of which search engine optimization is formed. Yes, the high role of SEO as an internet marketing tool is undeniable [1, 2].

We describe the approaches to the formation and management of start-up SEO optimization projects. At the beginning of the project, the main backbone of the team looks like this (Fig. 1):



Figure 1 – The initial team structure of the SEO-optimization start-up project

At the stage when the start-up project begins to scale [3], the number of tasks increases, new people are recruited. In this case, there are usually two options for

further development of the team - linear and hierarchical. An example of the linear development of the start up project team is shown in Figure 2.



Figure 2 – Linear development of the start-up team of the SEO-optimization project

The main disadvantages of the linear structure of the team are:

- Less time available per employee;
- It is difficult to scale;
- Team development is much slower.

Figure 3 shows an example of hierarchical development of the start-up team of the SEO-optimization project.



Figure 3 – Hierarchical development of the start-up team of the SEO-optimization project

The main idea of the hierarchical structure is to divide workers into subcommands. Thus, junior employees are assigned to a more experienced specialist.

The main advantages of this structure:

- Rapid team development – Middle Specialist hires a Junior Specialist and hone his skills;
- Backup – you always have people who know what is happening with a particular site, can replace it during the holidays, etc.;
- Delegation of tasks and responsibilities;
- Better individual development plans.

Regardless of the structure of the team (linear, hierarchical), in order to understand in which direction the team is moving, it is necessary to set adequate KPIs.

For an SEO startup, the following KPIs are best, which are easy to understand and easy to calculate:

- Site traffic;
- Site keywords;

- Search engine rankings;
- Number of external links.

For projects that already generate traffic and revenue, the best KPIs will be:

- Earned funds;
- Number of sales;
- ROI.

When the company has already gathered more than 5 full-fledged SEO teams, it is recommended to use the Tier Meetings Framework for coordination and management of these teams [4] (Figure 4).

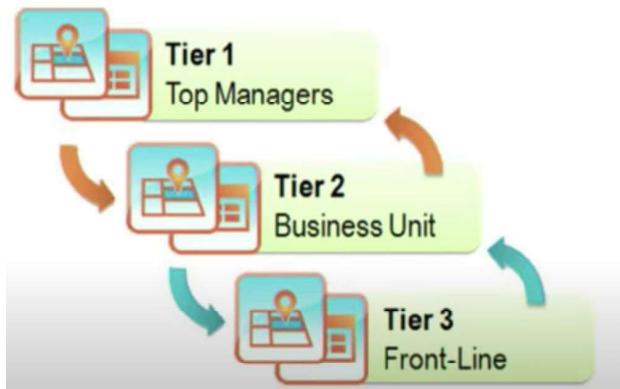


Figure 4 – Tier Meetings Framework structure

The main idea of the Tier Meetings Framework is regular rallies of specialists of the same level with specialists of the highest level. Thus, problems and tasks are transferred from one level to another. Consider the Tier Meetings Framework on the example of an SEO campaign:

#### Tier 1 Meeting.

Link builders hold a rally between link builders to discuss all the issues, problems and difficulties that may arise during the implementation of their KPI.

#### Tier 2 Meeting.

A rally between link builders, SEO specialists and Team Lead, at which top-level specialists (SEO and Team Lead in this case) analyze the problems and successes of lower-level specialists (link builders).

#### Tier 3 Meeting.

Rally between the Leads of various teams from the CMO (Chief Marketing Officer), where serious issues such as: sales, revenue, ROI, etc. are brought up for discussion. During such a rally, Team Lead can bring to the top management problems that prevent his team from achieving greater success.

#### Tier 4 Meeting.

Rally between senior positions in the company. CMO meets with CEO, Product

Manager, Chief Finance Officer. At such a meeting, participants discuss global issues that affect the business: what is happening with each area, whether any help or additional funding is needed.

#### Tier 5 Meeting.

Rally between business owners.

We define basic rules of Tier Meeting:

- 1 file in which everyone works, separate for each level;
- 1 rally for everyone;
- 1 hour time limit;
- Once a week, regularity.

Advantages of using the Tier Meetings Framework in SEO start-ups:

- Company synchronization;
- Transparency;
- Exchange of experience;
- Ability to take the problem to the next level;
- Interaction of different teams on a specific issue;
- Regular communication.

In general, the use of the proposed approaches (hierarchical team structure, application of KPIs, level meetings) in team management of start-up SEO-optimization projects is aimed at improving the efficiency of project teams, increasing their flexibility, simplifying the decision-making process.

The above helps to increase the probability of success of start-up projects, which are known to be implemented in a high-risk and poorly predicted market environment.

Future research in the chosen direction will be aimed at developing methods for managing start-up project teams. In particular, the current areas of further development include the following: digitalization of the competency model of project team members, digitalization of business processes of SEO optimization project management, development of separate algorithms for project management, project development models, technological maturity development models of the company implementing SEO optimization projects.

#### References:

1. Odden Lee, *Optimize: How to Attract and Engage More Customers by Integrating SEO, Social Media, and Content Marketing*, Wiley, 2012.
2. Dover Danny, Dafforn Erik, *Search Engine Optimization (SEO) Secrets*, Wiley, 2011.
3. Peter Thiel, *Zero to One: Notes on Startups, or How to Build the Future*, Currency, 2014.
4. Nidhi Shah, *Ways to make tiered meetings effective*, Kaizen Institute, 2014.  
<https://in.kaizen.com/blog/post/2014/01/23/ways-to-make-tiered-meetings-effective.html>.

<sup>1</sup> Volodymyr Yas

Magister

<sup>2</sup> Anna Kolomiets

PhD in Economics, Assistant Professor

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **IMPLEMENTATION OF PROJECTS IN THE MEDICAL FIELD USING BIG DATA AND WATERFALL METHODOLOGY**

**Abstract.** Digitalization of medicine is a promising direction that not only simplifies the work of clinics or laboratories, but can also save human lives. Data processing allows making diagnoses more accurate and helps using the accumulated information to predict possible "waves" of diseases. Big data, in addition to the already known and common tasks, can also be used to combat diseases and track the growth of epidemics.

**Keywords:** Big Data, healthcare, digital development, methodology, management.

Today, significant progress has been made in the tools and cost of data collection and storage. The most relevant problem in data management in the Big Data environment is the development of algorithms for complex analysis and interpretation of data in real time. Permanent collection and analysis of information at the level of advanced analytics not only allows you to notice at an early stage any deviations and anomalies in the readings, but also to identify hidden patterns.

The high relevance of the implementation of Big Data technologies in medicine is associated with new trends in the relationship between doctor and patient in the format of mobile medicine technologies. Medicine is becoming more patient-oriented, for which prognosis, disease prevention and personalization of treatment are important. Standard medical services lag behind the demands of patients who want to receive tools that allow you to control more and more physiological parameters and who are increasingly involved not only in the process of continuous monitoring of their health, but also in health management [1].

Images occupy a significant amount of information in medical information systems, which is an important source of data in the diagnosis, evaluation and planning of therapy. The amount of medical image data can range from a few megabytes per study to hundreds of megabytes per study. Storing such data for a long time requires powerful storage systems, as well as fast and accurate algorithms that provide the ability to automate decision-making processes that are performed using this data. In addition, if during the diagnosis, prognosis and treatment use other data sources obtained for each patient, there is a problem of developing effective methods to cover a wide range of data. Another area where digital development will stimulate returns for pharmaceutical companies - the development and familiarization of methods of advanced analytics, including predictive analysis, construction of simulators and variable models. In this environment, the winners will be those pharmaceutical companies that will be able to influence the algorithm of clinical decision-making,

offering doctors sound information about the best treatment options, based on advanced analytics [2].

The implementation of projects in the medical field requires clearly defined processes and precise deadlines, and therefore the cascade model is best suited here, as each stage of project creation continues the previous one and cannot begin until the previous one is completed. In this case, arbitrary transitions forward or backward are not allowed, and the stages do not overlap each other [3]. This model has a rigid structure. The Waterfall methodology originates in sectors such as construction, finance, and aircraft construction, where there is a clear sequence of work that can only begin after the completion of previous ones.

The main stages of product development: system and software requirements - these requirements are defined in the product requirements document; analysis - reproduced in models, schemes and business rules; design - the internal software architecture, ways of realization of requirements are developed; code writing - software integration; testing - checking the final product, finding errors and defects in the program code; operations - the product is adapted to different operating systems and updated to correct errors found [4].

In the field of health care, there are many people who are familiar with the cascade model approach and confirm its effectiveness. This is directly true for project managers who want to see a sequence of future tasks in the form of a Gantt chart, which provides a visual description of the project. Some health projects that may require final approval by the Food and Drug Administration (FDA) —for example, the design or development of a medical device — may benefit from the Waterfall methodology. This is due to the fact that the FDA, along with other government agencies, requires quality and design verification at various stages of development, which is easier to document using a cascade model [5].

Thus, Big Data technologies can provide (thanks to the increasingly advanced data processing apparatus with accurate conclusions) the desired tool to the doctor, which will take into account the characteristics of each unique organism. And the use of the "Waterfall" methodology will significantly increase the chance of successful completion of the project with a correspondingly high quality.

#### References:

1. M. Kozin. Application of Big Data in medicine [Zastosuvannya Big Data v medytsyni], 2018. // URL: <https://blog.mednote.life/articles/primenenie-big-data-v-medicine>.
2. D. Reinsel, J. Gantz, J. Rydning. Data Age 2025: The Evolution of Data to Life-Critical Don't Focus on Big Data; Focus on the Data That's Big, 2017. URL: <https://www.seagate.com/files/www-content/ourstory/trends/files/ Seagate- WP-DataAge2025-March-2017.pdf>.
3. E. Samoylova Бессмертная классика Waterfall [Bessmertnaya klassika Waterfall], 2017. URL: <https://worksection.com/ua/blog/waterfall.html>.
4. Winston W. Royce Managing the Development of Large Software Systems // URL:<http://static1.1.sqspcdn.com/static/f/447037/6486244/1270928531727/waterfall+royce.pdf>.
5. B. Simon Healthcare Project Management 101: Experts Share Techniques, Examples and Future Outlook, 2020. URL: <https://www.smartsheet.com/content/healthcare-project-management>.

<sup>1</sup> **Alina Zharikova**

Student of the Master's Project Management Program

<sup>2</sup> **Viktor Morozov**

PhD, Professor, Head of The Department of Technology Management

<sup>1,2</sup> Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

## PROJECT MANAGEMENT OF DEVELOPMENT BUSINESS MESSENGER FOR COMMUNICATION WITH FOREIGN CLIENTS

Usually busy people have little time. Therefore, business messenger will help them to conveniently manage business, communicate with foreign clients just holding the phone in their hands. The business idea is to create the first business messenger to communicate with people from all over the world, edit documents from the phone and conveniently manage your own business.

This messenger will have cloud storage, it will be available to the user without downloading files it to the computer. Thanks to this, several users will be able to edit documents directly in the chat. Each of the users will see who, when and what edited in the document. Also, if one of your colleagues or partners is in another country and the user does not know the language, you can turn on the automatic translator feature and receive messages from any country in the selected language. Messenger will have the function of making collective audio and video calls, in an urgent meeting the user will not need to find a computer, and will be able to communicate with colleagues only with a phone.

The project has three main objectives:

1. The possibility of increasing the free time of business owners.
2. The possibility of increasing the number of foreign customers.
3. Reducing the chance of losing or damaging documents.

The project team consists of a project manager, business analyst, marketer, sales manager, two developers, two testers and a designer.

Project limitations affect the capabilities of the project manager. Usually projects have three main constraints: time, scope of work and budget. The developed project has the following limitations:

1. Time limit (the project must be implemented in 1 year).
2. Budget restrictions (a total of 884 thousand hryvnias was allocated for the project implementation).
3. Scope of work (the project has an approved scope of work that must be performed to implement the project).

The approved scope of work to be performed for the project is made out in the WBS. The WBS of project is used to detail project results to clearly demonstrate ways to achieve them. The first level of detail indicates the project itself, the second – the main results of the project, and the third – the main work that will help achieve the goals (figure 1).

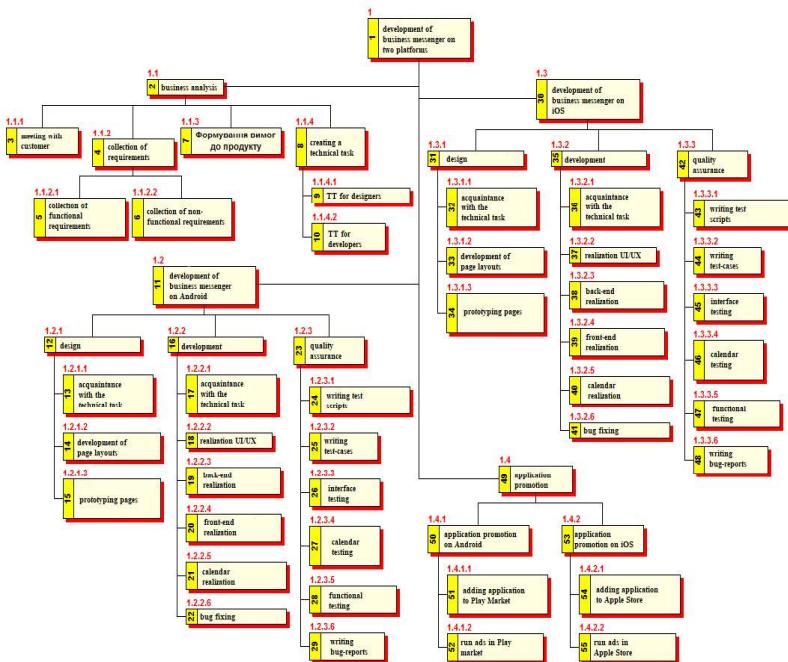


Figure 1 – WBS of the project

The organizational breakdown structure of the project (OBS) is a hierarchical structure of project management and shows the relationship between project participants (figure 2).

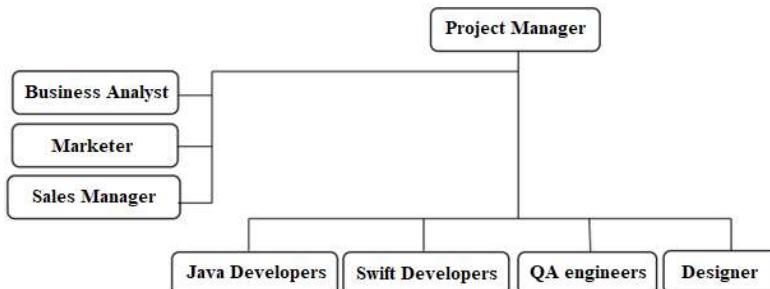


Figure 2 – OBS of the project

PM manages the development team, testers, designers, and is assisted in this by a business analyst, marketer and sales manager.

<sup>1</sup> Dmytro Zubets

VP of Business Development, IIBA Ukraine Chapter

<sup>2</sup> Grygorii Steshenko

P.h.D., associate professor

<sup>1</sup> Enterprise Solutions Principle, SoftServe

<sup>2</sup> Taras Shevchenko National University of Kyiv

## BUSINESS ANALYSIS IN UKRAINE

**Abstract.** The purpose of this article introducing the business analysis area in IT and product development. Authors cleared an ambiguity about the profession of a business analyst, which the standards and requirements are for this job, and what the organization has a corner office in this area of expertise.

**Keywords:** BABOK® v 3.0, business analysis, business analyst, The International Institute of Business Analysis™ (IIBA®)

**What is Business Analysis?** According to the latest version of the business analysis Bible – BABOK® v 3.0. Business Analysis is the practice of enabling change in the context of an enterprise by defining needs and recommending solutions that deliver value to stakeholders. Now, let's analyze this definition and translate it into simple words.

When we talk about practice, we mean the set of activities, tasks, processes, frameworks, tools, and so on. But there should be some reasons to use all this stuff – this magic question “Why?” All of these activities are about enabling change – getting from one state to the other. Let's look further at the definition. “By defining needs” – again, why? So, any change starts from defining why a person wants to move from the existing state of anything to the future state.

“Recommending solutions that deliver value” – this part of the definition answers the “What?” and “How?” questions. We have some needs to transform something from the as-is state to the future state. But how can we do it? Well, it's up to the business analysts to propose or recommend solutions. However, solutions should deliver value, because without value these solutions are useless. And the last part – “to stakeholders”. Stakeholders are people or organizations that will be affected by the change.

**Who is a Business Analyst (BA)?** The first answer that comes to mind is a professional who performs the business analysis. And it's correct!

According to the definition given in the BABOK®, a business analyst is any person who performs business analysis, no matter what their job title or the role in the organization is.

In fact, business analysts may have different titles: product owner, product manager, requirements manager, business architect, system analyst, etc. In many cases, these professionals perform not only business analysis activities and tasks. Almost any member of a project team can wear a business analyst's hat. We will dive deeper into

each of the six business analysis knowledge areas which are essential to the success of the business analyst's work.

In other words, a business analyst is a professional who can define and specify the business needs of the change, define the business case for the change, understand the as-is and the future states, propose the solutions, elicit and specify the requirements of all levels, and act as a liaison/a bridge/a translator between the business and the implementation team.

The big part of the BA skills and competencies are the so-called soft skills. A business analyst works with different stakeholders on both sides of the solution development and implementation processes. A business analyst communicates a lot with them, presents deliverables, drives meetings and workshops, and so on. That is why soft skills are a major part of the business analysis's competency model. A Business Analyst wears many hats: negotiator, diplomat, facilitator, skilled listener, motivational speaker, team leader, visionary, troubleshooter, and many more.

One of the main requirements for a business analyst is to be a leader, a person who is not afraid of problems and failures, and who takes responsibility. Another important thing, particularly for the Ukrainian market, which is mostly about outsourcing, is a good command of English.

A Business Analyst has to perform a huge range of tasks and to be able to do it in a correct and professional way. A business analyst should be familiar with many techniques. These techniques can be specific for BA and common for different business professionals. Here you can see the main techniques, which business analysts use in their daily work. For instance, requirements elicitation sessions can be conducted in the form of requirements workshops and focus groups, interviews, and brainstorming meetings. When a BA has to analyze the market, solutions, or customers, he/she can use benchmarking and SWOT analysis techniques. To analyze the as-is state, a BA uses observation, interface analysis, and document analysis. One of the core BA competencies – requirements specification – can be covered by such techniques as user stories, use cases, data modeling, and prototyping. To manage requirements the prioritization techniques are necessary. These are just a few core techniques. The more techniques a BA knows and uses the better. However, you do not need to be an expert in using each and every technique – it's enough to know where each technique can be used, and just read more about it in the book.

**Standards in a Business Analysis.** The International Institute of Business Analysis™ (IIBA®) is a non-profit professional association serving the growing field of business analysis. As the global thought leader and voice of the business analysis community, IIBA® actively supports the recognition of the profession and works to maintain global standards for the ongoing development of the practice and certifications [1].



Figure 1 – IIBA achievements

Since 2003, the International Institute for Business Analysis™ (IIBA®) has supported professionals working in business analysis and related business roles by maintaining global standards for the practice and certification, our Chapters, and through professional development and career resources, networking and volunteer opportunities. IIBA is a member organization of the Federation of Enterprise Architecture Professional Organizations (FEAPO)[2] [1].

#### IIBA Goals:

- Creating and developing awareness and recognition of the value and contribution of the business analyst
- Defining the *Business Analysis Body of Knowledge® (BABOK®)*
- Providing a forum for knowledge sharing and contribution to the business analysis profession
- Publicly recognizing and certifying qualified practitioners.

*A Guide to the Business Analysis Body of Knowledge® (BABOK® Guide)* is the globally recognized standard for the practice of business analysis. It includes such areas of business analysis practice:

- Business Analysis knowledge areas
- Tasks
- Underlying competencies
- Techniques and perspectives

The first version of BABOK was written by The Body of Knowledge Committee (IIBA) in 2005. In 2008 – 1.6, 2009 – 2.0. And the latest one was publicly redacted by teams of experts and practitioners and was officially published in 2015.

#### Certification opportunities:

In 2020 IIBA has 3 core levels of professional certifications:

- Entry Certificate in Business Analysis™ (ECBA™)
- Certification of Capability in Business Analysis™ (CCBA®)
- Certified Business Analysis Professional (CBAP®)

And 3 specialized:

- Certificate in Cybersecurity Analysis (IIBA®- CCA)
- Agile Analysis Certification (IIBA®-AAC)
- Business Data Analytics Certification (IIBA®-CBDA)

Business analysis skill ranks #6 in the top 15 skills companies need most in 2020 [4] and in future the demand on business analysis skills will be increased dramatically. Therefore, certification helps earn more money for the same job level. Jobs related to business analysis are projected to increase 14% from 2018 to 2028 [5].



Figure 2 – IIBA certification

**IIBA in Ukraine.** The IT industry has been strongly integrated in the worldwide tech ecosystem for the last ten years. More than 20 cities have created their own IT-clusters. The technology sector became the country's second export industry in 2018. Every year more than 1,000 events related to technology, start-ups and investments are held in Ukraine. Official launch of IIBA Kiev Chapter was in 2012, but the IIBA Ukraine chapter was launched in 2018.

Now we have 140 active members of IIBA Ukraine (ECBA -2, CCBA - 7, CABP -20). Participation and membership is free [3].

Certification results in IIBA Ukraine since 2012: ECBA -4, CCBA - 21, CBAP - 40

**Conclusion.** Business analysis evolved in independent area between management and product development. It helps better understand client need and deliver the best possible outcomes for stakeholders.

#### References:

1. <https://www.iiba.org/>
2. [https://en.wikipedia.org/wiki/Federation\\_of\\_Enterprise\\_Architecture\\_Professional\\_Organizations](https://en.wikipedia.org/wiki/Federation_of_Enterprise_Architecture_Professional_Organizations)
3. <https://ukraine.iiba.org/standard-chapter-membership>
4. CNBCTV18. <https://www.cnbc.com/business/linkedin-says-these-are-the-top-15-skills-companies-need-most-in-2020-5037841.htm>
5. United States Department of Labor, Bureau of Labor Statistics. <https://www.bls.gov/ooh/business-and-financial/management-analysts.html>

**E-COMMERCE, E-GOVERNMENT  
AND E-LEARNING  
TECHNOLOGIES**



<sup>1</sup> **Olena Bezlutska**

Doctor of Philosophy, Associate Professor, Deputy Head of the Department of Social and Humanitarian Disciplines and Innovative Pedagogy

<sup>2</sup> **Alona Leshchenko**

Doctor of Philosophy, Associate Professor, Professor, Director of the Science Park "Maritime Industry Innovations"

<sup>3</sup> **Alona Yurzhenko**

PhD in Pedagogy, Head of International Affairs Department

<sup>4</sup> **Alla Paziak**

Specialist of International Affairs Department

<sup>1-4</sup> *Kherson State Maritime Academy*

## **INFORMATIONAL VISUALIZATION ON E-COURSES OF HIGHER MARITIME EDUCATIONAL INSTITUTIONS**

**Abstract.** The article is devoted to the problem of informational visualization on e-courses of higher maritime educational institutions. It is highlighted that one of the effective means of visualization of information content is infographics in e-courses. The types of infographics are described. The actions to develop effective infographics are listed. It is concluded that the use of infographics contributes to an increase in the effectiveness of the educational process, forms the interest of students in studying the course.

**Keywords:** LMS MOODLE, informational visualization, e-course

Infographics as one of the effective means of visualization of information content is used in almost all spheres of public life. Higher maritime education is no exception. In our opinion, infographics is an integration model of presenting information as a result of high-quality data processing, which is subject to acquaintance and study by the interested audience. Teachers use infographics when explaining new learning material and to update basic knowledge.

Due to the rapid development of information and communication technologies, a number of special tools for information visualization have appeared. You can use special resource templates ([canva.com](http://canva.com), [piktochart.com](http://piktochart.com), [Easel.ly](http://Easel.ly), [Vizualize.me](http://Vizualize.me)) to create infographics. The positive aspects of using these services are the optional need to understand the basics of design, the ability to use professional templates in the library, customizing images and formatting, the ability to embed on your resource (platform). The most popular tools for creating an educational video are: [bandicam.com](http://bandicam.com) and [icecreamapps.com](http://icecreamapps.com), etc.

According to the classification of types of infographics and specifics of free economic zones, the most used among teachers are the following types:

1. flow chart (algorithm in the form of blocks that are interconnected in the form of lines and indicate the sequence);
2. the timeline (the cadet can trace the chronological sequence of the process, the phenomenon);

3. useful bait (set of useful information, reference material);
4. versus infographic (comparison of certain processes, models);
5. photo infographic;

To qualitatively create information infographics for future specialists in the maritime industry teachers can use an algorithm:

1. to determine the purpose of the infographic and choose the type according to the classification;
2. to collect data in accordance with the subject of the discipline;
3. to systematize information, organize the presentation of basic and additional information with the help of schematic sketches, choose the sequence of placement;
4. for visual perception it is also necessary to determine the color palette of the elements (not more than 6 basic colors);
5. to follow the process of creating infographics using a specially selected service

While using different services normally following actions should be done:

6. to create an account for the service;
7. to view and select a template by theme or create a new design;
8. to add text and images according to the content of the topic of discipline (or to download this data);
9. to format the text (choose the size, style) and image (size, placement);
10. -to save and copy the link for placement on the e-course of the discipline on the LMS Moodle (or to embed the infographic using another way);

Because of COVID-19 pandemic situation the educational process of all higher educational institutions of Ukraine including maritime ones came into distance format. Kherson State Maritime Academy (KSMA) uses Learning Management System (LMS) MOODLE to maintain its educational process. An important issue became the presentation of material of lectures, practical lessons, independent works etc. It became harder to motivate students because distance education requires student self-discipline. One of the tools to interest students is to visualize the information, present it in more interesting way.

Except those, all the activities of e-course (Assignment, Book, Chat, Choice, Database, External tool, File, Holder, Forum, Glossary, H5P, HotPot, IMS content page, Lesson, Quiz, SCORM package, Survey, Wiki, Workshop) has the option to display its description on e-course main page.

LMS MOODLE also allows to embed the images of gif animation to divide different types of activities in e-course.

The video can also be embedded in the course main page.

However, it should be noted that the disadvantage of information visualization may be in a simplification in understanding ambiguous interpretations of something, the presentation of information is too structured, without additional details. The cadet must understand that the visualization of information is an auxiliary element, and to study the discipline is not enough just pictorial, schematic forms of presentation.

The analysis of scientific knowledge and pedagogical practice demonstrates that

the infographics in e-courses is an effective instrument. We'll analyze the data on the example of English for professional purpose e-course for future marine transport specialist (first-year student). The final goal of e-course was to form the communicative competence of students. By analyzing the data of success after using the e-courses with infographics, one can observe that the current state of formation of the communicative competence of future marine transport specialists of the 2019-2020 at KSMA is better compared with 2018-2019 academic year.

According to the results, we see an increase in the success (by 21%) and knowledge quality (by 13%). Qualitative indicator of success was taken as the number of students by "good" and "very good" multiplied by 100% and divided by the total number of students. An absolute success indicator was taken as number of students by "good", "very good" and "sufficient" multiplied by 100% and divided by the total number of students. The total number of future marine transport professionals taking part in the research is 64 students.

The data was taken from the processing of control testing results on LMS MOODLE of KSMA e-course "English for professional purpose" (Stop and checks activities). Stop and check is testing conducted by the end of every module of e-course and created with the help of quiz activity. Quizzes comprised the questions of various types, including multiple choice, matching, short-answer and numerical.

The statistical results of research have proved the idea on the positive impact of infographics in English for professional purpose e-course of higher maritime educational establishment. The research has also showed low level of digital competence of certain teachers. To solve this problem e-course with module about infographics creation was proposed. The prospects of further research are seen by authors in researching the specifics of infographics in various fields of scientific knowledge, as well as in the possibilities of using it for conducting e-courses to improve the qualifications of teachers of higher educational institutions.

#### References:

1. G. M. Kvon, V. B. Vaks, A. M. Kalimullin, A. R. Bayanova, A. R. Shaidullina, A. V. Dolzhikova, N. I. Lapidus, Developing the informational and digital environment of a university: Problem analysis and assessment. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(10) (2019) doi:10.29333/ejmste/109503
2. E. R. Amit-Danhi, L. Shifman, Off the charts: User engagement enhancers in election infographics, *Information Communication and Society*, (2020) doi:10.1080/1369118X.2020.1761858
3. V. Kasyanov, E. Kasyanova, Information visualization on the base of graph models, *Sci. Vis.* 6(1), 31–50 (2014).
4. A. Peña-Ayala, *Learning Analytics: Fundaments, Applications, and Trends*. Springer International Publishing (2017).
5. A. Peña-Ayala, Learning analytics: A glance of evolution, status, and trends according to a proposed taxonomy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(3): e1243. (2018). <https://doi.org/10.1002/widm.1243>
6. T. Downer, M. Gray, P. Andersen, Three-dimensional technology: Evaluating the use of visualisation in midwifery education. *Clinical Simulation in Nursing*, 39, P.27-32 (2020). <https://doi.org/10.1016/j.ecns.2019.10.008>
7. R. Eiris, M. Gheisari, Research trends of virtual human applications in architecture, engineering and construction. *Journal of Information Technology in Construction*, 22(9), P.168-184.

(2017).

8. V. Vasenin, A. Itkes, M. Krivchikov, E. Yavtushenko, ChRelBAC data access control model for large-scale interactive informational-analytical systems, Journal of Computer Virology and Hacking Techniques, (2020) doi:10.1007/s11416-020-00365-9.
9. G. Di Battista, P. Eades, R. Tamassia, I.G. Tollis, Graph Drawing: Algorithms for Visualization of Graphs, Prentice Hall, Cambridge (2009).
10. T. Kühl, S. D. Navratil, S. Münzer, Animations and static pictures: The influence of prompting and time of testing. Learning and Instruction, 58, 201-209 (2018) doi:10.1016/j.learninstruc.2018.07.006
11. J. Chicca, K. Chunta, Engaging students with visual stories: Using infographics in nursing education. Teaching and Learning in Nursing, 15(1), P. 32-36. (2020). <https://doi.org/10.1016/j.teln.2019.09.003>
12. O. R. Shefer, S. V. Kraineva, I. I. Bespal, Visualization of the formation of undergraduate competencies, Espacios, 40(29), 1-9 (2019).
13. R. Tamassia, B. Palazzi, C. Papamanthou, Graph drawing for security visualization, LNCS 5417, 2-13 (2009).
14. M. Nowakowski, Analysis and evaluation of information usefulness and user experience for content presentation in electronic media, Procedia Computer Science, 176 3654-3664. (2020). doi:10.1016/j.procs.2020.09.021
15. H. Chen, S. Liu, Practical ability training on visualization teaching design based on information visualization technology, 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE, pp. 517-520. (2018). <https://doi.org/10.1109/icitbs.2018.00136>
16. V.N. Kasyanov, I.A. Lisitsyn, Hierarchical graph models and visual processing, Proceedings of 16th IFIP Congress, pp. 179–182 (2000)
17. W. Xu, M. Shehab, G. Ahn, Visualization based policy analysis: case study in SELinux, Proceedings of the 13th ACM Symposium on Access Control Models and Technologies, pp. 165–174 (2008).
18. J. Montemayor, A. Freeman, J. Gersh, T. Llano, D. Patrone, Information visualization for rule-based resource access control, Proceedings of International Symposium on Usable Privacy and Security (2006).
19. Anwar M., P.W.L. Fong, A visualization tool for evaluating access control policies in facebook-style social network systems, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pp. 1443–1450 (2012).
20. K. Vaniea, Q. Ni, L. Cranor, E. Bertino, Access control policy analysis and visualization tools for security professionals, USM'08: Workshop on Usable IT Security Management (2008).

**<sup>1</sup> Iryna Domanetska**

PhD in Engineering Science, Associate Professor

**<sup>2</sup> Oleh Ilarionov**

PhD in Engineering Science, Associate Professor

**<sup>3</sup> Olena Fedusenko**

PhD in Engineering Science, Associate Professor

**<sup>4</sup> Oksana Vlasenko**

Specialist

<sup>1,2,3,4</sup> Taras Shevchenko National University of Kyiv

## DYNAMIC ANALYSIS OF THE QUIZ COMPLEXITY IN MOODLE

Testing occupies a prominent place in the modern educational process, which is characterized by the widespread use of modern information technology. Quizzes became the predominant technology of knowledge control and formed the methodological basis of training during the COVID 2019 quarantine. Testing provides an opportunity to check a significant amount of educational material, ensures for feedback between a student and educational content of the discipline, allows individualization of work with students.

Previous research in the field of building knowledge control systems has shown the need to divide quiz assignments by the level of complexity [1,2]. Without such a division of assignments, the assessment of knowledge lacks objectivity and often does not correlate with the true level of knowledge of students. It means that if a better student receives only difficult tasks, while a weak student – only easy ones, the assessment shows the same level of knowledge of both students, which is not true. Also, if students with the same level of knowledge receive assignments of different levels of complexity, the knowledge test will reveal a different level of preparedness, which is not objective. The distribution of tasks by the teacher according to the level of complexity introduces subjectivity in the process of assessing students' knowledge as what is considered as an easy task by the teacher is not just as easy for students.

Modern Learning Management Systems (LMSs) are usually equipped with the tools for statistical evaluation of test quality and indirect indicators of the complexity of quiz questions. For example, one of the most widely used LMS, Moodle, has built-in tools for calculating the statistical characteristics of quiz questions. This system uses statistical indicators, which are calculated using the classical (STT - Classical Test Theory) and modern test theory (IRT - Item Response Theory) [3,4,5]. These indicators are calculated based on the final response of the user; they do not consider the dynamics of the process of its formation. In such a model, the evaluation of a question to which the student did not answer at all and a question to which he answered incorrectly will be equivalent. The second aspect to pay attention to is the dynamics of changing the complexity of the quiz questions. If the complexity has a clear tendency to decrease from test session to session, this indicates that the task loses its practical significance and cannot be used to assess the level of knowledge of students. Such a scenario is especially common when using remote testing, when the teacher cannot

control the services active on the student's computer. Capturing information from the screen nullifies the control function of the test, even a massive bank of questions does not help.

To monitor such situations, we suggest the following:

- 1) use low-level access to the LMS Moodle database to obtain more detailed statistical information on the formation of answers to questions;
- 2) to implement dynamic analysis of the received information by specialized means of data mining.

Free Moodle Adminer software is used to access LMS Moodle database information. The LMS Moodle database has a complex structure (421 tables, 3822 attributes) [6]. To solve the problem of dynamic analysis of the complexity of quiz questions, we will need only a limited set of tables, namely a group of tables that directly describe the questions themselves, quiz questions and the students' results.

The values obtained from a query to the Moodle database are a table. Columns of the table - questions from the bank of quiz questions of this discipline, and rows - depersonalized test results ("0" - incorrect answer; "1" - correct answer). Data processing was performed by Rapidminer [7], which is an integrated environment for data preparation, machine learning, in-depth learning, text mining and predictive analytics. A quick analysis of these data revealed "easy" quiz questions (questions that all students answer) and "difficult" (questions that are often not answered correctly). The developed model of data analysis in Rapidminer provides the ability to track trends in the complexity of test tasks over time.

#### **References:**

1. Zinkovskyy Yu.F., Mirskykh G.O. Methodology of assessment of complexity level of academic quizzes // Proceeding of the National Technical University of Ukraine "KPI" 163 Series – Radiotechnics, Radio device building. – 2010. - #41, P.157-163
2. Zhunusakunova A.D. Approaches to determination of the complexity level of quiz assignments / A.D. Zhunusakunova. – Text – direct // Young scientist. – 2017. - #4.1 (138.1). – P.55-58. – Available at: <https://moluch.ru/archive/138/39700/>.
3. Lord F.M. Application of Item Response Theory to Practical Testing Problems. Hillsdale N-J. Lawrence Erlbaum Ass., 1980. – 266 p. 11.
4. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. With a Foreword and Afterword by B.D. Wright. The Univ. of Chicago Press. – Chicago & London, 1980. – 199 p.
5. Quiz statistics report // MoodleDocs/ - Available at: [https://docs.moodle.org/32/en/Quiz\\_statistics\\_report](https://docs.moodle.org/32/en/Quiz_statistics_report)
6. Moodle database framework and description of its tables. – Available at: <https://moodle.org/mod/forum/discuss.php?d=319762>
7. Rapidminer official web-site // Docs.rapidminer/ Available at: <https://docs.rapidminer.com/latest/studio/>

**Oxana Gradinari**

Phd student

*Tiraspol State University, Chisinau, Republic of Moldova***ANALYSIS OF EXISTING MODELS OF INFORMATION COMPETENCE**

**Abstract.** This article provides a comparative analysis of existing models aimed at the formation and development of information competence. The article identifies their strengths and weaknesses. As a result of the analysis, the structural content of information competence was determined.

**Keywords:** information competence, didactic models.

Analysis of scientific literature in the languages of international circulation on the design of a model of the process of formation and development of information competence (IC) made it possible to single out didactic models, developed to a greater extent from the standpoint of the systemic and content-activity approaches [1, 2, 3, 4]. The systematic approach, as a rule, is based on the structural content of IC, which is modeled in accordance with the effective development of the information field of future graduates, and therefore is based solely on the teacher-researcher's own experience.

So, for example, Zaitseva E. M. [5] in the structure of IC distinguishes cognitive, active and motivational components. Undoubtedly, this model makes it possible to determine the main directions of work on the formation and development of IC for students of vocational schools, but only taking into account the identification of this phenomenon with competence in the field of information and communication technologies (ICT) [6, p. 85]. This is explained by the fact that in this case, attention is focused only on the knowledge and skills of working with information that are acquired using a computer. Let us list them: (a) knowledge of the principles of information processing and work with a computer and software products used in the educational process (for example, knowledge of methods for processing the results of industrial practice at an enterprise); (b) skills in using information technology means (for example, mastering computer methods of information processing in professional activities). We consider this judgment erroneous.

Argument: despite the fact that ICT competence provides an opportunity to make a high-quality educational service mass and accessible to various categories of students and affects the motivation for success in achieving educational results, it does not give the necessary answer to the question: "How to learn to understand, express, produce, share, collaborate, create and innovate using ICTs?" We believe that in this case, it is necessary to activate the universal thinking skills. These include observation, consistency, the ability to formalize and abstract, analyze, and use abstract models. That is, there is a need, first of all, in the formation and development of IC [7, p.34].

In the work of the authors Kotenko V. V. and Surmenko S. L. [8] there appears an extended version of the pedagogical model, designed from the point of view of a systems approach. These researchers supplement the IC structure with a new component - reflection. The importance of including this component in the content of

the IC is determined by the ability to determine the attitude of the individual to himself and to the world, to his own practical activity.

If we consider the pedagogical model of V. V. Kotenko and S. L. Surmenko through the prism of this study, then it can undoubtedly be applied as one of the options for modernizing the process in training students in vocational schools. The proposed model, in contrast to the previous one, is aimed not only at the formation of ICT competence, but also at mastering the personality's abilities to extract knowledge from information using information tools.

In the scientific work of Khodotova M. I. [9] it is noted that at the basis of any pedagogical system, two initial concepts are clearly traced: a didactic task and an algorithm for its solution. In the content of the didactic task, the goal is put forward in the first place, the achievement of which is due to the pedagogical conditions and the available information of the intended activity. On this basis, in order to model the educational process, taking into account the formation and development of IC in students of vocational schools, we consider it expedient to single out didactic models that are designed according to the principle of the organizational-activity approach (Vasilyeva L. D. [10] and others).

Considering the formation and development of IC from the perspective of personality-oriented pedagogy, we highlight the work of Bogdanova V. A. [11], which focuses on such activities as: cognitive, value-motivational, communicative, technical and technological and reflective. These types of activities contribute to the formation of skills that contribute to the development of the student's personality.

When designing the author's didactic model of the process of formation and development of IC in students of vocational schools by means of curricular and extracurricular activities, we relied on the proposed models of the above authors. But we took into account the specifics of their professional training, characterized by the following intensive changes - an increase in the intellectualization of professional activity, the complication of labor and professional functions, the development of automated and mechanized labor.

Therefore, the structural content of IC should be supplemented with a communicative component, in view of the fact that one of the main tasks of the future worker is interpersonal relationships in professional activity through constructive communication and cooperation using digital tools. We also consider it advisable to single out the creative component that requires from the future specialist in blue-collar professions such personal qualities as: initiative, individual freedom, independence and responsibility, willingness to take risks and independence of judgment.

Thus, the proposed models can serve as a basis for creating your own didactic model, which will allow for the purposeful design of curricular and extracurricular activities.

**References:**

1. M. V. Goryacheva, Model for training information skills in the process of independent work of students outside the classroom. In: *Izvestia of the Russian State Pedagogical University. A. I. Herzen*, 2008. p. 368-372 ISSN 1992-6464
2. N. A. Afanasyeva, Model of the formation of information competence of future teachers of vocational training. In: *Bulletin of the Bryansk State University*, 2011. 4 p. ISSN 2072-2087
3. G. A. Krutchinina, V. A. Shilova, Fostering the Informational Competence of Students in Engineering Profils. In: *The Education and science journal*. 2013, № 1(2). p. 86-96. URL: <https://doi.org/10.17853/1994-5639-2013-2-86-96>
4. V. Kyva, V. Yahupov, Pedagogical modelling of development of information and communication competence of professors of distance learning military education system. In: *ARTTE* Vol. 7, No. 3, 2019 c. 224-232 ISSN 1314-8788
5. E. M. Zaitseva, *Management technology for the development of information competence of students of radio engineering specialties*: дис. .... канд. пед. н. Ижевск, 2007. 189 с.
6. S. A. VORONOV, The structure and content of competencies in the use of information and communication technologies. In: *Vestnik ChGPU im. I. Ya. Yakovleva. No. 3 (95). h.1*. 2017. p. 83-89 ISSN: 2073-5499
7. C. S. DOYLE, *Information literacy in information society: a concept for the information age*. New York: Syracuse University, 1994. 83 p. ISBN 0-937597-38-4
8. V. V. Kotenko, S. L. Surmenko, Information and computer competence as a component of professional training of a future teacher of computer science. In: *Bulletin of the Omsk State Pedagogical University*. 2006. 4 p.
9. M. I. Khodotova, Models of formation of information competence of graduates of physical culture universities. In: *Pedagogy: Traditions and Innovations: Materials of the 1 Intern. scientific. conf. Chelyabinsk*, 2011, p. 97-100
10. L. D. Vasilieva, Model of the formation of information and functional competence of students in grades 5-9 in a rural educational school. 2011. №17. URL: <https://urok.1sept.ru/статьи/210223/2>
11. V. A. Bogdanova, Formation of information-functional competence of schoolchildren in the process of implementing a personality-oriented model of the pedagogical process. All-Russian Internet Pedagogical Council. 2007. URL: <[http://pedsovet.org/component/option,com\\_mtree/taskviewlink/link\\_id,3843/Itemid,0/](http://pedsovet.org/component/option,com_mtree/taskviewlink/link_id,3843/Itemid,0/)

**Iryna Horbas**

PhD in Economics, Associate Professor

*Taras Shevchenko National University of Kyiv*

**"A STATE IN A SMARTPHONE" CONCEPT BY UKRAINIAN GOVERNMENT**

**Abstract.** Ukraine's digitalization drive is currently in its early stages, but main potential benefits are already obvious. As well as reducing the human factor in some public services, digitalization can provide a boost to a range of structural reforms and rapidly raise income levels while cutting government waste.

**Keywords:** transformation, digitalization, e-governance, e-services, Diia.

President Zelensky and the parliament elected in early elections have launched a "turbo mode" of state changes. At the iForum on May 23, 2019, the Head of state presented a digital action plan for the implementation of "A state in a smartphone" program in order to limit bureaucracy, digitalize the country's economy, simplify and transfer of public services to the Internet [4, 5]. He noted that it is necessary to implement e-governance at 4 levels: information about the state, communication and transactions with the state, involvement in the state management [1].

Deputy Prime Minister Mykhailo Fedorov is responsible for the digitization process and "transformation of the Ukrainians' way of thinking" through heading the Ministry of Digital Transformation. The main goal for the Ministry is to turn Ukraine into a tech nation by 2024 through 100% of public services are available to citizens and businesses online, 95% of towns and their transport infrastructure, social facilities have access to high-speed Internet, 6 million Ukrainians are involved in the digital skills development program, 10% share of IT in the country's Gross Domestic Product [2]. In addition to him, the issues of digitization are also handled by the Minister of the Cabinet of Ministers Dmytro Dubilet. To guarantee that the digitalization will cover the whole of Ukraine the Cabinet of Ministers will establish the responsible persons for the implementation of digital transformation at all levels of government: from the Ministry to the district administration (CDTO - Chief digital transformation officer).

On June 2019, the first expert discussion of the concept was hosted. Representatives of the Coalition of the Electronic State (65 different organizations and experts in the field of IT) provided their recommendations which were related to the development of tools for e-state and democracy, digital infrastructures, digital identification, cybersecurity, restoring order in state registries, electronic democracy and the importance of supporting Ukrainian technological start-ups [4].

During a meeting with government officials in July 2019 the representatives of the European Commission gave their recommendations to Ukrainian initiative: to provide broadband internet access for the entire territory of the country, to develop the country's digital identification system of citizens, to introduce effective management structure for the sphere's development and organize the implementation of the ambitious tasks set [3].

Experts from abroad have already promised to help in solving these issues.

The Japanese (Hiroshi Mikitani — the founder of Rakuten company and co-owner of Viber) have declared their participation in the “State in a Smartphone” program. The Estonian E-Governance Academy is supposed to work on a system enabling data exchange between the existing national registers and on the modernization of these registers. The e-Governance Academy and the Swedish International Development Cooperation Agency (SIDA) have signed (2014) a cooperation agreement to support the Ukrainian government in the introduction of e-governance. The agreement covers the development of policy documents, planning of e-governance activities, development of legislation, enhancement of e-capacity, promotion of involvement in the development of e-governance, and development of e-services for the population and businesses. The e-Governance Academy will implement the agreement in collaboration with experts from Estonia, Sweden and Moldova [10]. IT companies from Poland also ready to play a part in the implementation of the wide digitization program pursued by its eastern neighbour [9]. UNDP in Ukraine has supported "a state in a smartphone" initiative through commissioning the assessment of government transparency and Ukraine's readiness to work with open data.

Nowadays major active projects of e-governance by “A State in a smartphone” concept are the follows [2]:

- a mobile application and a website “Diia” combine all services in one “electronic window” provided by the state and serve as the main channel of digital communication between the government and the country’s citizens. The users have access to a driver’s license and vehicle registration certificate, a digital student ID and the ability to test a biometric passport and ID card. By the end of May 2020, more than 3 512 000 Ukrainians had downloaded this app, by the end of September 2020, there were more than 5 000 000 users [10];

- free info platform of digital literacy “Diia. Digital Education” [6] where experts and celebrities explain Ukrainians basic Internet safety rules and help them to start using the Internet and smartphones. In June 2020, the Ministry of Digital Transformation, in partnership with the Ukrainian Parliament Commissioner for Human Rights and UNDP (United Nations Development Programme), has prepared a new series of educational videos “Personal data protection” for public officials to raise awareness of legislation on access to public information. National digital literacy program of Government should reach 6 million Ukrainians in 3 years [2, 12];

- a state portal “Diia. Business” [7] proposes different services for entrepreneurs, as follows: a guide, a list of legal documents, online and programs services, 39 types of free consultations etc. The online registration of LLCs is made much easier – instead of 88 lines, only 46 lines remained in the application. More than 50 000 private individuals have registered business online at the portal in 2020. Till the end of 2020, the savings due to online business registration will more than UAH 17 million. Due to the use of the electronic format of registration, liquidation and change of data by private individuals Ukraine is expected to save up to UAH 255 million per year [10];

- online service “Diia. E-malyatko” provides 10 services of parenthood by 1 application. By the end of summer 2020 it was available in 10 cities of Ukraine: Kyiv,

Odesa, Lviv, Kharkiv, Kryvyi Rig, Vinnytsya, Lutsk, Mariupol, Zaporizhya, Rivne;

- online service for foreigners "Diia. E-residency" gives the opportunity to remotely access public services and do business in Ukraine from abroad, open a visa or bank account under a simplified procedure etc.;

- investment program "Ukrainian Startup Fund" [8] supports Ukraine-based innovation projects and helps the most talented Ukrainian entrepreneurs create successful companies at scale. By October 2020 there were more than 1800 applications and more than 30 Startups have been funded;

- "Diia. Green cards" brings 5000 high-qualified international tech professionals from all around the world which are available in Kyiv, Lviv, Dnipro, Odesa, and Kharkiv. The app gives permits for work without annual prolongation; work according to the Ukrainian labour legislation; emigration with the family members; the possibility to open company as the individual entrepreneur;

- "Diia City" (at the moment is in the discussion phase) will be a virtual model of a free economic zone for representatives of the creative economy. It will allow Ukrainian and foreign tech companies to open businesses and R&D centres in Ukraine under simplified procedures, to work according to the English Law, and pay 10% of tax etc [10, 11].

In this way, "Diia" is going to be a service of the new generation, a new philosophy of providing e-services and involving citizens to e-governance. "Diia" have already saved UAH 495 million for Ukrainians and provided UAH 2 billion in additional revenues through the elimination of "tax cuts" [3, 11]. It is important to continue to focus on the implementation of such comprehensive and breakthrough electronic services countrywide.

It should be noted that on the way to the digitalization of Ukraine, the following difficulties may arise:

- it is necessary to analyze how much the changes will cost, what practical benefits they will bring and how long will it take,

- the «digital state» should coordinate the work of central, regional, and local authorities by dividing their work into four areas: information (public access to data from institutions and agencies), communication (exchange of data with the state and the citizens), transactional (public services and electronic document management), personalization (content targeting depending on citizens' tasks),

- there are serious problems in the identification of citizens – the existing electronic signature and digital identification systems are based on outdated technologies and have a limited scope of application,

- driven by increased civic activism, Ukrainian citizens now seek to have more control over their lives, which significantly raises e-participation in the country

The creation of a digital state is a complex yet important process accompanied by the resistance of the system, outdated legislation, scepticism and often disbelief, but all the worldwide revolutionary projects have passed this way [1, 11].

**References:**

1. V. Zelenskyy, I dream about a state in a smartphone, 2019. Official website “President of Ukraine”. URL: <https://www.president.gov.ua/en/news/ya-mriyu-pro-derzhavu-u-smartfonu-volodimir-zelenskij-55585>
2. Official website of the Ministry of Digital Transformation. URL: <https://thedigital.gov.ua>
3. M. Fedorov, EU supports the implementation of “A State in a Smartphone” concept. URL: <https://www.president.gov.ua/en/news/yevropejskij-soyuz-pidtrimaye-realizaciyu-koncepciyi-derzhav-56337>
4. President during the first expert discussion of “A state in a smartphone” concept: I want rapid changes and guarantee political will. URL: <https://en.ukraynahaber.com/2019/06/13/president-during-the-first-expert-discussion-of-a-state-in-a-smartphone-concept-i-want-rapid-changes-and-guarantee-political-will>
5. Overview of 2019: new government, weak reforms, and expectations for brighter future. URL: <https://www.unian.info/economics/10813331-overview-of-2019-new-government-weak-reforms-and-expectations-for-brighter-future.html>
6. Official website “Diia. Digital Education”. URL: <https://osvita.diia.gov.ua>
7. Official website “Diia. Business”. URL: <https://business.diia.gov.ua>
8. Official website “Ukrainian Startup Fund”. URL: <https://usf.com.ua/en/>
9. Official website “Obserwator Finansowy”. URL: <https://www.obserwatorfinansowy.pl/in-english/new-trends/ukraine-in-a-smartphone-is-that-possible-2/>
10. Official website “Government Portal”. URL: <https://www.kmu.gov.ua/en/news/mihajlo-fedorov-ponad-50-tisyach-fop-vidkrito-onlajn-na-portali-diya-u-2020-roci>
11. Starting from September 2021, no institution of the central executive body will require any paper from citizens, 2020. Official website “President of Ukraine”. URL: <https://www.president.gov.ua/en/news/z-veresnya-2021-roku-zhodna-ustanova-centralnogo-organu-viko-64273>
12. Official website “United Nations Development Programme in Ukraine”. URL: <https://www.ua.undp.org/content/ukraine/en/home/presscenter/articles/2016/08/04/ukraine-jumps-by-23-in-un-e-government-development-index-/>

**<sup>1</sup>Rusudan Makhachashvili**

Doctor Habilitated, Associate professor, Head of Department

**<sup>2</sup>Ivan Semenist**

PhD, Head of Department

**<sup>3</sup>Anna Bakhtina**

Graduate student, Assistant Professor

<sup>1,2,3</sup>Borys Grinchenko Kyiv University

## **ICT TOOLS FOR FINAL QUALIFICATION ASSESSMENT SURVEY STUDY FOR EUROPEAN AND ORIENTAL LANGUAGES PROGRAMS**

The global pandemic and subsequent quarantine measures and restrictions have posed a challenge to the structure and procedure of university summative assessment process. Qualification assessment for Foreign Languages major programs in particular is a strict regimen process that involves different stages (oral and written exams, final project viva, internal and external review). This study seeks to analyze the practices of Borys Grinchenko Kyiv University digital qualification assessment for students of European (French, Italian, Spanish, English, German) and Asian (Mandarin, Japanese) Languages major programs, employed in the year 2020 due to quarantine measures. The survey and analysis of different ICT tools is used to translate real life qualification assessment practices into online blended format. The investigation also seeks to identify various groups of applied digital skills and collaboration skills, utilized through qualification assessment process by all parties (students, faculty and referees). Comparative results of ICT tools and practices efficiency for respondents of European and Oriental languages programs are provided.

Based on the activity profile (Final Qualification Assessment) a survey was conducted among the participants of the Final Qualification Assessment at Borys Grinchenko Kyiv University Foreign European and Oriental languages programs (Spanish, French, Italian, English, Mandarin Chinese, Japanese major) in order to assess the efficiency of qualification assessment transfer into digital format via various ICT tools employed.

The survey comprised of 12 questions total (multiple choice and scoring), divided into such categories: 1) questions on overall experiences of Final Qualification Assessment participants in all procedures, conducted via ICT tools; 2) questions on digital literacy skills, required of Final Qualification Assessment participants; 3) questions on soft skills, required of Final Qualification Assessment participants; 4) questions, aimed to conduct Efficiency Ranking of most widely used Final Qualification Assessment ICT tools.

The overall digital qualification assessment experience on the scale of 1 to 5 was defined as mostly agreeable (4) by 50% of respondents, most agreeable (5) by 29% of respondents and less agreeable (3) by 17% of respondents across all foreign language programs surveyed. Digital activities got overall rankings of 5-4 from respondents of European languages programs and 4-2 from respondents of Oriental languages programs.

The respondents identified all the ICT digital tools that they have to employ the most in digital qualification assessment process. The highest scoring ICT tools by all the groups of respondents of both European and Oriental language programs were: e-mail (93% of respondents), Google services (76% of respondents), videoconferencing services (84% of respondents), social media platforms (77% of respondents), automated testing systems and learning management systems (31% of respondents).

Respondents from European languages programs identify Information sharing as and overwhelmingly prominent (59,1%) across all ICT tools employed for Final Qualification Assessment. For respondents of Oriental languages programs Speech quality assessment features as prominent as Information sharing across identified ICT tools. The following is inferred as being due to the phonetical and tonal features of Mandarin Chinese and Japanese languages being essential to meaning comprehension and decoding, which is hard to recreate and evaluate in a digital communicative environment.

*Respondents of the European languages program* have assessed the dominant ICT tools requirements being Intuitive interface (28,8%), elementary digital literacy (26,9) and specialized software (17,3%). *Respondents of the Oriental languages program* have assessed the prominent ICT tools requirements (Figure 8) being Intuitive interface (31,8%), elementary digital literacy (31,8) and bandwidth and advanced digital literacy (9,1).

Video conferencing services (Google Meet, Zoom, Webex) score the highest efficiency ranking for synchronous communication (62.5% for top score 5), but get a surprising ratio of lowest score as well (18,9% for lowest score 1). Learning management systems (29,8% for top score 1) and Google services (25%) get a proportional highest score 5 for efficiency in Synchronous communication in the framework of Final Qualification assessment across European and oriental languages programs.

This sample ranking testifies to the following suppositions: a) the specificity of ICT use for transference of Final Qualification assessment into digital mode for foreign languages programs that may not be encountered outside of this activity framework; b) the specificity of digital literacy, featured by participants of Final Qualification assessment for foreign languages programs.

All procedures and scenarios of the Final Qualification Assessment activities for foreign languages at Borys Grinchenko Kyiv university have been successfully transferred to digital remote format with the use of various sets of ICT tools in the framework of the COVID-19 pandemic adjustments. This transference could serve as a best practice model for other universities of Ukraine and European countries both as an adaptable measure for prolonged lockdown and as a way to further advance of blended learning and further digitalization and democratization of educational process.

**<sup>1</sup> Victoria Mironova**

Candidate of Technical Science, Associate professor

**<sup>2</sup> Mykola Pyroh**

Assistant professor

**<sup>3</sup> Iryna Harko**

Candidate of Physical and Mathematical Sciences, Assistant professor, Senior lecturer

<sup>1,2,3</sup> Taras Shevchenko National University of Kyiv

<sup>3</sup> National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

## **METHODOLOGY OF BUILDING AGILE-EDUCATION PROCESSES IN HIGHER EDUCATION INSTITUTIONS**

**Abstract.** In this paper the technology of using the popular flexible software development methodology – agile – is considered for the development and construction of courses and semantic links of higher education disciplines for training specialists in the field of information technology. The authors analyze the basic principles of the Agile-Manifest and their application for the formation of the curriculum program. The influence of Agile-methodology on the academic groups success is considered and a statistical comparison with the classical approach to teaching is made.

**Keywords:** agile-methodology, scrum, education, Information Technology, training of specialists, professional training

Training of specialists in the field of information technologies requires the formation of educational programs and courses using modern technologies, programming languages, libraries and frameworks. However, the rapid growth and changing trends in the IT sector makes it impossible to create such a program even for four years of bachelor's degree [1]. That is, even if at the beginning of the educational cycle, the program was designed taking into account the most relevant approaches and topics, at the time of graduation, it becomes obsolete. To understand the need to adapt to such changes, it is worth estimating the frequency of programming languages standards updates and releases, which are 1-3 years for Java, 3 years for C ++ and from six months to 3 years in C#. And even more, programming libraries and frameworks are released almost each month.

Agile management – an iterative method of planning and managing projects and processes [2].

If we consider the process of training a specialist in higher education in terms of project management, and the specialist – as a unique product, the training can be approached from the standpoint of project management and apply to this process all modern and successful project management methodologies. Due to the success of the Agile approach in industry, software design, marketing and business, foreign experts in the field of education have tried to apply similar principles to the field of education, thus forming the Agile Schools Manifesto [3, 4].

The introduction of flexible methodologies in the educational process optimizes approaches in the training of future IT professionals [5, 6]. The implementations of the

flexible methodology of software development – Agile – in the process of development of educational and methodical support and the proposed approaches in teaching in higher education contribute not only to better understanding and mastering by students of theoretical and practical educational material. This approach also helps students understand the production processes in IT projects that are set up according to the Agile methodology, set themselves to work within Scrum, gives an initial idea of the complexity of the task, mandatory and additional product requirements – their knowledge of the discipline.

Agile and Scrum methodologies [7] allow to develop and implement a modern, adapted to the requirements of the labor market training program for IT professionals and helps to develop skills to find the necessary information and interact with peers, which will also help further training and improve their “soft skills”.

#### **References:**

1. D. Semyonov, “Ukrainian higher education is dead. Stay away from it”, Medium, 2016 [Online]. Available: <https://cutt.ly/Mymj2R5>.
2. Agile-software development manifesto, 2001 [Online]. Available: <https://cutt.ly/SyEftwh>.
3. V. Seminozhenko, “Modern educational policy will cause a lack of IT specialists”, Euro Education, 2014 [Online]. Available: <https://cutt.ly/Jymzx4M>.
4. P. Salza, P. Musmarra, and F. Ferrucci, “Agile Methodologies in Education: A Review. In Agile and Lean Concepts for Teaching and Learning”, Springer, pp. 25–45, 2019. Available: <https://cutt.ly/Niu2nzk>.
5. V. Mironova, and M. Pyroh, "AGILE methodology in education as the main tool for building modern educational disciplines of higher education", Economics and Management, №4 (76), pp. 124–130, 2017.
6. V. Mironova, M. Pyroh, and I. Harko, "Agile-methodology and informatization of the educational process on the example of project work in the discipline "Algorithmization and basics of programming""", in VII International. scientific and technical conf. Problems of informatization, Cherkasy, 2019, P. 30.
7. J. Stewart, C. DeCusatis, K. Kidder, J. Massi, and K. Anne, “Evaluating Agile Principles in Active and Cooperative Learning”, in Student-Faculty Research Day, CSIS, Pace University, P. B3, 2009. Available: <https://cutt.ly/piuBBUI>.

**<sup>1</sup>Nataliia Morze**

Doctor Habilitated, Full Professor,

Corresponding Member of the National Academy of Pedagogical Sciences

**<sup>2</sup>Rusudan Makhachashvili**

Doctor Habilitated, Associate Professor, Head of Department

<sup>1,2</sup>Borys Grinchenko Kyiv University

**DIGITAL COMPETENCE IN E-GOVERNANCE EDUCATION: A SURVEY STUDY**

Factors of social change culminate in the development of the so-called networked society. Subsequently, networked society calls for networked governance. The development of e-government in Ukraine is impossible without appropriate training of relevant qualified professionals. Based on e-governance the activity profile a survey was conducted among the stakeholders of electronic government institutions – in-service government officials and students of government management programs.

The paper objective is to assess the survey as to the needs and possible avenues of E-governance curriculum development for higher educational institutions, in-service government officials and general public. The Master's Program in E-government, developed within the framework of a Ukraine-Estonia joint project on e-governance curriculum implementation, will promote the state policy in the field of information, e-governance, development and use of national electronic information resources, elaboration of the information society.

Based on the activity profile (e-governance) a survey was conducted among the stakeholders of electronic government institutions – in-service government officials and students of government management programs. The survey comprised of 13 questions total (multiple choice and scoring), divided into such categories: 1) questions on overall experiences in e-government; 2) questions on the needs and modes of e-government education; 3) questions on e-democracy as a social framework (to be disclosed fully in the upcoming studies). 70 respondents total took part in the survey.

Group 1 of survey questions - *overall experiences in e-government* - yielded the following results across the board. The prevailing understanding of e-governance by stake-holders (Figure 1) is disclosed by the such top scoring concepts: 1) Round the clock access to information and e-services (35,7%); ICT implementation for interaction of state and community (22,9%); 3) ICT technologies use for corruption surveillance (20%).

Group 2 of survey questions - *the needs and modes of e-government education* - yielded the following results across the board. The educational formats (Figure 5), efficient or sought after in the area of governance digitization is distributed as such: one-off trainings and workshops (40%); persistent online courses (34,3%); webinars (22,9%).

When asked to assess the knowledge needed or lacking to use digital technologies (Figure 7), in-service and in-training governance stakeholders identified the following top scoring priorities: Digital services development (67,1); Digital data

bases operation (60%); Digital literacy and digital skills (58,6%); Digital workplace tools proficiency (48,6); Re-engineering of government services (44,3%).

Digital competences, mandatory for any modern in-service government official (Figure 8) were assessed by respondents according to the following ranking: Digital transformations management (50%); Electronic democracy and social engagement (41,4%); Cybersecurity basics (41,4%); Terminology mastery in the realm of digital governance (34,3%)

The evaluated educational components, skills and practices provide a groundwork for the estimated structure of E-governance education curriculum project.

Survey results in the area of digital competence and structures of knowledge, sought after or in demand by e-governance stakeholders in specialized education, correspond directly to the comprehensive frameworks of digital competences, elaborated and tested in the recent decade.

The survey results corroborate the informed decisions behind elaboration and implementation of a comprehensive curriculum project of an integrated E-governance Master's program.

In the framework of globalization, information society development and social strife against corruption in government, traditional methods of interaction of state and local government with citizens and business become less sufficient. To avoid the emergence of corruption components and to create new and more convenient methods of access to information and services, the state is implementing e-governance - a form of government which provides a new level of open cooperation between the state and society, due to the widespread use of modern ICT, supplying a full range of public services for all categories of citizens and enterprises. The use of new ICT in public government determines the need for training highly qualified specialists in the field of public administration and management, in possession of the ultimate, up to date ICT skills.

Borys Grinchenko Kyiv University proposes a master's program of e-government under the auspices of the Ministry of Education of Estonia and in collaboration with the Tallinn University of Technology.

The curriculum project outline presupposes 4 stages of implementation: 1) Resources accumulation; 2) Theoretical premise of teaching e-democracy development; 3) Teaching e-democracy development methodology development; 4) E-democracy and e-governance studies dissemination.

E-governance curriculum development project corresponds fully to the National standard of e-government implementation in Ukraine. The 8th Framework project of European Commission Horizon 2020 academic collaboration incorporates a specific creed of Europe in the Changing World studies, which unfolds into a range of problematic issues open for project studies and development, including Understanding Europe - Promoting The European Public and Cultural Space, including civil society development as an operative foundation for e-democracy elaboration. E-governance comprehensive curriculum development is also in keeping with key priorities of the 9th Framework project of European Commission Horizon Europe.

**<sup>1</sup>Nataliia Morze,**

Doctor of Pedagogical Sciences, Professor, Corresponding Member of the National Academy of Pedagogical Sciences of Ukraine

**<sup>2</sup>Oksana Strutynska**

PhD in Pedagogy, Associate professor, Vice-Dean of Scientific Work and International Activities, Faculty of Informatics

**<sup>1</sup>Borys Grinchenko Kyiv University**

**<sup>2</sup>National Pedagogical Dragomanov University**

## **DEVELOPMENT OF THE DIGITAL TRANSFORMATION MODEL FOR HIGHER EDUCATIONAL INSTITUTIONS**

**Abstract.** The research is devoted to the analysis of the digital transformation processes that are currently taking place in the economy, production, education and society as a whole. There are analyzed approaches to developing a model of digital transformation of education. According to the results of the research, the model of digital transformation of the higher education institution was developed. In response to the impact of digital transformation, it can be used as a roadmap for solutions for the digital transition to an innovative model for the functioning of the modern university.

**Keywords:** digital transformation, digital transformation model, digital transformation of education, higher educational institution.

**Introduction.** The basis of modern society is digital technology: artificial intelligence, robotics, IoT, blockchain and 3D technologies, etc. The report of the consulting company Accenture (for 2017) identifies five new digital technologies that can transform global economic development [1, p. 4]: *Internet of Things* (IoT), *Artificial Intelligence* (AI), *Blockchain*, *Big Data*, *Robotic Process Automation* (RPA).

The use of digital technologies is transforming business models, resulting in new products and services; the format of works is changing (outsourcing, online platforms, improved automation, robotics, etc.). Real-time work with digital data fundamentally changes the ways of management, production, sale and use of products [2, p. 42].

Combined digital technologies are increasing their impact on business and social life. Fig. 1 shows the areas in which fundamental changes are expected due to the digital transformation:

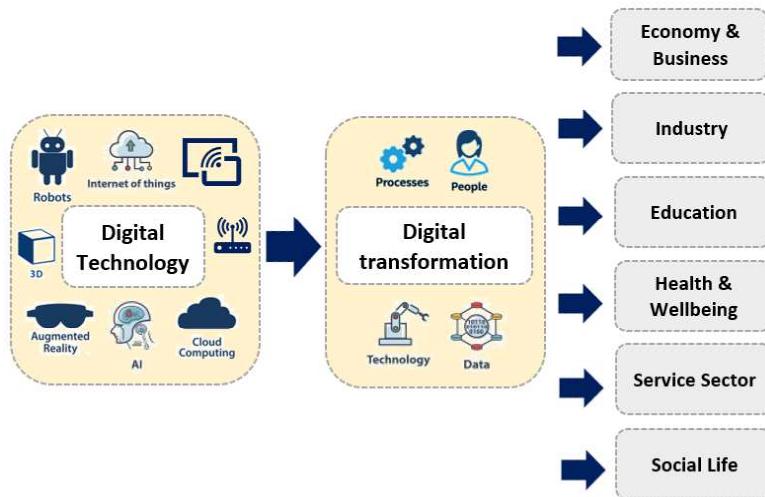


Figure 1 – Areas in which fundamental changes take place due to the digital transformation (Source: own work)

However, the education system is failing behind the general state of digital transformation in society. In our opinion, the main problem is the lack of understanding by the participants of the educational process of the institutions (higher, secondary and vocational) what is the difference between the use of digital technologies and innovations provided by the transformational changes that digital technologies bring to the educational process, and comprehension of concepts, structure, required and sufficient conditions and processes of digital transformation in general and in education in particular. That is why, the purpose of this paper is to define, analyze and develop models of digital transformation that can have a place in enterprises, businesses and educational institutions. The developed model of digital transformation of an educational institution will help educational politicians and leaders of different levels to build a strategy for the development of digitalization, taking into account the characteristics of each educational institution.

### Digital Transformation Model for Higher Educational Institutions

Digital transformation (DT) is the result of digitisation and digitalisation of economies and societies. DT is an ongoing process. The introduction of digital technologies creates both new opportunities and new challenges.

In the conditions of intensive development of digital technologies, digitalization, digital transformation of many branches of human activity, fast change of professions demanded in the labor market and, accordingly, professional requirements to competences of experts, educational activity needs updating of the maintenance and methods of training, search for innovative forms of training, expanding access to educational resources, and the implementation of learning opportunities without space-

and time-based restrictions, the introduction of new approaches to the organization of educational services in general. Thus, the digital transformation of education is an integral part of the processes taking place in society today.

The areas of developing the digital transformation of education are following [3]: transformation of goals, content and corresponding methods and forms of educational activities, which are associated with the penetration of new digital tools in various areas of human activity; educational institutions have to master new digital tools that increase the efficiency of the educational process; pupils/students should master new digital tools to increase the efficiency of their educational activities, and their digital competence needs developing; teachers should master: (a) new digital tools to increase the efficiency of their professional activities (b) content, methods and forms of educational activities that are transformed due to the impact of new digital tools on various areas of human activity; (c) new digital tools that increase the efficiency of the educational process, which is also changing; education leaders should master: (a) new digital tools that increase the efficiency of their professional activities; (b) digital tools that increase the efficiency of the organization of the changing educational process.

Thus, authors propose the components that will change within the process of digital transformation of higher education institution (Fig. 2).

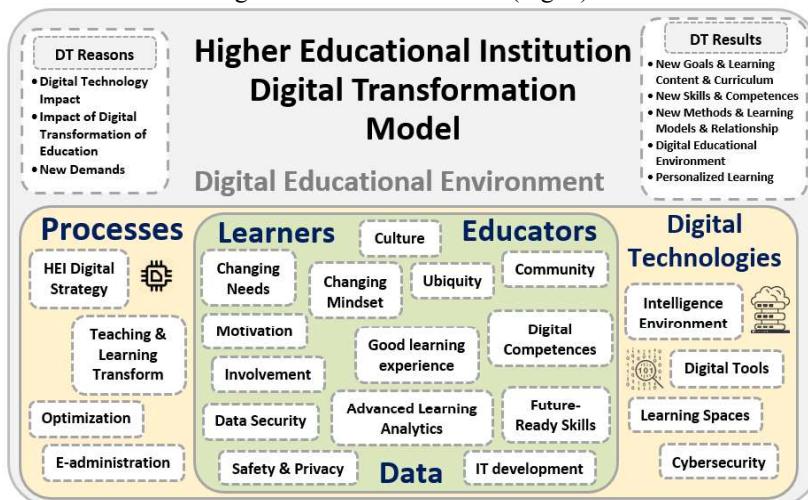


Figure 2 – Higher Educational Institution Digital Transformation Model (Source: own work)

The main components of the proposed model include: educational environment (taking into account conditions of wide use of digital technologies there will be a virtualization of educational process, processes of

communication, cooperation, and educational institution management); technology and tools used by teachers and students; conditions of teachers-students interaction within the digital environment; it is important to overcome the academic digital gap by developing professors' digital skills, as students are already highly motivated to use digital learning tools; management of university process and the internal process in general.

An envisioned model in response to the impact of digital transformation (Fig. 2) can be used as a roadmap for solutions for the digital transition to an innovative model for the functioning of the modern university.

**Conclusions.** Thus, the result of the digital transformation of education is: creating a modern digital educational environment to provide equal access to quality educational services and resources anywhere, anytime and in order to improve the quality of education; digitalization of all components of the educational process; effective use of modern digital technologies and data through the development of digital skills and competencies of all education stakeholders; formation of new competencies of the educational process participants, i.e., competencies which are necessary for a successful life in the digital society.

#### **References:**

1. Ford, F.R. & Lobo, I. (2017). Digital disruption: Development unleashed. Multiply innovation, collaboration and impact through digital in international development. Accenture. Retrieved from: [https://www.accenture.com/t20170601T083538Z\\_\\_w\\_\\_/usen/\\_acnmedia/PDF-40/Accenture-Digital-Disruption-Development-Unleashed.pdf](https://www.accenture.com/t20170601T083538Z__w__/usen/_acnmedia/PDF-40/Accenture-Digital-Disruption-Development-Unleashed.pdf) (accessed on 25.10.2020).
2. Digitalization of Ukraine's economy: transformational potential: monograph / V.P. Vyshnevskyi, O.M. Harkushenko, S.I. Kniaziev, D.V. Lypnytskyi, V.D. Chekina; NAN Ukrayiny, Instytut ekonomiky promyslovosti. Kyiv. Akademperiodyka [NAS of Ukraine, Institute of Industrial Economics. Kiev. Academic periodicals]. 2020. 188 p. <https://doi.org/10.15407/akademperiodyka.398.188>
3. Uvarov A. On the way to the digital transformation of the school. Education and Informatics. Moscow. 2018. 120 p. ISBN 978-5-906721-12-9.

<sup>1</sup> Roman Ponomarenko

PhP in Physics and Mathematics, assistant

<sup>1</sup> Taras Shevchenko National University of Kyiv

## KNOWLEDGE TEST SYSTEMS BASED ON TYPE 2 TAKAGI-SUGENO FUZZY INFERENCE

**Abstract.** In this article the problems of systems for assessing the quality of knowledge based on test control. The method of complex assessment of students' knowledge based on the Type 2 Tagaki-Sugeno fuzzy model are proposed.

In the modern educational industry, automated methods are increasingly being used to identify and test the quality of students' knowledge. In particular, knowledge testing systems are becoming more and more popular, moreover, gradually moving from an auxiliary tool to the main form of knowledge quality control. Knowledge testing systems have several advantages: the speed of knowledge testing, a unified approach to examiners, the ability of a student to take direct part in the examination process, and compare their results with similar results of their colleagues [1, 2].

Considering various grading scales (100-point, 5-point, 7-point, 12-point, etc.), we can note their common feature - not depending on the degree of graduation, most of them have a linguistic scale: "Excellent", "Good", "Satisfactory", "Unsatisfactory". Moreover, it is not always possible to accurately determine the transition boundary between two neighboring estimates. It can be argued that the process of assessing the quality of knowledge is intellectual in itself, and systems that automate these processes are humanistic systems [3], in which human judgments and the operation of quality indicators play a large role.

Takagi-Sugeno FLS T2 involve the use of interval type 2 (IT2) fuzzy sets (FS) [4] in the antecedents of IF-THEN fuzzy rules of the form:

$$\begin{aligned} R^k: & \text{If } x_1 \text{ is } \tilde{A}_1^k \text{ and ... and } x_m \text{ is } \tilde{A}_m^k \\ & \text{Then } g(x)^k = w_0^k + w_1^k x_1 + \dots + w_m^k x_m, \end{aligned} \quad (1)$$

where  $\tilde{A}_1^k \dots \tilde{A}_m^k$  – IT2 FS,  $k$  is the number of the rule. IT2 FS have the form

$$\tilde{A} = \int_X \frac{\mu_{\tilde{A}}(x)}{x} = \int_X \frac{\left[ \int_{J_x^u} \frac{1}{x} \right]}{x}, \quad J_x^u = \left\{ (x, u) : u \in [\bar{\mu}_{\tilde{A}}(x), \underline{\mu}_{\tilde{A}}(x)] \right\} \subseteq [0, 1]. \quad (2)$$

A method of fuzzy assessment of the quality of knowledge has been developed to obtain a comprehensive characteristic of a student for a training course (module). The Takagi-Sugeno fuzzy inference model with interval fuzzy membership functions of type 2 was taken as a basis. This model allows one to take into account the vague nature of the boundaries of linguistic estimates. Thus, giving at the output a more objective characteristic of knowledge (using the Karnik-Mendel fuzzy inference algorithm [5]).

The Takagi-Sugeno fuzzy model, the fuzzy rule consequents of which are presented in the form of functional dependencies, was not chosen by chance. Since this model allows you to form an expert opinion based on the numerical rating points given to the student during the course.

Figure 1 schematically shows the organization of the fuzzy rule calculations when using the proposed method. Every fuzzy rule  $R^k$  of rule block  $R = \{R^1, R^2, \dots, R^n\}$  as input parameters  $X = \{x_1, x_2, \dots, x_m\}$  in the antecedent are accepts the evaluation  $x_i$  for each lesson (topic), where  $m$  is number of lessons.

The block of rules for fuzzy inference is drawn up by an expert teacher and can take into account the nonlinear dependencies of a student's knowledge for individual lessons (topics). This may take into account the incompleteness of the student's knowledge, as well as the subjective methodology of teaching and assessing certain academic disciplines.

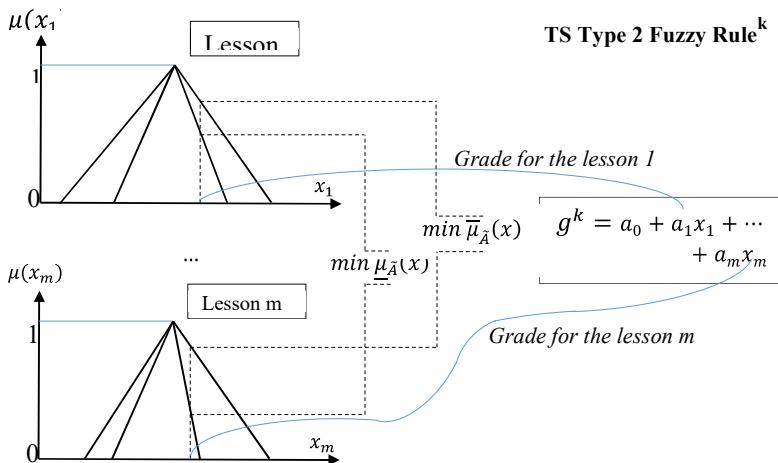


Figure 1 – Organization scheme of the Takagi-Sugeno Type 2 fuzzy rule in the complex assessment of student knowledge

#### References:

1. T. Gorbunova, Testing Methodology in the Student Learning Process. European Journal of Contemporary Education 6(2) (2017) 254–263.
2. O. Alekseev, Methodological Peculiarities of Diagnosing Students in Learning While Using Imitational Testing Models. Educational Discourse 1(3) (2011) 12–23.
3. L. Zadeh, The concept of a linguistic variable and its application to approximate reasoning. American Elsevier Publishing Company, 1973.
4. J.M. Mendel, Introduction to type-2 fuzzy logic control: theory and application. John Wiley & Sons, Inc., Hoboken, New Jersey, 356 p.
5. N.N. Karnik, Introduction to type-2 fuzzy logic system. Proc. 1998 IEEE FUZZ Conf 5, 1998, pp. 915–920.

**<sup>1</sup> Alexander Provtar**

Doctor of Physical and Mathematical Science, Head of the Department of Information Systems, Professor

**<sup>2</sup> Maksym Veres**

Candidate of Physical and Mathematical Science, Associate Professor

**<sup>3</sup> Maksym Samoilenko**

Postgraduate

<sup>1,2,3</sup>Taras Shevchenko National University of Kyiv

## USING EDUCATIONAL IoT SYSTEM

**Abstract.** This paper investigates using educational IoT in group of students. There is formal investigation of IoT system approach presented as transition systems and their composition which presented as Petri net with next liveness checking. Paper describes system architecture of IoT system which is the implementation of Petri net model.

**Keywords:** educational IoT, web server, transition system, Petri net

This paper investigates using educational IoT in group of students. There is formal investigation of IoT system approach presented as transition systems and their composition which presented as Petri net with next liveness checking.

Main purpose of education system is learning programming by coding behavior of IoT device, where each correctly solved task can improve programming skills. Therefore, the group of students can solve common group task where each student has IoT device assigned. Common task is one task for group of students where each student can take a part in task solving by means of IoT device.

IoT system model for education consists of three model: IoT device model itself which assigned to each student of group, task generating service which can create task for each device and can control task solutions from each student in group in context of solving common task and third model is a model of communicating device which is controlled by each of students which can exchange information between task service model and IoT device model.

Task service model, communication device model and IoT device model in the paper are presented as transition systems with sets of states and transitions accordingly. There is a described interaction model of system which is a result of multiplication of three transition systems which is presented as Petri net.

Also, liveness checking of Petri net by means of TSS algorithm is used, which guarantees correct work of IoT system which is implemented by Petri net model.

Paper describes system architecture of IoT system which is the implementation of Petri net model, which consists of three parts: IoT device system, which is controlled by mobile phone system via Wi – Fi, mobile phone as system, which can be controlled by a student which implements the communication between IoT device and task server via restful communication protocol, and task server which can create task for each IoT device and control solutions of student's task.

In conclusion, this paper investigates model of education IoT as Petri net with liveness checking which is multiplication of transition systems and next system implementations of transition systems. System implementations are presented as task service server, mobile device system and IoT system with their communication protocols accordingly.

**References:**

1. Karatrantou, Anthi & Panagiotakopoulos, Christos. (2008). Algorithm, Pseudo-Code and Lego Mindstorms Programming, Workshop Proceedings of SIMPAR 2008 Intl. Conf. on SIMULATION, MODELING and PROGRAMMING for AUTONOMOUS ROBOTS, Venice(Italy) 2008 November, 3-4.
2. Nielsen, Claus & Adams, Paul. (2015). Active learning via LEGO MINDSTORMS in Systems Engineering education. 489-495. 10.1109/SysEng.2015.7302802, 2015 IEEE International Symposium on Systems Engineering (ISSE).
3. Mylonas, Georgios & Amaxilatis, Dimitrios & Pocero, Lidia & Markelis, Iraklis & Hofstaetter, Joerg & Koulouris, Pavlos. (2019). Using an Educational IoT Lab Kit and Gamification for Energy Awareness in European Schools. This is a preprint version of a paper submitted toFabLearn Europe'18, Proceedings of the Conference on Creativity and Making in Education. DOI: 10.1145/3213818.3213823.
4. Lloret, Jaime & Sendra, Sandra & Gonzalez Ramirez, Pedro & Parra, Lorena. (2019). An IoT Group-Based Protocol for Smart City Interconnection. 10.1007/978-3-030-12804-3\_13. Springer Nature Switzerland AG 2019S. Nesmachnow and L. Hernández Callejo (Eds.): ICSC-CITIES 2018, CCIS 978, pp. 164–178, 2019. [https://doi.org/10.1007/978-3-030-12804-3\\_13](https://doi.org/10.1007/978-3-030-12804-3_13).
5. S. L. Kryvyi, S. D. Pogorilyy, M. S. Slinko, MODEL JUSTIFICATION OF GPU-BASED APPLICATIONS, DOI <https://doi.org/10.15407/usim.2018.04.0046>, ISSN 0130-5395, Control systems and computers, 2018, № 4.
6. Kryvyi S. L., Boyko Y. V., Pogorilyy S. D., Boretskyi O. F., Glybovets M. M. Design of Grid Structures on the Basis of Transition Systems with the Substantiation of the Correctness of Their Operation. Cybernetics and Systems Analysis.
7. Kryvyi S.L. 2015. Linear Diophantine constraints and their application. Chernivtsi: “Bukrek” Publishing House.

<sup>1</sup> Oleksandr Riabov

Master's degree student, Data Analyst

<sup>2</sup> Iulia Khlevna

Doctor of Technical Science, Associate Professor

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## RECOMMENDATION SYSTEM DESIGN IN PYTHON BY METHODS OF EMOTIONAL ANALYSIS AND MACHINE LEARNING

**Abstract.** The article is devoted to the design of an algorithm for providing accurate offers to online store customers based on their previous reviews. The article contains modeling methods explanations as well as practical use cases.

**Keywords:** modeling, emotional analysis, recommendation system, machine learning.

The development of an e-commerce activity is one of the Internet components [1]. This activity is a calling of our time. Simultaneously, these trends transform the user's decision-making process that is influenced by various forms of recommendations. Online marketing personalization is an obvious trend of the last decade [2]. According to McKinsey, 35% of Amazon revenue, and 75% of Netflix revenue comes from recommended products, and this percentage is likely to increase [3]. It is reasonable to form recommendations based on previously provided feedbacks and that is why an accurate algorithm design for product recommendations is an actual challenge.

Classification model is an underlying model for the sentiment detection. The dataset that has been used for the research contains 23 thousand reviews with 10 feature variables such as goods id, category, text, rating etc. Reviews that have 4 or higher rating were marked as positive (True) and reviews that have 2 or lower rating were marked as negative (False). Neutral reviews that are equal to 3 were not included to the research.

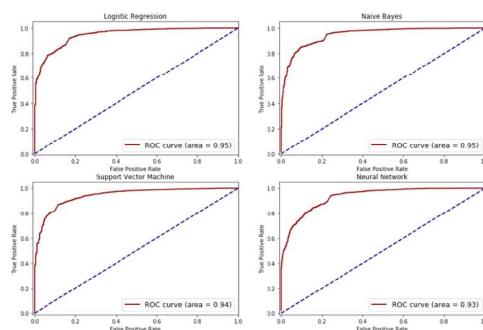


Figure 1 – Classification modeling results

On the other hand, Naïve Bayes takes less computation time and when we have a bigger dataset, this difference might be an important advantage.

The dataset has been divided as training and test samples and been prepared to classification model design using different methods: Logistic Regression, Naïve Bayes, Support Vector Machine and Neural Network. It's been concluded that Naïve Bayes and Logistic Regression show the best results - the area under ROC-curve is equal to 0,95 (Figure 1). Thus, both of them are very effective in sentiment prediction.

The next step is a use of collaborative filtering as a standard method for product recommendations. Pearson correlation has been used to identify customers similar to the current customer in terms of their ratings on the same set of goods. After that, we took their ratings of goods the current customer has not bought yet and then recommended goods with the highest average rating to the user.

The model accuracy evaluation has been conducted by the cross-validation and RMSE as a key indicator (Figure 2). Based on the results it's been concluded that suggested algorithm performs much better than random recommendations and RMSE decreases with the increase of the nearest neighbors number, that is why we recommend running algorithm based on at least 30 connected customers in a graph.

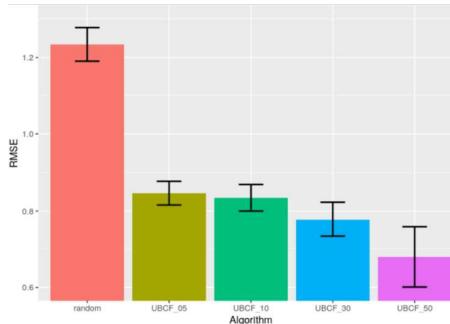


Figure 2 – Model evaluation using RMSE

Provided recommendation system can be implemented to the back-end part of online shops.

#### References:

1. A. Gupta. Role of e-commerce in today's business//International Journal of Computing and Corporate Research. URL: <https://www.ijccr.com/January2014/10.pdf>
2. M. Balan. Personalize, Summarize or Let them Read? A Study on Online Word of Mouth Strategies and Consumer Decision Process. Inf Syst Front (2020). URL: <https://doi.org/10.1007/s10796-020-09980-9>
3. How retailers can keep up with consumers. URL: <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers#>

**<sup>1</sup> Anna Selivanova**

Lecturer

**<sup>2</sup> Oleg Pursky**

Doctor of science, professor

**<sup>3</sup> Yurii Yurchenko**

Lecturer

**<sup>4</sup> Hanna Samoylenko**

Candidate of science, associate professor

**<sup>5</sup> Tatiana Dubovyk**

Doctor of science, professor

*Kyiv National University of Trade and Economics, Ukraine*

## **AGENT MODELING OF ONLINE STORE ACTIVITIES**

The current development of the economy is characterized by a rapid growth of the pace of informatization of economic processes, the expansion of the scale of e-business and, in particular, of retail e-trade on a large scale [1, 2]. Rapid development of e-commerce and online marketing are becoming an attractive alternative to media campaigns because they can be cost effective with a relatively small budget and specifically targeted to user profiles.

The aim of the study is to construct a model, describes in detail the work of an online store. The main task is the construction of a real object model and its functioning dynamics. For this purpose we used AnyLogic computing environment [3].

For fast tracking and responding to changes in the market, and to image dynamics increase the number of customers, it is advisable to use agent-based modeling technologies. The first step in building a model will be to determine the criteria and conditions under which the experiment will begin. A relatively small market with 5,000 people will be considered. To implement the model, each client will be an agent. Since it is determined that the conditional company is new, no one will be interested in the product at first, people's interest will appear under the influence of advertising. After that, the number of successful sales will be affected by the natural increase in customers, which will occur due to the fact that customers who have already bought the product will share information with their friends. The last to the model will be added indicators that can negatively affect the operation of the system, as they will change the conditions under which each purchase will be made.

The block scheme of the online store is very simple, step by step it can be described as follows:

- buyer goes to the website of the online store, looks for the product and sends it to the cart when placing an order;

- client indicates his contact details and method of payment;

- manager contacts the client to confirm the order and clarify contact information, delivery point;

- customer pays for the order at this stage, or does so upon arrival of the goods; the manager of the online store, or another employee, packs the goods and sends it by

courier service (if the customer did not prefer self-pickup);

- customer receives the order and makes the payment if he has not done so before
- after-sales interaction with the buyer (cross-sell, up-sell, advertising and email-marketing).

Potential buyer visits the online store, registers, places the selected order and:

1) the goods are delivered to the warehouse

2) collection of the order in a warehouse

3) change the status of the order in the database to the current state at the time

4) registration of documents to the order (check) and notification to the client about readiness

The process of building an agent model consists of successive steps.

*Step 0.* Analysis of the initial data.

*Step 1.* Create a new model.

*Step 2.* Creating drives.

*Step 3.* Add a product sales stream.

*Step 4.* Adding constants.

*Step 5.* Setting the initial values of drives.

*Step 6.* Creating dynamic variables.

*Step 7.* Configure model startup.

*Step 8.* Launch the model.

*Step 9.* Add charts.

In this model, the number of potential consumers is not reduced to zero, but is constantly replenished as consumers re-buy products instead of unusable ones. The intensity of product acquisition increases, decreases, and eventually takes on a value depending on the average suitability of the product and the parameters that determine the intensity of this flow. The presence of a product in the withdrawal model means that some part of the population will always remain potential consumers. The most effective option for an online store is one in which the time to collect the order is minimal. Thus, agent-based modeling technology makes it possible to conduct preliminary modeling of any enterprise in e-commerce to ensure the possibility of choosing the optimal market strategy and the required number of consumers of the product with minimal costs for their implementation.

#### **References:**

1. G. Schneider, Electronic Commerce, 12th ed., Cengage Learning, Boston, 2016.
2. S. Belew, J. Elad, Starting an Online Business All-in-One for Dummies, 5th ed., For Dummies, New Jersey, 2017.
3. The AnyLogic Company: Official Website of The AnyLogic Company, 2020. URL: <https://www.anylogic.com/features/cloud/>.

**<sup>1</sup> Artem Yurchenko**

PhD (Pedagogical Sciences), Associate Professor of the Department of Computer Science

**<sup>2</sup> Olena Semenikhina**

Dr. (Pedagogical Science), Professor, Chair of the Department of Computer Science

**<sup>3</sup> Volodymyr Shamonia**

PhD (Physical and mathematical Sciences), Associate Professor of the Department of Computer Science

**<sup>4</sup> Yurii Khvorostina**

PhD (Physical and mathematical Sciences), Associate Professor of the Department of Mathematics

<sup>1,2,3,4</sup> *Makarenko Sumy State Pedagogical University, Ukraine*

## OPEN EDUCATIONAL RESOURCES IN IT SPHERE

**Abstract.** The authors provide their quantitative analysis of IT courses on the following educational platforms: Coursera, EdX, Udemy, MIT Open Course Ware, OpenLearn, Intuit, Prometheus, UoPeople, Open Learning Initiative, Open University of Maidan. The authors propose to use open educational resources for organization of independent work within the same Ukrainian courses in Ukrainian realities; for organization of distance learning; for professional development of teachers.

**Keywords.** MOOC, Online Learning, online courses, online training courses, educational platforms, IT courses.

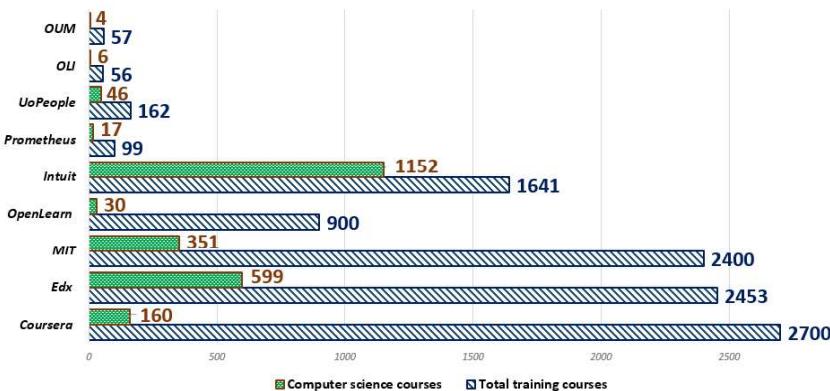
Due to the great popularity of specialties in the information technologies (IT) field in Ukraine there is a huge expansion of work in the “freelance” format with customers from abroad, with different standards for creating or presenting virtual content. Therefore the analysis of an open educational resources in the IT field on the different educational platforms becomes relevant and would contribute to resolving the contradiction between the demand of Ukrainian society for a competitive IT specialist and the established content, methods, approaches, etc., of the training for such specialists in Ukraine.

By definition, UNESCO's “open educational resources are educational and scientific resources that are openly accessible or licensed, allowing their free use and modification by third parties” [1]. In Ukraine, the term “mass open online courses” or the MOOC is commonly used, which is understood as an open platform, that after the help of the recognized specialist in certain industry provides not only the free access, and online resources but also the social networks for active communication between a large number of students who self-organize their participation according to their own learning goals and prior knowledge and skills [2]. Such course can be positioned as organized by a program, which defines certain terms of study and topics, but at the same time active communication during the course can lead its organizers beyond the chosen problem. Such courses are not accredited and are not intended to guarantee certain learning outcomes, in contrast to open educational resources, which include, in

particular, the acquisition of a certain set of knowledge / skills for testing them internally and / or at the end of a course study in the form of testing or performing certain tasks.

We conducted an analysis of ten open educational resources (ER) (Coursera, EdX, Udemy, MIT OpenCourse Ware, OpenLearn, Intuit, Prometheus, UoPeople, Open Learning Initiative, Open University of Maidan) regarding the availability of courses in the IT industry (analyzed as of March 2020). We conducted a similar analysis for open educational resources in mathematics [3]

Quantitative data on the number of courses in total and the number of courses in the IT industry are shown in Figure 1.



**Figure 1.** The proportion of computer science courses on various ER

According to fig. 1 analysis, it can be concluded that among well-known educational platforms the most filled with different courses are foreign Coursera, EdX, MIT OpenCourse Ware, OpenLean and Intuit. The number of courses on them exceeds the value of 2.4 thousand, while in the Ukrainian Internet space the most saturated courses platform Prometheus where about 100 courses are located.

Obvious (table 1) is the leadership position of the resource Udemy (in fig. 3 this resource is not displayed for a better visualization of the content of other resources) of the number of courses in general – 82943, and on IT – 35727. According to this resource in the number of IT courses is a platform Intuit, which offers 1152 courses, then go EdX (599 courses) and MIT OpenCourseWare (351 course). At least ten courses on IT-industry offer Open Learning Initiative (6 courses) and OUM (4 courses).

The results of quantitative analysis of the IT open in the IT field give grounds to formulate the following proposals for improvement of professional training of IT industry professionals in Ukraine.

Organization of independent work within the separate courses of training plan specialists can be connected with the passage of a non-commercial course on one of the mentioned platforms. It is worth considering the amount of independent work (the

amount chosen on the open platform of the course should be about the same or smaller). Obtaining a certificate of course on an open platform will be a condition of enrolling actual work.

We also consider it advisable to use the open non-profit ER for organization of remote or correspondence courses within the variation part of the training plan of the specialist.

The level of the teacher's professionalism directly flows to the level of future specialist training, and therefore the successful completion of courses on open platforms should be perceived as a training for teachers. The certificate of completion of the course of 3 or more ECTS credits for a long-term improvement of training and 2 credits ECTS [3] for short-term administration may be counted as a confirmation of advanced training for Teachers of educational institutions.

The analysis conducted by us can encourage teachers to research specific open OR depending on their preferences and educational or scientific tasks. We believe that this will positively affect the quality of teaching copyright courses, the dissemination of copyright techniques and the teacher's professionalism. We also note the possibility and feasibility of distributing copyright techniques in the development of their own courses and promoting them on open platforms.

#### **References:**

1. S. Touzé, Open Educational Resources in France: Overview, Perspectives and Recommendations. UNESCO Institute for Information Technologies in Education, 2014.
2. A. McAuley, B. Stewart, G. Siemens, D. Cormier, The MOOC Model for Digital Practice. Created through funding received by the University of Prince Edward Island through the Social Sciences and Humanities Research Council's "Knowledge Synthesis Grants on the Digital Economy", 2010.
3. O.V. Semenikhina, M.G. Drushlyak, Yu.A. Bondarenko, S.M. Kondratiuk, I.M. Ionova. Open Educational Resources as a Trend of Modern Education. Proceedings of 42 International convention on information and communication technology, electronics and microelectronics "MIPRO 2019", Opatija (Croatia), 2019, pp. 779–782.
4. Zakonodavstvo Ukrayny, Regulations on the improvement of qualification and internship of pedagogical and scientific-pedagogical workers of higher educational establishments, 2013. URL: <https://zakon.rada.gov.ua/laws/show/z0488-13>.

<sup>1</sup> **Sergiy Zagorodnyuk**  
PhD, associate professor

<sup>2</sup> **Bohdan Sus**  
PhD

<sup>3</sup> **Oleksandr Bauzha**  
PhD, associate professor

<sup>1-3</sup> Taras Shevchenko National University of Kyiv

## THE APPLICATION OF NETWORK COMMUNICATION FOR ORGANIZING A LABORATORY WORK

**Abstract.** This paper describes the developed network infrastructure of remote laboratories and discusses the effective mechanisms of remote sessions and technologies of network administration of laboratory equipment. The developed remote laboratory works and systems of visualization of the measured data, settings of a computer network, process of switching of the laboratory computer equipment and remote supervision are discussed in detail. Developed technical solutions, developed laboratory facilities, engineering and network solutions with a detailed description were analyzed in detail.

### Introduction

Computer support of the educational process provides additional opportunities for both the independent activity of the student and for his work in classrooms and laboratories. Handling the network infrastructure and e-learning materials, the student can learn independently from textbooks with interactive examples and demonstrations, perform tests, prepare for scientific research.

A remote lab is a distributed computing environment that allows students to conduct experiments by interacting with real devices, tools, and mechanisms through a customized telecommunications platform. [1, 2].

Today, remote laboratory work is an important topic for the educational community. Laboratory activity is an integral part of the educational process in higher education institutions in the study of natural sciences. By conducting laboratory work the student not only acquires fundamental knowledge but also acquires the necessary skills, which is a requirement for its formation as a competent specialist[3, 4]. In engineering education, practical work is an important addition to theoretical courses. Students come to the laboratories to conduct experiments and evaluate the differences between their observations and predictions according to theoretical courses. The need for physical access to laboratory equipment is not in doubt, but the implementation during quarantine is an urgent and complex problem, which is especially actual for laboratory work in natural sciences[5].

Remote laboratories can be used for academic, industrial and research purposes. Laboratory equipment connected to the Internet provides remote control of the laboratory work process. [6]. In particular, the remote control is supported by telescopes, large-scale installations operating on particle accelerators, devices for

recording ionizing radiation, etc. Distance laboratory work is one of the most important applications of IoT principles to education, especially in physics and other fields for which experiments are a key part of the learning process. The Internet of Things (IoT) is a network of physical devices connected by local and wide area networks, including the Internet. As technology has evolved, the importance of IoT has increased tremendously.

Despite the advantages and disadvantages, every year more and more universities start practicing remote laboratory works. Earlier, they were available in only a few universities that had research teams specializing in Internet technology. Now, many universities offer telecommuting in addition to traditional practice laboratories to present students of other universities the opportunity to operate with laboratory experiments that they would not be able to prepare in their own laboratories and to provide them extra laboratory time.

Students and teachers typically access remote lab work through specialized web applications that allow them to control lab equipment, conduct experiments, and retrieve real measurement data. [7, 8].

### **Practical implementation**

The laboratory work is based on a research setup. This work well demonstrates the implementation of the concept of remote laboratory work for natural and technical sciences. Remote access to devices is implemented using standard interfaces and a computer connected to the Internet.

The exponential method is used for calculation and analysis in this laboratory work. Its essence is the direct registration of the kinetics of photocurrent changes through a semiconductor sample under the influence of the latter on U-shaped pulses of light.

The system architecture of experimental setup consists of three parts.

The first part is designed for the generation of strobe-shaped pulses of light to the samples under investigation.

The second part of the experimental setup consists of a set of samples for measurement controlled by mosfet keys and digital oscilloscope.

Distance learning students can access the laboratory through a specialized authorized website, where each student will receive personal instructions on how to join the remote laboratory of the university and perform laboratory work.

### **Conclusions**

The development of remote laboratories is a promising direction for natural and engineering education, as they provide students with the means to design and conduct experimental research without the need for a physical presence near the laboratory equipment.

The suggested and assembled remote laboratory setups demonstrate that due to the use of standard hardware interfaces and common mechanisms of network infrastructure to create a similar remote laboratory work can any researcher or teacher interested in ensuring that the learning process in higher education continues without

interruption and without significant restrictions on the flexibility and functionality of the laboratory installation.

### References:

1. Khazri, Y., Toumi, H., Al Sabri, A., Moussetad, M., Sabir, B., Fahli A.: Remote Control Laboratory Experiments in Physics using LabVIEW. International Jounal of Information Science & Technology –iJIST, 1 (1), 11-16 (2017). ISSN : 2550-5114
2. Teng, M., Considine, H., Nedie, Z., Nafalski, A.: Current and Future Developments in Remote Laboratory NetLab. Int. J. Onl. Eng. 12, 4 (2016). <https://doi.org/10.3991/ijoe.v12i08.6034>.
3. Zapata Rivera, L.F., Larrondo Petrie, M.M.: Models of Collaborative Remote Laboratories and Integration with Learning Environments. Int. J. Onl. Eng. 12, 14 (2016). <https://doi.org/10.3991/ijoe.v12i09.6129>.
4. Khazri, Y., Rouane, M., Fahli, A., Mohamed, M., Khaldouni, A., Naddami, A.: Developing a Remote Practice for Laboratory Experiments on Measuring Instruments. Int. J. Onl. Eng. 10, 12 (2014). <https://doi.org/10.3991/ijoe.v10i5.3833>.
5. Gadzhanov, S., Nafalski, A., Nedic, Z.: Remote Laboratory for Advanced Motion Control Experiments. Int. J. Onl. Eng. 10, 43 (2014). <https://doi.org/10.3991/ijoe.v10i5.3835>.
6. Heradio, R., de la Torre, L., Dormido, S.: Virtual and remote labs in control education: A survey. Annual Reviews in Control. 42, 1–10 (2016). <https://doi.org/10.1016/j.arcontrol.2016.08.001>.
7. Colwell, C., Scanlon, E., Cooper, M.: Using remote laboratories to extend access to science and engineering. Computers & Education. 38, 65–76 (2002). [https://doi.org/10.1016/S0360-1315\(01\)00077-X](https://doi.org/10.1016/S0360-1315(01)00077-X).
8. Bencomo, S.D.: The Role of Interactivity in Control Learning. IFAC Proceedings Volumes. 36, 1–12 (2003). [https://doi.org/10.1016/S1474-6670\(17\)33646-7](https://doi.org/10.1016/S1474-6670(17)33646-7).

**<sup>1</sup> Viktoriia Zinchenko**

PhD in Pedagogical Sciences, Associate Professor

**<sup>2</sup> Anna Kyrypa**

PhD in Pedagogical Sciences, Associate Professor

**<sup>3</sup> Olena Stepanenko**

PhD in Philological Sciences, Associate Professor

<sup>1</sup> Donetsk National University of Economics and Trade named after Mikhail Tugan-Baranovsky

<sup>2,3</sup> Communal Institution of Higher Education «Dnipro Academy of Continuing Education» of Dnipropetrovsk Regional Council

## **INFORMATION AND COMMUNICATION TECHNOLOGIES WHILE FORMING NON-PHILOLOGICAL STUDENTS' PROFESSIONAL LANGUAGE AND SPEECH COMPETENCES**

**Abstract.** The research is devoted to analyzing the way of implementation of information and communication technologies in teaching process which could be helpful while forming professional language and speech competences of non-philological students. The tasks of information and communication technologies as well as psychological prerequisites for mastering the knowledge and forming language and speech competences, communication skills and abilities are described.

**Keywords:** teaching and learning model, information and communication technologies, professional competence, language and speech competence.

In modern conditions of developing and widespread using computer equipment the learning process in higher educational institutions must be arranged with applying information and communication technologies which promotes differentiation and humanization of education, intensive development and realization of students' spiritual and intellectual potential. That is why educational institutions review the approaches for teaching the humanitarian disciplines in a professional direction, as well as search for innovative methods taking into account the development of information and communication technologies. In order to transform the traditional teaching model based on the teacher's explanations, the use of new teaching and learning models connected with information and communication technologies is being actively implemented [1].

The high speech culture is a part of common human culture. Today we observe the heightened interest in the humanitarian disciplines as well as the state language; it is not enough for students to have knowledge acquired at school. Modern students are characterized by a predominantly imaginative perception of the world, so it is necessary to seek and use new opportunities of the human brains to form optimal methods of mastering knowledge. The graduate of higher educational institutions must perfectly know and use the Ukrainian language first of all. Developed language and speech as well as professional language competences encourage mastering the subject, increase working efficiency and influence positively on business relationships. The high speech culture and writing skills for a future professional is not just a reflection of his or her

professionalism, tolerance, good manners and intellectuality. It also determines work culture in all aspects. The search for innovative and effective methods of teaching at the courses in non-philological higher educational institutions determines studying peculiarities of forming language and speech competences of non-philological students with applying modern information and communication technologies. Also, the use of information and communication technologies makes it possible to diversify the work, to make the process of learning the humanitarian disciplines go beyond the teaching audience and to make it continuous, promotes the organization of students' independent work. Sometimes it is due to multimedia resources that students' desire and need to learn the humanitarian disciplines are activated, which significantly affects the learning performance. According to many scientists, the development of individual competences in modern conditions is significantly influenced by the introduction of new information and communication technologies in the educational process [2, 3, 4, 5, 6].

This article reveals psychologists' and pedagogs' views on introducing information and communication technologies into teaching humanitarian disciplines from professional point. It determines fundamental linguo-didactic problems which can be solved by introducing information and communication technologies into the teaching process of humanitarian disciplines.

Psychological and pedagogical conditions for forming professional language and speech competences of non-philological students with applying information and communication technologies are described. Changes in the model and components of teaching professional language with applying information and communication technologies are considered. We agree with the scientists who emphasize such tasks of applying information and communication technologies as forming information and communication culture of future specialists; ensuring the development of students' personal traits and skills, disclosing the creative potential of every student; increasing the efficiency of the educational process; providing creative research activities; intensifying the educational process.

Psychologists include necessary psychological prerequisites for mastering the knowledge and forming language and speech competences, communication skills and abilities. They underline positive attitude to the learning process (interest in the subject, constant attention, and high pace of educational work); processes of direct sensual familiarization with the material (from the language point it includes schematic visualization, interactive learning technologies); thinking process (activation of concrete and abstract, conceptual and artistic thinking, perception, comprehension and understanding of the material); memorizing and storing the information received, the ability to reproduce it performing individual speech abilities. These prerequisites contribute to creating learning motives, encouraging non-philological students to learn linguistic theory and to produce speech activity, ensuring the development of linguistic persona. Their implementation is significantly facilitated by the introduction of information and communication technology training [4, 7].

The use of information and communication technology while studying at the courses of humanitarian disciplines in higher educational institutions of non-philological profile helps to solve the problems of humanization of the educational

process and allows to significantly increase its effectiveness through intensification, differentiation (taking into account students' individual characteristics), effective combination of individual and team activities; prompt interrelation of the source of educational information and the student; timeliness, usefulness, appropriate dosage, availability (understandability) of educational information; adaptation of the rate of its presentation to the rate of assimilation; introduction of innovative teaching methods and tools.

The analysis of approaches to the methodology of teaching at the courses of humanitarian disciplines in higher educational institutions of non-philological orientation shows that there is an objective need to develop a special methodological system of language training for the professional direction of future professionals with applying information and communication technologies, which would contribute to the formation of language and speech skills among students and their professional communication skills.

The prospect of the research can be seen in the further study of the linguo-didactic potential of modern information and communication technologies, a detailed analysis of the peculiarities of applying each tool of specific information and communication technology while the process of teaching different aspects of language by professional direction and forming skills in different types of language activity.

#### **References:**

1. Concept of implementation of media education in Ukraine. URL: [http://www.ispp.org.ua/news\\_44.htm](http://www.ispp.org.ua/news_44.htm). (in Ukrainian).
2. P. V. Sysoyev, Information and communicative technologies in linguistic education. Moscow : The «Librocom» Publ., p. 264, 2013 (in Russian).
3. R. S. Hurevych, “Usage of Internet technologies in the preparation of specialists: aspects of education in networked communities”, Scientific journal of M. P. Drahomanov NPU. Series № 2. Computer-oriented systems of education, vol. 16 (23), pp. 8–12, 2015 (in Ukrainian).
4. H. B. Zakharova, “The use of information technology as an innovative vector of higher education didactics”, Scientific works. Pedagogy, pp. 63–66, 2013 (in Ukrainian).
5. O. V. Semenikhina, New Paradigms in the field of education in the conditions of the transition to the «smart» society. URL: <http://nvd.luguniv.edu.ua/archiv / NN23 / 13sov.pdfs.pdf>. (in Ukrainian).
6. A. Kyrpa, Interactive technologies and gamification in foreign language teaching, European Academy of Science, Bonn, Germany: «EAS», 2019, pp. 75-76.
7. G. F. Ponomaryova, New pedagogical technologies: teaching aids. Kharkiv, p. 282, 2013 (in Ukrainian).

# **MATHEMATICAL FOUNDATIONS OF INFORMATION TECHNOLOGY**



**<sup>1</sup> Oleksii Bychkov**

Doctor of Technical Science, Head of the Department of Software Systems and Technologies, Associate Professor

**<sup>2</sup> Oleksii Ivanchenko**

Candidate of Technical Science, Head of the Personnel Department

**<sup>3</sup> Kateryna Merkulova**

Candidate of Technical Science, Associate Professor of the Department of Software Systems and Technologies, Associate Professor

**<sup>4</sup> Yelyzaveta Zhabska**

Assistant of the Department of Software Systems and Technologies

*<sup>1,2,3,4</sup> Taras Shevchenko National University of Kyiv*

## **MATHEMATICAL METHODS FOR INFORMATION TECHNOLOGY OF BIOMETRIC IDENTIFICATION IN CONDITIONS OF INCOMPLETE DATA**

**Abstract.** The purpose of this research is to develop mathematical methods for information technology of biometric identification which will allow to recognize person's face in conditions of incomplete data such as wearing a medical face masks during the pandemic.

**Keywords:** biometric identification, face recognition, wavelet transform.

Nowadays wearing medical masks caused increasing of the probability of false identification results of recognition systems from 5 to 50%, as the US National Institute of Standards and Technology (NIST) concluded [1].

During the ongoing pandemic researchers focus on quick and effective solutions to develop technologies that handle this problem. This research concentrates on the analysis of the already existing solutions and proposes a mathematical method of face identification for information technology based on wavelet transform under the condition of wearing masks by people. Since there is no database which contains face images with masks, a new database was created. This database contains 820 images of 40 people, whose faces was limited only by top part of the face (forehead, eyes).

First set of experiments was performed with the use of standard Python library `face_recognition`, which allows to recognize and manipulate faces from Python or from the command line with the world's simplest face recognition library. It built using `dlib`'s state-of-the-art face recognition toolkit with deep learning. `dlib` model is a ResNet network with 29 conv layers.

In the second set of experiments FaceNet system was used. It is a system that, given a picture of a face, will extract high-quality features from the face and predict a 128-element vector representation these features, called a face embedding.

Third part of experiments was performed to analyze the efficiency of three popular facial recognition methods: eigenfaces, fisherfaces, and LBPH. Eigenfaces algorithm processes all the training images of all the people as a whole and tries to extract the components which are relevant and useful. Fisherfaces algorithm extracts principal components that differentiate one person from the others, so an individual's

components become more useful over the others. The purpose of LBPH algorithm is to find face local structure by comparing each pixel to the neighboring pixel, forming a list of local binary patterns, that can be converted into a decimal number.

Results of performed experiments are presented on the Figure 1.

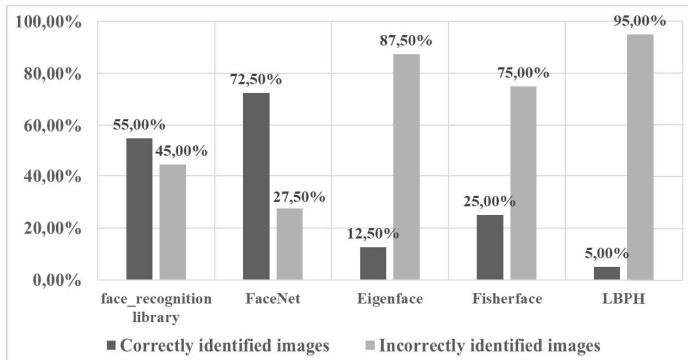


Figure 1 – Results of identification with the use of face\_recognition, FaceNet, Eigenfaces, Fisherfaces and LBPH methods on the dataset of masked images

Mathematical methods proposed in this research based on wavelet transform, that provides processing of patterns hidden in the data performing data analysis in general as well as in the detail. To compare the results of the commonly used algorithms with wavelet transform there was developed algorithm with the use of Daubechies wavelets and reverse biorthogonal wavelets.

Daubechies wavelets [2] are the type of basic wavelets, that orthonormal basis defined as:

$$\phi_{r,j,k}(x) = 2^{\frac{j}{2}} \phi_r(2^j x - k), j, k \in Z, \quad (1)$$

where function  $\{\phi_r(x - k) | k \in Z\}$ ,  $j$  is the scaling index,  $k$  is the displacement index, and  $r$  is the filter index.

Reverse biorthogonal wavelets [3] can be described with the following:

$$[\langle \phi(x) | \bar{\phi}(x) \rangle] = I, \quad (2)$$

$$[\langle \psi(x) | \bar{\psi}(x) \rangle] = I. \quad (3)$$

Figure 2 depicts results of the Daubechies wavelet transform and standard deviation calculation on the dataset of masked images.

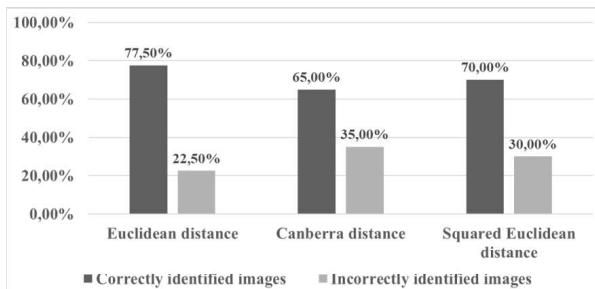


Figure 2 – Results of the Daubechies wavelet transform and standard deviation calculation on the dataset of masked images

Figure 3 depicts results of the reverse biorthogonal wavelet transform and standard deviation calculation.

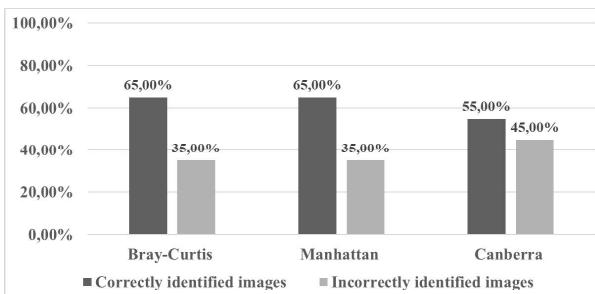


Figure 3 – Results of the reverse biorthogonal wavelet transform and standard deviation calculation on the dataset of masked images

The results of experimentation results analysis indicate that the popular and commonly used methods of face identification do not demonstrate high efficiency results. Proposed mathematical methods for information technology based on wavelet transform improves the face recognition and identification process under the condition of faces covered with mask. Specifically, the most accurate identification rate of 77,5% was obtained with the use of Daubechies wavelets.

#### References:

1. M. Ngan, P. Grother, K. Hanaoka. National Institute of Standards and Technology Interagency or Internal Report 8311 Natl. Inst. Stand. Technol. Interag. Intern. Rep. 8311 (2020). doi: 10.6028/NIST.IR.8311.
2. I. Daubechies. Ten lectures on wavelets. Society for Industrial and Applied Mathematics, USA, 1992.
3. C. Stolojescu, I. Railean, S. Moga and A. Isar. "Comparison of wavelet families with application to WiMAX traffic forecasting." 12th International Conference on Optimization of Electrical and Electronic Equipment, Basov, 2010, pp. 932-937. doi: 10.1109/OPTIM.2010.5510403.

<sup>1</sup> Hryhorii Hnatienko

PhD

<sup>2</sup> Artem Rimek

Student

<sup>1,2</sup> Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

## USE OF ALGEBRAIC APPROACH WHEN EVALUATING THE CORRECT SEQUENCE OF THE PRESENT LIST ELEMENTS IN TESTING TASKS

Testing is controversial, debatable, but reliable, powerful, effective and, for some branches of human life, irreplaceable tool. In particular, the testing procedure is fruitfully used in programming, engineering, medicine, psychiatry, and education. Moreover, for example, pedagogical testing simultaneously performs several functions: teaching, diagnostic, evaluative, incentive, developing, educational, etc.

Today there are many opinions about the appropriateness of using tests. On the one hand, tests are viewed as a means of positively improving the educational process towards its technologization, objectivity, and reducing labour intensity. On the other hand, tests are seen as a means of reducing the role of the teacher, and the test results are sometimes considered insufficiently reliable. However, it should be noted that, despite the different attitudes towards this area, testing itself becomes the main form of diagnostics of the level of students' training.

We only note that tests, at least, eliminate the drawbacks of empirical control. After all, a test consists of a number of tasks in a certain area and a standard known to the teacher, that is, a sample of complete and correct performance of the task.

**Types of test items.** Test items are traditionally divided into two large groups: closed and open. Closed-type tasks are considered in this work, namely: the problem of determining the arrangement of list elements in a certain sequence, that is, determining the order for arranging elements (objects, alternatives, entities), the sequence of actions, operations, the course of processes, the order of calculations, the chain of events, judgments and the like. In this case, respondents are offered a list of concepts, phenomena, dates, words, etc., which they must arrange in a correct sequence. Such test problems arise in various areas, for example:

- Establish the chronological sequence of events;
- Define some logical sequence;
- Formulate a definition from a set of randomly given words;
- Arrange some numbers in ascending or descending order;
- Restore the order of the proof of some theorem;
- Write a sequence of calculations when writing a program code that ensures the determination of the value of a given formula and the like.

Such tasks help to form students' algorithmic thinking, consolidate the relevant knowledge and skills.

**Formulation of a task.** Consider a formal description of the ordering of items in closed questions during testing. Note that models and methods for determining the competence of respondents based on the axiom of unbiasedness when ranking alternatives were studied, in particular, in the monograph [1].

Let the set of elements of a complete answer  $a_i \in A$ ,  $i \in I = \{1, \dots, n\}$  be given; the number of them is equal to  $n$ ,  $n = |A|$ . A respondent must build a linear (complete) order for this set, that is, a strict ranking of the given elements of the answer. We will denote the correct order of elements, which is known to the teacher, and for which the respondent receives the maximum score, through  $R^0$ ,  $R^0 = a_{i_1} \succ a_{i_2} \succ \dots \succ a_{i_n}$ ,  $i_j \in I$ ,  $j \in I$ . Thus, the testing procedure can be formalized in the class of ranking problems.

Note that the ability to guess the answer is the main reason for the negative attitude of teachers to the closed form of tasks. To eliminate this deficiency, a correction for guessing test points is even applied. Its essence lies in the fact that the number that can be guessed is subtracted from the total score received by each respondent, in accordance with the provisions of the theory of probability.

Since the number of possible answers to the tasks of ranking elements is equal to  $n!$ , then at  $n = 5$ , for example,  $n! = 5! = 120$  and the value significantly increases with an increase in the number of elements that need to be ordered. That is, the probability of guessing the correct answer is extremely low. In this regard, the warnings of teachers regarding the possibility of guessing are baseless.

**Solution algorithm.** When assessing the tasks of ordering a set of given elements, the dichotomous assessment of the task performance is most often used: "Yes" - "No", 0 or 1. Less often, some heuristic assessment rules are used. For example, a correctly completed task is assessed using 3 points, an error at the end of the task is assessed using 2 points, an error in the middle of a task is assessed in 1 point, and an error at the beginning of a task results in a zero value. It should be noted that sometimes in these tasks it is advisable to establish only a dichotomous, that is, a binary assessment. But a large number of test items allow for variations in grades over a wide range.

We will apply algebraic approach, where it is appropriate and justified. That is, the value of the respondent's score  $C(R^*)$  will proportionally depend on the distance for the answer provided by him  $R^*$  to the correct (or - ideal, reference) answer  $R^0$ , and will symbolically denote the following:

$$C(R^*) = B \cdot (1 - d(R^0, R^*) / d^M),$$

Where  $B$  – the maximum possible score for the answer,  $d^M$  – the maximum possible distance, that is, the distance from the correct answer to the completely incorrect which is opposite to the ideal one. For example, when the answer  $R^0 = a_1 \succ a_2 \succ a_3 \succ a_4$  is correct, then the "worst" answer being farthest from it is  $R^* = a_4 \succ a_3 \succ a_2 \succ a_1$ .

According to works [1, 2], distances in ordinal (rank) scales, in particular, ranking, are measured using various metrics, in particular:

- Cook's metric for mismatching ranks (places, positions) of list elements

$$d^K(R^0, R^*) = \sum_{i \in I} |r_i^0 - r_i^*|, \quad (1)$$

Where  $r_i^0$  – the rank of the  $i$ -th list element in the reference ranking of list elements  $R^0$ ,  $r_i^*$  – the rank of the  $i$ -th list element in the ranking of the elements specified by the respondent;

- Hamming metric

$$d^H(R^0, R^*) = \sum_{i \in I} \sum_{j \in I} |b_{ij}^0 - b_{ij}^*|, \quad (2)$$

Where  $b_{ij}^0 = 1$ ,  $i, j \in I$ , if and only if there is a relation  $a_i \succ a_j$ ,  $i, j \in I$  in the correct answer  $R^0$ ;  $b_{ij}^0 = -1$ ,  $i, j \in I$ , when the ratio  $a_i \prec a_j$ ,  $i, j \in I$  takes place in the correct answer  $R^0$ ;  $b_{ij}^0 = 0$ ,  $\forall i, j \in I$ ;  $b_{ij}^* = 1$ ,  $i, j \in I$ , if and only if there is a relation  $a_i \succ a_j$ ,  $i, j \in I$  in the respondent's answer  $R^*$ ;  $b_{ij}^* = -1$ ,  $i, j \in I$ , when the respondent specified the following order  $a_i \prec a_j$ ,  $i, j \in I$ ;  $b_{ij}^* = 0$ ,  $\forall i, j \in I$  in his/her answer  $R^*$

- Euclid's metric

$$d^E(R^0, R^*) = \left( \sum_{i \in I} (r_i^0 - r_i^*)^2 \right)^{1/2}, \quad i \in I;$$

- a vector of preferences, the elements of which are the number of alternatives that precede each alternative in the ranking.

Maximum possible distances between the benchmark and the worst ranking are:

- For the Cook metric of the form (1)  $d^{K^M} = n^2 / 2$  - for even ; and for odd ones it is calculated according to the formula  $d^{K^M} = (n^2 - 1)/2$ ;

- For the Hamming metric of the form (2) the maximum distance is  $d^{H^M} = n(n-1)/2$ .

It should be noted that the respondent's partial answers should also be perceived and fairly evaluated. It is clear that this procedure must be justified and formalized. That is, the approach described in this work can be generalized in the case of incomplete answers when the respondent could not answer for technical reasons, or did not have time to fill out the test, or does not know the full correct answer, but is sure of its fragments, or does not want to give a full ranking of a given set of elements, etc. This situation can be viewed as a case of incomplete rankings.

**Conclusions.** The paper proposes a new approach to calculating the score for testing using questions that involve compiling a list element. The approach is grounded and formalized; therefore, it can be applied in various subject areas. A positive feature of the proposed approach is the transparency of the rules that are a priori set by the organizers of testing and the absence of situations of uncertainty during the assessment procedure. In addition, the described approach allows for the possibility of further development and improvement.

## References:

1. Hnatienko H.M., Snityuk V.E. Expert decision making technologies: Monograph. - M.: LLC "McLaut", 2008. – 444p.
2. Voloshin A.F., Mashchenko S.O. Models and methods of decision-making: a textbook for students of higher educational institutions. - 2nd edition - M.: Publishing and Printing Centre "Kiev University", 2010. - 336 p.

**Dmitriy Klyushin**

Doctor of Physics and Mathematics, Professor  
*Taras Shevchenko National University of Kyiv*

**RANDOMNESS: OLD AND NEW IDEAS**

**Introduction.** Randomness is one of the most important and difficult notions in computer science. The importance of randomness is justified by the fact that we face randomness in the simulation of real processes (physical, chemical etc.), cryptography and many randomized algorithms. Difficulty of randomness lies in the fact that its nature is controversial. Some classics of computer science (e.g. Turing and von Neumann) considered randomness as a purely physical phenomenon. Other mathematicians (von Mises, Kolmogorov etc.) tried to construct a rigorous mathematical theory of randomness. Thus, now there exist two parallel worlds: the world of physicists, engineers, computer scientists and other investigators that work in the field of real random events and the world of pure mathematicians that attend to theoretical issues of randomness (e.g. consistency of axioms, interconnections between different definitions of randomness etc.). We propose a strongly valid mathematical theory of randomness that is easily applicable in practice. We consider three classical approaches to randomness and propose an alternative model.

**Frequency approach (von Mises).** The model proposed by von Mises uses a concept of an infinite binary sequence  $x_1, x_2, \dots$  (collective) that meet the following conditions: 1) if  $h_n$  is the relative frequency of units in first  $n$  elements of a sequence,

then there is a limit  $\lim_{n \rightarrow \infty} h_n(T, A) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n x_k = p$ ,  $0 < p < 1$  (global regularity); 2) every infinite subsequence  $x_{i_1}, x_{i_2}, \dots$  drawn from the sequence  $x_1, x_2, \dots$  using a rule of acceptable selection has the same limit  $p$  (local regularity). A sequence is random by von Mises if it has two properties: 1) every sequence of relative frequencies of units in the collective has the same limit; 2) the relative frequencies are invariant under the procedure of so-called acceptable selection, that is, the choice of a sequence in which the choice of an  $n$ -th element does not depend on its value.

**Randomness as computational complexity (Kolmogorov).** Kolmogorov complexity of a sequence  $x = (x_1, x_2, \dots, x_N)$ , or is algorithmic entropy, is the length  $K(x)$  of its shortest description, constructed using a Turing machine. If there is some additional information  $y$ , then we can consider conditional Kolmogorov complexity  $K(x|y)$ . A sequence is called Bernoullian by Kolmogorov if its complexity is close to  $\log_2 C_N^k$ , i.e.  $K(x|N, k) \approx \log_2 C_N^k$ . Also, Kolmogorov introduced the notions of chaotic sequence that satisfies the condition  $K(x|N, k) \geq \log_2 C_N^k - m$ . Is a set  $A$  contains finite number of elements  $x_1, x_2, \dots, x_N$ , then the complexity of its element is less or equal to  $\log_2 N$ . An element  $x$  of  $A$  is random if its complexity is close to maximal, i.e.

$K(x|A) \approx \log_2 N$ . The difference  $\log_2 N - K(x|A)$  is a defect of randomness of an element  $x$ . An infinite binary sequence  $x_1, x_2, \dots, x_n, \dots$  is random by Kolmogorov if for any natural  $n$  there is a constant  $c$  so that  $K(x_1, x_2, \dots, x_n) \geq n - c$ .

**Randomness as typicality (Martin-Löf).** An infinite binary sequence  $x_1, x_2, \dots, x_n, \dots$  is called random with respect to uniform measure, if for an arbitrary natural  $n$  there is a constant  $c$  such that  $K(x_1, x_2, \dots, x_n) \geq n - c$ . Obviously, a sequence that is random by Martin-Löf is random by von Mises. From the other side, there are sequences that do not satisfy the conditions of Martin-Löf.

**Alternative model (Petunin–Klyushin).** Consider the trial  $T$  having two consequences:  $A$  and  $\bar{A}$ . Introduce the value  $x_k$  such that  $x_k = 1$  if under  $k$ th repetition of the trial  $T$  the consequence  $A$  is realized, and 0 otherwise. The sequence of bits  $x_1, x_2, \dots$  is called a Bernoullian sequence of order  $p$ ,  $0 \leq p \leq 1$ , if  $\lim_{n \rightarrow \infty} h_n(T, A) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n x_k = p$ , where  $h_n(T, A)$  is the frequency of the event  $A$  under the  $n$  repetition of the trial  $T$ .

The main point in our theory is that for correct definition of the concept of random trial we must use an infinite set of series  $X_1, X_2, \dots$ . For convenience, let arrange these series as indefinite characteristics matrix  $\Theta(T) = \{x_{ij}\}_{i,j=1}^\infty$ .

Let  $X_i = (x_{i1}, x_{i2}, \dots, x_{in}, \dots)$  be rows and  $X_j^* = (x_{1j}, x_{2j}, \dots, x_{nj}, \dots)$  be columns of characteristics matrix  $\Theta(T)$ . It is easy to see that every row  $X_n$  and every column  $X_n^*$  of the matrix  $\Theta(T)$  generate real numbers  $\alpha_n$  and  $\alpha_n^*$  from the segment  $[0,1]$ . Really, put  $\alpha_n = 0, x_{n1} x_{n2} \dots x_{nn} \dots$  and  $\alpha_n^* = 0, x_{1n} x_{2n} \dots x_{nn} \dots$  and consider these expressions as binary ratios. Denote as  $M$  and  $M^*$  the sets of numbers  $\alpha_n$  and  $\alpha_n^*$ , respectively. A trial  $T$  is said to be random if: 1) every row  $X_n$  and column  $X_n^*$  ( $n=1, 2, \dots$ ) of  $\Theta(T)$  is a Bernoullian binary sequence of the same order  $p \in [0,1]$ ; 2) sets  $M$  and  $M^*$  corresponding to rows and columns of  $\Theta(T)$ , respectively, are everywhere dense in  $[0,1]$ . A random experiment  $E$  consists of infinite series of the random trial  $T$ . A random event  $R_E$  is an outcome of  $T$  inducing the experiment  $E$ . The result  $R$  can be outcome of a random event  $A$  or  $\bar{A}$ . A probability  $p(E, A)$  of a random event  $(E, A)$  is an order  $p \in [0,1]$  of a Bernoullian sequence consisting of results of a random trial  $T$  generating the random experiment  $E$ .

<sup>1</sup> **Igor Kovalenko**

Dr.Sc., Professor of the Department of Software Engineering

<sup>2</sup> **Yevhen Davydenko**

Ph.D., Associate Professor of the Department of Software Engineering

<sup>3</sup> **Alyona Shved**

Ph.D., Associate Professor of the Department of Software Engineering

<sup>1,2,3</sup> *Petro Mohyla Black Sea National University*

## STRUCTURING OF GROUP EXPERT JUDGMENTS FORMED UNDER COMPLEX FORMS OF IGNORANCE

**Abstract.** The technology for structuring of group expert assessments formed under complex (combined) types of ignorance caused by uncertainty, inaccuracy, inconsistency, conflict (contradiction) and their combinations has been proposed, based on the mathematical apparatus of the theory of evidence and the theory of plausible and paradoxical reasoning. Different quantified measure of uncertainty and conflict in evidence theory has been considered.

**Keywords:** information technology, decision-making, expert evidence, combination rule, ignorance.

The problem of choosing and making managerial decisions is often solved in the face of an increasing number of external and internal factors that affect the achievement of the goal. In this case a person cannot, at the heuristic level, guarantee synthesis of optimal strategy for decision-making taking into account all the conflicting factors. The situation is aggravated in the presence of various types of ignorance, which are inextricably linked with the processes of formalization and presentation of the existing system of knowledge, and have a direct negative impact on the processes of extraction, analysis, presentation and processing of data and expert knowledge. This, in turn, creates the preconditions for the synthesis of a complex of formalized mathematical models focused on the intellectual support of the processes of preparation and decision-making in conditions of complex forms of ignorance, multi-criteria and risk.

The aim of the work is to present and formalize the main ideas of the technology for structuring expert knowledge, in the context of complex (combined) types of ignorance generated by uncertainty, inaccuracy, inconsistency, conflict, contradiction of expert knowledge. It is based on a systematic approach to the identification of various forms and types of ignorance, and its combinations that creates conditions for reasonable choice and application of methods for expert knowledge analysis.

Let's consider the main ideas of the proposed technology. Let  $A = \{A_i | i = \overline{1, n}\}$  be a set of alternatives, on which certain limitations can be imposed: mutually exclusive and / or mutually exhaustive elements, which determines the type of model in frame of which expert evidence will be formed.

Let a group of experts  $E = \{E_j | j = \overline{1, t}\}$  have formed profiles of expert preferences  $B = \{B_j | j = \overline{1, t}\}$  on the set  $A = \{A_i | i = \overline{1, n}\}$ . The profile  $B_j$  formed by the expert  $E_j$  reflects

his preferences regarding all analyzed elements of the set A (depending on the chosen analysis model).

For each expert the same instruction has been presented, which prescribes what they should do with the set A. The instruction contains information about a scale measurement type, within which experts express their preferences, which in turn affects the information received from experts (words, conditional gradations, numbers, rankings, binary relations or other objects of non-numerical nature).

The profile  $B_j$  formed by the expert  $E_j$  reflects his preferences, expressed within a given scale, with respect to the elements of the set A. The expert himself decides which elements (or selected groups of elements) of the set A will be evaluated. Thus, the profile of preferences  $B_j$  formed by  $E_j$  may contain: estimates expressed with respect to all elements of the set A; the assessments expressed regarding the preferred elements of the set A; the estimates expressed regarding the selected groups of preferred elements of the set A.

Next, the set of expert assessments  $B = \{B_i | i = \overline{1, n}\}$  is fed to the input of the block of ignorance  $nfi$  identification,  $nfi \in NF$ , in this case we are talking about such types of ignorance  $nfi$  as uncertainty, inaccuracy, inconsistency, conflict or their combinations that can be simultaneously present in the knowledge system. In the ignorance identification block, a system of identification criteria  $C_i = \{c_j^{(i)} | j = \overline{1, z}\}$ ,  $i = \overline{1, p}$ , of the analyzed forms of ignorance  $NF = \{nfi | i = \overline{1, p}\}$  is formed. On the basis of formed  $C_i = \{c_j^{(i)} | j = \overline{1, z}\}$  a system of decision rules  $SR_i = \{R_l^{(i)} | l = \overline{1, h}\}$  for  $nfi$  identification is developed. For  $nfi$  identification it can be used one or combination of features, which allows to unambiguously establish the presence of  $nfi$  in the initial data (knowledge) set. The absence of  $nfi$  is recognized if for all set of proposed ignorance identification criteria  $C_i$  confirmed absence of  $nfi$  ( $\forall j : c_j^{(i)} \rightarrow \text{absence of } nfi$ ); the presence of  $nfi$  is recognized if there is at least 1 criterion  $c_j^{(i)}$  (from a given set  $C_i$ ) signals the presence of  $nfi$  ( $\exists j : c_j^{(i)} \rightarrow nfi$ ).

First, it is necessary to form a set of criteria  $C_i = \{c_j^{(i)} | j = \overline{1, z}\}$ , which, in turn, are considered as indicators of the presence of ignorance in the analyzed data (knowledge) set.

For identification the above forms of ignorance (and their combinations), it is proposed to use the following features:

1. The structure of expert evidence.

The level of conflict.

Indicators of the quality of the received evidence: level of auto-conflict; the degree of specificity of the generated evidence, etc.

The degree of inconsistency of the formed expert evidence.

Limitations which are imposed on the frame of discernment A.

The next step is to form a system of decision rules  $SR_i = \{R_l^{(i)} | l = \overline{1, h}\}$  for analyzed forms of ignorance identification.

Based on the formed decision rules  $R_l^{(i)}$ , a rule for choosing a method for modeling the above forms of ignorance (and their combinations) can be obtained:

$$B_j \in \begin{cases} P_1, & \text{if } \forall l: R_l^{(i)} \rightarrow \text{absence } nf_i; \\ P_2, & \text{if } \exists l: R_l^{(i)} \rightarrow nf_i; \end{cases}$$

where  $P_1$  indicates that expert evidence are no contradict, have a high (acceptable) quality, and consistent;  $P_2$  indicates that expert evidence have a high (not acceptable) level of conflict.

If  $B_j \in P_1$ , then it is assumed that the expert evidence are consistent (they are characterized by close evidence, the presence of a low / insignificant level of conflict), and may indicate a high (acceptable) quality of expert evidence.

In this case, evidence combination rules can be used to find the aggregate expert assessments formed in frame of DS or DSm model [1–4]. The algorithm for the complex use of evidence combination rules for aggregation of expert assessments formed in frame of DS model has been proposed.

Let a set of  $P = \{P_i | i = \overline{1, k}\}$  potential combination rules be given. It is proposed to choose a rule  $P \in P$ ,  $m_{combP} = m_i P m_j$ , that minimizes the value of the total uncertainty of the combined  $bpa$ 's  $\min(T(m_{combP}))$ .

Formally, the procedure for choosing a combination rule can be represented as next successive stages. At the first stage, from the set of available combination rules  $P = \{P_i | i = \overline{1, k}\}$ , a subset  $P' \subseteq P$  is selected that satisfies a set of specified criteria  $C = \{c_i | i = \overline{1, q}\}$ .

The analysis model (DS model, DSm model), information about expert's competence, information about conflicts, degree of interaction and the structure of expert evidence can be considered as criteria for choosing a combination rule. Recommendations for the selection of combination rules based on the analysis of a number of criteria are given in [1, 5]. As a result, the initial set  $P = \{P_i | i = \overline{1, k}\}$  will be narrowed down to subset  $P' = \{P_i | i = \overline{1, z}\}$ ,  $z \leq k$ , that obtained by excluding from the set  $P$ , rules that do not satisfy the formed set of criteria

The choosing the combination rule based on the analysis of quantitative characteristics of uncertainty has done on the second stage.

At first, the combination rule  $P_l \in P'$  is selected that maximizes the value of measure that reflects the degree of specificity of the combination result  $\max(\delta_s(m_i P_l m_j))$ ,  $\delta_s(m_i P_l m_j) \neq 1$ .

Next, the combination rule  $P_r \in P'$  is selected that minimizes the value of the measure that reflects the degree of contradiction of the combination result  $\min(Contr(m_i P_r m_j))$ ,  $Contr(m_i P_r m_j) \neq 0$ .

If  $P_l \neq P_r$ , then a combination rule is selected that satisfies the following condition:

$$P = \begin{cases} P_l, & T(m_i P_l m_j) < T(m_i P_r m_j); \\ P_r, & T(m_i P_l m_j) > T(m_i P_r m_j), \end{cases}$$

where  $T$  is a measure of global uncertainty.

If  $B_j \in P_2$ , then it was revealed that there is inconsistency (conflict) of expert evidence, which indicates the presence of several subgroups of experts with similar

assessments, or the presence of so-called dissident experts (one or more experts with estimates significantly different from those of the main group).

As a result, three tasks arise:

1. Identification and exclusion of conflicting (contradictory) evidence (experts).

Partitioning (clustering) of the original set of expert evidence into homogeneous (with an acceptable level of conflict) subgroups.

Aggregation of conflicting (contradictory) evidence in order to develop a group assessment.

The next step is to determine the nature of the task of expert assessments structuring and the choice of the mathematical formalism for the analysis of expert assessments. If the absence of ignorance type  $nfi$  (or its combinations) is confirmed, then the procedure for structuring the set of expert assessments is reduced to solving the problem of finding an aggregated (generalized) expert assessment. If the analysis reveals the presence of  $nfi$  (the set of expert assessments is characterized by low consistency, or inconsistency of assessments), then the procedure for structuring the set of expert assessments is reduced to solving the problem of partition the expert group into several subgroups (clusters) of experts with close (agreed, consistent, homogenous) assessments, for their subsequent analysis and search for an aggregated estimate within each of the selected subgroups.

The final stage is the interpretation of the results of structuring and the development of a group solution.

The proposed technology is designed to solve the problem of analysis (ranking, clustering, ranking clusters) of group expert assessments under multi-criteria, multi-alternativeness and complex forms of ignorance (uncertainty, inaccuracy, inconsistency, conflict) in order to synthesis a final (generalized) assessment that takes into account opinions (arguments) of each expert.

## References

1. K. K. Annamdas, Evidence based uncertainty models and particles swarm optimization for multiobjective optimization of engineering systems, Ph.D. thesis, University of Miami, Coral Gables, Florida, 2009.
2. A. Shved, I. Kovalenko, Y. Davydenko, Method of detection the consistent subgroups of expert assessments in a group based on measures of dissimilarity in evidence theory, in: N. Shakhovska, M. Medykovskyy (Eds.), Advances in Intelligent Systems and Computing IV, CCSIT 2019, Advances in Intelligent Systems and Computing 1080, Springer, Cham, 2020, pp. 36–53.
2. F. Smarandache, J. Dezert, Representation of DSmT, in: Advances and Applications of DSmT for Information Fusion, volume 1, American Research Press, Rehoboth, 2004, pp. 3–35.
3. F. Smarandache, J. Dezert, Advances and applications of DSmT for information fusion. Collected works, volume 2, American Research Press, Rehoboth, 2006.
4. F. Smarandache, Unification of Fusion Theories (UFT). International Journal of Applied Mathematics and Statistics 2 (2004): 1–14.

**<sup>1</sup> Borys Kredentser**

Doctor of technical science, professor, professor of the Department of "Construction of telecommunication systems"

**<sup>2</sup> Dmytro Mogylevych**

Doctor of technical science, professor, professor of the Department of Information and Telecommunication Networks

**<sup>3</sup> Iryna Kononova**

Candidate of technical sciences, associate professor of the Department of Information and Telecommunication Networks

**<sup>4</sup> Vadym Mohylevych**

Student

<sup>1</sup> Military Institute of Telecommunications and Informatization named after Heroes Krut

<sup>2,3</sup> The Institute of Telecommunication Systems of the National Technical University of Ukraine "Kyiv Polytechnic Institute"

<sup>4</sup> Taras Shevchenko National University of Kyiv

## **ANALYTICAL MODEL WITH INTERRUPTION OF SERVICE OF SHORT-TERM OBJECTS WITH TEMPORARY RESERVATION**

**Abstract.** The article presents an analytical model for the maintenance of short-term facilities with temporary redundancy, which are used not continuously, but occasionally. These systems perform tasks that arrive at random times and take some time to complete. Examples of such systems are communication systems as well as various automated control systems. In the developed model, various factors of the real functioning of this class of systems are taken into account.

**Keywords:** information and communication networks, refusal, failure, readiness index, reliability index.

The article a method is proposed for determining the optimal maintenance frequency and extreme values of the indicators of the systems under consideration: a complex reliability indicator – the coefficient of technical utilization and a cost indicator – the average costs per unit time of the object's stay in a working condition (average unit costs).

The need for maintenance arises at the stage of operation of any technical system. The main purpose of maintenance is to ensure the maximum efficiency from the usage of the system during operation. This goal can be achieved by purposeful intervention in the functioning of the system through a rational choice of the type of maintenance, justification of the optimal timing and content of maintenance. Therefore, the solution to the problem of optimal maintenance, in which the extreme values of the selected indicators are provided, is relevant.

Consider a system with a replenishment reserve of time (system "object-time"), in which an object is represented by one generalized structural element. Let the operating time of an object to failure  $t_0$  be distributed according to an arbitrary law

$F(t) = P\{t_0 < t\}$  with a finite mathematical expectation  $\bar{t}_0$ , and a failure manifests itself at the moment of its occurrence (a system with instant indication of failures).

The system performs tasks arriving at random moment of time, and the time intervals between the times of arrival are distributed exponentially with a parameter  $\gamma_f$ . We will assume that the duration of the task is so short in comparison with the average operating time of the object to failure that it can be practically neglected (short-term system). This means that if an incoming task finds the system in a working state, then it is executed with a probability of one.

Suppose that at the initial moment of time  $t = 0$  the object is operational and a scheduled preventive maintenance is assigned at a deterministic time  $T$ , which determines the frequency of maintenance. Before the start of service execution, one of two independent events can occur: object failure or task arrival. Let the object fail first. In this case, the restoration of its performance immediately begins, the duration of which is a random variable  $t_R$ , distributed according to an exponential law with a parameter  $\mu$  and a finite mathematical expectation  $\bar{t}_R = 1/\gamma_z$ . If a task is received during the recovery process, then a delay (delay) in its execution for a time  $\tau_a$  is allowed, which can be a random variable, exponentially distributed with a parameter  $\gamma$ , or a deterministic value  $\tau_a = \text{const}$ . If the restoration of the object's operability is completed before the use of the time reserve  $\tau_a$  (or  $t_a$ ) that determines the permissible lag time, then the task is executed with probability one, otherwise, at the moment the condition  $t_R > \tau_a$  (or  $t_R > t_a$ ) is fulfilled, the system fails (task execution failure).

Let the object fail until the moment  $T$  (an event  $t_0 > T$  has occurred). Then, at the appointed time  $T$ , maintenance starts, the duration of which is exponentially distributed with a parameter  $\Theta$ . If a task arrives in the process of servicing, then servicing is terminated and the object is transferred to the main (operating) mode in a random time  $t_z$ , exponentially distributed with a parameter  $\gamma_z$ . At the same time, the system provides for a permissible time for bringing an object to readiness to perform a task  $\tau_{a1}$ , which can be a random variable with an exponential distribution law with a parameter  $\gamma_1$  or a deterministic value  $\tau_{a1} = \text{const}$ . If  $t_z < \tau_{a1}$  (or  $t_z < t_{a1}$ ), then the received task is executed with probability one, after which the object is transferred to the maintenance completion mode. Otherwise (at  $t_z > \tau_{a1}$  or  $t_z > t_{a1}$ ) at the moment when the time reserve is used up, the system malfunctioning (refusal to complete the task) occurs.

Thus, the system uses two components of the replenished reserve of time: one component is provided in the system itself ( $\tau_a$  and  $\tau_{a1}$  or  $t_a$  and  $t_{a1}$ ), and the other is due to the random nature of the arrival of tasks (the time from the moment of object failure or the start of maintenance to the moment the task is received).

It is necessary to obtain analytical expressions for the service quality indicators: technical usage coefficient  $K_u(T)$  and average unit costs  $\bar{C}(T)$  (objective functions) and determine the optimal values of the service frequency  $T^*$  and  $T_1^*$ , at which the

quality indicators take extreme values: the maximum value  $K_{uu}(T^*)$  for the technical utilization factor and the minimum value  $\bar{C}(T_1^*)$  for the average unit costs.

The studies carried out and the results obtained allow us to draw the following conclusions:

For the same values of the time for performing maintenance and restoring the object's operability  $\mu/0=1$ , the considered maintenance strategy provides a sufficiently high efficiency of the types of restoration work used, and the value of the technical utilization factor depends significantly on the ratio of parameters  $\gamma_z t_{a1}$ ,  $\gamma_f/\mu$ ,  $\gamma_f/\gamma_z$ , that determine the amount of time reserve and the efficiency of its use.

In particular, it can be seen that an increase in the intensity  $\gamma_z$  of the transfer of the object from the maintenance mode to the main mode (with an increase in  $\gamma_z t_{a1}$  and  $t_{a1} = \text{const}$ ) by only two times leads to a significant increase in the efficiency of maintenance. This is due to the fact that with an increase in  $\gamma_z t_{a1}$  system downtime due to maintenance, they are more and more compensated for by the time reserve existing in the system, while the amount of downtime compensation when restoring the object's operability does not change.

In the case when the value  $\lambda T \rightarrow \infty$  (at  $\lambda = \text{const}$ ), in the system, only the restoration of operability after failures is carried out and coefficient of technical usage asymptotically tend to the value of the stationary availability factor of the system  $K_h$  in which maintenance is not carried out.

The value of the technical utilization coefficient is also significantly influenced by the failure rate  $\lambda$  of the object under consideration.

#### References:

1. NSU 2860-94, Reliability of technology Terms and definitions, 1995. URL: [http://ksv.do.am/publ/dstu/dstu\\_2860\\_94/3-1-0-1102](http://ksv.do.am/publ/dstu/dstu_2860_94/3-1-0-1102).
2. W. Ahmad, O. Hasan, S. Tahar, Formal reliability and failure analysis of ethernet based communication networks in a smart grid substation, Formal Aspects of Computing, (2020) 71–111. doi: 10.1007/s00165-019-00503-1.
3. D. Mogilevich, I. Kononova, Improved Estimates for the Reliability Indicators of Information and Communication Network Objects with Limited Source Information, Springer Nature Switzerland (2019) 101–117. doi:10.1007/978-3-030-16770-7\_5.
4. Hall, P., Jin, Y. and Samaniego, F. J. Nonparametric estimation of component reliability based on failure-time data from systems of varying design, Statistica Sinica (2015) 1313 – 1335. doi: 10.5705/ss.2014.092
5. T. Zin, J. Lin, J.-S. Pan, P. Tin, M. Yokota Reliability Specification of Telecommunication Networks Based on the Failure Influence by Using Evolutional Algorithm, Genetic and Evolutionary Computing, Springer International Publishing Switzerland (2016) 115-126. doi: 10.1007/978-3-319-23207-2\_12.

**Alexander Makarenko**

Doctor of Science, Professor

*Institute of Applied System Analysis at National Technical University Igor Sikorski  
Kiev Polytechnic Institute*

## CELLULAR AUTOMATA MODELS WITH RIEMANN SURFACES

**Abstract.** New class of mathematical models for computation theory is considered – namely cellular automata (CA) with branching transition functions. The key point is possible multivaluedness of transition functions. Some cases with complex-value functions had been considered. The case of continuous-valued CA and their finite-valued approximations are discussed.

**Keywords:** Cellular automata, complex functions, Riemann surfaces, continuous-valued CA, branching, computations

### 1. Introduction

Typically, classical cellular automatic machines (CA) have a cell assembly structure with a certain set of possible states. The states change at discrete points in time according to certain rules. Up to now, as far as we know, these dynamic rules have been defined using unambiguous transition functions. Recently, there has been a need to investigate CA with multi-valued transition functions [1]. Further study of such new CA (with multivaluedness) depends on examining specific examples of such objects. In this paper, we propose complex-valued cellular automata with a branched (multi-valued) structure with values on the Riemann surface. CA with the location of cells on the Riemann surface are also considered.

### 2. One-dimensional case

In this paragraph we will illustrate the simplest cases.

2.1 One branching point where branching conditions are not met for any point (e.g. all values during evolution are on the same branch of a function, e.g. on a function  $\sqrt{\varphi}$  branch. Then the branching point is the only point, but let the CA values be outside the vicinity of the branching point. This corresponds to the class of unambiguous CA, but with certain generalizations. Namely, in our general case, CA looks like

$$x_i^{n+1} = \varphi(\{x_i^n\}_N; \vec{\alpha}) \quad (1)$$

$x_i^n$  - state value of  $i$  cell at  $n$  moment of time ( $n = 0, 1, 2, \dots$ ). In the case of classic CA, such as a game of life, the state of the cell is 0 or 1. In our case, the complex function  $\varphi$ ,  $x_i^n \in M, M \subset C^1$  where  $M$  - some subset of the space of complex numbers.

Here we represent the evolution of the values of one cell. Note that just in such case we have essential difference from the classical case. We can assume a whole set of values  $M$ , including  $\text{card}M = \infty$ , when a continuum of values is also allowed. Note that in general case it is possible to consider approximations of a continuum set  $M$  using a discrete set  $M_K \subset M$  with a finite number  $K$  of possible cell values. An interesting task

is the rigorous mathematical study of the applicability of such approximation, as well as the crystallization of the approximation.

2.2 A more complicated case is when the branching point of the transition function appears in the value range for only one cell. A simpler case appears to be implemented for the function  $\sqrt{\varphi}$  presumably with no jumps between branches, which requires further research. In this case, the idea of approximation with a finite set of values by branches can also be applied.

However, for arbitrary analytical functions with a large number of branches (or even an infinite number of branches) the problem is more complex. In this case, all studies on the limit behavior of CA remain valid. But, limit behavior must be investigated for a multi-valued case, for which the theory of multi-valued dynamic systems should be involved. Along the way, we will point out another new class of research problems. For traditional cellular automata (including the game of Life) we can set the following problem: according to the well-known rules of CA, to find CA with a complex-valued transition function such that traditional CA is an approximation of complex-valued CA. In this case, it is possible to raise the problem of studying traditional CA. Probably, the function  $\sqrt{\varphi}$  can only evolve over two branches independently of each other (without jumping from branch to branch).

This creates a whole new set of problems associated with finding special function  $\varphi$ . The first class of problems is related to the construction related to a specific CA (including traditional ones). Then it is necessary to simplify problem definition and achieve only the proximity of a given CA solution with a certain complex function  $\bar{\varphi}$  at a certain time range  $[0, T]$ . It is also possible to use functions  $\varphi$  with an infinite number of branching points. It is also important to construct a function with given branching points so that these branching points set a fixed behavior of the solution or at least approximate the required behavior of a multi-valued solution on a certain range.

2.3 Set out in pp. 2.1-2.2, it is easy to generalize for other problems as well. Thus, nothing limits the transfer from the case of 1D space with only one branching point in one cell to the case of possible branching points in each cell of the cell space. Everything said above is carried over to the case of 2D, 3D, ..., ND spaces. But it is also possible to generalize the very concept of cellular automata. Thus, it is possible to consider cell automata on Riemann surfaces. A particular problem is the study of the dependence of the dynamics of CA on the topology of the space in which CA is given.

By the way, the transition from CA with a continuum of possible values to approximations of finite sets correlates with the study of continuous dynamical systems with the methods symbolic dynamics in ergodic theory.

Thus in proposed paper we had been introduced new class of cellular automata with presumable multivaluedness. New research problems are proposed.

#### **References:**

1. Makarenko A. Multivaluedness in cellular automata with strong anticipation and prospects for computation theory. WSEAS Transactions on Information Science and Applications. 2020. Vol. 17. Pp. 69-79.

**<sup>1</sup> Volodymyr Polishchuk**

PhD, Associate professor of department of Software Systems

**<sup>2</sup> Mykola Malyar**

Doctor of Engineering Science, Professor, Professor of department of cybernetics and applied mathematics

**<sup>3</sup> Andriy Polishchuk**

PhD student of department of cybernetics and applied mathematics

<sup>1-3</sup> Uzhhorod National University, Narodna Square, 3, Uzhhorod, 88000, Ukraine

## **THE TECHNOLOGY FOR DETERMINING THE LEVEL OF PROCESS CONTROL IN COMPLEX SYSTEMS**

**Abstract.** A study of the current problem of developing technology for determining the level of process controllability in complex systems, with different modes of operation was done. In this investigation for the first time was proposed a fuzzy mathematical model based on expert hybrid data, using linguistic and quantitative variables.

**Keywords:** process controllability, fuzzy set, decision-making, modes of system operation.

Information systems and technologies are increasingly replacing intellectual ones, but the desire to represent the future does not disappear. At the present stage of human development, there is an increasing desire to control the processes in the world. Tools for analyzing massive data sets, today allow you to get new, high-quality knowledge from various information. The amount of data is growing rapidly in all areas, with them there is a need for processing, which is reduced to obtaining knowledge, on the basis of which further decisions are made. Such technologies make life more comfortable, more stable, smarter, and most importantly safer.

Today, decision support systems are increasingly using data mining tools. But most of them are designed to make decisions in the safe mode of operation of systems. For conditions where the system is rapidly changing modes from safe operation, emergency to disaster, most decision support models are not able to adequately assess the situation. Proof of this is the work of the municipality, region, state in the conditions of, for example, a pandemic of coronavirus infection (COVID-19).

Every day, due to various circumstances, extraordinary situations occur that lead to material destruction, threat to health or life. Often through the fault of decision-maker, making the wrong management decisions. Management decisions directly affect the safe state of the system environment. Sometimes decision-maker try to control processes in complex systems without suspecting that process control is very low or non-existent. There are circumstances that do not depend on people. However, there is our desire to know whether we can influence certain processes. There are events that we cannot change, but we must work to anticipate them.

At a time when a complex system is moving from a safe mode of operation to a catastrophe, the situation is changing rapidly, the controllability of processes is

declining, the data influencing decision-making are becoming increasingly vague. Any emergency or catastrophe is the end result of a consistent transition of the normal mode of operation of the system, respectively, in an emergency or catastrophic situation.

Confirmation of the above is illustrated by the following example. The investigation of the plane crash shows what factors and influences accompanied the events of the crash. The conclusion of the causes of the crash indicates whether the accident situation depended on the pilots, the technical condition of the vessel, weather conditions and the possibility of avoiding the accident. The factors of the internal condition of the aircraft, the influence of the external environment, the actions of pilots in an emergency situation and human factors in the management of air traffic are indicated. In other words, the assessment of process controllability in a complex ship piloting system is indicated. And most importantly, the International Civil Aviation Organization is taking clear steps to prevent similar situations in the future.

The above, argues and confirms the relevance of our study on the application of intelligent analysis, systems approach, processing fuzzy data to develop technology to assess the level of process control in complex systems from normal to disaster. The relevance of this study proves the need to understand the controllability of processes in different objects of study and different modes of operation, for sustainable operation of systems, achieving its goals, formalization of such processes, especially in a pandemic COVID-19.

The logic of the study is as follows: if the overall assessment of the system is high, the factors influencing the control processes for the appropriate mode of operation, then we can talk about a high level of process control in the system, and competent management decisions will achieve the goals of the system. ensure the appropriate level of security of the system operation environment.

The logic of the study is as follows: if the overall assessment of the system is high, the factors influencing the control processes for the appropriate mode of operation, then we can talk about a high level of process control in the system, and competent management decisions will achieve the goals of the system. ensure the appropriate level of security of the system operation environment.

Let it be known some object of study that we will consider as a complex poorly structured system  $S$ . There are many known system goals and many factors that affect the controllability of complex systems. Also known are the indicators of the system that allow to quantify or qualitatively assess the property of the system. Fuzzy models of system evaluation are built on the basis of known indicators. Within this, it is necessary to assess the controllability of processes in the object of study for quality decision-making depending on the modes  $C$ : regular situation, out of the regular mode, critical situation, emergency, accident situation, accident, catastrophic situation, catastrophe.

Suppose we have a set of indicators  $K$ , according to which we will assess the level of process control in a complex system  $S$ . Indicators can be a whole system of criteria, factors and models, based on which a single aggregate assessment is derived. For example: the level of control of the aircraft is influenced by indicators that depend on the factors of the technical condition of the aircraft, external conditions, human factors and risk-oriented situations; the level of safe financing of innovative projects is

influenced by factors of the microenvironment of the project (strength of the idea), environmental factors (competitors, market, policy), factors of risk management and anticipation, and the main level of project developers.

Formally, we can present a fuzzy model for determining the level of process control in complex systems, taking into account the different modes of operation as follows:

$$A(I; M; C) \rightarrow R(\mu), \quad (1)$$

$A$  – an operator that matches a set of output values  $R$ , with input variables  $I; M; C$ . The input data of the model are:  $I$  – expert indicator or quantitative assessment of the level of process control in the system, or a combination thereof;  $M$  – taking into account the reasoning of the decision-maker on the scenario of unfolding events;  $C$  – system operation mode. At the output of the evaluation model we have:  $\mu$  – assessment of process control in complex systems taking into account different modes of operation on the basis of which the level is determined  $R$ .

As a result of the study, the following results were obtained for the first time:

- ✓ a fuzzy mathematical model for determining the level of process control in complex (weakly structured) systems taking into account different modes of operation, based on expert hybrid data, using linguistic and quantitative variables has been proposed. The estimates of the indicators of a complex system are aggregated in relation to the decision-maker reasoning about the scenario of events. One aggregate estimate of the level of process control in a complex system with respect to different modes of operation is derived. Based on the obtained result, can be determined the level of safe operation of the system to prevent negative consequences or confidence in achieving the goals of the system. All this allows to reveal the uncertainties of expert opinions and data obtained, justifying the degree of decision-making and to draw adequate conclusions, taking into account the different modes of operation;
- ✓ experimental testing of the conducted research in the task of determining the level of process control in the airport management's information systems, with data security threats, taking into account different modes of operation was done. A web application was created for the developed model, with the help of which you can configure models and conduct experimental research for various application tasks.

The rationality of the obtained aggregate estimates, the level of safe state to prevent negative consequences of the system, or the confidence of achieving the goals of the system in different modes of operation, proves the advantages of the developed model. The reliability of the obtained results is ensured by the correct use of intellectual analysis of knowledge, system approach, fuzzy set theory, which is confirmed by research results.

<sup>1</sup> Volodymyr Rusyn

PhD, Assistant Professor of Department of Radio Engineering and Information Security

<sup>2</sup> Aceng Sambas

Lecturer of Department of Mechanical Engineering

<sup>1</sup> Yury Fedkovych Chernivtsi National University

<sup>2</sup> Universitas Muhammadiyah Tasikmalaya, Jl. Tamansari No. KM 2,5, Mulyasari, Kec. Tamansari, Tasikmalaya, Jawa Barat, 46196, Indonesia

## CIRCUIT REALIZATION OF THE PULSE TRANSFORMATION OF THE ANALOG NONLINEAR SIGNALS BASED ON CHUA'S GENERATOR

Chaotic signals have the following characteristics: sensitive dependence on the initial conditions, unpredictability, similarity to the noise, and difficulty to be deciphered. Therefore, it is especially suitable to be applied to the secure communication field [1-3]. In this paper, the pulse transformation of analog nonlinear signals is presented. Classical Chua's chaotic generator as a main part that generate nonlinear signals was used for demonstrate of this process. Electronic circuit and components with nominal values are presented. MultiSim software environment was selected for analysis and demonstrate of computer modelling results.

The electronic circuit is displayed in Fig. 1, with component values: capacitors C1 = 100 nF, C2 = 10 nF, diodes VD1, VD2 – 1N4148, DA1, DA2 – operational amplifier TL082, powered by a 9 V, VCC1 – +5 V, one transistor VT1 – 2N2222A, inductor L1 = 18 mH, voltage sources V1 = V2 = 9 V, potentiometer R1 = 2 k $\Omega$  (1.7 k $\Omega$ ), resistors R2 = 1 k $\Omega$ , R3 = 3.3 k $\Omega$ , R4 = 47 k $\Omega$ , R5 = 3.3 k $\Omega$ , R6 = 1 k $\Omega$ , R7 = 5.6 k $\Omega$ , R8 = 1 k $\Omega$ , R9 = 47 k $\Omega$ , R10 = 2.2 k $\Omega$ , R11 = R12 = 290  $\Omega$ , R13 = 1.2 k $\Omega$ .

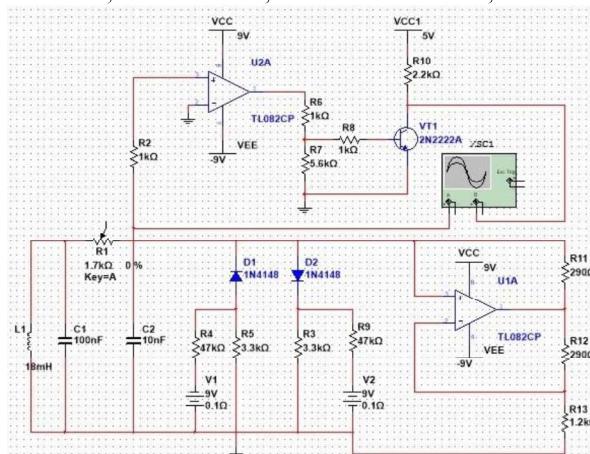


Figure 1 – Circuit realization of the process of pulse transformation

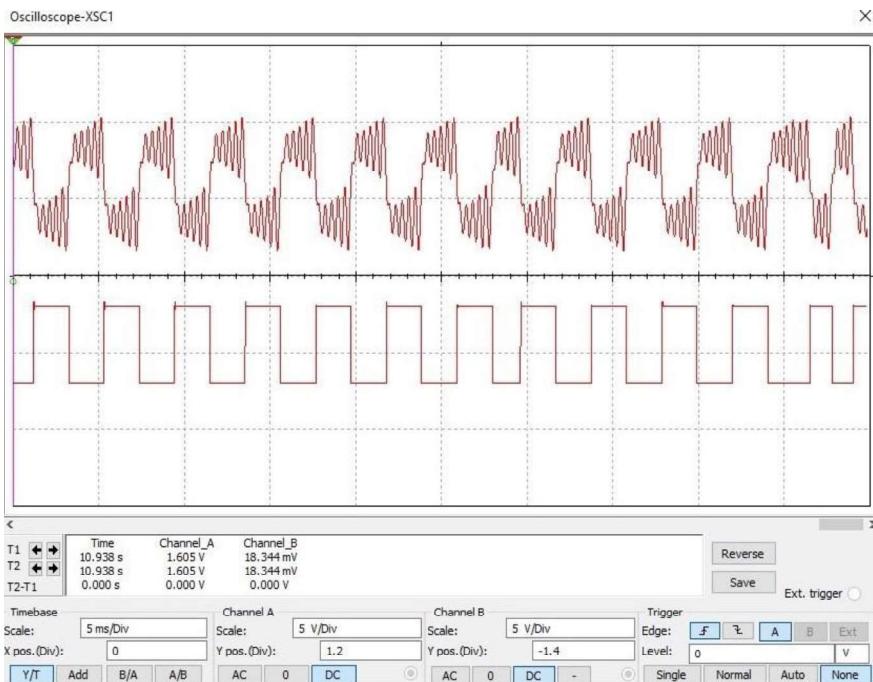


Figure 2 – Time series and pulse transformation

Fig. 2 shows time series and pulse transformation for chaotic coordinate X. The simulation parameters for Fig. 2: U1 = 5 V/div, U2 = 5 V/div, time scale 2 ms/div.

#### References:

1. Sambas A., Sanjaya WS. M., Mamat M., Putra Prastio R., Azar A.T. (2016) Mathematical modelling of chaotic jerk circuit and its application in secure communication system. Studies in Fuzziness and Soft Computing. Vol. 337, 133-153.
2. Sambas, A., Sanjaya, M. WS, Mamat, M., Diyah, H. (2013) Design and analysis bidirectional chaotic synchronization of rossler circuit and its application for secure communication. Applied Mathematical Sciences, Vol 7(1), 11-21.
3. Rusyn, V., Mohamad, M.A., Purwandari, D., Mamat, M., Titaley, J., Pinontoan, B. (2020) Chaotic and Controlling Regimes of a New Modified Chua's Generator. Journal of Advanced Research in Dynamical & Control Systems, Vol. 12, Issue 02, 556-561.

<sup>1</sup> **Vladimir Semenov**

D. Sc., Professor

<sup>2</sup> **Yana Vedel**

PhD student

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## CONVERGENCE OF ADAPTIVE METHODS FOR EQUILIBRIUM PROBLEMS IN HADAMARD SPACES

**Abstract.** We study and describe two new adaptive two-stage proximal algorithms for the approximate solution of equilibrium problems in Hadamard spaces. For pseudomonotone bifunctions of Lipschitz type, theorems on the weak convergence of sequences generated by the algorithms are proved. A new regularized adaptive extraproximal algorithm is also proposed and studied.

**Keywords:** Equilibrium problems, Hadamard space, adaptability, convergence, regularization, extraproximal algorithm.

In the talk based on works [1-3], we study and describe two new adaptive two-stage proximal algorithms for the approximate solution of equilibrium problems in Hadamard spaces. The proposed rules for choosing the step size do not calculate the values of the bifunction at additional points and do not require knowledge of the Lipschitz constants of the bifunction. For pseudo-monotone bifunctions of Lipschitz type, theorems on the weak convergence of sequences generated by the algorithms are proved. The proofs are based on the use of Fejer properties of algorithms with respect to the set of solutions to the problem.

A new regularized adaptive extraproximal algorithm is also proposed and studied. To regularize the adaptive scheme, the classical Halpern scheme was used, a version of which for Hadamard spaces was studied in [4].

Let  $(X, d)$  be a Hadamard space. For a non-empty convex closed set  $C \subseteq X$  and bifunction  $F : C \times C \rightarrow \mathbb{R}$  we consider an equilibrium problem:

$$\text{find } x \in C: F(x, y) \geq 0 \quad \forall y \in C. \quad (1)$$

We assume that the following conditions are satisfied:

- $F(x, x) = 0$  for all  $x \in C$ ;
- functions  $F(x, \cdot) : C \rightarrow \mathbb{R}$  convex and lower semicontinuous for all  $x \in C$ ;
- functions  $F(\cdot, y) : C \rightarrow \mathbb{R}$  are weakly upper semicontinuous for all  $y \in C$ ;

- bifunction  $F : C \times C \rightarrow \mathbb{R}$  pseudomonotone, that is for all  $x, y \in C$  from  $F(x, y) \geq 0$  it follows that  $F(y, x) \leq 0$ ;
- bifunction  $F : C \times C \rightarrow \mathbb{R}$  Lipschitz type, that is, there are two constants  $a > 0$ ,  $b > 0$  such that for all  $x, y, z \in C$

$$F(x, y) \leq F(x, z) + F(z, y) + ad^2(x, z) + bd^2(z, y).$$

We denote set of solutions for (1) by  $S$ . In what follows, we will assume that  $S \neq \emptyset$ .

Let's describe one of the algorithms.

### **Algorithm 1. Adaptive version of the extraproximal algorithm.**

**Initialization.** Choose element  $x_1 \in C$ ,  $\tau \in (0, 1)$ ,  $\lambda_1 \in (0, +\infty)$ . Assume  $n = 1$ .

**Step 1.** Calculate

$$y_n = \text{prox}_{\lambda_n F(x_n, \cdot)} x_n = \arg \min_{y \in C} (F(x_n, y) + \frac{1}{2\lambda_n} d^2(y, x_n)).$$

If  $x_n = y_n$ , then stop and  $x_n \in S$ . Otherwise, go to step 2.

**Step 2.** Calculate

$$x_{n+1} = \text{prox}_{\lambda_n F(y_n, \cdot)} x_n = \arg \min_{y \in C} (F(y_n, y) + \frac{1}{2\lambda_n} d^2(y, x_n)).$$

**Step 3.** Calculate

$$\lambda_{n+1} = \begin{cases} \lambda_n, & \text{if } F(x_n, x_{n+1}) - F(x_n, y_n) - F(y_n, x_{n+1}) \leq 0, \\ \min \left\{ \lambda_n, \frac{\tau}{2} \frac{d^2(x_n, y_n) + d^2(x_{n+1}, y_n)}{F(x_n, x_{n+1}) - F(x_n, y_n) - F(y_n, x_{n+1})} \right\}, & \text{otherwise.} \end{cases}$$

Set  $n := n + 1$  and go to step 1.

Algorithm 1 converge weakly. To ensure the convergence of the approximating sequences in the metric of the space to the solution of the equilibrium problem (1), we considered algorithm 1, regularized using the Halpern scheme [4], with adaptive choice of the step size.

### **Algorithm 2.**

**Initialization.** Choose elements  $a \in C$ ,  $x_1 \in C$ , values  $\tau \in (0, 1)$ ,  $\lambda_1 \in (0, +\infty)$  and sequence  $(\alpha_n)$  such that  $\alpha_n \in (0, 1)$ ,  $\lim_{n \rightarrow \infty} \alpha_n = 0$ ,  $\sum_{n=1}^{\infty} \alpha_n = +\infty$ . Set  $n = 1$ .

**Step 1.** Calculate

$$y_n = \text{prox}_{\lambda_n F(x_n, \cdot)} x_n.$$

**Step 2.** Calculate

$$z_n = \text{prox}_{\lambda_n F(y_n, \cdot)} x_n.$$

**Step 3.** Calculate

$$x_{n+1} = \alpha_n a \oplus (1 - \alpha_n) z_n.$$

**Step 4.** Calculate

$$\lambda_{n+1} = \begin{cases} \lambda_n, & \text{if } F(x_n, z_n) - F(x_n, y_n) - F(y_n, z_n) \leq 0, \\ \min \left\{ \lambda_n, \frac{\tau}{2} \frac{d^2(x_n, y_n) + d^2(z_n, y_n)}{F(x_n, z_n) - F(x_n, y_n) - F(y_n, z_n)} \right\}, & \text{otherwise.} \end{cases}$$

Set  $n := n + 1$  and go to step 1.

We plan to consider more special versions of algorithms for variational inequalities and minimax problems on Hadamard manifolds (for example, on the manifold of symmetric positive definite matrices).

This work was supported by Ministry of Education and Science of Ukraine (project “Mathematical modeling and optimization of dynamical systems for defense, medicine and ecology”, 0219U008403).

**References:**

1. Y.I. Vedel, G.V. Sandrakov, V.V. Semenov, L.M. Chabak, Convergence of a Two-Stage Proximal Algorithm for the Equilibrium Problem in Hadamard Spaces, Cybernetics and Systems Analysis. 56 (2020) 784–792. doi:10.1007/s10559-020-00299-6.
2. Y.I. Vedel, V.V. Semenov, Adaptive algorithms for equilibrium problems in Hadamard spaces, Dopov. Nac. akad. nauk Ukr. (2020). doi:10.15407/dopovidi2020.08.026.
3. Ya.I. Vedel, S.V. Denisov, V.V. Semenov, Algorithm for variational inequality problem over the set of solutions the equilibrium problems, Journal of Numerical and Applied Mathematics. (2020) No. 1 5–17.
4. M. Bacak, Convex Analysis and Optimization in Hadamard Spaces, De Gruyter, Berlin, 2014.

**<sup>1</sup> Valentyn Sobchuk**

Doctor Engineering Sciences, Professor of Department of Higher Mathematics

**<sup>2</sup> Yuliya Olimpiyeva**

Senior Lecturer of Department of Higher Mathematics

**<sup>3</sup> Andrii Musienko**

Doctor Engineering Sciences, Associate Professor Department of Design Automation for Energy Processes and Systems

**<sup>4</sup> Andrii Sobchuk**

graduate student of the Department of Information and Internet Technologies

*<sup>1,2</sup> State University of Telecommunications*

*<sup>3</sup> National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"*

*<sup>4</sup> Taras Shevchenko National University of Kyiv*

## **ENSURING THE PROPERTIES OF FUNCTIONAL STABILITY OF MANUFACTURING PROCESSES BASED ON THE APPLICATION OF NEURAL NETWORKS**

Most of the systems that modern science explores are complex. They form multilevel structures. The functioning of such systems is not described by the usual sum of interactions of their elements. Complex technical systems (CTS) are built to perform certain special tasks. The development of artificial systems leads to the complication of their functioning and the emergence of new properties, such as functional stability. [3-5]. Many modern studies in the field of functional stability of complex technical systems and in the field of artificial intelligence require in-depth analysis of the results and the possibility of their implementation in order to build functionally stable information systems. These information systems use artificial intelligence algorithms to diagnose the state of the system and maintain its functioning in accordance with its main purpose for the required period of time.

Of particular interest for our research are the properties of systems that ensure the possibility of their functioning when internal and external parameters change. The property of functional stability is the ability of the system to adapt to new and not always taken into account situations, to withstand any internal or external influences, while realizing its target function. The main task is to increase functional stability and, as a consequence, reliability through the use of various types of redundancy (functional, algorithmic, technical, topological, and temporary), and organizing it in the form of duplicate or majority structures. One of the methods the reliability theory offers to improve the quality of technical systems functioning is to ensure fault tolerance.

Technological production processes to ensure the achievement of parameters  $x(i)$  at each stage require external influences on the production process –  $u(i)$  (active effect, energy effect, chemical or other technological impact at each stage).

It is clear that the final quality of the product as well as the intermediate quantity at each stage depend on tight adherence to technology and ensuring control over the necessary parameters at each previous step. Next, we assume that this a priori requirement is met.

Denote:

$A(i)$  – matrix of product quality indicators dependence on  $(i + 1)$ -st stage on indicators on  $i$  stage, actually matrix of production process;

$C(i)$  – a matrix that determines the structure of influence on the production process  $u(i)$ .

Then the mathematical model of the technological process, provided by the information systems of the production enterprise, can be written as follows

$$\begin{aligned} x(i+1) &= A(i)x(i) + C(i)u(i), \quad i = 1, 2, \dots, N \\ x(i) &\in \mathbb{R}^n, \quad A(i) \in \mathbb{R}^{n \times n}, \quad C(i) \in \mathbb{R}^{n \times m}, \quad u(i) \in \mathbb{R}^m. \end{aligned} \quad (1)$$

Where  $x = (x_1, x_2, \dots, x_n)^T$  – state vector of dimension  $n$ ,  $u = (u_1, \dots, u_m)^T$  – the control vector has dimension  $m$ ,  $A(t)$  –  $n \times n$  – matrix,  $C(t)$  –  $n \times m$  – matrix,  $t = 0, 1, \dots, N - 1$ . Denote by  $I_N = \{0, 1, \dots, N\}$ ;  $x(t, x_0, u)$  system solution,  $t \in I_N$  when managing  $u(t)$ ,  $t \in I_{N-1}$ .

When automating such processes in practice, it is necessary to set certain management tasks that actually describe the design conditions of the control function  $u$ , which provides controlled purposeful execution of the process. The main task is to consider the search for the control function  $u$ , which ensures the execution of the process, so that the result of the process provides the final production in  $x(N)$  products that meet all the quality characteristics required by current standards. If at the end of the process the product has deviations from the specified standard parameters, then such deviations are guaranteed to fall into the set of permissible tolerances, which are defined by the current standards for such products. Mathematically, this means that there is a desired final state  $x_N$  and a positive parameter  $\varepsilon > 0$  such that

$$\|x(N) - x_N\| < \varepsilon.$$

Let  $\bar{x} = (\bar{x}(0), \bar{x}(1), \dots, \bar{x}(N))^T$  — reference process. The reference process guarantees full compliance with the set of parameters  $x(k)$ ,  $k = 0, 1, \dots, N$ , which should be followed in the ideal execution of the production process at all stages and at each of the links. This is a certain median value, which simultaneously assumes the presence of an a priori set of permissible deviations of the system parameters. The parameter  $\varepsilon > 0$  is set, which determines the set of permissible deviations (tolerances) from the reference values.

**Definition.** If for the data of matrices  $A$ ,  $C$  and vector  $u$  there exists a solution  $x = \bar{x} + e$  of system (7) such that  $\|e\| \leq \varepsilon$ , then such a technological process will be called functionally stable.

A valid theorem.

**Theorem.** Let the condition be fulfilled

$$u^T Qu = 0,$$

where  $Q = C^T Z(A^T)C$ ,  $Z(A^T) = E - AA^+$  – projector on a matrix core  $A^T$ ,  $A^+$  – pseudo-inverted matrix. With

$$\|A^+(Cu - A\bar{x})\| \leq \varepsilon.$$

Then the technological process described by equation (1) is functionally stable.

Ensuring the functional stability of production processes is one of the most important tasks. The implementation of technological processes requires constant

monitoring of product quality with simultaneous diagnostics of units and assemblies of production equipment. To solve this problem, it is proposed to use neural network technologies that will solve this problem.

Mathematically, a neuron is an adder of weights, the only output of which is determined through its inputs and a weight matrix in this way

$$y = f(u), \quad u = \sum_{i=1}^n w_i x_i + w_0 x_0.$$

Where  $w_i$  and  $x_i$  – respectively, the signals at the inputs of the neuron and the weights of the inputs, function  $u$  - induced local field,  $f(u)$  – activation function. A neural network (NN) is defined as an interconnected set of neurons. The variety of neural network models is determined by the presence of various activation functions and the topology of their connection and interaction.

Through training, the network acquires the ability to correctly respond not only to training data, but also to process other data from the valid set well. Moreover, this set can be constructed in such a way that the conditions of the theorem are satisfied. That is, the neural network has the ability to generalize. One of the advantages of neural networks is the ability to approximate a given function using an appropriate network architecture. Thus, on the basis of a neural network, an effective model of a dynamic system and a network observer can be built. In studies [1,2] various neural networks (with one hidden layer, multilayer feed forward network, radial basis network, Hopfield network, self-organizing neural network) are used to build various systems for test diagnostics of equipment to ensure functionally stable implementation of technological processes.

This paper investigates the use of numerous classes of neural networks for diagnosing the state of equipment, ensuring the functionally stable functioning of the corresponding technological processes of industrial enterprises. The expediency of using artificial neural networks as an effective toolkit for ensuring the functional stability of production processes has been substantiated.

### **References:**

1. Karpenko M., Sepehri N., Scuese D. Diagnosis of Process vavle Actuator Faults Using a Multilayer Neural Network // Control Engineering Practice. – Vol. 11, 2003, N 11: pp. 1289–1299.
2. Osokowski S. Neural networks for processing information. Finance and Credit, 202: 344p.
3. Sobchuk V. Method of creating a single information space in a production enterprise with a functionally sustainable technological process// Control, navigation and communication system. Academic Journal. pub.6(58), 2019: pp. 84 – 91.
4. Sobchuk V.V. Methodology for the establishment of a single information space in a computerized enterprise with a functionally stable computerized process. Science periodically seeing "Management systems, navigation and communication". Poltava, 2019, 6 (58): pp. 84 – 91.
5. Sobchuk V.V., Barabash O.V., Musinko A.P., Laptv O.A. Analysis of the main approaches and stages of the safety of the power of the functional efficiency and information systems of the enterprise. Sciences of Europe, Praha: Sciences of Europe, 2019. Vol 1, No 42: pp. 41 – 44.

**<sup>1</sup> Mykhailo Solomko**

PhD, Associate Professor

**<sup>2</sup> Liudmyla Zubyk**

PhD, Associate Professor

**<sup>3</sup> Yaroslav Zubyk**

Senior Lecturer

**<sup>4</sup> Anastasiia Ivanytska**

PhD, Assistant

*<sup>1,3</sup> National University Water and Environmental Engineering*

*<sup>2,4</sup> Taras Shevchenko National University of Kyiv*

## **MODIFIED ALGORITHM FOR TRANSFORMATION OF BOOLEAN FUNCTIONS**

**Abstract.** Classic methods of constructing of logic circuits need to be generalized and optimized that can be implemented by making changes to existing processing algorithms. One of the options for minimizing logical functions is based on the combining sequences of logical operations. The method of image transformations is used to minimize completely and incompletely defined Boolean functions, minimize Boolean functions in different views, minimize Boolean function of the system etc. Attention focusing of potential expansion and modernization of future technical base of computing elements makes our researches relevant.

**Keywords:** Boolean functions, models for adding of binary codes, optimization methods, codes figurative transformations

**Introduction.** Basic methods of constructing of logic circuits were described early in various sources [1, 2, 3]. However, they constantly need to be generalized and optimized that can be implemented by making changes to existing processing algorithms. One of the options for minimizing logical functions is based on the combining sequences of logical operations. The method of image transformations is used to: minimize completely and incompletely defined Boolean functions, minimize Boolean functions in different views (DNF, CNF), minimize Boolean function of the system, minimize Boolean functions in Schaeffer, Webb (Pierce) monobases, minimize random Boolean function (algorithm Blake-Poretsky) [4].

Focusing on potential futures expansion and modernization of the technical base of computing elements, research aimed at optimizing algorithms for minimizing logical functions remains relevant.

The purpose of this research is to modify algorithms for minimizing logical functions through image transformations. To achieve this goal, the following tasks were forming:

- generalize the known patterns of minimization of logical functions;
- to improve the algorithm for minimizing logical functions based on the method of image transformations.

**Modifying of algorithm trough figurative transformations.** The transformation of the Boolean functions using different mathematician methods was

described early in [5, 6, 7]. The process of the minimization of the interval structure is connected with using of the super-gluing logic operation for internal variables. The result of transformation can be acquired with the help of the 2-(n, b)-design system search algorithm in the truth table's structure. The specified algorithm aimed at the automation search for the intervals on the combinatorial systems 2-(n, b)-design with using of the truth table structure and is as a tool for automating the transformations of the logical functions by a method that was described earlier [5, 8, 9, 10].

The algorithm for the recognition of intervals (or combinatorial systems 2-(n, b)-design) and to search for its boundaries is shown on Figure 1 part a. As a part of the general scheme for the transformation of any Boolean function was formed algorithm that shown on Figure 1 part b.

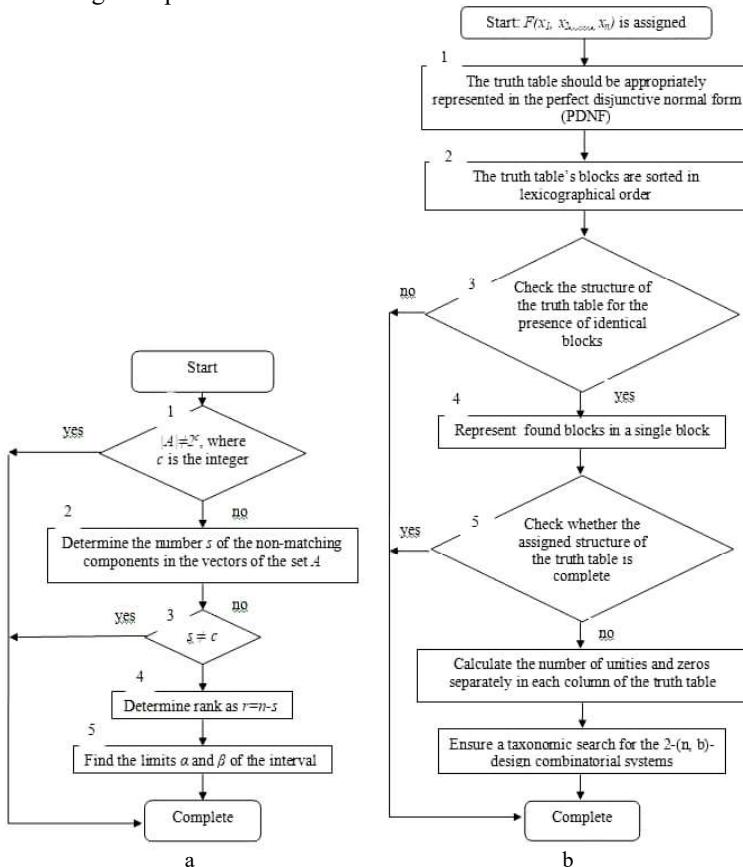


Figure 1 – a. The algorithm for the recognition of intervals (or combinatorial systems 2-(n, b)-design) and searching of it's boundaries; b. The main part of general scheme of Boolean function transformation

**Conclusions.** The algorithm for transformation Boolean functions using the figurative processing is similar to the procedure of searching for the intervals  $I(\alpha, \beta)$  in the Boolean space  $n$ .

The high speed of calculus for any of Boolean functions may be achieved by using of algorithm, based on the primary application of the operation of super-gluing the variables within the truth table.

Modified algorithm aims to extension the set of Boolean function representation, that are not always represented by the truth table, and ensuring of productive calculations for different logical functions view.

#### References:

1. D. A. Pospelov, Logicheskie metody analiza i sinteza shem. Moscow: Energiya, 1974, 368.
2. A. D. Zakrevskiy, Logicheskiy sintez kaskadnyh shem. Moscow: Nauka, 1981, 416.
3. B. E. Rytzar, Metod minimizatsii bulevyh funktsiy. Problemy upravleniya i informatiki, 2, 1997, pp. 100–113.
4. M. Solomko, N. Khomiuk, Y. Ivashchuk, V. Nazaruk, et al, "Implementation of the methods of image transformations for minimizing the Sheffer functions", Eastern-European Journal of Enterprise Technologies. Mathematics and cybernetics – applied aspects. Vol 5/4(107), 2020: 19-34, doi: 10.15587/1729-4061.2020.214899.
5. V. Riznyk, M. Solomko , Y. Sagic, Minimization of conjunctive normal forms of boolean functions by combinatorial method. Technology Audit and Production Reserves, 5 (2 (43)), 2018. pp. 42–55. doi:10.15587/2312-8372.2018.146312
6. J. Huang, Programing implementation of the Quine-McCluskey method for minimization of Boolean expression. arXiv, 2014. URL: <https://arxiv.org/ftp/arxiv/papers/1410/1410.1059.pdf>
7. M. Nosrati, R. Karimi, An Algorithm for Minimizing of Boolean Functions Based on Graph DS. World Applied Programming, 1 (3), 2011, pp. 209–214.
8. K. Y. Kabanian, A. El-Hajj, S. Fakhreddine, W. S. Smari, Computer tool for minimizing logic functions. Computer Applications in Engineering Education, 3 (1), 1995. pp. 55–64. doi: 10.1002/cae.6180030108
9. K. G. Samofalov, A. M. Romlinkevich, V. N. Valuyskiy, V. N., Kanevskiy, et al. Prikladnaya teoriya tsifrovyh avtomatov. Kyiv, 375, 1987. URL: [http://stu.scask.ru/book\\_pta.php?id=62](http://stu.scask.ru/book_pta.php?id=62)
10. Bonal, D. Karnaugh and Mahoney, Map Methods for Minimizing Boolean Expressions, 2013. URL: <http://davidbonal.com/karnaugh-and-mahoney-map-methods-for-minimizing-boolean-expressions/>

<sup>1</sup> **George Vostrov**

Ph.D., Associate Professor, Department of Applied Mathematics and Information Technologies

<sup>2</sup> **Andrii Khrinenko**

Ph.D. student

<sup>1,2</sup> *Odessa National Polytechnic University*

## MATHEMATICAL MODELS OF PSEUDORANDOM PROCESSES BEHAVIOR FOR NONLINEAR DYNAMICAL SYSTEMS

Modeling of random and pseudorandom processes in nonlinear dynamical systems based on randomized algorithms of cyclic trajectories of fixed points requires formation of information technologies that would include a wide range of mathematical methods for predicting occurrences of such trajectories and pattern recognition for their behavior that depends on previous states [1]. Algorithms for analyzing such behavior increase the range of approaches that can be used to effectively solve various problems in areas such as modeling of dynamical systems, functional analysis, function theory, cryptography and others.

To analyze above-formulated problem, this paper considers the processes in maps, which are examples of nonlinear dynamical systems. Nonlinear maps of the following classes are considered: "Tent", "Asymmetric tent", "Sawtooth" and multiplicative order map [2]. The choice of these maps allows us to consider and analyze the chaotic processes that are also observed in complex dynamical systems. Despite the simplicity of the above maps, their iterative cycles have properties that confirm the above statements. According to them, the structure of iterative cycles is determined not only by the properties of the maps themselves, but also by the properties of the numbers on which these maps are based and which have a significant impact on the structure and can significantly change it. The presented nonlinear maps allow to divide the set of primes  $p$  into a system of classes based on the length of the iterative process for given prime numbers. Note that there are an infinite number of prime numbers for which the length of the period is significantly less than the dimension of the number.

### References

1. Crandall R., Pomerance K. Prime numbers: cryptographic and computational aspects. – 2<sup>nd</sup> edition. – Springer, 2005. – 597 p.
2. Hirsch M., Smale S., Devaney R. Differential equations, dynamical systems, and an introduction to chaos. – Elsevier, 2013. – 423 p.
3. Vostrov, G., Opiata, R. Effective computability of dynamic system structure of prime number formation. ELTECS, 2017. – 244 p.

# **NETWORK AND INTERNET TECHNOLOGIES**



<sup>1</sup>**Volodymyr Barannik**

<sup>2</sup> **Yurii Babenko**

<sup>3</sup> **Sergii Shulgin**

<sup>4</sup> **Maksym Parkhomenko**

<sup>1,3,4</sup> *Kharkov National University of Radio Electronics*

<sup>2</sup> *Taras Shevchenko National University of Kyiv*

## **VIDEO ENCODING TO INCREASE VIDEO AVAILABILITY IN TELECOMMUNICATION SYSTEMS**

**Abstract.** Article shows presence of the imbalance caused by insufficient level of productivity of modern and perspective information communication technologies concerning information intensity of bit streams. It is described how imbalance level reducing can be formed as a result of increasing of information processing technologies efficiency and that JPEG platform is the basic concept for construction of technologies of compression representation. Therefore it is proposed to provide further development of video processing methods using individual components of the JPEG platform for improving the integrity of information in terms of ensuring the required level of its availability.:

**Keywords:** video image, integrity of information, areas of coherence, local-position coding.

Nowadays the development of information and communication technologies is accelerating as never before. Brand new technological concepts are being created such as quantum computing, 5G generation telecommunications systems, robotic systems, computer vision, intelligent technology, VR / AR applications. On the other hand, such technological breakthrough in the creation of IT concepts, in turn, provides background for the development of information services, increasing their popularity and quality. As a result, the level of imbalance between the capacity of existing and upcoming IT systems and the level of information load is not reduced. Moreover there is an increase in such an imbalance which level reaches 99%.

Reducing the level of imbalance is achieved by increasing the efficiency of information processing technologies [1 - 5]. At the same time, we have to face a contradiction due to the presence of an inversely proportional relationship between the amount of compressed representation of the video frame and the level of its information integrity [1-5]. Thus, increasing the efficiency of video data compression technologies is an urgent research and applied research issue.

Based on the assessment of the shortcomings of the existing methods of reducing the bit volume, it follows that to reduce the time delays for the delivery of the video it is necessary [5]:

1) to organize a reduction of the bit volume (RBV) of the video with a controlled level of visual assessments of based on the identification and approximation of the description of the elements for the areas of coherence;

2) to provide reduction of the bit volume relative to the code description of the characteristics of the areas of coherence should be carried out not only on the basis of reducing statistical redundancy, but also on the basis of reducing structural redundancy.

Accordingly, the purpose of the research in this article is to create a method of reducing the bit size of video frames based on the detection and processing of their areas of coherence.

Achieving the required level of availability of information in terms of maintaining indicators of its integrity is proposed to organize on the basis of ensuring the reliability of the structural components of the relevant objects of interest. To do this, it is proposed to identify the segments of the video frame by the degree of their importance from the standpoint of maintaining the required level of integrity of objects of interest. Identifying segments by component brightness is proposed. This can be explained by the fact that the brightness component carries the main information load among other color components in the color difference model of the video frame. It is proposed to carry out identification for localized areas of the video frame, which are characterized by greater homogeneity of their structural and statistical properties. Therefore, from a methodological point of view, the assessment of the information contribution of the segment should be based on the results of structural and statistical processing of its four microsegments,  $S(X)_{i,j}^{(u)}$ ,  $u = \overline{1, 4}$  from the point of preserving the semantic integrity of the video resource. Depending on the informational significance, microsegments  $S(X)_{i,j}^{(u)}$  are classified into three types, namely: microsegments with a high level of structural-statistical saturation (SSN) at the syntactic level of the description; microsegments with an average level of SSN; microsegments with low levels of SSN.

#### References:

1. Ian Richardson, H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia / Ian Richardson, pp. 368, 2005.
2. Pratt W. K., Chen W. H., Welch L. R. Slant transform image coding. Proc. Computer Processing in communications. New York: Polytechnic Press, 1969. P. 63-84.
3. Sindeev M., Konushin A., Rother C. Alpha-flow for video matting. Technical Report. 2012. P. 41–46.
4. Barannik, V.V., Ryabukha, Yu.N. and Kulitsa, O.S.: The method for improving security of the remote video information resource on the basis of intellectual processing of video frames in the telecommunication systems. In: Telecommunications and Radio Engineering. Vol. 76. No 9. pp. 785-797. (2017). doi: 10.1615/TelecomRadEng.v76.i9.40.
5. Barannik, V. and Barannik, N. and Ryabukha, Yu. and Barannik, D.: Indirect Steganographic Embedding Method Based On Modifications of The Basis of the Polyadic System. In: 15th IEEE International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET'2020), pp. 699-702 (2020) DOI: 10.1109/TCSET49122.2020.235522.

## Roman Belfer

Postgraduate Student at the Computer Engineering & System Programming Department

*Khmelnitsky National University*

## THE ARCHITECTURE OF THE LAYERED PEER TO PEER NETWORK

To design a blockchain network according to the Proof-of-Activity [1] consensus protocol principles, some clarifications regarding the architecture should be added. The classic blockchain network, for example Bitcoin which uses the Proof-of-Work consensus protocol, is based on the Peer-to-Peer network architecture. In that case, each node is an equal unit in the global system workflow. But, to design the network to handle the PoA protocol, the node's type or parametrization is required. It's necessary to follow the stages of the defined consensus algorithm to select the validator nodes.

Designing a peer to peer blockchain network, which is going to follow the Proof of Activity consensus protocol, requires additional details which is why multiple layers are suggested for handling the different types of parameters of the network. Each layer (or level) contains a subset of the nodes that are grouped by some specific condition. Nevertheless, the network still keeps the properties of the classic P2P network of nodes being equal. We refer to this enhanced clarified architecture by the name Layered Peer-to-Peer (LP2P).

According to the quantity and the quality of its activity in the network, a node may be transferred through the layers, promoting itself and increasing the possibility of being selected as the validator. Note that the location of a node in the network is not static and may change.

If the condition:

$$\text{activityIndex}(\text{node}) \geq \text{index}_1, \quad (1)$$

for  $\text{node}_0K$  is true, the transformation happens:

$$\text{node}_0K \rightarrow \text{node}_1L+1, \quad (2)$$

If the node located on layer  $N$  gets the level of activity, that's at least equal to activity index  $N+1$ , this node moves to the next layer and becomes a layer  $N+1$  node. Nodes will be grouped in layers based on their activity and will be able to participate in selecting nodes for validation roles.



Figure 1 – Multilayered Peer-to-Peer network and layered nodes

The LP2P network architecture becomes the structural pattern for the network designing to follow the Proof-of-Activity protocol and operate as blockchain. The multi layered structure allows to separate responsibilities of the nodes but not to limit the possibilities and opportunities for the separate ones. The following validation algorithm could be applied for different nodes, which belong to specific network layers, with different predefined conditions and parameters. That allows to scale the number of layers depending on input requirements for different number of predicted behaviors.

#### References:

1. Belfer R. Proof-of-Activity Consensus Protocol Based on a Network's Active Nodes / R. Belfer, A. Kashtalian, A. Nicheporuk, A. Sachenko, G. Markovsky // Proceedings of the 1st International Workshop on Intelligent Information Technologies & Systems of Information Security. - Khmelnytskyi, Ukraine, June 10-12, 2020. - Pp.239-251.
2. Bandara, H. M. N. D; A. P. Jayasumana, "Collaborative Applications over Peer-to-Peer Systems – Challenges and Solutions". Peer-to-Peer Networking and Applications., 2012, pp. 257–276 doi:10.1007/s12083-012-0157-3
3. Steinmetz, Ralf; Wehrle, Klaus, "2. What Is This "Peer-to-Peer" About?". Peer-to-Peer Systems and Applications. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 9–16 doi:10.1007/11530657\_2

<sup>1</sup> Serhii Buchyk

Doctor of Science, Professor

<sup>2</sup> Danil Palageychenko

Student

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## INFORMATION TECHNOLOGIES IN UKRAINIAN JUDICIAL SYSTEM

**Abstract.** In centure of repid introduction development of the latest technologies in all public spheres, the evaluation of its effectiveness is relevant. Accordingly, it becomes necessary to carry out research on implementation issues, which can only be carried out taking into account the practical achievements of Ukraine and other leading countries in this field. Electronic document has the right to exist, taking into account the implementation of the advances in technologies, and having in mind the need for a thorough assessment of its impact on the judicial system.

**Keywords:** Electronic document, management, network, information technologies, Ukrainian judicial system.

The development of information technology is an extremely important component of modern society. It is now difficult to find an area in which electronic media is not used (in 2014 the number of Internet users in Ukraine was 280 thousand, and today it is about 22 million [1]).

Innovations are actively introduced into the justice systems of a number of advanced European countries, as the legislation must meet the modern requirements of society. Ukraine, in the context of integration processes, is also trying to introduce electronic document management to ensure the efficiency of the judiciary and the quality of justice in general.

Electronic document is the creation of an appropriate information base on the media for the purpose of realization of conditions for using this system of the appropriate management in the process of performing its functions [2].

Thus, it is possible to distinguish its basic principles of functioning: one time registration of the document; when working with the document is guaranteed its continuity of movement; the document base should be the only one - it makes duplication of documents impossible; search engine is well organized.

The following requirements must be observed when storing electronic documents: the information contained in electronic documents must be available for further use; it must be possible to restore the electronic document in the format in which it was created, sent, received; if available, information should be stored that can establish the origin and purpose of the electronic document, as well as the date and time of its sending and receipt [3].

Every year, the realization of services that can be done through the Internet or using certain online resources is gaining in popularity. Often, you need to create your own electronic signature in order to identify a user or enter personal information. A paper document in its usual traditional form is information certified by a signature and in some cases a seal. However, the historical path of development of the paper document was so rapid that now there is its original new electronic form, which has

become a relevant element in business culture. However, in order to authenticate the document, ie to ensure the veracity of the information contained in it, it is necessary to take certain actions. It is through this that a natural or legal person seeks to verify the validity or authenticity of specific information. Thus, we can distinguish two main functions of electronic signatures in legal practice: (1) the connection between the document and the party who created it is checked; (2) Ensuring that the content of the document has been revised and that it meets certain standards.

Keeping a wealth of information, including the power of personal data, does not exclude the possibility of cyberattacks for the purpose of unauthorized access to it. To prevent this, proper development of the necessary protection of the automated system, as well as of all registers and databases that will be widely used, should be carried out by qualified personnel.

In this regard, the following principles of electronic signature in practice arise: 1) Identity: a person must use the method to identify themselves and indicate their intention. 2) Reliability: the method of identification should be as reliable as appropriate, taking into account the purpose of the message. 3) Consent: the person to whom the signature is given must consent to the use of electronic communication to comply with the signature requirement and the method of identification.

Problems with the transition of introduction of electronic document can arise from various reasons: financial, inadequate preparation to work with the system of electronic document flow of staff, lack of experience, etc. Quarantine has introduced some difficulties in the process of notifying persons and has significantly increased the time of consideration of the case, as quite often people do not have the opportunity to receive a mail (by postal). Therefore, taking into account the recommendations provided by the Council of Judges of Ukraine, including the Supreme Court, Ukrainian citizens started to use an electronic correspondence. Sufficient practice has not yet been established on this issue: there are not many cases, and the issue of admissibility or inadmissibility of electronic evidence has not been considered by the court of cassation. Technically, any citizen registered in the Electronic Court can send a document to the court, and this document will be registered - a more difficult issue is the ambiguous position of the courts on the possibility of using this system and obtaining documents electronically.

Thus, today the preconditions are gradually being created for the courts to work with electronic documents, but the decision on whether to accept them or not is still up to them.

#### **References:**

1. Article "E-Justice is a big step forward" from - 19.05.2018 // [Electronic resource] // - Access mode: <https://yaizakon.com.ua/elektronnoe-pravosudie-eto-bolshoj-shag-vpered-sudya-aleksandr-sasevich/>.
2. Regulation on the automated system of court circulation (decision of the Council of Judges of Ukraine No. 30 of November 26, 2010) in the version dated 02.03.2018 // [Official web-portal of the VRU] // - Access: <https://zakon.rada.gov.ua/rada/show/vr030414-10>
3. Вимоги щодо зберігання електронних документів // [Офіційний веб-портал Оболонської районної в місті Києві Державної адміністрації] // - Access mode: <https://obolon.kyivcity.gov.ua/content/vymogy-shchodo-zberigannda-elektronnyh-dokumentiv.html>
4. Стаття «Як отримати електронний підпис у 2019 році», // – Access mode: <https://ursambo.com/ukr/zahyst-yurosib-i-fop/iak-otrymaty-etsp-elektronnyj-tsyfrovyy-pidpys.html>

<sup>1</sup> **A. Cherevativ**

Student

<sup>2</sup> **S. Paliy**

PhD, Associate Professor of the Information Systems and Technologies Department

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## THE USE OF ARTIFICIAL INTELLIGENCE IN THE INTERNET OF THINGS SYSTEM

Internet of Things, IoT - is a global infrastructure of information area, which provides advanced services by interconnecting physical real-world objects with centers of control, management, and information processing. Communication is carried out using existing information technologies for the intellectual recognition, administration, tracking, monitoring and positioning.[1]

IoT allows us to look at everyday life from the other side, but only AI allows us to fully unleash the entire huge potential of the IoT system.

According to the Business Insider Intelligence, there will be approximately 55 billion IoT devices by 2025, compared to 9 billion in 2017. It is more than 4 devices per person.

Gartner forecasts that in the nearest future an AI feature will be used in more than 80% of enterprise IoT ventures, up from a mere 10% today.

Companies who can combine applying IoT data and AI in real-time could gain a major business edge, including a 38% boost in profitability by 2035, according to Forbes.

Usually a human has five senses: vision, sound, taste, scent, and touch.

The main problem is that a person has a limited number of sense organs, which is also limited by the distance to the object, which gives us a certain sensation. In other words, we cannot feel what is outside the limits of our perception.

However, an AIoT system can likely have an infinite number of sensors that capture even more complicated data over greater distances. [2]

It is clear that as the number of these devices increases, the amount of data that needs to be processed will also increase like an avalanche. This is where artificial intelligence reveals its potential, providing its learning capabilities for devices connected to the IoT.

These interdependent technologies change the way we interact with our devices all around, thus generating Artificial Intelligence of Things.

There are main sections where we can see the influence of AI on IoT:

- Smart Industry
- Smart Home
- Smart City
- Wearable

*Smart Industry* is a branch of industry that rely on AIoT technologies from supply-chain sensors to real-time data processing, smart technologies help stop risky market failures and minimize an effect of human's mistakes.

*Smart Home* is a type of house that is constantly connected with the user, his requests, wishes. Smart homes can use control over various household items to optimize consumption, safety, taking into account the user's characteristics, his wishes

*Smart City* is a system of interconnected IoT things that makes the life of citizens as comfortable as possible, while improving the efficiency of using natural resources. In smart cities, there is also the possibility of using IoT with AI to improve the quality of life in various aspects by analyzing data that comes in every day.

*Wearables* are devices that make everyday life easier and allow you to keep a constant record of the state of the body, using AI, they also have the ability to adjust to the human biorhythm to optimize effective work[3]

In general, artificial intelligence training algorithms can be divided into three categories, which are based on the level of human intervention in the process: training under full supervision, training under semi-supervision, training without supervision.

*Supervised learning* involves a fixed closed set of labels, and only certain selected data are available for processing. This type is only possible for small-scale applications, but cannot be used when working with a scale that covers a wide range of behaviors and contexts demonstrated by a large community of users.

*Semi-supervised learning* describes an approach in which only part of the data is labeled, which reduces the need for labeled examples, in which case algorithms study the mechanism of recognizing similarities or differences, identifying certain anomalies in the processing of a data set.

*Unsupervised learning* does not require labeling of data, which provides a much wider range of applications, but a more demanding learning process. Uncontrolled learning methods can be applied to a variable context based on behavioral similarities, but they can potentially lead to behaviors that do not correspond to activities that can be identified as useful for data processing. [4]

Comparing Iot and AIoT we can say that:

*IoT*: requires hardware that would be connected, while often such a system does not analyze the data collected during the operation of the system, such a system requires minimal human intervention.

*AIoT*: in addition to everything that a conventional IoT system requires, it also requires the development and implementation of a software product in order to obtain the highest quality data processing, with the aim of further training the system, and the next stage is the maximum reduction in the need to involve a person as a link that controls the learning process of the system [5]

*Conclusions*: Combining the use of IoT and artificial intelligence significantly expands the functionality of the system by analyzing and processing the collected data. This is a great solution to improve system performance and optimize many processes that would be much more difficult to implement without using AI. This is a breakthrough decision that will change the lives of many people in the nearest future.

#### **References:**

1. Carlos Salazar, Keyur Patel, Sunil Patel, P. Scholar. Internet of Things-IOT: Definition, Characteristics, Architecture, Enabling Technologies, Application & Future Challenges. 2016. [Електронний ресурс] Режим доступу: [https://www.researchgate.net/publication/330425585\\_Internet\\_of\\_Things-IOT\\_Definition\\_Characteristics\\_Architecture\\_Enabling\\_Technologies\\_Application\\_Future\\_Challenges](https://www.researchgate.net/publication/330425585_Internet_of_Things-IOT_Definition_Characteristics_Architecture_Enabling_Technologies_Application_Future_Challenges)
2. 2018.[Електронний ресурс] Режим доступу: [https://www.sas.com/ru\\_ru/explore/resources/aiot.html](https://www.sas.com/ru_ru/explore/resources/aiot.html)
3. Iman Ghosh. AIoT: When Artificial Intelligence Meets the Internet of Things. 2020. [Електронний ресурс] Режим доступу: <https://www.visualcapitalist.com/aiot-when-ai-meets-iot-technology/>
4. Artur Miguel Arsénio, Hugo Serra, Rui Francisco, Fernando Nabais Internet of Intelligent Things: Bringing Artificial Intelligence into Things and Communication Networks. 2014. [Електронний ресурс] Режим доступу: [https://www.researchgate.net/publication/265248986\\_Internet\\_of\\_Intelligent\\_Things\\_Bringing\\_Artificial\\_Intelligence\\_into\\_Things\\_and\\_Communication\\_Networks](https://www.researchgate.net/publication/265248986_Internet_of_Intelligent_Things_Bringing_Artificial_Intelligence_into_Things_and_Communication_Networks)
5. Dinesh Soundararajan. AIoT: The powerful convergence of AI and the IoT. 2020. [Електронний ресурс] Режим доступу: <https://www.iot-now.com/2020/04/10/102236-aiot-the-powerful-convergence-of-ai-and-the-iot/>
6. Copyright © 2020. All rights reserved

**<sup>1</sup> Andriy Dudnik**

Doctor of Technical Sciences, Associate Professor of the Department of Network and Internet Technologies

**<sup>2</sup> Mykyta Kobylchuk**

Student of the Department of Network and Internet Technologies

**<sup>3</sup> Daria Pokutnia**

Student of the Department of Network and Internet Technologies

<sup>1,2,3</sup> Taras Shevchenko National University of Kyiv

## **ANALYSIS OF THE CURRENT STATE OF TECHNOLOGY "SMART HOME"**

**Abstract.** The main task of this article is to analyze the current state of the "Smart Home" system. This paper analyzes the prototype of the system "Smart Home" with the mandatory content of the review of the concept of automation systems, the specifics and main aspects of the implementation of such systems. Theoretical and methodological bases of house automation systems are studied.

**Keywords:** smart home, Internet of Things, system parameters, methodology.

Today, there are many different systems and protocols designed to provide full automation of houses, houses or apartments. There are more than twenty technologies aimed at creating a so-called "smart home". These include some categories and subcategories, such as video surveillance systems, for ventilation, heating, air conditioning, entertainment, centralized and distributed systems, and so on. The object of research was the so-called "intellectual house". Hi-tech in all its manifestations has firmly taken its place in our lives, and one of its manifestations in everyday life is the smart home system.

One of the main advantages of intelligent buildings is the comfort they provide to their residents. Control of lighting and infrastructure of the house allows you to create different lighting options for the room, any combination, depending on the time of day and mood, at the touch of a button. The climate control system allows you to simultaneously reproduce the conditions of different climatic zones in different rooms. To do this, simply set the desired temperature on the touch control panel.

Another undeniable advantage of a smart home is its security system. Moreover, automation systems are designed to provide protection against any emergency. First, they provide protection against intrusion by CCTV cameras, automation of doors, gates, the role of blinds, burglar alarms. Secondly, there is almost no chance of a fire - the iron, tongs or oven, which are left on, will be turned off in time, and in case of fire or smoke, the fire alarm will go out. The system controls the flow of water, electricity and heat. This is achieved through the most rational use.

The relevance is that the modern market of IT technologies offers interested customers a wide range of so-called "smart" systems. Such systems are designed to automate the home use of a number of tasks without human intervention. As well as providing information to the user about the state of the system and the environment, such as temperature and humidity, the state of the devices.

The challenge was to find and develop effective systems for the comfort and safety of the owners. In the process of work the goal was set - to develop and determine the optimal Smart Home system, to calculate the main methods of studying the parameters of the Smart Home system (Fig. 1).

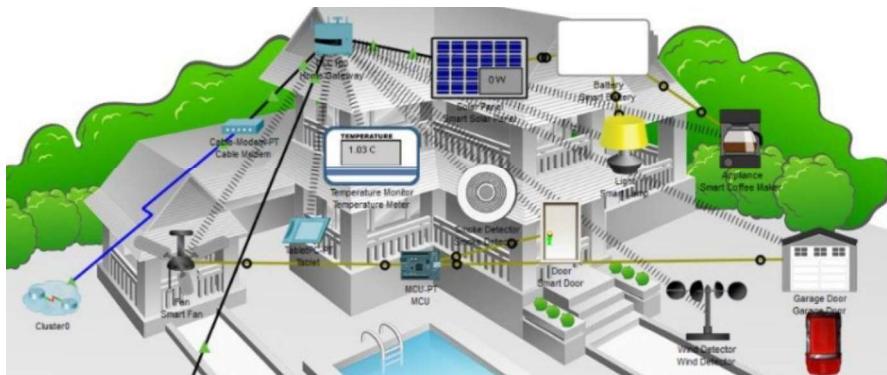


Figure 1 – Smart Home system

**Conclusion.** Also, today, hundreds or even thousands of different process automation systems have been developed in the world. But they are designed with Western monitoring standards in mind, so they are of little use to middle- or low-income countries without adaptation. Automation also constantly needs some support with the constant involvement of a specialist or developer, as it is quite a complex and lengthy process.

Another problem is that many of our potential users are not sufficiently informed about the purpose of automated systems. Most perceive this not as an opportunity to save resources and, consequently, financial resources, but as an indicator of prestige.

For this reason, manufacturers for many countries, such as Ukraine, do not seek to develop too much the technological level of production of systems, on the contrary, greatly simplifying the technical characteristics, not so much energy efficient as aesthetically attractive.

#### References:

1. Adriano D. B. et al. Iot-based Integrated Home Security and Monitoring System //Journal of Physics: Conference Series. – IOP Publishing, 2018. – T. 1140. – №. 1. – C. 012006.
2. Amri Y., Setiawan M. A. Improving Smart Home Concept with the Internet of Things Concept Using RaspberryPi and NodeMCU //IOP Conference 64Series: Materials Science and Engineering. – IOP Publishing, 2018. – T. 325. – №. 1. – C. 012021.
3. Ashton K. et al. That ‘internet of things’ thing //RFID journal. – 2009. – T. 22. – №. 7. – С. 97-114.
4. Hardeniya N. et al. Natural Language Processing: Python and NLTK. – Packt Publishing Ltd, 2016.
5. Kenter T., Borisov A., De Rijke M. Siamese cbow: Optimizing word embeddings for sentence representations //arXiv preprint arXiv:1606.04640. – 2016.

6. Kodali R. K., Yerroju S. Energy Efficient Home Automation Using IoT // 2018 International Conference on Communication, Computing and Internet of Things (IC3IoT). – IEEE, 2018. – C. 151-154.
7. Korobov M. Morphological analyzer and generator for Russian and Ukrainian languages //International Conference on Analysis of Images, Social Networks and Texts. – Springer, Cham, 2015. – C. 320-332.
8. Le Q., Mikolov T. Distributed representations of sentences and documents //International conference on machine learning. – 2014. – C. 1188-1196.
9. Lobaccaro G., Carlucci S., Löfström E. A review of systems and technologies for smart homes and smart grids //Energies. – 2016. – T. 9. – №. 5. – C. 348.
10. Meng Y. et al. Securing Consumer IoT in the Smart Home: Architecture, Challenges, and Countermeasures //IEEE Wireless Communications. – 2018. – T. 25. – №. 6. – C. 53-59.
11. Mobile Operating System Market Share Worldwide. – URL: <http://gs.statcounter.com/os-market-share/mobile/worldwide>.
12. Nath R. K., Bajpai R., Thapliyal H. IoT based indoor location detection system for smart home environment //2018 IEEE International Conference on Consumer Electronics (ICCE). – IEEE, 2018. – C. 1-3. 6566
13. Patel V., Shah A. Android Based Smart Home Controlled System Using Raspberry Pi //Journal of Control & Instrumentation. – 2018. – T. 9. – №. 2. – C. 5-11.
14. Perevalov A. et al. Question Embeddings Based on Shannon Entropy Solving intent classification task in goal-oriented dialogue system //Proceedings of the 7th International Conference on Applied Innovations in IT. – 2019.
15. Reina D. G. et al. The role of ad hoc networks in the internet of things: A case scenario for smart environments //Internet of things and inter-cooperative computational technologies for collective intelligence. – Springer, Berlin, Heidelberg, 2013. – C. 89-113.

**<sup>1</sup> Myroslava Gladka**

Teaching Assistant

**<sup>2</sup> Rostyslav Lisnevskyi**

Ph.D., Associate Professor, Docent

**<sup>3</sup> Mykola Kostikov**

Ph.D., Associate Professor, Docent

<sup>1,2,3</sup> Taras Shevchenko National University of Kyiv

<sup>3</sup> National University of Food Technologies

## USING THE INTERNET OF THINGS WHEN INTRODUCING CRM SYSTEMS IN THE BANKING SECTOR

**Abstract.** The Internet of Things (IoT) concept becomes more popular, in particular in the banking sector. Each customer needs special attention and an individual approach to service. Wireless communication between banking objects and the customer's mobile phone reduces the number of customer's actions and improves the protection of operations. Individual algorithms for interaction between customers and banking objects that work within modern customer relationship management (CRM) systems increase the loyalty of customers.

**Keywords:** Bank, customer, online banking, customer relationship management, Internet of Things, geolocation

### 1. Introduction

As the banking sector develops, the customers' requirements for its services increase. To attract and retain customers, banks must not only widen the range of services provided but also select which services should be offered to which customers. Introducing Customer relationship management (CRM) systems helps banks to analyze their interaction with each customer individually and to deliver efficient management based on this analysis [1].

One of the main areas of work with customers is using online banking that allows customers to work with their accounts from any point of the globe, given the availability of Internet connection [2]. Virtually all online banking customers use not only their PCs but also smartphones.

Internet of things (IoT) is a technology for wireless connection between devices via the Internet that allows them to exchange information and interact with minimal human intervention [3]. The development and use of IoT in the financial and banking sector for work with customers will allow increasing the quality and speed of service [4, 5].

## 2. Using the IoT for CRM Systems in Banking

### 2.1. Devices Interaction when Performing Banking Operations

All customer bank accounts are linked to a mobile phone number, which provides binding the IoT smartphone to the person who is the account user. When connecting to control devices, it is the smartphone that is identified by the personal number and, if necessary, by the password [6].

When a bank customer uses services of ATM, POS terminals, PC, etc., the authentication is possible through using the customer's smartphone geolocation [7]. The customer who uses an additional authentication for banking operations may use his smartphone that will function as a part of IoT and perform this action instead of the customer himself (figure 1).

To perform such actions, it is necessary to meet the next requirements: the customer permits to perform authentication by using a smartphone; the smartphone must have geolocation enabled [8]; the device and the object performing the operations must be within the strictly defined distance [9].

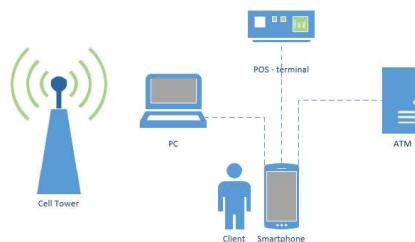


Figure 1 – IoT devices interaction for performing online banking operations

### 2.2. Determining the Distance for Interaction

We have all the needed conditions to determine the distance between objects: ATMs are installed at stationary positions that clearly define their geolocation; POS terminals have the functionality to determine their location; PCs and other devices are also able to show their location. Thus everything we need is to calculate and indicate the maximum allowable deviation distance between the action object and the customer's smartphone. Let the distance between objects be  $D$ . Then the allowable distance will be determined by the formula 1:

$$D \leq |(M \pm \Delta\beta) + (O \pm \Delta\gamma)|, \quad (1)$$

where  $M$  is the customer's smartphone location;

$\beta$  is the accuracy of determining the smartphone location;

$O$  is the location of the object performing banking operations;

$\gamma$  is the accuracy of determining the location of the object performing banking operations.

It is worth noting that for different objects of interaction, the distance D may be defined individually, as each object has its features for work. Thus it is necessary to define the maximum allowable distance between objects when writing interaction algorithms in a CRM system. We should also consider the possible location error. Each scenario must take into account the features of work with each object (figure 2).

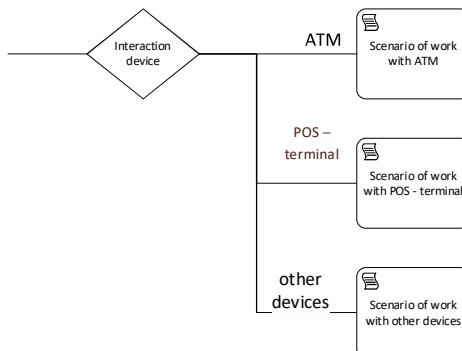


Figure 2 – Choosing a scenario for interaction between customers and objects in a CRM system

Due to using these functions at the device connection layer, a customer doesn't need to go through additional verifications for actions and operations.

### 2.3. Conclusions

Using modern IoT for work and customer support in banking helps to reduce the time for operations of two-factor authentication while providing an additional protection layer. However, there are some disadvantages. Not all bank customers are using modern smartphones, which is necessary to determine the customer's location accurately. Besides, the smartphone's geolocation must be on at all times for providing the functions needed.

But a smartphone with enabled geolocation can provide an additional access layer when performing operations. At this level, devices will exchange information without further actions by the customer. Additional actions for authorization are required from the customer only in case of lacking the listed conditions.

Introducing IoT in the banking sector of Ukraine will provide [10]:

- reducing human impact on operations;
- increasing protection of banking operations due to additional verification and authentication that are performed at the device level;

- optimizing and minimizing the message traffic of mobile network operators;
- gaining and maintaining leadership in the banking sector through the introduction of global trends in digitalization.

### **3. Acknowledgments**

The paper has been prepared based on data provided by the company Terrasoft which introduces and develops CRM systems in the banking and financial sector.

#### **References:**

1. M. Bardicchia, Digital CRM: Strategies and Emerging Trends: Building Customer Relationship in the Digital Era, Independently published, Milan, 2020.
2. O. P. Rishi, A. Sharma, Maximizing Business Performance and Efficiency Through Intelligent Systems, IGI Global, Hershey, PA, 2017.
3. M. Milenkovic, Internet of Things: Concepts and System Design, Springer, New York City, NY, 2020.
4. L. Lazepko, Interview: OTP Bank Digital Transformation of Corporate Banking, 2020. URL: <https://www.creatio.com/node/18139>.
5. IoT-Enabled Banking Services: White Paper, Infosys, Bengaluru, India, 2018. URL: [infosys.com/industries/financial-services/white-papers/documents/iot-enabled-banking.pdf](http://infosys.com/industries/financial-services/white-papers/documents/iot-enabled-banking.pdf).
6. Е. Петрова. Проблемы IoT в банках, 2017. URL: <https://bis-expert.ru/articles/56278>.
7. J. Shin, S. Kancharlapalli, M. Farcasin, E. Chan-Tin, SmartPass: a smarter geolocation-based authentication scheme, in: Security and Communication Networks 2015, Wiley, Hoboken, NJ, 2015, vol. 8, pp. 3927–3938, doi: 10.1002/sec.1311.
8. D. Nield, How Location Tracking Works on Your Phone in 2020, Gizmodo, 2020. URL: <https://gizmodo.com/how-to-turn-off-location-tracking-on-your-phone-1841695041>.
9. P. A. Shary, Models of Topography, in: Q. Zhou, B. Lees, G. Tang (Eds.), Advances in Digital Terrain Analysis, Springer, Berlin, Heidelberg, 2008, pp. 29–57. doi: 10.1007/978-3-540-77800-4\_3.
10. В. И. Гербич, Л. А. Нетребчук, Интернет вещей в банковской сфере, in: Актуальные проблемы маркетинга XXI века: матер. Междунар. науч.-практ. конф., Минск, 21–22 марта 2019, Ин-т бизнеса БГУ, Минск, 2019, с. 102–104.

<sup>1</sup> **Hrygorii Hnatienko**

PhD in Physical and Mathematical Sciences, Associate Professor

<sup>2</sup> **Volodymyr Kudin**

Doctor of Technical Sciences

<sup>3</sup> **Oleh Ilarionov**

PhD in Engineering Science, Associate Professor

<sup>4</sup> **Oksana Vlasenko**

Specialist

<sup>1,2,3,4</sup> Taras Shevchenko National University of Kyiv

## FUZZY DEFINITION OF RELATIVE ESTIMATES OF ALTERNATIVES BASED ON PAIRWISE COMPARISONS USING PSEUDOBASIC MATRICES

A significant number of papers in domestic and foreign sources are devoted to the study and analysis of methods for solving the problem of determining the structure of preferences on set of objects. It is generally accepted that it is difficult to obtain direct consistent information from an expert about the numerical values of the relative estimates of objects.

We will consider one of the most common ways to represent the values of relative estimates in the form of real numbers, taking into account the condition of normalization:  $\sum_{i \in I} \rho_i = 1, \rho_i > 0, i \in I$ .

But in many real-life cases, the values of the weights are fuzzy. In particular, fuzzy constraints and fuzzy criteria are used in multicriteria problems in different subject areas [1]. System of expert preferences in most cases is an extremely important heuristic when calculating an adequate solution of multicriteria optimization. Therefore, a promising area of research is to determine relative estimates in the form of a function of belonging to a fuzzy set.

It is known that an expert can not accurately determine the weights of alternatives, so for such cases indirect methods are used. In particular, one of the methods of solving the problem of determining the weight of objects by an incomplete metricized multiplicative matrix of pairwise comparisons between the importance of alternatives is described and investigated in [2, 3]:

$$M = (\mu_{ij}), i, j \in I. \quad (1)$$

At the initial stage of this method, a rectangular matrix of size  $(n \times N)$ ,  $N = n(n-1)/2$ ,

$$P = (\pi_{ij}), i \in H = \{1, \dots, N\}, j \in I, \quad (2)$$

is constructed from a matrix of the form (1).

Its elements are determined as follows:

$$\pi_{ij} = \begin{cases} 1, & \text{if } i = (2-n) + \sum_{l=1}^s (n-l), \quad s = 1, \dots, n-1, \\ \mu_{ij}, & \text{if } j = s+i \quad \text{for } i = 1, \dots, n-1, \\ j = s+i - \sum_{l=1}^s (n-l) \quad \text{for } i \geq l \quad \text{or } s \geq 2, \\ 0 & \text{in all other cases.} \end{cases}$$

At the next stage of the method, a brute force is organized, as a result of which all possible combinations of the  $(n-1)$ -th line are selected from the matrix (2) and are supplemented by a line of  $n$ , length consisting of single elements. The matrix constructed in this way is denoted by  $A^{(l)}, l \in \Lambda$ ; the resulting system of linear algebraic equations is as follows:

$$\begin{aligned} A^{(l)} \rho &= e, \\ \rho_i &> 0, \quad i \in I, \end{aligned} \tag{3}$$

where  $\Lambda$  – set of indices of form (3) systems, for which the matrix  $A^{(l)}, l \in \Lambda$ , is nondegenerate,  $e$  – vector of  $n$  length with elements  $(0, \dots, 0, 1)^T$ ,  $T$  – transposition sign. The components of the unknown variables (3) vector of are positive.

It is obvious that matrices  $A^{(l)}, l \in \Lambda$ , in systems of the form (3) are very sparse, so special methods can be applied to them, which allow to simplify the calculations when solving the problem.

Simplification can be achieved by:

- checking the conditions of nondegeneracy of the constraint matrix when replacing one line in the system (3);
- the positive condition of the components of the new solution when replacing one line of constraints (3).

It is easy to see that the number of compatible systems of the form (3), according to Kelly's theorem on the number of undirected trees in the graph [4], is  $n^{n-2}$ . With a large number of objects, a large number of systems of equations are incompatible. In particular, for 7 objects the number of compatible systems is 31% of the total number of systems of type (3), for 8 – 22%, for 9 – 16%, for 10 – 11%.

Note that the sequence of calculation of systems of the form (3) can be "organized" in such a way that each successive matrix of the form  $A^{(l)}, l \in \Lambda$ , differs from the previous one by only one line.

In the case of a significant increase in the speed of calculations the method of pseudobase matrices [5] can be successfully used for solving systems of the form (3). This method consists "natural" checking of the conditions of nondegeneracy of the matrix of constraints type (3) when changing one line, and the connection of the components of the solution vector and the elements of the inverse matrix as a result of such changes. That is,

before the basic calculations, to identify "not promising" in terms of fulfilling the conditions of matrix nondegeneracy constraint, the positiveness of the solution components, and compatibility.

Since the described method of determining the weights in the form of functions of belonging to a fuzzy set is computationally complex, it is proposed to apply the method of pseudobasic matrices, which significantly speeds up the calculation.

Here are the results of a computational experiment for six alternatives.

Given a matrix of pairwise comparisons of the form (1):

1	0,333	8	3	3	7
3	1	9	3	3	9
0,125	0,111	1	0,167	0,2	2
0,333	0,333	6	1	0,333	6
0,333	0,333	5	3	1	6
0,143	0,111	0,5	0,167	0,167	1

A rectangular matrix of the form (2) is constructed:

1	-0,333	0	0	0	0
1	0	-8	0	0	0
1	0	0	-3	0	0
1	0	0	0	-3	0
1	0	0	0	0	-7
0	1	-9	0	0	0
0	1	0	-3	0	0
0	1	0	0	-3	0
0	1	0	0	0	-9
0	0	1	-0,1667	0	0
0	0	1	0	-0,2	0
0	0	1	0	0	-2
0	0	0	1	-0,3333	0
0	0	0	1	0	-6
0	0	0	0	1	-6

The number of combinations in this problem is equal to 3003. And the number of possible compatible systems in this case is 1296.

As a result of the computational experiment, we obtain the frequency of the values of the weights of the alternatives, which is shown in the figures. When aggregating the obtained solutions, we use the method described in the monograph [3] for constructing the function of belonging to a fuzzy set based on the analysis of the frequency of values.

Horizontally, the graphs show the values that each of the coefficients of importance of alternatives acquires, vertically - the frequency of occurrence of these values in the respective ranges.

Several options for interpreting the results can be offered.

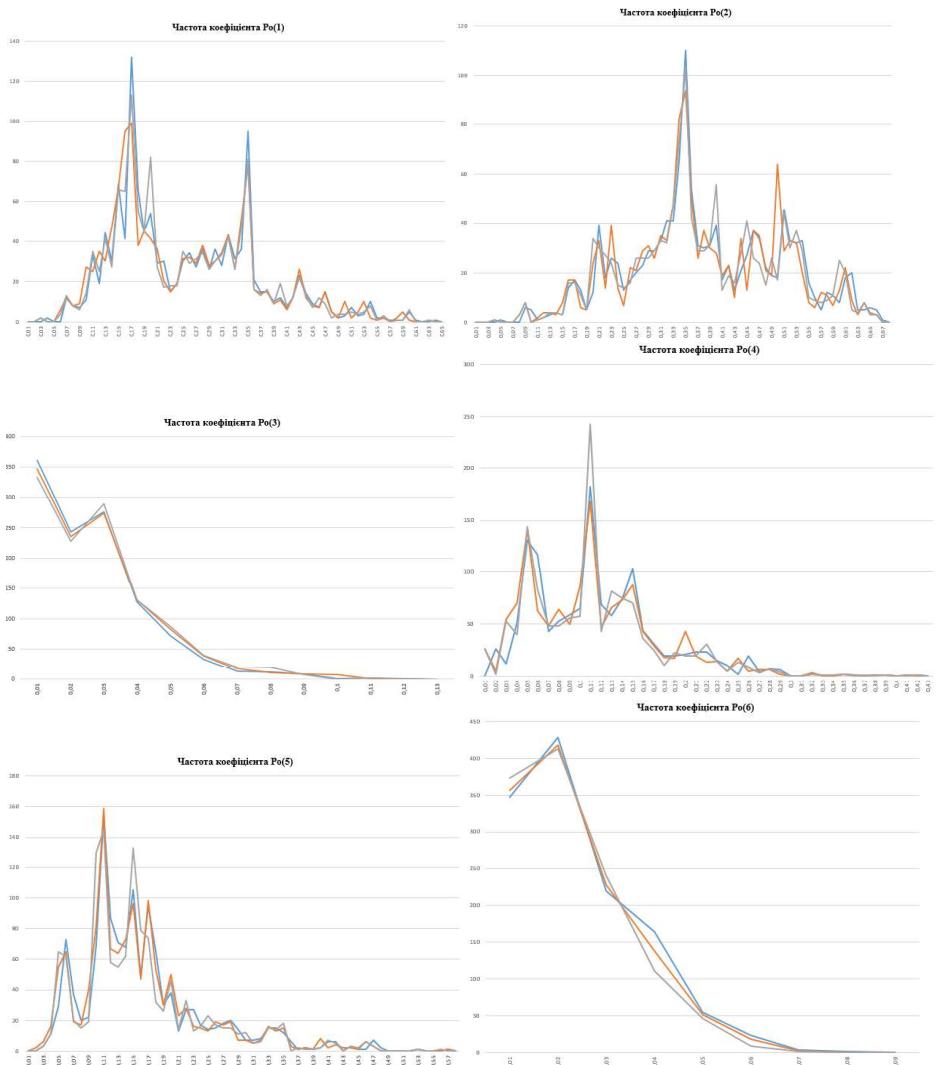
1. Define clusters on the hyperplane  $\sum_{i \in I} \rho_i = 1$ . The most powerful of the clusters

will contain solutions of the initial problem and it can be used to construct membership functions of the coefficients of importance of alternatives.

2. Determine the clusters on the set of points for each of the n dimensions found

by solving equations of the form (3), and build membership functions on the most powerful clusters on each axis.

3. From the set of obtained frequencies presented in the graphs, select the membership functions of a given type (triangular, trapezoidal, Z-shaped, S-shaped, Gaussian, etc.) and automatically generate logical expressions.



Thus, this paper proposes an approach that allows to find the whole set of possible values of the weights of alternatives, and at the same time to interpret the obtained solutions as a function of belonging to a fuzzy set. In order to increase the efficiency of calculations in this problem, which has great computational complexity, it is proposed to use the method of pseudobasic matrices. In the case of incomplete matrices of pairwise comparisons, this method can be used to calculate fuzzy weights of more than 10 alternatives.

**References:**

1. Zgurovsky M.Z., Zaichenko Y.P. Models and methods of decision making in fuzzy conditions, Naukova Dumka. K .; 2011.- 275 p.
2. Gnatienko G.N. Setting preferences on a set of criterion functions in multicriteria optimization problems // Proceedings of the Kyiv University. Modeling and optimization of complex systems. - 1990. - Issue. 9. - P. 87–92.
3. Gnatienko G.M. Expert technologies of decision making: monograph / G.M. Gnatienko, V.E. Snytyuk. - K.: Maclaut Ltd., 2008. - 444 p.
4. Boz'oki, S. and Tsyganok, V. (2019). The (logarithmic) least squares optimality of the arithmetic (geometric) mean of weight vectors calculated from all spanning trees for incomplete additive (multiplicative) pairwise comparison matrices. International Journal of General Systems, 48(4):362–381.
5. Kudin V.I., Lyashko S.I., Khritonenko N.V., Yatsenko Y.P. Analysis of properties of a linear system by the method of pseudobasic matrices // Cybernetics and system analysis. - 2007. - N 4. - P. 119–127.

**<sup>1</sup> Illia Kondratiuk**

Student

**<sup>2</sup> Sophia Vlasiuk**

Student

**<sup>3</sup> Sergiy Paliy**

Docent, Associate Professor of the Information Systems and Technologies Department

*<sup>1,2,3</sup>Taras Shevchenko National University of Kyiv*

## CURRENT PROBLEMS OF INFORMATION SECURITY OF IOT SYSTEMS

The Internet of Things is a system that integrates real things into a virtual network, reducing the need for human involvement in a particular process, or eliminating the need for it completely. Obviously, such a thing not only can't bypass information technologies around the world, but it is the next step in their development, and therefore, we need to identify and realize all current issues of cybersecurity of the Internet of Things.

Any IoT system is a complex technology, which means that its problems have to be considered at several layers. The hierarchy of the Internet of Things architecture [1] looks like this:

Perception layer is the physical level of collecting information about the environment, which means obtaining certain parameters (temperature, humidity, the presence of a certain object, etc.) using different sensors.

Network layer — the level of connection to the network, connection of smart devices to the server, to other devices or to each other. Its task is data processing and transmission.

Application layer — the level that's responsible for providing the user with specialized services of narrow use, so it processes specific tasks (smart home, smart parking, smart subway, etc.)

At every layer, there are a lot of issues that have inherited from already existing network device vulnerabilities, but there are also IoT-only issues. Let's consider each layer separately[2].

*Security at the perception layer.* At this layer, the threats have more physical character: breakdowns of sensors due to improper interaction with them, power outages, poor exposure to too low or high temperatures. If we are talking not about emergencies, then the important role is played by the accuracy of the data received by the sensors. In the future, IoT technologies can be used in medicine, in the physical or chemical industry, and in other fields which require high accuracy numbers, and therefore all indications must be microscopically accurate. Another problem is the maintenance of elements of the Internet of Things. Although they are scalable, they themselves have a fixed configuration and they rarely receive any internal firmware updates, which will limit possibility of their serviceability. Until these elements become physically universal, their distribution and use will be delayed.

*Security at the network layer.* Any Internet of Things system is built on basic communication networks, therefore, their problems extend to the system itself. This

includes CIA violations, DoS/DDoS attacks, man-in-the-middle attacks or any attempts of attacking a system with viruses or other malware. But IoT also has a list of specific problems, which worsen the existing issues(so they require more attention) or which are completely new(so no one is prepared for them). If we are talking about the already existing ones, they are usually connected with system overloads. Devices must compile clusters of information from thousands of sources, and then make some conclusions. This amount of work makes them vulnerable to DoS and DDoS attacks, because it's much easier to cause the system overload, and a tiny delay in data transmission can break all the system work. Speaking of IoT-specific problems, we should mention the insecurity of certification[3]. Networks should be made of trusted devices, but, as we said earlier, their amount is extremely big, and certificates for them are being sold for low prices and in large quantities. Any hacker can easily obtain a trusted device and then compromise a network, send corrupted data into them and even take control of them. Therefore, the approach to these issues needs to be reconsidered.

*Security at the application layer.* These include the vulnerabilities in embedded software, as well as all the security details of each individual application that will be used in the network. Though there is a large amount of devices in an IoT system, inaccurate data of just one of them may affect the operation of other devices. That means, that every single one of them must complete only those tasks, for which it was programmed, and access to the executable code should be limited. Any changes in executable files and programs have to be signed and confirmed by trusted people, and anything connected with device configuration has to be strongly encrypted.

*Conclusion.* All the actual problems of IoT security will continue to worsen, slowing down the process of these technologies being spread to the broad masses. Further analysis of the obstacles to Internet of things development and the search for a way to eliminate them should be intensified. All mentioned vulnerabilities have to be revised, and we have to take action on them, developing a new way of protecting information, which would specify on this kind of systems.

#### **References:**

1. Pallavi Sethi, Smruti R. Sarangi, Internet of Things: Architectures, Protocols, and Applications, 2017. URL - <https://www.hindawi.com/journals/jece/2017/9324035/>
2. Sokolov M.N., Smolianinova K.A., Iakusheva N.A., Problems of IoT security, 2015. URL - <https://cyberleninka.ru/article/n/problemy-bezopasnosti-internet-veschey-obzor/viewer>
3. Oreshkina Daria, Internet of Things Security Reference Architecture, 2017. URL - <https://www.anti-malware.ru/practice/solutions/iot-the-reference-security-architecture-part-1#part5>

<sup>1</sup> Yuliia Kovbas

Student

<sup>2</sup> Olga Izmailova

PhD, Associate professor

<sup>1,2</sup> Kyiv National University of Construction and Architecture

## SCENARIO FORMATION CONSTRUCTION OF A LOCAL CORPORATE NETWORK OF THE ENTERPRISE

**Abstract.** Stages of construction of a corporate network of the enterprise, ways of construction of a network, formation of scenarios of construction of a local enterprise network are considered.

**Keywords:** architecture, local area network, design, client, server, technologies.

With the growing number of enterprises, the problem of designing and creating local enterprise computer networks. This paper proposes a system for forming a scenario for building a local enterprise computer network (LECN)

The purpose of this system is to study the scenario of optimizing the structure of LECN taking into account the situational conditions of a particular subject area of the problem, user requirements based on the application of a set of models of system analysis, system design and multicriteria evaluation [1].

The scenario includes the following steps [2]:

1. Formation of system architecture. For corporate systems, a client/server architecture is recommended, which is mainly based on a template (Fig. 2), where the variables are the number of users and the type of computing resources.

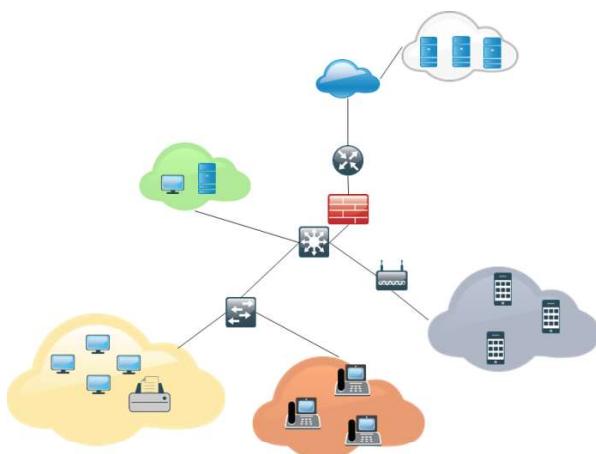


Figure 1: An example of LECN architecture

2. Forming a model of the script in the form of a diagram of precedents (Fig. 3), where the basic precedent is the choice of the necessary software (SW) and hardware (HW).

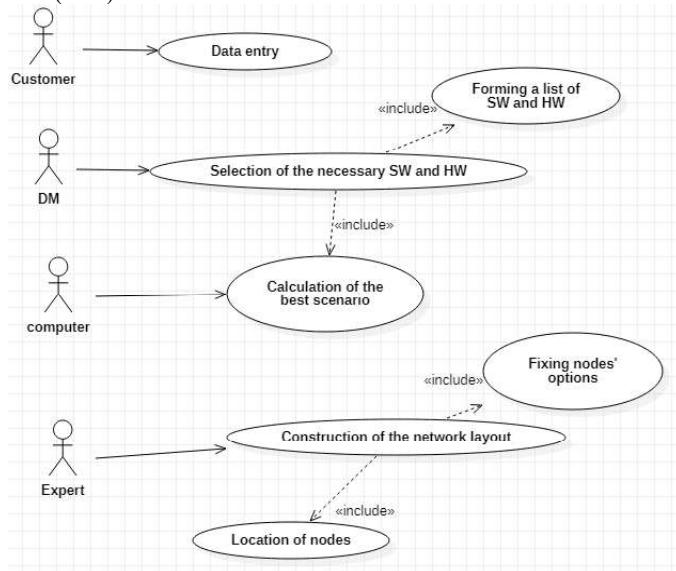


Figure 2: Diagram of system precedents.

The basis of the construction of options for determining the structure of computing resources are the following components:

- Computing resources. The computing resource of the network is a server, but it can be actually a corporate server, which from the company's point of view is HW, or a dedicated place on a remote server (cloud computing), which from the company's point of view is SW A mixed solution - a hybrid calculation - is also possible.
- Corporate database management system. The use of a database depends on the specifics of the company's information activities.
- Electronic document management system. EDMS systems implement a simplification of the company's document management mechanism.
- Services. A service is a web-identified software system with standardized interfaces that can communicate with each other and with third-party applications through messages based on certain protocols.
- Software. Each company has its own specifics, which is determined by the nature of its activities, which in turn determines the software.

Using statistics [3] to find popular solutions, in terms of the functions to be performed by the system, the method of morphological analysis created a possible set of variations (Tab. 1).

Table 1  
The result of morphological analysis

Computing resources	OS	Database	EDMS	Services			Software
				General	Security	Fault tolerance	
Cloud computing	Windows	MySQL	Zoho Docs	Proxy	With Backup	General	
Corporate server	Linux	MongoDB	Logical Doc	Log	Without Backup	MacOS - oriented + general	
Hybrid calculation	MacOS	Microsoft SQL Server	SharePoint Online	Proxy + Log		Windows - oriented + general	
		PostgreSQL		None			
		SQL					

Due to the compatibility of couples most appropriate components are the following variations:

- Z1: Windows + Microsoft SQL Server + SharePoint Online + Cloud (Mail + File) + corporate server (Web + Backup + Log) + Proxy + Software (Windows-oriented + general);
- Z2: Linux + MySQL + Logical Doc + Cloud (Mail + File + Web) + Software in general);
- Z3: Linux + MongoDB + Zoho Docs + corporate server (Mail + File + Web + Backup + Log) + Proxy + software (general);
- Z4: MacOS + PostgreSQL + Zoho Docs + Cloud (Mail + File + Web) + corporate server (Backup + Log) + Proxy + Software (MacOS - oriented + general);
- Z5: Linux + PostgreSQL + Zoho Docs + Cloud (Mail + File + Web) + corporate server (Backup + Log) + Proxy + software (general).

#### 4. Scenario choice

Choosing a scenario decision maker (DM) is guided by the following decision-making methods:

- Method of analysis of hierarchies [4], which allows to assess the composition and relative priority of the established evaluation criteria. The basic ones are: performance, scalability, security, manageability, compatibility and cost.
- Method of direct expert evaluation of the values of the criteria for each variant of the structure [5] (Tab. 2)

Table 2

Normalized values of the criteria of each scenario from 1 to 10 points

Scenario	Performance	Scalability	Security	Manageability	Compatibility	Cost
Z1	7	7	3,25	7	10	1
Z2	1	4	10	1	5,5	10
Z3	4	10	1	1	5,5	8,663
Z4	4	1	5,5	7	1	8,697
Z5	10	1	7,75	7	5,5	8,775
Z1	7	7	3,25	7	10	1

- Method of linear convolution of criteria, based on which the utility function of each scenario is determined.

The selected scenario is the basis for further work of ATS formation of the network layout and activity diagram (Fig. 2), which describes the logic of its implementation. When entering data, the customer provides information about:

1. Type of activity of the company,
2. Number of employees,
3. Plan of the room where the network will be deployed,
4. Specifics of information data flows.

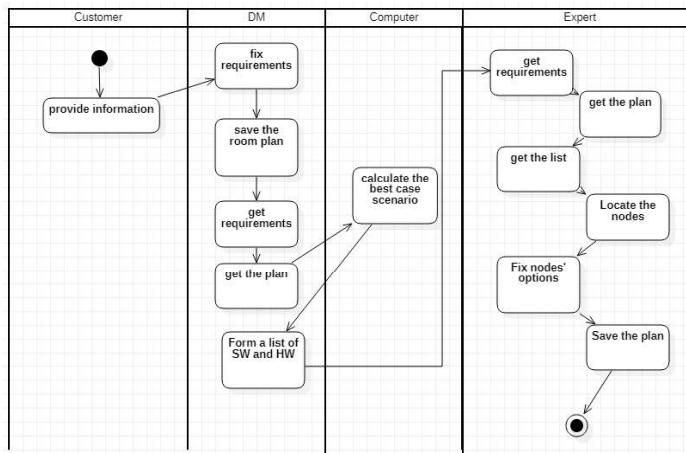


Figure 2: Diagram of activities at the level of precedents.

**References:**

1. T. Hamdi, Introduction to Operations Research, 6th. ed, Williams Publishing House, 2001.
2. Complete Cisco Handbook, 3rd ed, 2009.
3. Global Business Data Platform URL: <https://www.statista.com/>.
4. B. Bogush, O. Dovidkov, Design of secure information systems and networks, DUICT, 2006.
5. O. Izmailova, G. Krasovska, Approach to building an open database of DSS models for the assessment of investment projects of man-made safety, Management of complex systems development, Kyiv, 2018.

**<sup>1</sup> Yuri Kravchenko**

Doctor of Technical Sciences, Professor

**<sup>2</sup> Natalia Dakhno**

Doctor of Philosophy, Associate Professor

**<sup>3</sup> Olga Leshchenko**

Doctor of Philosophy, Associate Professor

**<sup>4</sup> Anastasiia Tolstokorova**

Student

<sup>1,2,3,4</sup> Taras Shevchenko National University of Kyiv

## MACHINE LEARNING ALGORITHMS FOR PREDICTING THE RESULTS OF COVID-19 CORONAVIRUS INFECTION

**Abstract.** The paper analyzes data collected from around the world on patients with COVID-19. A binary classifier has been developed that considers data on a person's health, symptoms, patient's age, and other properties and determines the patient's disease outcome by assigning it to one of two categories: fatal or not. The work's practical value is to help hospitals and health facilities decide who needs care in the first place when the system is overcrowded and to eliminate delays in providing the necessary care.

**Keywords:** supervised learning, classification problem, model fitting, feature selection, feature engineering, data normalization, model validation, confusion matrix, logistic regression, Naive Bayes, Decision tree, random forest.

In March 2020, the World Health Organization officially declared the Covid-19 coronavirus a global pandemic. COVID-19 coronavirus disease is an infectious disease caused by the recently discovered coronavirus SARS-CoV-2. By visualizing the development of the disease in other countries where the outbreak has passed, it is possible to build a truly effective behavior strategy that will save lives while minimally harming the economy, if possible, in these circumstances.

In this work, a DSS model was developed to help identify patterns between the characteristics of the patient (sex, age, types of symptoms, and chronic diseases) who contracted COVID-19 and mortality. This study offers a model of artificial intelligence that can provide hospitals and medical facilities with the information they need to address congestion. It will also allow developing a patient sorting strategy to address hospitalization priorities and eliminate delays in providing the necessary care.

Before the study, relevant data were found that meet the criteria for the work [1]. This dataset collects data from more than 920,000 patients from around the world of all ages, with various chronic diseases and symptoms, including men and women.

After obtaining the required data, several algorithms were selected that are most suitable for achieving this goal [2]. In this study, five classification algorithms were selected that had previously proven to be the best in this type of problem. These are the following algorithms [3]: Logistic regression,  $K$ -nearest neighbors algorithm, Decision trees, Reference vector method, Naive Bayesian classifier.

To assess the quality of classification models that will solve this problem, the following metrics were chosen [4]: accuracy, precision, recall, f-measure, logarithmic loss (logloss), area under the ROC curve.

The results are shown in Table 1 and Table 2.

Table 1

<b>Classifier model</b>	<b>Accuracy score</b>	<b>Precision score</b>	<b>Recall</b>
Logistic regression	0.85185185185	0.956521739130	0.55
K-nearest neighbors	0.86419753086	0.931034482758	0.675
Decision tree	0.90123456790	1.00	0.8
Method of reference vectors	0.777777777778	0.95652173913	0.55
Naive Bayesian	0.81481481481	0.757575757575	0.625

Table 2

<b>Classifier model</b>	<b>logloss</b>	<b>F-measure</b>	<b>ROC</b>
Logistic regression	0.78	0.22168696855	0.824782324771
K-nearest neighbors	0.58	0.3694783949	0.712064649542
Decision tree	0.89	0.1182328674	0.9
Method of reference vectors	0.47	0.428594828689	0.662064649542
Naive Bayesian	0.68	0.339922231006	0.810758598171

After analyzing each of the algorithms and comparing the results, we can say that in this study, among all the proposed methods of machine learning to solve problems of binary classification, the algorithm of decision-making trees best coped.

The developed classifier and its application in DSS can help hospitals and health facilities decide who needs attention in the first place when the system is overcrowded, as well as eliminate delays in providing the necessary care.

#### References:

1. B. Xu, B. Gutierrez, S. Mekaru and et al. "Epidemiological data from the COVID-19 outbreak, real-time case information", Scientific Data. 2020 Mar;7(1):106. DOI: 10.1038/s41597-020-0448-0.
2. D. T. Larose, C. D. Larose, "Discovering Knowledge in Data: An Introduction to Data Mining", John Wiley & Sons, 2014. ISBN: 978-0-470-90874-7.
3. T.M. Mitchell, "Machine Learning", 1997.
4. Y. Sasaki, "The truth of the F-measure", School of Computer Science, University of Manchester, 2007.

**<sup>1</sup>Ruslan Kucherenko**

master

**<sup>2</sup>Olha Kravchenko**

Ph.D., Associate Professor of Information Systems and Technologies

*<sup>1,2</sup>Taras Shevchenko National University of Kyiv*

## **IoT SOLUTIONS SYSTEM FOR CLIMATE CONTROL PROCESS OF MAKING CHEESE**

**Abstract:** The problem of industrial application of IoT of things is considered. A system of IoT solutions for climate control of the cheese-making process at a dairy plant is being developed. The structural scheme of the climate control system is proposed, which includes hardware and means of communication. Sensor data is transmitted to the data processing system.

**Keywords:** IoT solution, climate control system, production technology, block diagram, information technology

The aim of the research is the process of designing and practical implementation of the IoT solution system for climate control of the dairy production process on the example of the cheese ripening chamber.

The object of the study is industrial systems that contain industrial climate control of the process of making dairy products: the manufacture of hard cheeses [5]. Following the reviewed information sources, it was found that the development of IoT solutions in dairy production is poorly represented [2-4]. Given the visible advantages, it was decided to choose for the development of climate control system cheese ripening chamber Wi-Fi module ESP8266-01. The Arduino Nano V3 board and various sensors are used, which provides automatic climate control. The main requirements for the system of IoT solutions for climate control of the cheese-making process are ease of implementation, availability of hardware and software technologies, and price.

In the operation of the climate control system for cheese storage chambers, there is a comparison of defined and current values. According to the results, a signal is given to activate the required process. An indication is used to display all active processes.

Following this goal, we describe the algorithm of the climate control system for cheese storage chambers.

1. Display of nominal values of temperature and humidity depending on the type of cheese. According to the norms of DSTU 6003:2008, raw materials are stored indoors at temperatures from minus 4 to 6 °C and relative humidity - from 80% to 90% [5].

2. Get indicators that characterize the environment. The main ones are temperature and humidity. Additionally, the CO<sub>2</sub> concentration in the air and illuminance are read.

3. Compare nominal and real values. According to the result, give a signal to perform the processes: heating, cooling, humidification, dehumidification, ventilation, lighting control.

Structure diagram for automated climate control of image storage cameras in [1].

According to the described algorithm, the device should consist of an Arduino microcontroller, light sensor, temperature sensor, humidity sensor, CO<sub>2</sub> level sensor, relay set control system, WI-FI module [6-10].

The strength of the IoT solution system for climate control of the dairy production process is its structure. Due to the branched structure, the program simultaneously receives and processes data from light, pressure, and humidity sensors.

The disadvantage of this software is the initial stage of implementation of the system, which will require additional funds for the purchase of sensors and installation of the communication network as hardware for the IoT solution for climate control of dairy products, but it will pay off during operation.

In the future, this system allows you to add modules for the analysis of manufactured products or storage system.

Threats that will have negative consequences for the system of IoT solutions for

- climate control of the dairy production process can be considered insufficient funding;
- staff training costs;
- the need to introduce a specialist to the plant staff who will be responsible for maintaining the software product.

The developed software product takes into account the advantages of analogs on the market, but has a cheaper cost and has the advantages of ease of use required for cheese production.

It is planned to implement the system for production.

## References:

1. Kravchenko O. V., Kucherenko R. Y., Danchenko E.B., Besedina S.V. Розробка системи iot рішень для клімат контролю процесу виготовлення молочної продукції// Sciences of Europe, Vol 2, No 51. Praha, Czech Republic. 2020. Pp. 69-75.
2. Perera C., Chi H. L. M. A survey on internet of things from industrial market perspective // IEEE Access. – Volume 2, 2014, pp. 1660–1679.
3. Debasish Mondal, The Internet of Thing (IOT) and Industrial Automation: a future perspective// World Journal of Modelling and Simulation (2019) [Electronic resource] - Resource access mode: <https://cutt.ly/JgDW30c>
4. Shopan Dey, Ayon Roy and Sandip Das. Home Automation Using Internet of Thing// IRJET (2016), 2(3) pp. 1965-1970,
5. ДСТУ 6003:2008 Сири тверді. URL: <https://cutt.ly/8gDWVOD>
6. Arduino nano v3 datasheet URL: <https://cutt.ly/4t2cjeu>
7. GY-302 datasheet URL: <https://cutt.ly/Qt2czHG>
8. AM2320 datasheet URL: <https://cutt.ly/Dt2cQq5>
9. MQ-135 datasheet URL: <https://cutt.ly/Zt2cR1L>
10. ESP8266-01 datasheet URL: <https://cutt.ly/9gSxH1y>

**<sup>1</sup> V.Kudin**

Doctor of Technical Sciences, Associate Professor

**<sup>2</sup> A.Onyshchenko**

Doctor of Economic Sciences, Associate Professor

**<sup>3</sup> Oleh Ilarionov**

PhD in Engineering Science, Associate Professor

<sup>1,2,3</sup> Taras Shevchenko National University of Kyiv

## MODELING OF DYNAMIC ECOLOGICAL-ECONOMIC INTERACTION

The emergence and intensification of global environmental problems and the growing destabilization of the climate prove the need to change the established economic paradigm of development. It is becoming increasingly clear that the traditional model of economic growth, which ignores the importance of natural factors, cannot prevent the aggravation of global environmental problems, including further climate change; it has exhausted itself in the current historical period of civilization. Thus, there is a need to build a concept of balanced interaction of economic and environmental spheres.

Balance implies mutual coherence of the system of proportions of production at the macro and micro levels. It characterizes the degree of proportionality, i.e. the degree of conformity of production, needs, and environmental standards. More progressive proportions make it possible to achieve a higher level of balance.

The balance of the economy as a form of systematic interconnection of production and consumption implies the correspondence of the volume and structure of output of products and services to the volume and structure of social needs in kind and value, in time and space, in qualitative and quantitative characteristics.

In [1] it is proposed to consider the costs of meeting the emission limits of greenhouse gases in the structure of the main production industries in the form of a balance ecological and economic model:

$$\begin{cases} x_1 = A_{11}x_1 + A_{12}x_2 + Cy_2 + y_1, \\ x_2 = A_{21}x_1 + A_{22}x_2 - y_2, \end{cases} \quad (1)$$

The first equation of the proposed model reflects the economic balance – the distribution of gross industrial output for production consumption of primary and secondary production, final consumption of primary production and costs associated with the implementation of environmental obligations.

An important step in the study of ecological and economic interaction is to preserve the structural balance of the model in terms of its directed transition from one state to another. The balance of the economy as a form of systematic interconnection of production and consumption implies the correspondence of the volume and structure of products and services produced to the volume and structure of social needs in kind and value, in time and space, qualitative and quantitative characteristics. The transition to a high organization and efficiency economy leads to the achievement of a comprehensive dynamic balance of production and ecology, as this is a necessary prerequisite for proportional, deficit-free development of ecological and economic system, optimal functioning of all its structural subsystems. Without a stable balance of economic development, which requires the entire correspondence between the volume and structure of production and the volume and structure of needs, the optimal proportions of all reproduction processes, it is impossible to achieve planned environmental parameters, rational use, gradual intensification of economic development.

Successful solution of the process of interaction between ecology and economy is possible under the conditions of using the mathematical apparatus as an effective tool for formalizing the abstract field, i.e. mathematical modeling.

According to model (1), the original output vector  $x = (x_1, x_2) = (\underbrace{x_{11}, x_{12}, \dots, x_{1p}}_{x_1}, \underbrace{x_{2p+1}, x_{2p+2}, \dots, x_{2p+q}}_{x_2})$  can be interpreted as follows:

The components of the sub-vectors respectively form the resulting interaction of the components of the output of useful products of the main  $x_1 = (\underbrace{x_{11}, x_{12}, \dots, x_{1p}}_{x_1})$  and components of the auxiliary production of the “output” of pollutants  $x_2 = (\underbrace{x_{2p+1}, x_{2p+2}, \dots, x_{2p+q}}_{x_2})$  among themselves, i.e. the lower level of interaction.

Together, the components of the vector  $(x_1, x_2)$  form an aggregate interaction of sub-vectors, i.e.  $x$ . This vector will determine the resulting gradient of “exit” of the ecological and economic system. It is easy to see that changes in the ecological and economic model will cause a change in the vector of output – the exit of the system and vice versa.

The issue of determining the priority of development of economic and environmental components can be determined by the involvement of experts – specialists in this field. That is, to form the coefficients of decomposition of vector  $x$  by vectors  $x_1, x_2$ . The decision to give them a specific meaning can be made considering a system of criteria. In such conditions, the construction of the vector of advantages

according to the system of criteria is solved with the involvement of a group of experts (methods of expert evaluation [2]).

On the other hand, the result of the work of the team of experts (application of expert evaluation methods) also provides a mechanism for the formation of weights for the system of criteria, i.e. their priority. What for the specified model (1) forms the direction of the main development and improvement of the model. A separate (inverse), important task is to make targeted changes in the model, which will “fill” the content of the strategy of changes in the technological elements of the model, to achieve the set of development indicators. This is the task of building a given mainstream of development (directed change) of the ecological and economic model [3]. The complexity of the task increases with the need to consider the limitations in the system for conducting such organizational “innovations”.

Thus, there is a pressing issue of constructing mathematical algorithms and their application to the balance ecological and economic models of type (1) in order to work out ways to transfer the system from one state to another by given economic and environmental criteria.

#### **References:**

1. Onyshchenko A.M. Methodology of mathematical modeling of economic and ecological interaction in terms of implementation of the Kyoto Protocol / I.M. Onishchenko. Lyashenko, AM Onishchenko // Economic Cybernetics. - 2011. - 64-6 (70-72) - P. 17-26.
2. H. Hnatienko, V. Snytyuk. Expert decision-making technologies. K.; Ltd. “McLaut”, - 2008.-144 p.
3. V. Kudin, A. Onyshchenko, I. Onyshchenko. Algorithmizing the methods of basis matrices in the study of balance intersectoral ecological and economic models. Eastern-European Journal of Enterprise Technologies. ISSN 1729-3774, 3/4(99)2019, p. 45–55 DOI: 10.15587/1729-4061.2019.170516

<sup>1</sup> **Olga Leshchenko**

PhD in Technical Sciences, Associate Professor

<sup>2</sup> **Natalia Dakhno**

PhD in Technical Sciences, Associate Professor

<sup>3</sup> **Oksana Herasymenko**

PhD in Technical Sciences, Associate Professor

<sup>4</sup> **Vladislav Lavrinovich**

student Department of Networking and Internet Technologies

<sup>1-4</sup> Taras Shevchenko National University of Kyiv

## APPLICATION PECULIARITIES OF GRADIENT DESCENT ALGORITHMS IN NEURAL NETWORKS

The steepest descent method (gradient descent) is a method of numerical optimization to find the minimum or maximum of a function by moving in the direction of the gradient or anti-gradient. It has the easiest implementation compared to other methods of optimization, has a fairly weak convergence conditions, but very low speed (linear). The gradient step is very often used in other methods, such as Fletcher-Reeves.

The gradient descent method is used in various machine learning algorithms where we need to find the extremum of the function - neural networks, SVM, k-means, regression.

In neural networks, the gradient descent method with some modifications is very widely used for perceptron training and is also known as error backpropagation.

The problem is that for neural network training, it is necessary to change the weights in order to minimize the average loss at the output (use optimization algorithms on the loss function to reduce loss as quickly as possible). In other words, in order to formally make one step on the gradient, it is necessary to submit all data to the input sequentially, calculate the error for each value and calculate the required weight correction, and after submitting all data calculate the value in the adjustment of each coefficient (gradient sum) and correct the coefficients for this step, which takes a lot of operations to perform[1].

For instance, for linear regression, the loss function will be defined as

$$S(\theta) = \sum_{i=0}^n (y_i - \hat{y}_j)^2$$

where  $\theta$  is a free regression parameter. By definition, a gradient is a vector of the following form:

$$\nabla f = \frac{\partial f}{\partial x_1} e_1 + \dots + \frac{\partial f}{\partial x_n} e_n$$

and  $\frac{\partial f}{\partial x_n} e_n$  is a partial derivative.

Let us define the point  $a$ , around which the function is defined and differentiated. Then the vector of the antigradient will indicate the direction in which the function will fall the fastest. From this we can determine that in a certain point  $b = a - a\nabla f(a)$  for

some small  $a$ , the value of the function will be less than or equal to the value of the function at point  $a$ . Based on the afore mentioned concepts, we can derive a formula for determining the unknown parameters:

$$\theta_j = \theta_j - \alpha \frac{\partial S(\theta)}{\partial \theta_j}$$

where  $\alpha$  is a step of the method. From here we can derive a formula for calculating free regression parameters:

$$\frac{\partial S(\theta)}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} (y - \sum_{i=0}^p \theta_i x_i)^2 = -2(y - \sum_{i=0}^p \theta_i x_i) x_j$$

Let us rewrite the formula for a sample of  $n$  elements:

$$\theta_j = \theta_j + \alpha \sum_{i=0}^n (y^{(i)} - \hat{y}^{(i)}) x_j^{(i)}$$

In other words, for  $v$  iterations and  $n$  elements for each  $j$  it is necessary to repeat the above algorithm  $v$  times.

The requirement of simultaneity means that the derivative must be calculated with the old  $\theta$  values, i.e. we should not initially calculate the first parameter, then the second, etc., because after changing the first parameter, the individual derivative will also change its value.

Suppose we have a three-dimensional figure, and if we calculate the parameters one by one, then this process can be described as movement in coordinates (one coordinate at a time) instead of movement along the vector. This variant of the algorithm is called batch gradient descent. The number of iterations also means that the parameters will be adjusted until the previous and given values of the function become equal. In practice, such equality is impossible to achieve and the criterion  $\varepsilon$ , is introduced, which characterizes the convergence threshold, i.e. if the difference between the previous and current function is less than or equal to  $\varepsilon$  iterations cease.

There is another version of the algorithm - stochastic gradient descent. Let's compare the formulas of stochastic and batch gradient descent in the pseudocode of iterations:

Batch:     for  $i$  in train\_set: {

$$\theta_j = \theta_j + \alpha \sum_{i=0}^n (y^{(i)} - \hat{y}^{(i)}) x_j^{(i)}$$

}

Stochastic:   for  $i$  in train\_set: {

$$\theta_j = \theta_j + \alpha (y^{(i)} - \hat{y}^{(i)}) x_j^{(i)}$$

}

Although the formulas are similar, there is a significant difference between the algorithms: batch gradient descent calculates the step using the entire set of input parameters, and stochastic uses only one element per step.

The disadvantage of the batch algorithm is that when passing through the curved flat valley, the method performance is very slow, because the steps become very small [2]. However, the stochastic algorithm does not follow the exact direction of anti-gradient but deviates each time in random direction, although it almost always reaches an extreme.

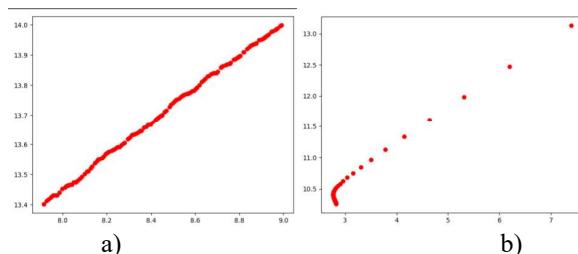


Figure 1. Comparison of stochastic (a) and batch (b) descents for 200 elements

As we can see in Fig.1 b) batch descent falls into the the curved valley, so the steps along it become very small, however, sooner or later the batch descent will reach the minimum. Therefore, there is no noticeable difference between algorithms in the sample of 200 elements, but if you increase the sample to 2000 elements, the stochastic descent shows much higher efficiency.

There is also a third type of algorithm: mini-batch. Unlike the batch and stochastic algorithms, the mini-batch algorithm uses not one or all values, but a number of values, such as 100 random ones. This algorithm is more often used in practice, because it is more efficient than the batch algorithm, and less random than stochastic descent [3].

Thus, the batch algorithm is well suited for strictly convex functions, stochastic works better with functions that have a large number of local minima (random values allow to come out of curved valleys), stochastic algorithm is generally faster, because it does not use the entire sample, although it performs more steps. The mini-batch method allows you to vectorize the stochastic algorithm, which speeds up execution and increases accuracy.

## References:

1. A. El Mouatasim (2018). Implementation of reduced gradient with bisection algorithms for non-convex optimization problem via stochastic perturbation. *Journal of NumericalAlgorithms*, 78(1):41 - 62.
2. Y. Yao, L. Rosasco, and A. Caponnetto. On early stopping in gradient descent learning. *Constructive Approximation*, 26(2):289–315, 2007
- 3.Cotter,A,Shamir,O.,Sridharan, K.,“Better mini-batch algorithms via accelerated gradient methods,” *Advances in Neural Information Processing Systems*,2011,pp.1647-1655.

**<sup>1</sup> Yehor Myroshnychenko**

Master degree Student of Information Systems and Technologies.

**<sup>2</sup> Sergiy Palij**

PhD, Associate Professor of Information Systems and Technologies.

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## ROAD TRAFFIC OPTIMIZATION BY IOT

**Abstract.** The article discusses the principles of monitoring road traffic. The basis of the traffic monitoring system is visual observers and sensors of traffic parameters. Existing systems and their traffic sensors show unsatisfactory results. In order to optimize road traffic in cities, it was decided to develop a microprocessor-based system for determining the traffic intensity. The basis is Hall sensors and the Arduino Pro Mini microprocessor controller.

**Keywords:** vehicle traffic monitoring, traffic optimization, IoT.

**Purpose and objectives.** Given the growing rate of increase in traffic, the issue of traffic monitoring is relevant for effective assessment of problematic sections of the route, obtaining operational traffic information, traffic management in cities, regulating the duration of traffic lights and timely activation of information signs. The purpose of this work is to develop a zonal controller that could perform all these functions. The task is to develop such a block so that it consists of a minimum number of functional units and provides accurate information about the condition of the road.

**Object and subject of research.** The object of research is the basis of information systems for traffic monitoring - visual observers and sensors of traffic parameters. The efficiency of the entire information system largely depends on their functionality and reliability of data. Modern road traffic control systems are based on zonal controllers (ZC), which consist of sensors to obtain primary information about the movement of vehicles and a microprocessor unit for data processing [1]. The subject of the study is the simplification of the design of the ZC while improving their characteristics.

**Scientific novelty.** Optical and ultrasonic sensors, magnetic gates, video surveillance systems are most often used in ZC [2]. Existing traffic sensors of foreign production show unsatisfactory results in the conditions of intensive traffic of vehicles (vehicles), parking of vehicles in the lane, visual overlap of vehicles of the traffic sensor and when moving vehicles between lanes. This leads to a premature transition of the ZC to the functional state of "congestion", and a significant reduction in the informativeness of the system. I proposed the implementation of the ZC, which eliminates all the above shortcomings.

Research results. To find the optimal solution for recording the movement of

cars, the currently existing traffic control systems [3,4] are analyzed, which use different physical effects and appropriate sensors to collect primary information. Analysis of existing monitoring systems shows that the weak point of such systems are the sensors / methods by which motion is detected. Therefore, the main task in creating a new traffic control system is to choose an effective method of recording traffic

means. Below we will consider the method based on the Hall effect and show the possibility of its application in traffic monitoring systems. Each vehicle has metal elements in its body and, when moving, it will in one way or another affect the Earth's magnetic field [5]. To record this change, we use magnetic field sensors. The data read by the sensors during the perturbation of the magnetic field is processed by a microcontroller with a pre-loaded program (in our case, the microcontroller performs the functions of determining the speed and direction of movement). The system includes the following components and blocks:

- two sensors - Hall sensors for removing primary information;
- data processing unit, consisting of a reading unit and a calculation unit;
- the reading unit, implements the reading of information from the Hall sensors with a certain reading step  $t_s$ .
- computing unit, that implements the algorithm of the system, performing the necessary speed calculations movement on the basis of information received from the reading unit;
- the data visualization unit implements the user interface of the system with the visualization of the monitoring results.

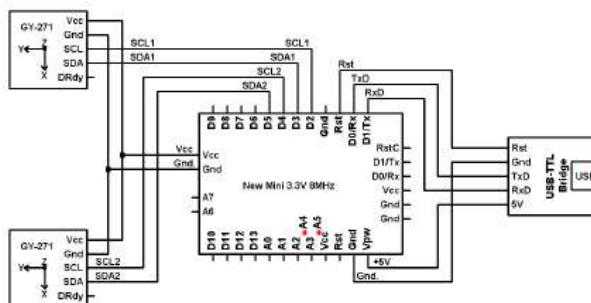


Figure 1 - Schematic of the electrical principle of the zonal controller

**Research methods and tools.** To test the efficiency of the above method of measuring the intensity of movement, a model of the experimental stand was designed. Magnetic field sensors were installed at a distance of 10 cm from each other. To emulate the movement of the vehicle, at a height of 2-3 cm from the sensors moved a piece of steel, oriented perpendicular to the direction of movement. The piece of steel was chosen as an object that to some extent resembles the front / rear axle of the car.

**Conclusions:**

1. The existing traffic control systems are analyzed.
2. The analysis of physical methods of control of movement of motor transport is carried out.
3. The analysis of methods of information processing and traffic monitoring is carried out.
4. To register the movement of vehicles, it is proposed to use sensitive magnetic field sensors - digital compasses.
5. Developed a project of a microprocessor system with magnetic sensors to record the movement of vehicles.

**References:**

1. Ju R. S. Transportation quarterly / Eno Foundation for ... v.41 [Електронний ресурс] / R. S. Ju, A. R. Cook, T. H. Maze // Westport, Conn. : The Foundation. – 1987. – Режим доступу до ресурсу: <https://goo.su/2scr>.
2. Ahmed S. A. Urban freeway traffic management technology [Електронний ресурс] / Ahmed // Journal of Transportation Engineering / Volume 112 Issue 4 – Режим доступу до ресурсу: <https://goo.su/2scs>.
3. Jovanis P. P. Coordination of Actuated Arterial Traffic Signal Systems [Електронний ресурс] / P. P. Jovanis, J. A. Gregor // Journal of Transportation Engineering / Volume 112 Issue 4. – 1986. – Режим доступу до ресурсу: <https://goo.su/2SCS>.
4. Monitoring avtomobilnyih dorog [Електронний ресурс] // ONIL SAPR GIS. – 2012. – Режим доступу до ресурсу: <https://goo.su/2sCu>.
5. Sistemyi dorozhnogo monitoringa. Intellektualnaya transportnaya sistema [Електронний ресурс] // MK Grupp. – 2015. – Режим доступу до ресурсу: <https://goo.su/2scV>.
6. Datchyky mahnitnoho polia. Novi zastosuvannia i tekhnolohii vymiriuvannia rukhu i strumu [Електронний ресурс] // Ecworld. – 2017. – Режим доступу до ресурсу: <https://goo.su/2SCY>.
7. ARDUINO VY'MIRYuYeMO MAGNITNE POLE ZEMLI MAGNITOMETROM HMC5883L [Електронний ресурс] // Mikrotik KPI – Режим доступу до ресурсу: <https://goo.su/2SCZ>.
8. Sbor i obrabotka dannyih IoT. Stoit li vnedryat i kak eto sdelat effektivno [Електронний ресурс] // DIS Group. – 2018. – Режим доступу до ресурсу: <https://goo.su/2Sd0>.
9. Biloschitsky A. A. Optimizatsiya dorozhnogo trafika s ispolzovaniem sredstv IoT [Електронний ресурс] / A. A. Biloschitsky, Y. M. Myroshnychenko // «Molodiy vcheniy» # 5 (81). – 2020. – Режим доступу до ресурсу: <https://goo.su/2Sd2>.

**<sup>1</sup>Volodymyr Nakonechnyi**

Doctor of Technical Sciences, Professor at the Department of Cyber Security and Information Protection, Faculty of Information Technologies

**<sup>2</sup>Oleksandr Pliushch**

Candidate of Technical Sciences, Docent at the Department of Mobile and Videoinformation Technologies

**<sup>3</sup>Andrii Bielikov**

Student

<sup>1,3</sup> Taras Shevchenko National University of Kyiv

<sup>2</sup> State University of Telecommunications, Kyiv, Ukraine

## **DEVELOPMENT AND ANALYSIS OF ALGORITHMS FOR RECOGNIZING MOVING OBJECTS IN THE DATA STREAM**

**Abstract.** The problem of detecting the movement of objects in a video stream is relevant for the development of systems based on computer vision. Computer vision is a new scientific direction of robotics and artificial intelligence, associated with image processing and the selection of real-world objects. This direction is closely related to such areas as computer vision, image processing, analysis and pattern recognition.

**Keywords:** motion recognition, algorithm, stream, frame, image.

### **Introduction:**

With an ever-expanding scope of information systems and computer technologies, the task of processing video images in the video stream occurs quite often.

Today, there are a lot of works that overviewed recognition of objects and multimedia data by such scientists as D. Forsyth, A. Yuriev, V. Pratt, J. Furman, etc. The authors of these works mainly provided a theoretical basis for the algorithms for motion detector and object tracking [1]. They mathematically described the principles of their operations as well. Therefore, it became necessary to develop a program that would automatically track objects along the entire "field of view" of the camera.

### **The problem of identifying moving objects and methods of its solution:**

The task of detecting the movement of an object was first implemented in motion sensors. Motion sensor is a special device that tracks the movement of some objects. This term is most often understood as an electronic infrared sensor that detects the presence and movement of a person, switching the power supply of electrical appliances (most often, lighting). The principle of operation of motion sensors is based on tracking the level of infrared radiation in the sensor's field of view. When a person (or other massive object with a temperature higher than the background temperature) appears, the voltage rises at the output of the pyroelectric sensor. Typically, motion sensors are used in two cases: either for automatic control of lighting devices, or for burglar alarms [2].

Nowadays several algorithms have been developed to solve the problem of detecting moving objects in two-dimensional images. Most of them are based on

separating the foreground from the background of the image. Background subtraction methods make a pixel-by-pixel comparison of the current frame with the model and associates pixels with the foreground or background.

The main problem that arises when processing data from a video stream is the speed of the algorithms. It is necessary to have time to process the current frame before the next one is received. The idea of the algorithm is to compare the location of all objects in the "field of view" of the web-camera on two adjacent frames, and determine which of them has changed its position. In other words, you need to identify and recognize a moved object.

Object recognition theory develops the theoretical foundations and methods for classifying and identifying objects, phenomena, processes, signals, situations and other objects characterized by a finite set of certain properties and attributes. Such tasks have to be solved quite often. Example is crossing a street at traffic signals. Recognizing the color of the traffic light and knowledge of the traffic rules allow you to make the right decision about crossing the street at the moment [3].

There are two main directions in object recognition that can be distinguished. First one is the knowledge of properties of recognition. Second one is their explanation and modeling, as well as the development of the theory and methods of constructing devices designed to solve individual problems in applied problems.

Typically tasks such as tracking and detecting motion can be complicated by foreign objects in the frame and obstacles. Typical tasks of computer vision are [4]:

- Recognition (the task of determining a specific characteristic object, feature or activity containing video data).
- Motion understanding (a sequence of images is analyzed to find an estimate of the velocities of each image point).
- Scene reconstruction (reproduction of a three-dimensional model of a circuit with two or more images).
- Image restoration (removing noise, sharpening motion blur, etc.).

During the research, a webcam was used to receive a continuous stream of video data. To analyze an image from a webcam, it was split into frames. A frame is a picture in .jpg format, which is a snapshot from a video stream.

#### References:

1. Zellner A., Ando T. Bayesian and Non-Bayesian Analysis of the Seemingly Unrelated Regression Model with Student-t Errors, and its Application for Forecasting. International Journal of Forecasting, 2010. P. 413-434.
2. T. Bouwmans and E. H. Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. Computer Vision and Image Understanding, 2014.
3. C. Guyon, T. Bouwmans, and E.H. Zahzah. Robust principal component analysis for background subtraction: Systematic evaluation and comparative analysis, 2014.
4. Zoran Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In Pattern Recognition, ICPR 2004, IEEE Proceedings of the 17th International Conference, P. 28–31.

<sup>1</sup> **Kateryna Nemchenko**

Bachelor degree Student of Information Systems and Technologies

<sup>2</sup> **Sergiy Paliy**

PhD, Associate Professor of Information Systems and Technologies

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## **STATEMENT OF THE TASK OF BUILDING AN ADAPTIVE SYSTEM OF ENERGY-EFFICIENT LIGHTING FOR ADMINISTRATIVE BUILDINGS BASED ON THE INTERNET OF THINGS**

**Abstract.** The relevance of developing adaptive lighting systems based on the Internet of Things has been described. The requirements for the system's capabilities have been formulated.

**Keywords:** Internet of Things, lightning, administrative buildings.

Every year, the use of systems based on the Internet of Things (IoT) is becoming more common in various industries. The use of such systems for controlling technological processes in modern office premises, buildings, and structures provides them unique advantages, which positively affect the level of competitiveness. IoT is widely applied for automating business processes, managing the operation of climate systems, controlling object security, and accessing premises and data in Information Systems. An example of IoT such systems is a system for controlling the lighting of administrative buildings.

*The advantages of developing and using a smart lighting system are the improvement of energy efficiency and reducing energy costs; the possibility of centralized and individual devices management; monitoring the state of the lighting system; creating comfortable working conditions; managing system with a smartphone; ensuring security for people in the building.*

It is proposed to develop an adaptive system for energy-efficient lighting of administrative buildings using the Internet of Things technology.

For the convenience of users, the system should provide the ability to control each lamp individually or to control the formed group of lamps simultaneously. The process of controlling indoor lighting can be simplified by creating basic scenarios for typical situations that do not require configuration, as well as the ability to create personalized scenarios depending on the user's needs.

For optimal IoT operation, the system must be equipped with light sources that allow you to change the color temperature of lighting, brightness, and color of light. Small changes, such as increasing or decreasing the brightness or setting warmer or colder lighting, as in Figure 1, will contribute to better creativity of employees, or will help direct their attention, focus, and thus increase their performance [1, 2].

The lighting system must interact with the security system and amplify alerts through color lighting and distinct scenarios that signal a dangerous situation.



Figure 1 – Different light temperatures in the Herstedlund school

Such a system implementation will increase the level of energy efficiency, reduce energy consumption, and create a controlled, comfortable production environment. The introduction of scenarios based on the use of presence sensors and light level sensors, in the absence of people, allows you to either completely turn off the lights or illuminate the room with lower brightness, taking into account the brightness of daylight, if the room has windows [3, 4].

While implementing an IoT system for lighting control potential problems should be taken into account. *Examples of such issues* are cyber threats to the system or insufficient digital competence of personnel.

**Conclusions.** The feasibility of the development lies in the fact that the use of the Internet of Things (IoT) system for administrative buildings such as lighting control systems will optimize business processes, conditions for their implementation, reduce energy and financial costs. In particular, the implementation of such a system will expand the capabilities of devices management and help to improve processes by analyzing the data obtained while monitoring the usage of premises and devices. Moreover, it will have a positive impact on the employees' performance and reduce energy consumption.

#### References:

1. Effect of Light Color Temperature on Human Concentration and Creativity. URL: <https://pubmed.ncbi.nlm.nih.gov/26098084/>
2. School invests in ‘concentration’ lighting. URL: <https://www.luxreview.com/2019/03/08/schools-invests-in-concentration-lighting/>
3. Saving Energy with Internet of Things. URL: <https://www.schoolofthefuture.sg/saving-energy-with-internet-of-things>
4. Creating Smarter Schools: Benefits and Applications of IoT in Education. URL: <https://www.igor-tech.com/news-and-insights/articles/creating-smarter-schools-benefits-and-applications-of-iot-in-education>

**<sup>1</sup>Peter Nikolyuk**

Doctor of physics and mathematics sciences, professor of computer sciences and information technologies department, professor

**<sup>2</sup>Tatiana Neskorodieva**

PhD, associate professor of computer science and information technologies department

**<sup>3</sup>Eugen Fedorov**

Doctor of technical science, associate professor, professor of department robotics and specialized computer systems

**<sup>4</sup>Esta Chioma**

Student of computer sciences and information technologies department

<sup>1,2,4</sup> Vasil' Stus Donetsk National University

<sup>3</sup>Cherkasy State Technological University

## **INTELLECTUAL ALGORITHM IMPLEMENTATION FOR MEGACITY TRAFFIC MANAGEMENT**

**Abstract.** This investigation is aimed solving traffic problem in a big city. Laying of an optimal route for each vehicle in metropolis has been performed. As a result of application of urban traffic regulation an optimal regime of vehicle movement is achieved.

**Keywords:** urban traffic, optimal route, A<sup>\*</sup>-algorithm, multigraph, java program

The research represents technology that implements an optimal route for each vehicle in a large city [1]. As a criterion of an optimality, time travel vehicle on the pointed route is selected. Let us call this criterion as t-optimal one.

Transport network of any city is presented as a weighted oriented graph [2]. The laying of t-optimal route in such graph is based on: 1). fixing vehicles that cross each intersection; 2). control of all traffic lines between intersections and as a result, the optimal route is selected; 3) use of A<sup>\*</sup>-algorithm.

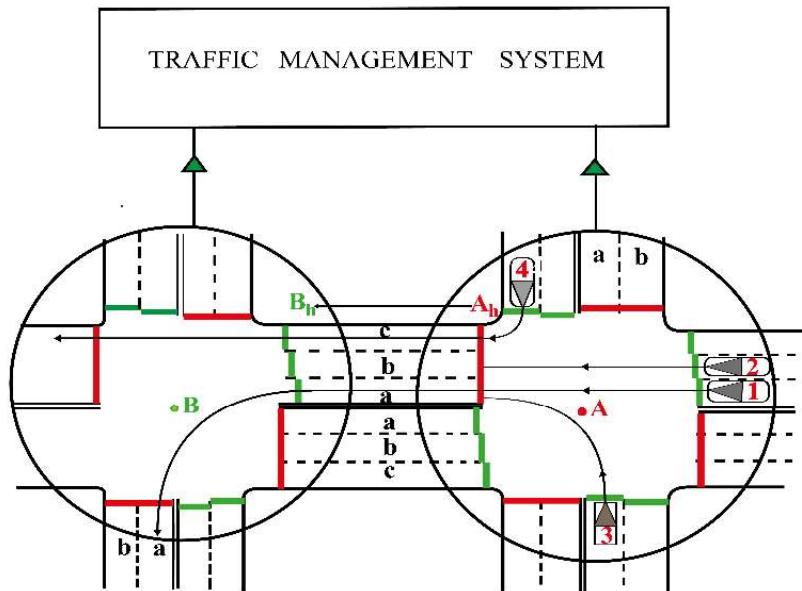
Thus, the entire transport network of the city is controlled by an intelligent traffic management system. By plotting the optimal route in the graph, we can technically perform a procedure for regulating vehicle flows due to a dynamic interaction in real time between a traffic management system (TMS) and each vehicle that has set its position (Fig. 1).

The TMS transmits to each driver a voice commands concerning a route to a destination declared by each driver as with ordinal GPS navigation. The peculiarity consists in the fact that program puts the t-optimal route, but not geometrically optimal (g-optimal) one as modern GPS navigation does. Because of g-optimal routes traffic jams in cities take place. In a case of applying a criterion of t-optimality, there is an opportunity not only to improve a travel through individual intersection, but also to organize optimal routes for all vehicles in the metropolis, taking into account a traffic situation at each particular time moment.

The proposed technology focuses not only on laying t-optimal routes but also is extremely dynamic working with

renewable data. The integrated effect of the intelligent regulation leads to a new quality – synchronization of traffic flows, and consequently, to the disappearance of the traffic jams. An ultimate goal of the study is the algorithm that allows synchronizing the flows of vehicles, optimally using all city transport routes and accompanying each vehicle to the final destination in a shortest possible time. With this approach traffic jams in cities will not occur in general!

Figure 1 – Two adjacent cross-shaped intersections – A and B – and vehicles that enter



the lane between these intersections are shown. The red stripes are input sensors. Output sensors are represented as adjacent green stripes.

#### References:

1. D. Boguto, V. Volynets, P. K. Nikolyuk, P.P. Nikolyuk, Automated system management of vehicle movement in city, Bulletin of Kharkiv University, series “Mathematical Modeling. Information Technology. Automated Control Systems” is.35, pp. 3-9, (2017).
2. D. Boguto, V. Komarov, P.K. Nikolyuk, P.P. Nikolyuk, Intellectual urban transport traffic management algorithm, Bulletin of Kharkiv University, series “Mathematical Modeling. Information Technology. Automated Control Systems” is.38, pp. 4-13, (2018).

**<sup>1</sup>Vadym Paiuk**

Postgraduate Student at the Computer Engineering & System Programming Department

**<sup>2</sup>Olena Heidarova**

Candidate of Economical Sciences (PhD), Associate Professor at the Economic & Management Department

<sup>1,2</sup> Khmelnytskyi National University

## **DETECTING SOFTWARE MALICIOUS IMPLANT BASED ON ANOMALIES RESEARCH ON LOCAL AREA NETWORKS**

The development of information technology in various fields continues to be accompanied by a relentless desire of attackers to benefit from shortcomings in their protection. The most relevant for the benefit from the point of view of attackers are organizations and enterprises in which information technology operates. There are many ways to penetrate the local computer networks of enterprises (organizations) for the purpose of unauthorized access to information in them. One way for attackers to access information resources of enterprises (organizations) is to use undocumented capabilities in software and hardware of personal computers and peripherals, which allow for covert unauthorized access to system resources, usually through a local network. A program implant is a secretly implemented program that poses a threat to the information contained in a computer [1]. The main purpose of implant of software is to provide unauthorized access to confidential information.

As an object of study we will consider program implant of software used in local computer networks of enterprises (organizations).

Program implant software bookmarks can take part in the creation of botnets, the implementation of Trojans, and so on. Therefore, the problem of detecting malicious software, in particular, undocumented software bookmarks, remains relevant.

One of the tasks that needs to be solved is to develop models of program implant software in local computer networks.

Consider the types of threats from program implant that can be implemented in the local network. Their analysis is related to the security requirements of computer systems on the network: confidentiality, integrity, availability and authenticity.

Implant in program models in LAN computers: 1) the model of "interception"; 2) the "observer" model; 3) the model of "compromise"; 4) model "distortion or initiator of errors"; 5) model "garbage collection".

Research of implant in program by professional specialists in cybersecurity allows to establish their presence on the following grounds: the presence of software modules that do not meet the purpose of the process; the presence of operating system objects that are opened by the process that do not meet the purpose of the process; high intensity of input-output operations from a certain process; a high percentage of CPU or internal memory usage from a particular process; the similarity of the file name to the file name related to the operating system; the executable file of the operating system process is not placed in the conventional directory; the process related to the operating

system is performed on behalf of the local user; the code execution protection system in the data area, which is enabled for all processes, is disabled for the considered process; for a process related to the operating system, other directories are involved, different from what should be for such a process; there is no digital signature in the executable files of the software; high network activity of the process, which must work locally; etc. However, to improve the effectiveness of implant in program detection, tools are needed to establish the existence of implant in program without the intervention of a network administrator who may not handle some of the features for various reasons. Program implant can use masking tools in the system, which complicates their detection.

A distributed multilevel detection system was used for experiments [2]. implant in program were developed as part of each of the typical botnets. Then, the purpose of the experiments was to test the application of the method of botnet detection, the work of the classifier in the structure of the distributed system and determine the dependence of the percentage of detected botnet nodes on their representation by vectors containing program implant. The experiment was performed for the classifier without adding copies of the created botnets and with them, ie the test was performed without training the classifier on the created samples and with the preliminary assignment of samples by class. The duration of monitoring of the computer of the local network was 350 hours for each instance of the botnet of each of the two classifiers. The attack from the nodes of the botnet was not carried out. Bot network nodes worked only in the mode of computer control and support of the bot network structure through sent messages. Thus, for the distribute system components, the objects of study were launched in the computer processes and, accordingly, the construction of vectors on them. To conduct the experiment, botnets were selected that use the strategy of obtaining full control in the computer by activating their components implant in program. To perform the experiment by means of API monitoring in the com, vectors were obtained, which were alternately processed by the component classifier.

Implant in program of software used on local computer networks can cause significant harm to personal computer users, and especially to companies that operate computer networks and use specialized software.

Models of Implant in program software allow to include them in detection means after the corresponding formalization. The application of the developed models of implant in program in the distribute system [2] made it possible to improve the efficiency of detection of botnets of which they were part.

#### **References:**

1. DSTU 3396.2-97 Zaxy'st informaciyi. Texnichny'j zaxy'st informaciyi. Terminy' ta vy'znachennya [DSTU 3396.2-97 Information protection. Technical protection of information. Terms and definitions]
2. Nicheporuk A. Detecting Software Malicious Implant Based on Anomalies Research on Local Area Networks / A. Nicheporuk, V. Paiuk, B. Savenko, O. Savenko, O. Geidarova // Proceedings of the 1st International Workshop on Intelligent Information Technologies & Systems of Information Security. - Khmelnytskyi, Ukraine, June 10-12, 2020. - Pp.196-207.

<sup>1</sup> **Roman Ponomarenko**

PhP in Physics and Mathematics, assistant

<sup>2</sup> **Roman Tkachenko**

Student

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## METHOD OF PROCESSING COMPLEX OBJECTS BASED ON OBJECT-ORIENTED PROXY SYSTEM

A new method of processing objects of complex object-oriented systems is proposed based on objects' proxying technology.

Object proxying technology is designed to mediate access between the object and its environment [1]. It provides the use of a special Proxy object to wrap the proxied object and further intercept actions on it, to delegate the organization of access to it and its behavior Proxy-object.

Proxying technology is widely used in the design of browser applications and solutions (frameworks).

The syntax for using the Proxy object proxying class [2–4] (JavaScript programming language):

```
let proxy = new Proxy (target, handler);
```

The target parameter is an object (the object can also be a function, an array of objects, or a Proxy object) to be proxied, and the handler parameter is a wrapper function that specifies the behavior for the proxied target object.

The main idea of using the Proxy class is to create special "traps", each of which allows you to intercept the execution of internal methods (low-level methods that describe most actions on the object) over the original object, substituting each of their behavior. The internal methods include (the full list is given in [2, 3]):

- [[Get]] - read property;
- [[Set]] - record property;
- [[Delete]] - delete the property (call the delete operator);
- [[Call]] - call the object method;
- [[Construct]] - call the new operator, etc.

Accordingly, by extending the configuration of the Proxy object in the handle parameter, you can "set traps" for basic actions on the object, but subject to certain restrictions on their implementation.

The paper proposes a new method of processing objects of complex object-oriented systems based on the application of object-proxying technology. A library has been developed that organizes the transmission of messages by objects to basic Proxy-objects for further analysis of the received information. The purpose of this development is to simplify the processing of complex objects, tracking their condition at any time, to capture memory leaks.

Based on the developed method, is offered an algorithm for counting the number of duplications of objects of a complex system. The proposed algorithm also can be used for efficient and quick check objects for uniqueness under conditions of complex inhomogeneous system structure.

In fig. 1 presents a complex system with a heterogeneous structure of relationships between objects. Dotted arrows indicate the properties of objects (connections such as aggregation), solid arrows - the direction of transmission of messages by objects.

When using the proposed method, to successfully verify the uniqueness of a particular object in the system, it is sufficient to generate a request "up" to the appropriate Proxy with the appropriate message. Accordingly, the object Obj4 (Fig. 1), being non-unique, will generate a message twice. Proxy-object is enough to count the number of non-unique objects based on the received messages (without using the operations of direct analysis and comparison of system objects).

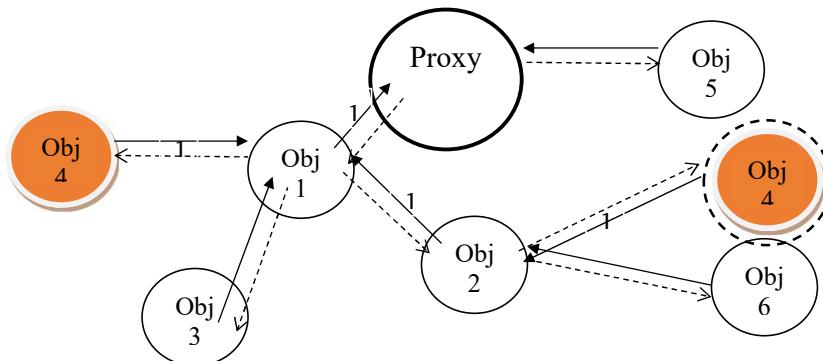


Figure 1 - Example of a complex system based on Proxy-objects

### References:

1. Van Cutsem T. and Miller MS Trustworthy Proxies. In: Castagna G. (eds) ECOOP 2013 - Object-Oriented Programming. ECOOP 2013. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, 2013, vol. 7920, pp.154-178. doi: 10.1007 / 978-3-642-39038-8\_7.
2. Proxy and Reflect. The Modern JavaScript Tutorial. Available: <<https://learn.javascript.ru/proxy>>.
3. Proxy. MDSN web docs. Available: <[https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global\\_Objects/Proxy](https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/Proxy)>.
4. Ferguson R. and Cirkel K. Working with Proxies. In: JavaScript Recipes. Apress, Berkeley, CA., 2017, pp.277-280. doi: 10.1007 / 978-1-4302-6107-0\_16.

<sup>1</sup>**Dmytro Sakharov**

MA in programming technologies of IoT

<sup>2</sup>**Olga Kravchenko**

Associate Professor at the Department of informational systems and technologies

<sup>1,2</sup>*Taras Shevchenko National University of Kyiv*

## **IoT SEISMOLOGICAL SITUATION MONITORING SYSTEM DEVELOPMENT WITH ONE OF THE REGIONS OF UKRAINE AS AN EXAMPLE**

**Keywords:** IoT, earthquake, early-warning system, MQTT.

Over the past few years, IoT has become one of the most important technologies of the 21st century. Nowadays it is possible to connect household items - kitchen appliances, machines, thermostats to the Internet to let the communication between people and things (devices) to be seamless.

With low-cost cloud computing, big data, analytics and mobile technology, physical devices can share and collect data with minimal human intervention. In this hyperconnected world, digital systems can record, monitor and regulate every interaction between connected things. The physical world corresponds to the digital world, and they tightly cooperate[1].

An earthquake is an accidental natural disaster that damages property and sometimes causes human deaths. We cannot stop or fight against earthquakes, but we could be notified in advance. Currently, there are numerous approaches that can be used to detect small earthquakes and vibrations. Precautions could be made before major oscillations occur inside the earth surface. Lifes can be saved by warning people in advance.

Significant seismic processes exist in western Ukraine, particularly in Zakarpatska oblast, where there have been earthquakes occurred with an intensity of 6-7 points on Richter scale. The epicenters of these earthquakes were in the areas of Svalyava, Dovhy, Teredasva, Mukachevo, Uzhhorod[2]. Therefore, the Zakarpatska oblast can serve as an example of the region for the creation of a system for monitoring seismic activity.

The most common type of data driven IoT system that reports a natural disaster is a tsunami detection system. Early warning of a disaster helps prevent many lost lives and reduces economic and material damage. Sensors are installed in coastal regions and transmit data to a remote data center. Such data centers perform data aggregation and operational procedure for detecting destructive events. Methods include GIS (geographic information system) for capturing, storing, analyzing and sending notifications.

In a seismic monitoring system, seismographs should be placed in the form of a grid covering the area of potential earthquake occurrence. The distance between the sensors will depend on the measurement accuracy of each sensor. The location in the form of a grid will allow to use the method of triangulation to determine the hypocenter of the earthquake. The implementation of the sensor is a moderately powerful

microcontroller, with GSM or satellite communication module. Two accelerometers can be used to measure seismic events i.e. to measure longitudinal and transverse acceleration. The design of the sensor must comply with standard of operation of the device outdoors with the influence of any atmospheric factors in temperate and cold macroclimatic areas.

The communication protocol between the sensors and the processing center is MQTT - a standardized, lightweight protocol that distinguish all devices as publishers and subscribers [3]. The publishers are sensors, the subscribers is a processing and analysis center and an smartphone or a web app for results representation.

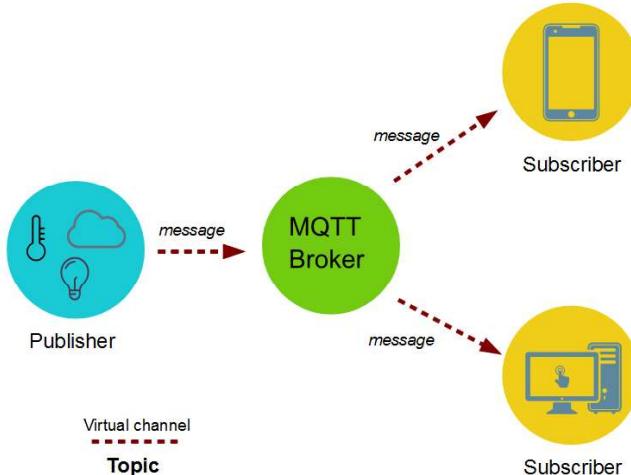


Figure 1 – протокол MQTT [4]

As of today, there is no earthquake early warning system in Ukraine. But with increasing seismic activity in some regions of our country, the implementation of such systems may become necessary to prevent deaths and significant economic and material damage.

#### References:

1. Oracle, What is the Internet of Things, 2019. URL: <http://bit.ly/2qadArR>
2. Kyiv Institute of Geophysics, Seismic activity of Ukraine, 2019. URL: [http://bit.ly/2qadArR](#)
3. Mqtt.org, MQTT Protocol, 2019. URL: <http://mqtt.org/>
4. F. Azzola, MQTT Protocol Tutorial: Technical description, 2019. URL: <http://bit.ly/2Y9iSQL>

<sup>1</sup> **Tatyana Selivorstova**

Candidate of technical sciences, Associate Professor

<sup>2</sup> **Serhii Kyrychenko**

Master's degree, specialty 122 "computer science"

<sup>3</sup> **Volodymyr Brodskyi**

Master's degree, specialty 122 "computer science"

<sup>4</sup> **Nataliia Tarkovska**

Master's degree, specialty 122 "computer science"

<sup>1-4</sup> *The National Metallurgical Academy of Ukraine*

<sup>2-4</sup> *EPAM Dnipro*

## RESEARCH OF APPLICATION METRICS DEPLOYED IN MONOLITHIC AND MICROSERVICE ARCHITECTURES

The rapid growth in the number of complex and extensive web-applications has led to the introduction of the latest communication technologies within the project. The monolithic architecture is logically replaced by microservices [1]. In this regard, the task of determining the technical and technological features of this technology becomes relevant.

It is known that the monolithic architecture involves the process of developing the application as a whole. Any changes, even the smallest ones, require significant restructuring and deployment of the entire application. Over time, it becomes increasingly difficult to maintain a clear and understandable modular structure, as changes in the logic of one module tend to affect the code of other modules. Unlike a monolithic architecture, a microservice architecture is an approach to building a server application as a set of almost unrelated services (Figure 1). Services are developed and deployed separately and independently of each other [2]. A separate process is required to run each service. Communication between processes is implemented using HTTP / HTTPS, WebSockets or some other protocols (Figure 2).

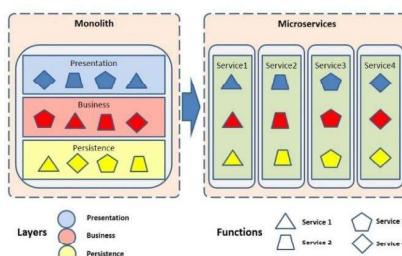


Figure 5 – The typical transition from a monolith to microservices [3]

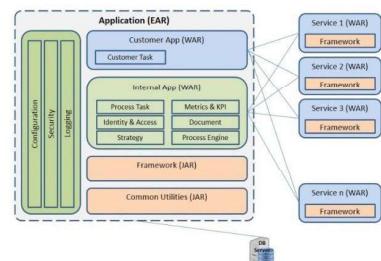


Figure 2 – Microservices landscape for the refactored monolith [3]

Implementation of microservice architecture involves the use of a stack of information technology and related software (Figure 3, [4]). In particular, one of the

options for deploying the Kubernetes cluster requires the use of Minikube, Kubectl, Docker. Minikube implements the local launch of a single-node Kubernetes cluster on a virtual or physical machine. Kubectl is a command line tool for managing Kubernetes clusters. Docker is a container runtime.

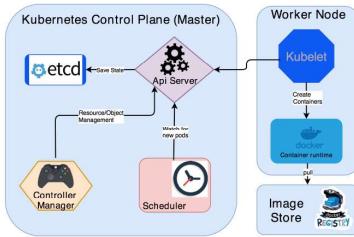


Figure 3 –Kubernetes architecture

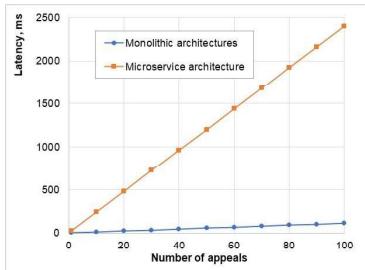


Figure 4 – Response time to the request

To test the technical and architectural parameters of the considered implementations, the application was deployed [5] on a virtual machine AWS EC2 – monolithic architecture and Kubernetes cluster on AWS EKS – microservice architecture. Consider the example of deploying a Kubernetes cluster on a Linux machine.

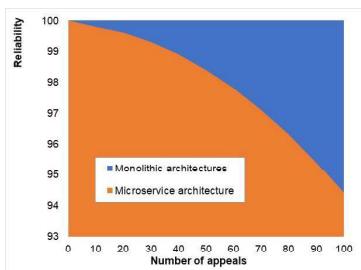


Figure 5 – Reliability

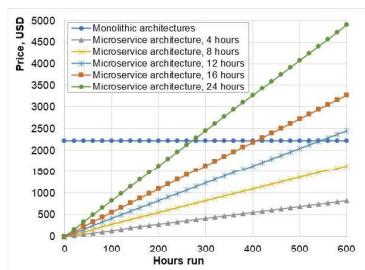


Figure 6 – Price scaling

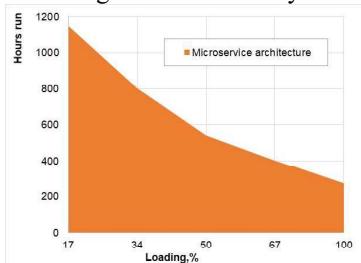


Figure 7 – Time for which the cost of operating microservice architecture becomes more expensive than operating monolithic

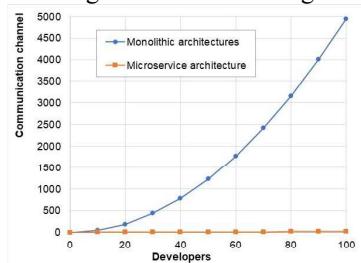


Figure 8 – Dependence of the number of communication links depending on the number of project participants

As part of the study, studies were conducted on the following characteristics [5, 6]:

- 1) Speed of response to a request (Figure 4);
- 2) Reliability (Figure 5);
- 3) Scaling (Figure 6, 7);
- 4) Team communication (Figure 8).

We investigated the configuration of the microservice and monolithic application architecture. We concluded that the microservice architecture has much greater fault tolerance and flexibility compared to the monolithic architecture. Applications deployed in a microservice architecture are much easier to scale and update than applications deployed in a monolithic architecture. Application developers with a microservice architecture communicate much more efficiently in smaller teams compared to the large teams required to develop applications with a monolithic architecture.

However, the monolithic application deployment architecture is simpler and great for deploying an application with unallocated workloads. Monolithic architecture responds much faster to requests and has fewer moving parts. Therefore, it may be easier to administer and maintain applications with monolithic architecture.

#### **References:**

1. Deploy to Kubernetes. Docker Documentation. Retrieved 1 November 2020, from <https://docs.docker.com/get-started/kube-deploy/>.
2. Kubernetes Certified Administration. Retrieved 2 November 2020, from <https://github.com/walidshaari/Kubernetes-Certified-Administrator>.
3. John, S. (2020). A Transition From Monolith to Microservices – DZone Microservices. dzone.com. Retrieved 1 November 2020, from <https://dzone.com/articles/a-transition-from-monolith-to-microservices>.
4. Abdelghani, R. (2019). Building and deploying a spring boot application to kubernetes. Medium. Retrieved 1 November 2020, from <https://medium.com/faun/building-and-deploying-a-spring-boot-application-to-kubernetes-12da35d1e2fc>.
5. WalidShaari/Kubernetes-Certified-Administrator. GitHub. Retrieved 1 September 2020, from <https://github.com/walidshaari/Kubernetes-Certified-Administrator>.
6. Kainz, A., Smith, T., & Fiskeaux, C. (2020). Microservices vs. Monoliths: An Operational Comparison – The New Stack. The New Stack. Retrieved 1 September 2020, from <https://thenewstack.io/microservices-vs-monoliths-an-operational-comparison/>.

**<sup>1</sup> Oleksandr Turovsky**

Candidate of Technical Sciences, Associate Professor

**<sup>2</sup> Valerii Kozlovskyi**

Doctor of Technical Sciences, Professor

**<sup>3</sup> Yurii Balaniuk**

Candidate of Technical Sciences, Associate Professor

**<sup>4</sup> Yuliia Boiko**

Candidate of Technical Sciences, Associate Professor

<sup>1</sup> State University of Telecommunications

<sup>2,3</sup> National Aviation University

<sup>4</sup> Taras Shevchenko National University of Kyiv

## **MINIMIZATION OF PHASE ERROR DISPERSION IN CLOSED TYPE PHASE SYNCHRONIZATION SYSTEMS IN CARRIER FREQUENCY TRACKING MODE**

Phase synchronization systems are widely implemented in various radio engineering devices of communication, radar and control technology, as well as in the device of accurate magnetic recording. It should be noted that the dynamics of the synchronization system, as an automatic control system, is directly related to the level of its astatism. That is, increasing the level of astatism of the synchronization system to the second and higher orders while ensuring a sufficient level of its speed is one of the ways to ensure the efficient operation of the system as a whole [1]. Research of the construction scheme of a closed synchronization system, composed of a combination of logic devices, in the direction of assessing the possibility of increasing the order of astatism of the system, reducing the variance of constant and transient errors in the process of carrier frequency tracking in the presence of noise in the communication channel. the problem to which this work is devoted.

Due to the fact that it is practically impossible to create an absolutely invariant system, in practice selective absolutely invariant systems are used, which allow to provide zero constant error under certain types of external influences.

In fact, the condition of selective absolute invariance is to require the equality of zero of the first few error rates of the system or, in other words, to require a certain order of astatism of the system [2].

On the other hand, ensuring a minimum phase filtering error is realized by automatically adjusting the specified process due to the presence in the control circuit of the feedback synchronization system. The specified feedback is a proportional controller, the task of which is to increase the accuracy of control commands, ie to increase the astatism of the control system of the carrier frequency synchronization scheme.

A characteristic feature of astatism is the presence of the structural scheme of the system K integrating links [1].

The order of astatism of a closed system in relation to the control effect is equal to the number of integrating links included in the feedback circuit between the points of application of this influence (input) and the error measurement point (output) and does

not depend on the number of integrating links included in the direct signal conversion circuit between these points [1].

In this case, as shown in [3], the choice of system parameters must be made under the condition of compromise adjustment.

A similar situation arises in the closed synchronization systems (CSS) of a higher order, the approach of the filter in a closed loop to the integrating second order also worsens the transition process [2].

In addition, as follows from the relations and conclusions given in [4], for CSS, by switching from PIF to IF it is possible to reduce the permanent error, but it is not possible to eliminate it completely, and at  $r = 2$  the system remains inoperable .

Obviously, the further development of the process of minimizing the phase error requires solving problems to change the scheme of construction of the synchronization system. One of the ways to solve this is the synthesis of combined synchronization systems in the direction of the introduction of open communication. This issue is sufficiently covered in the works [5].

The research was carried out in relation to the variant of the scheme of construction of the system of phase synchronization of the closed type at tracking of the carrier frequency in the conditions of presence of noise in the communication channel offered in article. It is shown that taking into account the additive Gaussian noise and instability of generators, the desire to minimize the phase error variance in the class of closed synchronization systems causes a deterioration of the system dynamics and does not increase the order of astigmatism. A further direction of research is to work on the synthesis of broken communication in combined synchronization systems against the background of additive Gaussian noise, taking into account the phase instability of the generators.

#### References:

1. Voronov A.A. Theory of automatic control. Ch. 1, 2. - M.: Energiya, 1986, 1987.
2. Misrihaiov M.Sh. Invariant control of multidimensional systems. Moscow: Energotomizdat, 2003. 236 p.
3. Kanatchikov A.A., Kulikov G.V. Research of possibilities of construction of system of clock synchronization on the basis of the autocorrelation demodulator for reception of signals with the minimum frequency manipulation, Scientific herald of MSTU GA. Series Radiophysics and Radio Engineering. 2010. № 152. P.11-15.
4. Oleksandr Turovsky, Yurii Khlaponin, Muhi-Aldin Hassan Mohamed et al, "Combined system of phase synchronization with increased order of astigmatism in frequency monitoring mode", CEUR Workshop Proceedings, 2020, Vol-2616, Session. 1, pp. 53–62. <http://ceur-ws.org/Vol-2616/paper5.pdf>
5. Turovsky O.L. Estimation of the possibilities of the combined synchronization system with open-link to minimize the dispersion of the phase error when tracking the carrier frequency under the conditions of the influence of additive noise. Technology audit and production reserves. 2020. Vol 3, No 4, pp. 16-22,. DOI: 10.15587/2312-8372.2020.210242.

<sup>1</sup> **Solomiia Tymoshchuk**

Bachelor degree Student of Information Systems and Technologies

<sup>2</sup> **Roman Ponomarenko**

Candidate of Physical and Mathematical Sciences, Assistant Professor of the Department of Information Systems and Technology.

<sup>1,2</sup> Taras Shevchenko National University of Kyiv

## THE RESEARCH AND DEVELOPMENT OF THE SOFTWARE TO SUPPORT THE EDUCATIONAL PROCESS IN HIGHER EDUCATION INSTITUTIONS

**Abstract.** The learning process management system has been developed using the cross-platform Java programming language in the Netbeans environment with a connection to the XAMPP web server. The efficiency of this system and the advantages of its use for teachers were investigated in order to speed up the learning process and its better organization in educational institutions with a large number of students.

**Keywords:** software, managing education system, web application.

With the increase in the number of students in the universities every year, there is a problem of convenient organization of data about students, their grades and the list of disciplines. That is why the question was raised about the development of a convenient learning management system for teachers with a special interface and database connection. The ability to adjust the data at any time according to changes in real time is another advantage of the system, which will significantly reduce time spent on the organizational processes.

This will allow the teacher's ability to directly access the graphical interface of the system and, if necessary, the database in which the information is stored. Such architectural decisions were made on the basis of an open-layer architecture of a multilevel model.[1]

Previously, in these systems, such users could not directly access the database, which slowed down the learning process and required more staff. The developed system solves this problem and during the work process the teacher does not need to apply to such bodies of the educational process as the dean's office or senior management, because all the information can be obtained directly in the system.

The aim of the work is to create an effective software system for managing the educational process in educational institutions, which will be a very effective multifunctional mechanism, in order to rationally use resources and improve the work of teachers and students in general.

The implementation of this system were possible only due to use of the Xampp multi-platform web server and the phpMyAdmin web application to administer the MySQL database.

Figure 1 shows the system's ability to adapt to each teacher who has their own username and password in advance. To add grades to the system, the teacher only needs to choose the discipline taught by the teacher and the student from the list. In Figure 1, you can also see that when entering data, the user ensures the confidentiality of information by transferring data automatically to the database table.

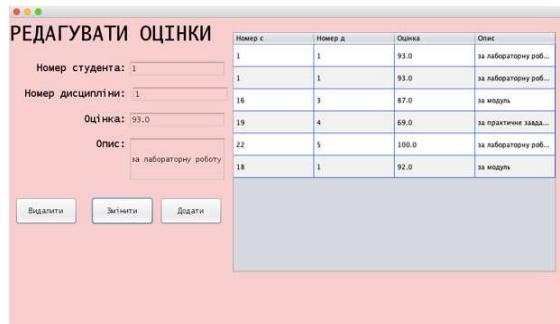


Figure 1 – Page of adding marks

Then, by connecting to the Xampp web server and the phpMyAdmin web application, that are extremely easy to install and low in memory, the teacher will be able to see the interface for accessing student data, their grades for the subject and the number of exact hours, dedicated to the particular subject.

Below in Figure 2 you can see how it looks like in the phpMyAdmin web application (i.e. in the database table):

Figure 2 – The view of structure of the database with the table names

**Conclusions.** The implementation of this system is very successful, because it simplifies the learning process and saves time for the teachers and staff in educational institutions. The development of software systems of this type guarantees the improvement of education in educational institutions as a whole, by solving organizational problems in this way.

#### References:

1. S. Ershov, R. Ponomarenko, The Architecture of the Software System of the Hierarchical Fuzzy Inference, National Academy of Sciences of Ukraine, Institute of Software Systems, 2018, pp. 99-108.
2. Modern Java Tools, The Software Engineering, Paradigms, Technologies. URL: [https://stud.com.ua/164909/informatika/suchasni\\_zasobi\\_java](https://stud.com.ua/164909/informatika/suchasni_zasobi_java)
3. The Content Management and Content Processing in the Intelligent Learning Systems. URL: <http://www.setlab.net/?view=dissertation-1-2>

## Authors

- |   |  |   |
|---|--|---|
| <p>Abduramanov Z. 13<br/>         Andrushchenko Y. 65<br/>         Antonevych M. 17<br/>         Astakhov A. 21<br/>         Babenko Y. 323<br/>         Bakhtina A. 260<br/>         Balaniuk Y. 378<br/>         Barannik V. 323<br/>         Bauzha O. 53, 281<br/>         Belfer R. 325<br/>         Bezlutska O. 247<br/>         Bielikov A. 363<br/>         Bila T. 126<br/>         Bobicev V. 29<br/>         Boiko A. 139<br/>         Boiko Y. 378<br/>         Bokan V. 121<br/>         Bondar T. 25<br/>         Bondarenko B. 40<br/>         Bondarenko M. 86<br/>         Braïlovskyi M. 107<br/>         Brodskyi V. 375<br/>         Buchyk S. 65, 68, 99, 110, 327<br/>         Buhriv A. 227<br/>         Buhyl B. 115<br/>         Bura Y. 124<br/>         Burmistenko O. 126<br/>         Bychkov O. 289<br/>         Bystrov A. 77<br/>         Cherevatov A. 329<br/>         Chioma E. 367<br/>         Dakhno N. 357, 350<br/>         Davydenko Y. 297<br/>         Dehtiarova Y. 191<br/>         Derevianchenko O. 27<br/>         Didyk A. 17<br/>         Dolgikh S. 128<br/>         Domanetska I. 251<br/>         Dubovyk T. 276<br/>         Dudnik A. 332<br/>         Dulich O. 57<br/>         Dvoretska S. 131<br/>         Dvoretskyi M. 131, 139<br/>         Fedorchenko I. 135<br/>         Fedorov E. 36, 367<br/>         Fedusenko O. 251</p> | <p>Fesenko A. 79<br/>         Gamotska S. 194<br/>         Gladka M. 335<br/>         Glavchey M. 29<br/>         Gnatyuk S. 79<br/>         Gradinari O. 253<br/>         Harko I. 262<br/>         Heidarova O. 369<br/>         Herasymenko O. 357<br/>         Hlavcheva Y. 29<br/>         Hnatiienko H. 25, 292, 339<br/>         Horban D. 227<br/>         Horban H. 131, 139<br/>         Horbas I. 256<br/>         Ilarionov O. 21, 33, 251, 354, 339<br/>         Ivanchenko O. 289<br/>         Ivanov I. 200<br/>         Ivanytska A. 317<br/>         Izmailova O. 346<br/>         Kadomskyi K. 31<br/>         Kambur M. 197<br/>         Kandyba I. 139<br/>         Kanishcheva O. 29<br/>         Kashtalian A. 71<br/>         Kazantsev A. 92<br/>         Kerdun N. 231<br/>         Kharchenko A. 135<br/>         Khlevna I. 124, 150, 274<br/>         Khlevnyi A. 121, 143, 172<br/>         Khrinenko A. 320<br/>         Khvorostina Y. 278<br/>         Kiktev N. 145<br/>         Klimenkova N. 229<br/>         Klochko V. 83<br/>         Klyushin D. 295<br/>         Kobylchuk M. 332<br/>         Kolesnikova K. 172, 181<br/>         Kolomiiets A. 212, 237<br/>         Koltsov D. 74<br/>         Kondrashov K. 174<br/>         Kondratuk I. 344<br/>         Kondruk N. 148<br/>         Kononova I. 301<br/>         Kopp A. 163<br/>         Kostikov M. 335<br/>         Kostiv O. 115</p> | <p>Kotov M. 113<br/>         Koval B. 143, 150<br/>         Kovalenko A. 200<br/>         Kovalenko I. 297<br/>         Kovbas Y. 346<br/>         Kozlovskyi V. 378<br/>         Krasnovidov S. 49<br/>         Krasovska H. 33<br/>         Kravchenko O. 352, 373<br/>         Kravchenko Y. 350<br/>         Kredentser B. 301<br/>         Kucherenko R. 352<br/>         Kucherova H. 42<br/>         Kudin V. 339, 354<br/>         Kulko A. 110<br/>         Kuzminova K. 153<br/>         Kyrychenko S. 375<br/>         Latysheva T. 202<br/>         Lavrinovich V. 357<br/>         Lendiel T. 145<br/>         Leshchenko A. 247<br/>         Leshchenko O. 350, 357<br/>         Linder Y. 153<br/>         Lisnevskyi R. 335<br/>         Loik O. 204<br/>         Los V. 42<br/>         Lub P. 204<br/>         Lukova-Chuiko N. 77, 79, 83, 95<br/>         Makarenko A. 304<br/>         Makhachashvili R. 260<br/>         Makhachashvili R. 264<br/>         Malyar M. 306<br/>         Merkulova K. 289<br/>         Mezentseva O. 157, 161, 165, 184<br/>         Mikhieiev V. 157<br/>         Minaeva J. 159<br/>         Mironova V. 262<br/>         Mogylevych D. 301<br/>         Mohylevych V. 301<br/>         Morozov V. 191, 200, 206, 208, 239,<br/>         Morze N. 264, 266<br/>         Mudra A. 161<br/>         Mulesa O. 128</p> |
|---|--|---|

- Musienko A. 314  
 Myroshnychenko Y. 360  
 Nakonechnyi V. 86, 88, 113, 363  
 Naumenko A. 212  
 Nazarchuk I. 33  
 Nemchenko K. 365  
 Neskorodieva T. 36, 367  
 Nezdoliy Y. 131  
 Nicheporuk A. 92, 105  
 Nikolaiev A. 27  
 Nikolyuk P. 367  
 Oberemok I. 214  
 Oberemok N. 214  
 Ocheretin D. 42  
 Olinyuk A. 135  
 Olimpiyeva Y. 314  
 Onyshchenko A. 354  
 Orlovskyi D. 163  
 Osypenko V. 145  
 Paiuk V. 369  
 Palageychenko D. 327  
 Paliy S. 329, 344, 360, 365  
 Papirna H. 79  
 Parkhomenko I. 74, 107  
 Parkhomenko M. 323  
 Patsyuk M. 233  
 Paziak A. 247  
 Petrychko M. 170  
 Pliushch O. 363  
 Pokutnia D. 332  
 Polishchuk A. 306  
 Polishchuk V. 306  
 Ponomarenko R. 270, 371, 380  
 Ponomarov S. 95  
 Proskurin M. 208  
 Provotar A. 272  
 Pursky O. 276  
 Pyroh M. 262  
 Raichuk I. 217  
 Revenchuk I. 53  
 Riabov O. 274  
 Rimek A. 292  
 Rudenko A. 225  
 Rudenko V. 165  
 Rusyn V. 97, 309  
 Sakharov D. 373  
 Saman M. 135  
 Sambas A. 97, 309  
 Samoilenco M. 272  
 Samokhvalov Y. 40  
 Samonenko A. 221  
 Samoylenko H. 276  
 Savchenko R. 51  
 Savenko B. 105  
 Savenko O. 92  
 Sazonov A. 223  
 Seidametova Z. 13  
 Selivanova A. 276  
 Selivorstova T. 375  
 Semenikhina O. 278  
 Semenist I. 260  
 Semenov V. 311  
 Semerikov S. 42  
 Shabatskaya S. 143  
 Shamonia V. 278  
 Sharkadi M. 46  
 Sharybura A. 204  
 Shelest T. 168, 225  
 Shestak Y. 110  
 Shpur O. 115  
 Shtovba S. 170  
 Shulgın S. 323  
 Shved A. 99, 297  
 Slipachuk L. 102  
 Smishchenko D. 202  
 Snytyuk V. 17  
 Sobchuk A. 314  
 Sobchuk V. 314  
 Solodovnyk V. 113  
 Solomko M. 317  
 Soroka P. 49, 51, 194  
 Statsenko D. 126  
 Statsenko V. 126  
 Stepanenko A. 135  
 Stepanenko O. 284  
 Steshenko G. 227, 241  
 Stetsiuk M. 105  
 Strutynska O. 266  
 Suprun O. 229  
 Sus B. 53, 281  
 Symonychenko A. 68  
 Symonychenko Y. 68  
 Taborovskiy A. 172  
 Tarkovska N. 375  
 Tereshchenkova O. 174  
 Timinsky A. 231, 233  
 Timrova Y. 227  
 Tkachenko R. 371  
 Tmienova N. 57  
 Toliupa S. 102, 107, 110, 113  
 Tolstokorova A. 350  
 Tregubenko I. 59  
 Triska M. 204  
 Tsykun V. 121  
 Tukalo S. 115  
 Turowsky O. 378  
 Tymoshchuk S. 380  
 Valieva N. 13  
 Vavilenkova A. 178  
 Vedel Y. 311  
 Veres M. 153, 272  
 Vlasenko O. 251, 339  
 Vlasiuk S. 344  
 Voitenko I. 88  
 Vostrov G. 320  
 Yakymenko D. 59  
 Yas V. 237  
 Yefremov H. 181  
 Yehorchenkov O. 186, 197, 221  
 Yehorchenkova N. 223  
 Yeremieieva V. 168  
 Yeshchenkov V. 184  
 Yurchenko A. 278  
 Yurchenko Y. 276  
 Yurzhenko A. 247  
 Zagorodnyuk S. 281  
 Zhabska Y. 289  
 Zharikova A. 239  
 Zhovtukhin D. 186  
 Zhurakovskiy B. 107  
 Zinchenko V. 284  
 Zubets D. 241  
 Zubyk L. 317  
 Zubyk Y. 317

# PARTNERS OF THE CONFERENCE

## AM INTEGRATOR



AM INTEGRATOR LLC is one of the leading TOP 3 IT operators in Ukraine, being a supplier and integrator of a wide range of solutions in the field of modern IT infrastructure, cloud systems, multimedia solutions, engineering infrastructure, business applications and complex business security systems for medium and large enterprises, and also for the public sector.

<https://amintegrator.com>

**GlobalLogic**

**GlobalLogic®**

GlobalLogic is a digital product engineering company. They integrate experience design and complex engineering to help our clients imagine what's possible and accelerate their transition into tomorrow's digital businesses.

For nearly 20 years, GlobalLogic has partnered with businesses across every major industry to make amazing products and connect the dots between people, products, and business opportunities. Now, in an increasingly digital world, they place the consumer at the center of every interaction. The company show businesses how they can better engage their consumers, innovate within predictable budgets, and bring the next generation of digital products and services to market in the shortest possible time.

#### GlobalLogic Company Facts At-A-Glance

- Founded in 2000
- Global Headquarters in Silicon Valley (San Jose, California)
- 16,000+ Employees in 14 countries
- 300+ Amazing Active clients
- 70+ Private label customer labs
- 27 Product Engineering centers
- 9 sales offices
- 3 Method Design Studios (San Jose, New York, London)
- 200+ UX/UI designers and developers
- Privately owned and funded by CPP Investment Board and Partners Group.

<https://www.globallogic.com/>

## Skyglyph



Skyglyph is a dynamic start-up consisting of creative and industry savvy aerial survey enthusiasts based in Sofia, Bulgaria encompassing an established business network in North America, Eastern Europe, and Southeast Asia.

Skyglyph renowned specialization is the development of cloud-based software and industrial surveillance solutions that transform drones into concise instruments for deployment in a wide range of business use cases.

Skyglyph's cloud-based platform simplifies the scouting process and management of flight control, asset mapping, site monitoring, data transfer and visual result interpretation. We excel in the professional interpretation of images via the generation of comprehensive analytical and recommended actions reports that can be applied to a variety of survey use case requirements.

Our solutions are designed to bridge the gap between drone technologies and user requirements that deliver the master results that you expect.

Skyglyph eliminates the gap between the technical aspects of aerial surveying and the “plug and play” needs of its users. The results are simple and concise while at the same time providing an intrinsically valuable asset for site analysis.

<https://skyglyph.com/>  
Telegram: Skyglyph

## S&T Ukraine



S&T Ukraine was established in 1993 and today they are a leading Ukrainian integrator of telecommunication and corporate information systems, provider of IT services.

S&T Ukraine is a company of the S&T AG corporation. With approximately 4,300 employees, S&T AG is the leading provider of comprehensive consulting, outsourcing, systems integration and IT services in Central and Eastern Europe as well as in the German speaking region. The S&T AG operates in more than 25 countries. S&T's core business focuses on the consulting, development, implementation and operation of IT solutions, processes and systems.

The company's business areas:

- IT Infrastructure
- Networks
- Information Security
- IT Services
- Specialized Solutions

S&T Ukraine became the first company in Ukraine to meet all Cisco requirements for Cisco Gold Certified Partner status (2003). The company has successfully recertified for this status in 2019, sixteenth straight time. S&T Ukraine also became the first company in Ukraine to receive all new Cisco Advanced Specializations (2008) and all new Cisco Advanced Architecture Specializations (2013). The company provides technical support in 24x7x365 mode within the framework of partnership with Cisco.

The Quality Management System of S&T Ukraine is certified against the ISO 9001-2015 International Standard.

S&T AG serves companies along the entire IT value chain in the fields of manufacturing, commerce, telecommunications, public services, financial services as well as energy and utility companies. They have completed hundreds of complex projects for customers that are leading Ukrainian companies in financial, telecommunication, manufacturing, retail and utility industries.

## Наукове видання

Інформаційні технології та взаємодії (сателітна): Матеріали конференції, 04 грудня 2020 р., Київ, Україна

Information Technology and Interactions (Satellite): Conference Proceedings, December 04, 2020, Kyiv, Ukraine

В авторській редакції

Редактор: Снитюк В.С.

Верстка: Гамоцька С.Л.

Обкладинка: Дубницька А.С.

Підп. до друку 18.12.2020 р. Формат 60x84<sup>1</sup>/16.  
Папір офсетний. Друк офс. Гарнітура Times New Roman  
Умовн. друк. арк. 24,8. Обл.-вид. арк. 23,8

Видавництво «Стилос»  
04070, Київ-70, Контрактова пл. 7  
Свідоцтво Держкомінформу України  
(серія ДК №150 від 16.08.2000 р.)  
Тел. (044) 467-51-89