

Харківський національний
університет радіоелектроніки

Kharkiv National
University of Radio Electronics

Державне підприємство
"Південний державний
проектно-конструкторський
та науково-дослідний інститут
авіаційної промисловості"

State Enterprise
"Southern National Design
&
Research Institute
of Aerospace Industries"

**СУЧАСНИЙ СТАН
НАУКОВИХ
ДОСЛІДЖЕНЬ
ТА ТЕХНОЛОГІЙ
В ПРОМИСЛОВОСТІ**

**INNOVATIVE
TECHNOLOGIES
AND
SCIENTIFIC SOLUTIONS
FOR INDUSTRIES**

№ 4 (26), 2023

No. 4 (26), 2023

*Щоквартальний
науковий
журнал*

*Quarterly
scientific
journal*

Харків
2023

Kharkiv
2023

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор
Бодянский Євгеній Володимирович,
д-р техн. наук, професор

Заступник головного редактора
Айзенберг Ігор Наумович,
канд. техн. наук, професор (США);
Шекер Серхат,
д-р техн. наук, професор (Туреччина)

Члени редколегії:

Артиух Роман Володимирович, канд. техн. наук, професор;
Бабенко Віталіна Олексіївна, д-р екон. наук, канд. техн. наук, професор;
Безкоровайний Володимир Валентинович, д-р техн. наук, професор;
Гасімов Юсіф, д-р мат. наук, професор (Азербайджан);
Гопсінко Віктор, д-р техн. наук, професор (Латвія);
Го Цян, д-р техн. наук, професор (КНР);
Джавад Хамісабаді, канд. техн. наук, доцент (Іран);
Зайцева Єлена, д-р техн. наук, професор (Словаччина);
Зачко Олег Богданович, д-р техн. наук, доцент;
Коваленко Андрій Анатолійович, д-р техн. наук, професор;
Костін Юрій Дмитрович, д-р екон. наук, професор;
Левашенко Віталій, д-р техн. наук, професор (Словаччина);
Лемешко Олександр Віталійович, д-р техн. наук, професор;
Малєєва Ольга Володимирівна, д-р техн. наук, професор;
Момот Тетяна Валеріївна, д-р екон. наук, професор;
Музыка Катерина Миколаївна, д-р техн. наук, професор;
Назарова Галина Валентинівна, д-р екон. наук, професор;
Невлюдов Ігор Шакірович, д-р техн. наук, професор;
Опанасюк Анатолій Сергійович, д-р фіз.-мат. наук, професор;
Павлов Сергій Володимирович, д-р техн. наук, професор;
Перова Ірина Геннадіївна, д-р техн. наук, доцент;
Петленков Едуард, канд. техн. наук (Естонія);
Петришин Любомир, д-р техн. наук, професор (Польща);
Рубан Ігор Вікторович, д-р техн. наук, професор;
Семенець Валерій Васильович, д-р техн. наук, професор;
Семенов Сергій, д-р техн. наук, професор (Польща);
Сетлак Галина, д-р техн. наук, професор (Польща);
Терзіян Ваган Якович, д-р техн. наук, професор (Фінляндія);
Телетов Олександр Сергійович, д-р екон. наук, професор;
Тімофєєв Володимир Олександрович, д-р техн. наук, професор;
Філатов Валентин Олександрович, д-р техн. наук, професор;
Чумаченко Ігор Володимирович, д-р техн. наук, професор;
Чухрай Наталія Іванівна, д-р екон. наук, професор;
Юн Джин, канд. фіз.-мат. наук, професор (КНР);
Ястремська Олена Миколаївна, д-р екон. наук, професор.

ЗАСНОВНИКИ

Харківський національний університет радіоелектроніки,
Державне підприємство "Південний державний
проектно-конструкторський та науково-дослідний
інститут авіаційної промисловості"

АДРЕСА РЕДАКЦІЇ:

Україна, 61166, м. Харків, проспект Науки, 14
Інформаційний сайт: <http://itssi-journal.com>
<https://journals.uran.ua/itssi>
E-mail редколегії: journal.itssi@gmail.com

EDITORIAL BOARD

Editor in Chief
Bodyanskiy Yevgeniy,
Dr. Sc. (Engineering), Professor, Ukraine

Deputy Chief Editor
Igor Aizenberg,
PhD (Computer Science), Professor (United States)
Serhat Seker,
Dr. Sc. (Engineering), Professor (Turkey)

Editorial Board Members:

Artiukh Roman, PhD (Engineering Sciences) (Ukraine);
Babenko Vitalina, Dr. Sc. (Economics); PhD (Engineering Sciences), Professor (Ukraine);
Bezkorovainyi Volodymyr, Dr. Sc. (Engineering), Professor (Ukraine);
Gasimov Yusif, Dr. Sc. (Mathematical), Professor (Azerbaijan);
Gopeyenko Victors, Dr. Sc. (Engineering), Professor (Latvia);
Guo Qiang, Dr. Sc. (Engineering), Professor (P.R. of China);
Javad Khamisabadi, PhD (Industrial Management), Associate Professor (Iran);
Zaitseva Elena, Dr. Sc. (Engineering), Professor (Slovak Republic);
Zachko Oleh, Dr. Sc. (Engineering), Associate Professor (Ukraine);
Kovalenko Andrey, Dr. Sc. (Engineering), Professor, (Ukraine);
Kostin Yuri, Dr. Sc. (Economics), Professor (Ukraine);
Levashenko Vitaly, Dr. Sc. (Engineering), Professor (Slovakia);
Lemeshko Oleksandr, Dr. Sc. (Engineering), Professor (Ukraine);
Malyeyeva Olga, Dr. Sc. (Engineering), Professor (Ukraine);
Momot Tetiana, Dr. Sc. (Economics), Professor (Ukraine);
Muzyka Kateryna, Dr. Sc. (Engineering), Professor (Ukraine);
Nazarova Galina, Dr. Sc. (Economics), Professor (Ukraine);
Nevliudov Igor, Dr. Sc. (Engineering), Professor (Ukraine);
Opanasyuk Anatoliy, Dr. Sc. (Physical and Mathematical), Professor (Ukraine);
Pavlov Sergii, Dr. Sc. (Engineering), Professor (Ukraine);
Perova Iryna, Dr. Sc. (Engineering), Associate Professor (Ukraine);
Petlenkov Eduard, PhD (Engineering Sciences) (Poland);
Petryshyn Lubomyr, Dr. Sc. (Engineering), Professor (Poland);
Ruban Igor, Dr. Sc. (Engineering), Professor, (Ukraine);
Semenets Valery, Dr. Sc. (Engineering), Professor, (Ukraine);
Semenov Serhii, Dr. Sc. (Engineering), Professor (Poland);
Setlak Galina, Dr. Sc. (Engineering), Professor (Poland);
Terziyan Vagan, Dr. Sc. (Engineering), Professor, (Finland);
Teletov Aleksandr, Dr. Sc. (Economics), Professor (Ukraine);
Timofeyev Volodymyr, Dr. Sc. (Engineering), Professor (Ukraine);
Filatov Valentin, Dr. Sc. (Engineering), Professor (Ukraine);
Chumachenko Igor, Dr. Sc. (Engineering), Professor (Ukraine);
Chukhray Nataliya, Dr. Sc. (Economics), Professor (Ukraine);
Yu Zheng, PhD (Physico-Mathematical Sciences), Professor (P.R. of China);
Iastremaska Olena, Dr. Sc. (Economics), Professor (Ukraine).

ESTABLISHERS

Kharkiv National University of Radio Electronics,
State Enterprise "National Design & Research Institute
of Aerospace Industries"

EDITORIAL OFFICE ADDRESS:

Ukraine, 61166, Kharkiv, Nauka Ave, 14
Information site: <http://itssi-journal.com>
<https://journals.uran.ua/itssi>
E-mail of the editorial board: journal.itssi@gmail.com

Журнал включено до Переліку наукових фахових видань України, в яких можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора і кандидата наук наказом Міністерства освіти і науки України від 16.07.2018 №775 (додаток 7).

Свідоцтво про державну реєстрацію журналу Серія КВ № 22696-12596Р від 04.05.2017 р.

© Харківський національний університет радіоелектроніки,
Державне підприємство "Південний державний проектно-конструкторський
та науково-дослідний інститут авіаційної промисловості"

ЗМІСТ

Інформаційні технології

- 5 **Бакурова А. В., Савранська А. В., Терещенко Е. В., Широкопад Д. В., Шевчук М. В.**
Аналіз задачі вибору критеріїв у моделях диверсифікації (eng)
- 16 **Барковська О. Ю., Гаврашенко А. О.**
Нейромережна модель у задачах оброблення та аналізу аудіофайлів (eng)
- 24 **Гавриленко О. В., Мякий М. Ю.**
Прогнозування курсу криптовалюти на основі ранжування експертних оцінок. (eng)
- 33 **Гринченко М. А., Роговий М. А.**
Модель ідентифікації завдань спринту проєкту на основі їх опису (eng)
- 45 **Ляшик В. А., Шубін І. Ю.**
Метод логічних мереж для моделювання систем адаптивного тестування знань (ua)
- 58 **Порошенко А. І., Коваленко А. А.**
Метод передачі аудіосигналу у мережній системі аудіоаналітики (eng)
- 68 **Романків Н. Д., Ситніков Д. Е.**
Аналіз і вибір методів кастомізації *SaaS*-рішень,
побудованих за допомогою технологій *cloud-native* (ua)
- 78 **Федорченко В. М., Шматко О. В., Михайленко І. В., Третяк В. Ф., Коломійцев О. В.**
Інтеграція аналітичних статистичних моделей, послідовного аналізу закономірностей
та теорії нечітких множин для розширеного оцінювання надійності мобільних застосунків (eng)
- 87 **Чернишов Д. В., Ситніков Д. Е.**
Бінарна класифікація на основі поєднання теорії приблизних множин і дерев рішень (ua)
- 95 **Шаповалов В. Б., Стрижак О. Е.**
Трансдисциплінарна інтеграція знань у середовищі віртуального STEM-центру (eng)
- 108 **Шовковий Є. І., Гриньова О. Є., Удовенко С. Г., Чала Л. Є.**
Система автоматичного сурдоперекладу
з використанням нейромережних технологій та 3D-анімації (eng)
- 145 **Бондаренко К. О.**
Аналіз і вибір релевантної метрики виявлення мережних аномалій (ua)

Інженерія та промислові технології

- 122 **Шкіль О. С., Рахліс Д. Ю., Філіпенко І. В., Корнієнко В. Р.**
Проектування та самодіагностика кіберфізичних пристроїв керування на платформі *SoC* (ua)

Телекомунікаційні системи та комп'ютерні мережі

- 135 **Невлюдов І. Ш., Вжеснєвський М. О., Ромашов Ю. В., Чала О. О.**
Математичне моделювання мехатронних шатлів як об'єктів автоматизації
для багаторівневих систем внутрішньоскладської логістики (eng)
- 158 **Алфавітний покажчик**

За достовірність викладених фактів, цитат та інших відомостей відповідальність несе автор

CONTENTS

Information Technology

- 5 **Bakurova A., Savranska A., Tereschenko E., Shyrokorad D., Shevchuk M.**
Analysis of the criteria selection problem in diversification models (eng)
- 16 **Barkovska O., Havrashenko A.**
Analysis of the influence of selected audio pre-processing stages
on accuracy of speaker language recognition (eng)
- 24 **Gavrilenko O., Miahkyi M.**
Forecasting the cryptocurrency exchange rate based on the ranking of expert opinions (eng)
- 33 **Grinchenko M., Rohovyi M.**
A model for identifying project sprint tasks based on their description (eng)
- 45 **Liashyk V., Shubin I.**
The Method Of Logical Networks For Modeling Adaptive Knowledge Testing Systems (ua)
- 58 **Poroshenko A., Kovalenko A.**
Audio signal transmission method in network-based audio analytics system (eng)
- 68 **Romankiv N., Sytnikov D.**
Analysis and selection of methods for customizing *SaaS* solutions built
using *Cloud-Native* technologies(ua)
- 78 **Fedorchenko V., Shmatko O. Mykhailenko I., Tretiak V, Kolomiitsev O.**
Integrating analytical statistical models, sequential pattern mining,
and fuzzy set theory for advanced mobile app reliability assessment (eng)
- 87 **Chernyshov D., Sytnikov D.**
Binary classification based on a combination of rough set theory and decision trees (ua)
- 95 **Shapovalov V., Stryzhak O.**
Transdisciplinary integration of knowledge in the environment of a virtual STEM center (eng)
- 108 **Shovkovyi Y., Grinyova O., Udovenko S., Chala L.**
Data structures for deductive simulation of HDL conditional operators (eng)
- 145 **Bondarenko K.**
Analysis and selection of relevant network anomaly detection metrics (ua)

Engineering & Industry Technologies

- 122 **Shkil A., Rakhlis D., Filippenko I., Korniienko V.**
Design and self-diagnostics of cyberphysical control devices on *SOC* platform (ua)

Telecommunication Systems & Computer Network

- 135 **Nevliudov I., Vzhesnievskiy M., Romashov Y., Chala O.**
Mathematical modeling of mechatronic shuttles as automation objects
for multilevel systems of intra-warehouse logistics (eng)
- 158 **Alphabetical index**

The author is responsible for the accuracy of the facts, quotations and other information

A. BAKUROVA, A. SAVRANSKA, E. TERESCHENKO, D. SHYROKORAD, M. SHEVCHUK

ANALYSIS OF THE CRITERIA SELECTION PROBLEM IN DIVERSIFICATION MODELS

The digitalization of the economy reduces the cost of doing business by automating the relevant processes, but any transformation creates new risks and economic instability. Economic instability leads to a drop in the standard of living and, as a result, negatively affects the activities of trade enterprises. Small and medium businesses are especially sensitive to any changes. The decrease in demand for most everyday goods has a painful effect on the activities of small and medium-sized businesses and leads to the emergence of new risks. These risks have a significant impact on reducing the profitability of enterprises. Therefore, it is important for each enterprise to diversify the activities of the enterprise, which includes the expansion of the product range, the reorientation of sales markets and the optimal distribution of goods between divisions of one enterprise. **The subject** of the article is multi-criteria models of a diversified portfolio that minimize the risks that arise in the era of the digital economy when managing retail chains. To formalize the problem, five models are proposed that differ in vector objective functions, both in the quantity and quality of the selected criteria. **The aim of the work** is to analyze the problem of choosing criteria in the corresponding multicriteria or vector diversification problems. The article examines the advantages of introducing an additional criterion of entropy maximization into the criteria of the classical two-criteria model of portfolio theory, which characterizes the degree of diversity of the portfolio composition. A complex combination of methods of classical portfolio theory and multicriteria optimization is applied. **The results** include a comparison of three methods for solving the following problems: criteria convolution, successive concessions, and computer simulation of the Pareto set. **Conclusions:** the results obtained will be useful for automating the risk management of retail chains. The practical value is that the obtained results of real data for the network have demonstrated the possibility of using the developed tool for automatic allocation of resources in the form of pareto-optimal portfolios in order to minimize risks.

Keywords: computer simulation; multicriteria problem; optimal portfolio problem; convolution of criteria; method of successive concessions; Pareto set; entropy.

Introduction

The transition to a digital economy and digital trade in the world has had a significant impact on the Ukrainian economy. Large-scale digital transformation was accelerated by the COVID-19 pandemic, which affected consumer behavior and changed the rules for business operations. The digitalization of the economy reduces the cost of doing business by automating relevant processes, but any transformation generates new risks and economic instability. Economic instability leads to a decline in living standards and, as a result, negatively affects the activities of trading companies, especially in the context of Ukraine's recovery [1]. Small and medium-sized businesses are very sensitive to any changes. The decline in demand for most everyday goods has a painful impact on the activities of small and medium-sized businesses and leads to the emergence of new risks. These risks have a significant impact on the decline in the profitability of enterprises. Therefore, it is important for every enterprise to diversify its activities, which involves expanding the product range, reorienting sales markets, and optimizing the distribution of goods between structural units.

The portfolio theory has high potential in many areas. Currently, different approaches to choosing optimality criteria and solution algorithms are used to determine the best strategy. Both exact algorithms for finding optimal solutions and approximate algorithms of three categories have been developed: heuristic, metaheuristic, and hyperheuristic [2]. The introduction of machine learning technologies is relevant [3, 4]. Paper [3] uses modern advances in neural network architecture for efficient convex optimization for risk-based portfolio performance: minimum variance, maximum diversification, and equal risk contribution [4]. Further development of the use of neural networks for optimal portfolio construction tasks is covered in [5-7]. For example, paper [5] presents an approach to portfolio construction strategy based on a hybrid machine learning model that combines a convolutional neural network and bidirectional long-term short-term memory with reliable input characteristics derived from the Huber location for stock forecasting and the Markovian mean-variance model for the proposed optimal portfolio construction. Paper [6] considers maximizing portfolio returns using *Reinforcement Learning*, taking into account dynamic risks corresponding to market conditions, by dynamically

rebalancing the portfolio. Recurrent networks, namely the *Deep Belief-Recurrent Neural Network*, together with a hybrid algorithm called *HH-DHO*, are also used for portfolio forecasting tasks [7].

Several studies contain the results of extending methods for formalizing optimal portfolio problems [8, 9]. In [9], a method is proposed to reduce the problem to a linear minimization program subject to a linear constraint of an arbitrary positively homogeneous convex functional, the dual set of which is given by linear inequalities, which allows us to expand the class of linear problems of portfolio theory

$$F = (F_1(x), F_2(x), \dots, F_N(x)), \quad (1)$$

which is defined on the admissible set X . One of the options for formalizing the set of alternatives is the Pareto set.

The Pareto set \tilde{x} consists of non-dominant solutions \tilde{x} , for each of which there is an unacceptable solution $x^* \in X$ that satisfies the inequality

$$F_i(x^*) \leq F_i(\tilde{x}), \quad (2)$$

where $i = 1, 2, \dots, N$, among which at least one is strict.

It is important to distinguish between two types of formulations of multicriteria problems, namely, an individual problem and a mass problem [8]. An individual problem has fixed parameters of the vector objective function $[F = (F_1, F_2, \dots, F_N)]$, and a system of constraints. In the formulation of a mass problem, which has a common name, some parameters are not fixed and are given by signs. For example,

$$\Lambda_N = \left\{ \lambda = (\lambda_1, \dots, \lambda_N) : \sum_{v=1}^N \lambda_v = 1, \lambda_v = 1, \lambda_v > 0, v = 1, 2, \dots, N \right\}.$$

Consider an individual problem with N criteria to be maximized and defined on the set of admissible solutions $s \in X = \{x\}$. Let us denote the set of alternatives to this problem by X^* , $X^* \subseteq X$. If for each element $x^* \in X^*$ there is a vector $\lambda^* \in \Lambda_N$ corresponding to the equality $F^{\lambda^*}(x^*) = \max_{x \in X} F^{\lambda^*}(x)$, then the problem of finding $SA \ X^*$ is considered to be solved using a linear convolution algorithm. If the solution determined in this way is characteristic of all individual problems of the mass problem, then for each of them it is possible to find the MA using convolution algorithms. This problem is unsolvable by convolution algorithms if, for the problem under consideration, there is an individual problem from $SA \ X^*$ containing such an element $x^* \in X^*$

the classical two-criteria Markowitz portfolio problem with a vector objective function $F = (R, D)$ is a mass problem.

Methods for solving multi-criteria (vector) problems are based on different approaches. One of them is the construction of a generalized criterion that aggregates the vector of criteria of the VOF (1). For example, the method of linear or multiplicative convolution of criteria, the majority criterion, the geometric criterion based on immersion in a metric space. Another approach is to determine the lexicographic order of the criteria. Thus, attempts are being made to move from a multi-criteria problem to a single-criteria problem or a sequence of single-criteria problems with certain constraints. The choice of a solution strategy has an impact on the solution obtained, since the previous constraints on the solution of the problem are changed and new constraints are added. Not all methods can guarantee an acceptable solution. In particular, for certain problems, the linear convolution method does not allow to obtain a Pareto set. In this regard, the problem of solving multi-criteria problems using linear convolution of criteria (LCC) is considered separately [14].

Let's consider this algorithm. The linear convolution algorithms are based on the fact that, given a positively definable VOF, the element $x \in X$ maximizing (minimizing) the linear convolution of the criteria

$$F^\lambda(x) = \sum_{v=1}^N \lambda_v F_v(x), \quad (3)$$

is pareto-optimal. Here the vector $\lambda \in \Lambda_N$, where

on which the convolution extremum $F^\lambda(x) \forall \lambda \in \Lambda_N$ is not reached, i.e., for any $\lambda \in \Lambda_N$ the inequality $F^\lambda(x^*) < \max_{x \in X} F^{\lambda^*}(x)$ is clearly hold.

Another method for solving multicriteria optimization problems is the method of sequential concessions, which requires a preliminary ranking of the criteria by importance. At each step k , a single-criteria problem with an objective function of rank k is solved. New constraints are also added to the system of constraints, which ensure that the value of criteria from rank 1 to $(k-1)$ deviates by the amount of the permissible concession $\delta_i > 0, i = 1, 2, \dots, k-1$. The problem is solved when N single-criteria conditional optimization problems with criteria $F_i(x)$,

where $i = 1, 2, \dots, N$ are solved. The end result is the optimal value of the least important criterion, provided that the values of the previous criteria are guaranteed. Paper [19] analyzes the effectiveness of using the method of successive concessions to solve multicriteria problems of diversification of a centralized pharmacy network of different sizes and identifies zones of stability in the space of parameters of the method of concessions.

Solving multicriteria optimization problems is a non-trivial task due to the conceptual uncertainty of vectors that are incommensurable. The final decision is always made by the decision maker. To justify such a choice, it is necessary to assess the properties of the solutions obtained by applying different approaches.

Therefore, the purpose of this article is to analyze the problem of choosing a set of criteria and the effectiveness of solving the multicriteria problem of diversification of a retail network using different methods: sequential concessions, linear convolution and computer modeling.

Statement of the problem and results

Let's formulate the mathematical problem of diversifying a retail network in more detail and recall the basic definitions. The mathematical formulation uses the apparatus for describing multidimensional random variables. The profitability of the network is estimated as

$$R = \sum_{i=1}^n r_i x_i,$$

where the vector $x^* = \begin{pmatrix} x_1^* \\ \dots \\ x_n^* \end{pmatrix}$ is a share of retail chain units

in the chain's asset portfolio;

r_i – profitability of the network unit, $i = \overline{1, n}$.

Risk D , estimated using the variance matrix $W = |\omega_{ij}|$, $\omega_{ij} = \text{cov}(x_i, x_j)$ – covariance, $i, j = \overline{1, n}$.

$$D = \sum_{i=1}^n \omega_{ii} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j,$$

if

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^n x_i = 1.$$

As noted above, the classical optimal portfolio model is a two-criteria problem with a vector objective function $F = (R, D)$.

Let's apply the classical model to the formal description of the problem of optimal distribution of goods of a wholesale trade network among branches. It is necessary to determine the share of goods for each branch in order to maximize the profit R of the entire network under the condition of minimal risk D . To achieve the best possible ratio of expected profitability and risk, it is important to carry out diversification measures, the effectiveness of which requires research. The peculiarity of the problem is the presence of mutual influence between the network units.

In such a formulation, it is necessary to define the concepts of "profit" and "risk" in order to determine the factors of influence and quality criteria for evaluating possible alternatives.

Consider a chain with n outlets (branches). Let us denote:

v_i is the expected value of the goods sold in sales prices of the i -th branch for the year (the average value of sales for each branch for m years), $i = \overline{1, n, \dots}$;

v_{si} is the expected value of sales of goods in purchase prices of the i -th branch for the year (average for m years), $i = \overline{1, n, \dots}$;

v_{zi} is the expected amount of expenses of the i -th branch for the year (average for m years), $i = \overline{1, n}$.

Then $v_{0i} = v_{si} + v_{zi}$ – cost price of goods sold.

The profitability of the i -th branch will be as follows:

$$r_i = \frac{\mathcal{G}_i + \mathcal{G}_{0i}}{\mathcal{G}_{0i}}. \quad (4)$$

Let us denote the share of the distributed resource of the i -th branch by

$$x_i = \frac{\mathcal{G}_{0i}}{\sum_{i=1}^n \mathcal{G}_{0i}}, \quad (5)$$

then the profitability of the entire enterprise will be as follows

$$R = \sum_{i=1}^n r_i x_i. \quad (6)$$

Really,

$$R = \sum_{i=1}^n \frac{\mathcal{G}_i + \mathcal{G}_{0i}}{\mathcal{G}_{0i}} * \frac{\mathcal{G}_{0i}}{\sum_{i=1}^n \mathcal{G}_{0i}} = \frac{\sum_{i=1}^n \mathcal{G}_i - \sum_{i=1}^n \mathcal{G}_{0i}}{\sum_{i=1}^n \mathcal{G}_{0i}} = \frac{\mathcal{G} + \mathcal{G}_0}{\mathcal{G}_0},$$

where \mathcal{G} – the expected cost of goods sold during the year throughout the enterprise;

\mathcal{G}_0 – the expected cost of goods sold during the year across the enterprise.

Risks in the process of forming an assortment portfolio are calculated using the variance:

$$D = \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j = \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j, \quad (7)$$

where $\omega_{ij} = \text{cov}(x_i, x_j)$ is a covariance, $\omega_{ij} = \omega_{ji}$, $i, j = \overline{1, n}$.

The solution is the vector $x^* = \begin{pmatrix} x_1^* \\ \dots \\ x_n^* \end{pmatrix}$. Knowing

x_i^* , $i = \overline{1, n}$, we can calculate the number of distributed resources by branch. From formula (5) we have

$$x_i^* = \frac{v_{si} + v_{zi}}{\sum_{i=1}^n v_{si} + \sum_{i=1}^n v_{zi}}.$$

Hence

$$v_{si} = \left(\sum_{i=1}^n v_{si} + \sum_{i=1}^n v_{zi} \right) x_i^* - v_{zi},$$

where $\sum v_{si}$ – total distributed resource;

$\sum v_{zi}$ – the average value of total expenses for a certain period;

v_{zi} – the average value of expenses for each division for a particular period.

The level of diversification is assessed by determining the value of entropy according to the method of K. Shannon, which characterizes the degree of diversity of the system. The introduction of entropy as the third criterion will allow to influence the level of diversification, as well as the assortment structure of the portfolio

$$E = - \sum_{i=1}^n x_i \ln x_i.$$

Below are five models for diversifying the portfolio of a retail chain with different composition of VOFs (2).

MODEL 1 corresponds to a two-criteria optimization problem with a vector objective function Φ_1 containing the criteria of risk D and entropy E .

It is necessary to find the vector $x^* = \begin{pmatrix} x_1^* \\ \dots \\ x_n^* \end{pmatrix}$ in case

of known $W = |\omega_{ij}|$, $\omega_{ij} = \text{cov}(x_i, x_j)$, that

$$\Phi_1 = (D, E), \quad (8)$$

where

$$D = \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j \rightarrow \min,$$

$$E = - \sum_{i=1}^n x_i \ln x_i \rightarrow \max.$$

Subject to restrictions on the level of profitability r_p , which is chosen by the expert

$$R = \sum_{i=1}^n r_i x_i \geq r_p,$$

and

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^n x_i = 1.$$

MODEL 2 formalizes a three-criteria optimization problem with a vector objective function Φ_2 containing the criteria of network profitability R , risk D , and entropy E .

It is necessary to find the vector $x^* = \begin{pmatrix} x_1^* \\ \dots \\ x_n^* \end{pmatrix}$ in case

of known $W = |\omega_{ij}|$, $\omega_{ij} = \text{cov}(x_i, x_j)$ that

$$\Phi_2 = (R, D, E), \quad (9)$$

where

$$R = \sum_{i=1}^n r_i x_i \rightarrow \max$$

$$D = \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j \rightarrow \min,$$

$$E = - \sum_{i=1}^n x_i \ln x_i \rightarrow \max.$$

Subject to restrictions on the level of profitability r_p , which is chosen by the expert

$$R = \sum_{i=1}^n r_i x_i \geq r_p,$$

and

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^n x_i = 1.$$

MODEL 3 is a single-criteria problem derived from the classical problem with a vector objective function $F_3 = (R, D)$ by convolution of criteria in the form R/D .

The solution of the problem is the vector $x^* = \begin{pmatrix} x_1^* \\ \dots \\ x_n^* \end{pmatrix}$,

if $W = |\omega_{ij}|$, $\omega_{ij} = \text{cov}(x_i, x_j)$ are known that

$$\Phi_3 = R/D \rightarrow \max, \quad (10)$$

where

$$R = \sum_{i=1}^n r_i x_i,$$

$$D = \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j.$$

Subject to restrictions on the level of profitability r_p – , which is chosen by the expert

$$R = \sum_{i=1}^n r_i x_i \geq r_p ,$$

and

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^n x_i = 1 .$$

MODEL 4 is a two-criteria optimization problem with a vector objective function Φ_4 , which contains the convolution criterion from model 3, i.e., Φ_3 , and the entropy criterion E .

It is necessary to find the vector $x^* = \begin{pmatrix} x_1^* \\ \dots \\ x_n^* \end{pmatrix}$,

if $W = |\omega_{ij}|$, $\omega_{ij} = \text{cov}(x_i, x_j)$ are known, that

$$\Phi_4 = (F_3, E) , \quad (11)$$

where

$$\Phi_3 = R/D \rightarrow \max ,$$

where

$$R = \sum_{i=1}^n r_i x_i ,$$

$$D = \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j ,$$

$$E = - \sum_{i=1}^n x_i \ln x_i \rightarrow \max .$$

Subject to restrictions on the level of profitability r_p – , which is chosen by the expert

$$R = \sum_{i=1}^n r_i x_i \geq r_p ,$$

and

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^n x_i = 1 .$$

MODEL 5 is a modification of Model 2 and formalizes the problem of two-criteria optimization with a vector objective function Φ_5 containing the criteria of network profitability R and risk D .

It is necessary to find the vector $x^* = \begin{pmatrix} x_1^* \\ \dots \\ x_n^* \end{pmatrix}$ subject

to known $W = |\omega_{ij}|$, $\omega_{ij} = \text{cov}(x_i, x_j)$, that

$$\Phi_5 = (R, D) , \quad (12)$$

where

$$R = \sum_{i=1}^n r_i x_i \rightarrow \max ,$$

$$D = \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j \rightarrow \min .$$

Subject to restrictions on the level of profitability r_p – , which is chosen by the expert

$$R = \sum_{i=1}^n r_i x_i \geq r_p ,$$

where

$$0 \leq x_i \leq 1, \quad \sum_{i=1}^n x_i = 1 .$$

The five models described above are mass tasks. In the process of working with real numerical indicators, corresponding individual tasks are formed.

Experiments

The experiments were carried out for individual problems of models 1–5 based on the indicators provided by the decision maker in the retail network.

To substantiate the choice of the final solution, we will perform the solution using the methods related to the construction of a common criterion and the method of concessions for a different set of criteria, as well as apply different software and analyze the results.

Numerical experiments were conducted on the same indicators using different software: 1) using the method of the reduced gradient in the "Solution Finder" service *MS Excel* and 2) using the developed software in the *Matlab* package [10].

First, let us consider the application of the solution approach using the generalized criterion of linear convolution of criteria. It is necessary to construct an optimization integral criterion with an objective function of the form

$$C = \sum_{i=1}^N a_i C_i \rightarrow \text{extr} , \quad (13)$$

where C_i – are the normalized values of the VOF components (2), i.e. $F = (C_1, C_2, \dots, C_N)$;

$\sum_{i=1}^n a_i = 1, 0 \leq a_i \leq 1$ – is a constant denoting the degree of importance of each partial criterion C_i .

We build an optimization problem based on model 1 with the objective function

$$\Phi_1' = -a^{D/D_{\max} E/E_{\max}^{\max}} , \quad (14)$$

where $0 \leq a_i \leq 1$ – is a constant denoting the degree of importance of each partial criterion.

Optimization problem based on model 2 with the objective function:

$$\Phi_2' = a_1^{R/R_{\max} D/D_{\max} E/E_{\max}^{\max}} , \quad (15)$$

where $0 \leq a_i \leq 1$ – a constant indicating the degree of importance of each partial criterion, $\sum_{i=1}^n a_i = 1$.

Optimization problem based on model 5 with an objective function:

$$\Phi'_5 = -a^{D/D_{max} R/R_{max}} \quad (16)$$

The problem will be solved by linear convolution of the criteria (3) using the method of reduced gradient in the "Solution Finder" service of MS Excel.

We formulate an individual task of creating an effective investment portfolio for a trading company

$$\omega_{ij} = \begin{pmatrix} 0,000663 & 0,0003 & 0,000091 & -0,000214 & 0,000152 \\ 0,0003 & 0,000011 & -0,00033 & -0,00001 & 0,000024 \\ 0,000091 & -0,00033 & 0,000151 & 0,000004 & -0,000139 \\ -0,000214 & -0,00001 & 0,000004 & 0,000043 & 0,000028 \\ -0,000152 & 0,000024 & -0,000139 & 0,000028 & 0,000135 \end{pmatrix}.$$

Standard deviation based on 2017–2021 data:

$$\sigma = (0,0257; 0,003; 0,0123; 0,0086; 0,0116).$$

We find the efficient portfolio using the linear convolution method under the condition $a_1 = a_2 = 0,5$.

$$x^* = (0,13; 0,29; 0,25; 0,19; 0,14),$$

$$R = \sum_{i=1}^n r_i x_i = 0,018,$$

$$D = \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j = 1,12 E - 05,$$

$$E = 0,67819.$$

Next, let's consider the method of sequential concessions, which is used to solve multi-criteria problems with a preliminary ranking of the criteria by importance. At each step, a single-criteria conditional optimization problem is solved. At the first step, the objective function is the first-ranked optimization criterion. The constraints coincide with the constraints of the original problem. At each subsequent step k , a single-criteria problem with an objective function of rank k is solved and new constraints are added to ensure that the values of the criteria from rank 1 to $k - 1$ deviate by the amount of the permissible deviation $\delta_1 > 0, i = 1, 2, \dots, k - 1$.

Let's demonstrate the work of the method of sequential concessions on the example of model 2 with a vector objective function $F_2 = (R, D, E)$ under the condition of ranking the criteria: entropy > risk > profit, i.e. $E > D > R$. With the chosen ranking, we have the following sequence of single-criteria conditional optimization problems.

First step.

$$E = -\sum_{i=1}^n x_i \ln x_i \rightarrow \max ,$$

$$\begin{cases} R = \sum_{i=1}^n r_i x_i \geq r_p \\ 0 \leq x_i \leq 1, \sum_{i=1}^n x_i = 1. \end{cases} \quad (17)$$

with five branches. Based on the data on sales and expenses of this enterprise for five years (2017–2021), the vectors of resource allocation \mathcal{G}_s and profitability r are compiled:

$$\nu_s = (88228,15; 189947; 170569; 141857; 99669),$$

$$r = (0,0050; 0,0393; 0,0123; 0,0085; 0,0116).$$

Covariance coefficients:

The value E^* is the optimal value according to the first rank criterion.

Second step.

The objective function is to minimize the risk. The condition of deviation $\delta_1 > 0$ of the optimal value of E^* by the amount of the permissible concession is added to the constraints of the original problem:

$$D = \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j \rightarrow \min ,$$

$$\begin{cases} \left| -\sum_{i=1}^n x_i \ln x_i - E^* \right| \leq \delta_1 \\ R = \sum_{i=1}^n r_i x_i \geq r_p \\ 0 \leq x_i \leq 1, \sum_{i=1}^n x_i = 1. \end{cases} \quad (18)$$

The risk value D^* is the optimal value according to the second-ranked criterion.

Third step.

The objective function is to maximize the profit R . To the constraints of the second step problem, we add the condition that the optimal value of D^* deviates by no more than the amount of acceptable deviations $\delta_2 > 0$:

$$R = \sum_{i=1}^n r_i x_i \rightarrow \max ,$$

$$\begin{cases} \left| -\sum_{i=1}^n x_i \ln x_i - E^* \right| \leq \delta_1 \\ \left| \sum_{i=1}^n \omega_{ij} x_i^2 + 2 \sum_{i=1}^n \sum_{j=1}^n \omega_{ij} x_i x_j - D^* \right| \leq \delta_2 \\ R = \sum_{i=1}^n r_i x_i \geq r_p \\ 0 \leq x_i \leq 1, \sum_{i=1}^n x_i = 1. \end{cases} \quad (19)$$

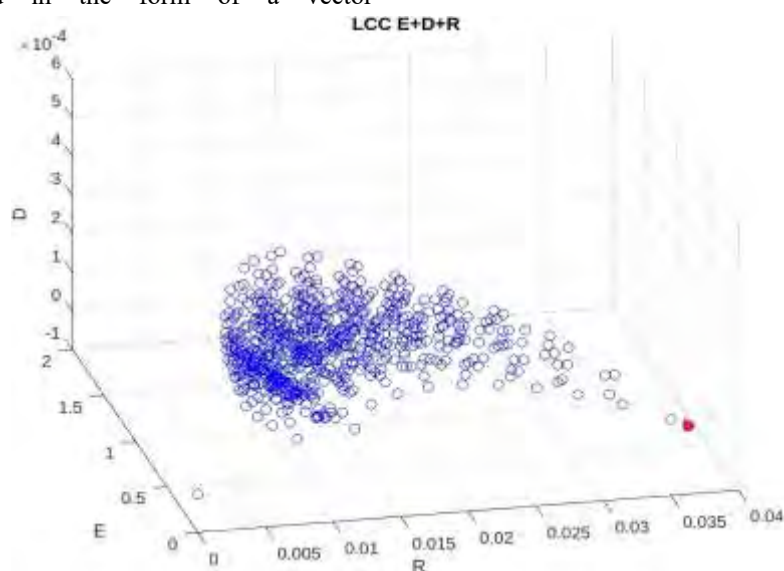
Eight experiments were conducted, the content of which is presented in Table 1.

Table 1. Content of the experiments conducted

№ Experiment	Model	Contentofcriteria	Solutionmethod
1	Model 1, $\Phi'1$	E, D	LCC (14)
2	Model 2, $\Phi'2$	E, D, R	LCC (15)
3	Model 5, $\Phi'5$	D, R	LCC (16)
4	Model 1, $\Phi1$	$E > D$	concession (8)
5	Model 2, $\Phi2$	$E > D > R$	concession (9)
6	Model 2, $\Phi2$	$E > R > D$	concession (9)
7	Model 3, $\Phi3$	R, D	multiplicativeconvolution (10)
8	Model 4, $\Phi4$	$E > \Phi3$	concession (11)

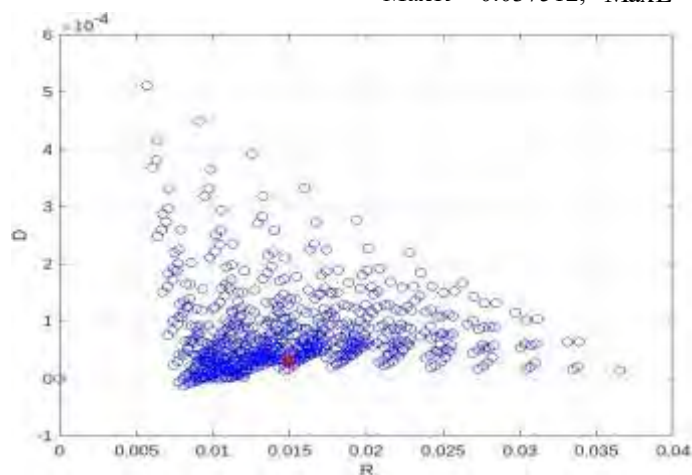
The results of experiment 1 on model 1 with the solution by the method of linear convolution of criteria and computer modeling are shown in Fig. 1. The optimal solution is obtained in the form of a vector

$X = (0.124; 0.201; 0.032; 0.445; 0.197)$, at which the optimal values of the criteria: $\text{Min}D = 2.0467 \times 10^{-5}$, $\text{Max}R = 0.015$, $\text{Max}E = 1.3729$ are achieved.

**Fig. 1.** Experiment 1: model 1(LCC) F1(6) entropy + risk

The second experiment consisted of solving the problem according to model 2 by the method of linear convolution of criteria. The results of computer modeling are shown in Fig. 2.

The optimal solution is obtained in the form of a vector $X = (0.001; 0.939; 0.001; 0.035; 0.023)$, at which the optimal values of the criteria: $\text{Min}D = 0.000013$, $\text{Max}R = 0.037512$, $\text{Max}E = 0.2772$ are achieved.

**Fig. 2.** Experiment 2: model 2(LCC) F2(7) entropy + risk + profit – projection on the plane (profit, risk) (R, D)

The results of experiment 5 were atypical, since the solution by the method of concessions according to model 2 led to an invalid solution $X = (0.095; 0.277; 0.069; 0.325; 0.233)$, at which the

values of the criteria: $\text{Min}R = 0.000030$, $\text{Max}E = 1$ are achieved. The results of computer modeling for this case are shown in Fig. 3.

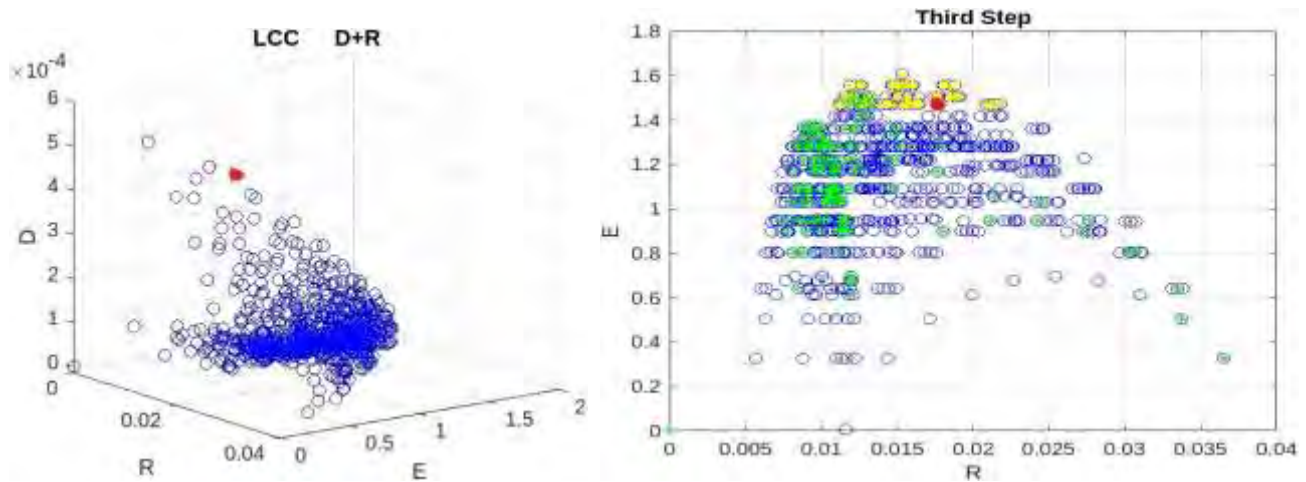


Fig. 3. Experiment 5: Model 2 (concession method): $E > D > R$.

Yellow color – the area of concession in terms of entropy E , green color – the area of concession in terms of risk D

The results of experiment 7 are related to model 3 and are shown in Fig. 4. The optimal solution has the form $X = (0.038; 0.206; 0.063; 0.679; 0.013)$, at which

the optimal values of the criteria: $\text{Min}D = 0.000028$, $\text{Max}R = 0.015$, $\text{Max}E = 0.9425$ are achieved.

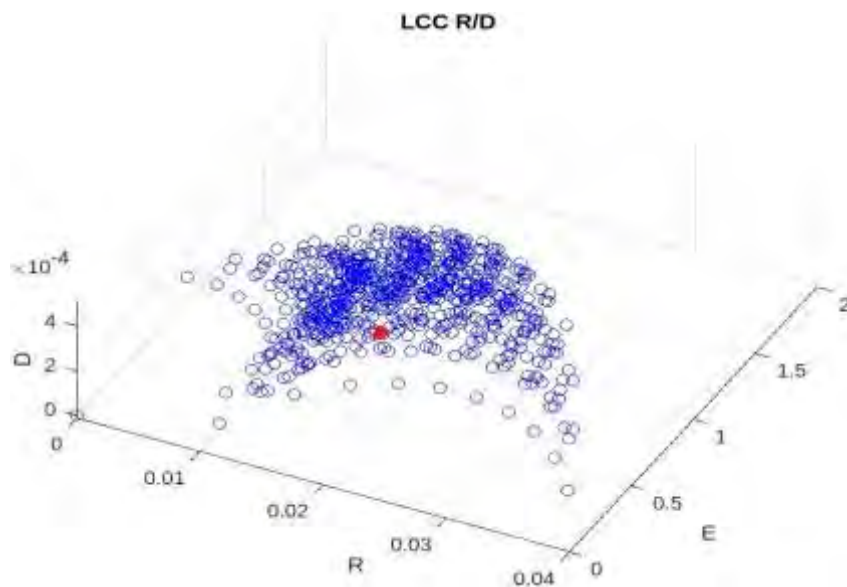


Fig. 4. Experiment 7, model 3 with R/D convolution

The results of the experiments conducted on the same indicators using different software are shown in Table 2.

The analysis of Table 2 shows that the results obtained belong to the Pareto set in all experiments and are non-dominant and non-comparable, except for the non-Pareto result of Experiment 7. A comparison of

results 7 and 8 shows that the risk is reduced when the entropy criterion is introduced. The fact that experiment 8 dominates the result of 7 proves the importance of adding entropy. Comparison of the results of experiments 1 and 2 proves the necessity of introducing the profitability criterion. The analysis of the results when applying the method of linear convolution of criteria proves that

only the application of all three criteria allows to obtain an adequate result (experiments 1–3, 7). Experiments 4 and 8 demonstrated tolerance to the results of the entropy and risk criteria when using the concession method, and in the process of implementing the LCC, the result improves in terms of entropy and deteriorates in terms of risk (experiments 1 and 4, 8). The result of

experiment 5 is that in the concession method we set 10% of the possible deviation. The numerical method did not allow us to find this result. The deviation in the second step is 40%. That is, by analogy with the concept of intractability by the convolution method, we can consider this example as intractable by the concession method.

Table 2. Results of solving the problem of enterprise diversification

MS Excel				MATLAB			
№exp	Max E	Max R	Min D	№exp	Max E	Max R	Min D
1	0,69088	0,015	4,09E-05	1	1,3729	0,015	2,05 E-05
4	0,62907	0,015	2,39E-05	4	1,456153	0,015122	2,50 E-05
8	0,62907	0,015	2,39E-05	8	0,973765	0,015	1,40 E-05
5	0,62907	0,01575	2,63E-05	5	1,469858	0,017697	3,40 E-05
6	0,62907	0,02067	4,45E-05	6	1,452399	0,017933	3,30 E-05
2	0,61663	0,02301	5,98E-05	2	0,27729	0,037512	1,30 E-05
3	4,05E-08	0,0393	1,10E-05	3	0,000011	0,0393	1,45 E-05
7	3,80E-08	0,0393	1,10E-05	7	0,942531	0,015	2,80 E-05

Different results in *MS Excel* and *Matlab* indicate the peculiarities of the numerical solution and the problem of software selection. Built-in *MS Excel* services and built-in *Matlab* functions apply numerical methods, which produces standard features of numerical solutions, such as the accumulation of numerical method errors and calculation errors.

Discussions and Conclusions

To formalize the problem, five models are proposed, which differ in vector objective functions both in terms of the number and quality of the selected criteria.

Two methods of solving multicriteria problems are considered. The first is the construction of a generalized criterion based on the components of the objective function vector of a multicriteria problem. The second method is a step-by-step solution based on the lexicographic order of the criteria in terms of importance to the decision maker. The results of the experiments allow us to draw general conclusions: the use of the entropy criterion reduces risk; model 2 helps to obtain the highest profitability when solving both criteria and concessions by the linear convolution method; the multiplicative convolution produces a non-Pareto solution.

References

1. "The Digital Economy: A Look Into the 2022 Digital Frontier, U.S. Chamber of Commerce's 2nd Annual Global Forum". URL: <https://www.uschamber.com/on-demand/technology/digital-economy-the-global-competition-to-write-the-rules> (last accessed 22.03.2023).

Thus, using the conclusions of the classical portfolio theory, which have been tested by experience and time, the authors of the article have developed a methodology for the efficient allocation of resources between branches of a retail network, which takes into account the expected profitability and diversification of the distribution and minimizes risks.

The scientific novelty of this work lies in the formalization of diversification models for a wholesale trading network based on portfolio theory and multicriteria optimization methods.

The practical value is that the results were obtained on real indicators for the network and they demonstrated the possibility of using the developed tool for automatic resource allocation in the form of Pareto optimal portfolios to minimize risks. The areas for further research include conducting a number of experiments with different ways of formalizing risk in portfolio models and finding appropriate analytical dependencies.

The research was carried out within the framework of the research work "Development of methods for studying complex socio-economic systems based on intelligent technologies", No. 0121U113264, at the Department of System Analysis and Computational Mathematics of the National University of "Zaporizhzhia Polytechnic".

2. Zanjirdar, M., (2020), Overview of Portfolio Optimization Models. *Advances in mathematical finance and applications*. 5(4). P. 419–435. DOI: 10.22034/amfa.2020.1897346.1407
3. Ghandehari, M., Azar, A., Yazdani, A., Golarzi, Gh. (2019), "A Hybrid Model of Stochastic Dynamic Programming and Genetic Algorithm for Multistage Portfolio Optimization with Glue VaR Risk Measurement". *Industrial Management Journal*. No. 11 (3). P. 517–542. DOI: 10.22059/IMJ.2019.278912.1007579
4. Kwon, R., Butler, A. (2021), "Covariance Estimation for Risk-Based Portfolio Optimization". *An Integrated Approach. Journal of Risk*. No. 24 (2). P. 11–41. DOI: 10.21314/JOR.2021.020
5. Chaweewanchon, A., Chaysiri, R. (2011), "Markowitz Mean-Variance Portfolio Optimization with Predictive Stock Selection Using Machine Learning", *International Journal of Financial Studies*. No. 10 (3), P. 64–73. DOI: <https://doi.org/10.3390/ijfs10030064>
6. Lim, Q.Y.E., Cao, Q. Quek, C. (2022), "Dynamic portfolio rebalancing through rein for cement learning". *Neural Computing and Applications*. Vol. 34, P. 7125–7139. DOI:10.1007/s00521-021-06853-3
7. Sharma, M., Shekhawat, H.S. (2022), "Portfolio optimization and return prediction by integrating modified deep belief network and recurrent neural network". *Knowledge-Based Systems*. Vol. 250, P. 1–19. DOI:10.1016/j.knsys.2022.109024
8. Escobar-Anel, M., Kschonnek, M., Zagst, R. (2022), "Portfolio optimization: not necessarily concave utility and constraints on wealth and allocation". *Mathematical Method sof Operations Research*. Vol. 95. P. 101–140. <https://doi.org/10.1007/s00186-022-00772-2>
9. Grechuk, B., Hao, D. (2022), "Individual and cooperative portfolio optimization as line ar program". *Optimization Letters*. Vol.16. P. 2569–2589. DOI:10.1007/s11590-022-01901-w
10. Mazin, A. M. Al Janabi (2021), "M.A.M.: Multivariate portfolio optimization under illiquid market prospects: a review of theoretical algorithms and practical techniques for liquidity risk management". *Journal of Modelling in Management*. No.16(1). P. 288–309. DOI:10.1108/JM2-07-2019-0178
11. Ahmadi-Javid, A., Fallah-Tafti, M. (2019), "Portfolio optimization with entropic value-at-risk". *European Journal of Operational Research*. No. 279(1). P. 225–241. <https://doi.org/10.1016/j.ejor.2019.02.007>
12. Markowitz, H. M., Blay, K. "Risk–Return Analysis. The Theory and Practice of Rational Investing (a four-volume series), McGraw-Hill". 2014. 208 p. URL: https://books.google.com.ua/books/about/Risk_Return_Analysis_The_Theory_and_Prac.html?id=_GknVPOReYoC&redir_esc=y
13. Xidonas, P., Steuer, R. Hassapis, C. (2020), "Robust portfolio optimization: a categorized bibliographic review". *Annals of Operations Research*. Vol. 292. P. 533–552. DOI: 10.1007/s10479-020-03630-8
14. Perepelitsa, V. A., Kozin I. V., Tereshchenko, E. V. (2012), Classification tasks: approaches, methods, algorithms [Zadachi klassifikatsii i formirovanie znaniy. – Saarbrücken, Germany] LAP LAMBERT Academic Publishing GmbH&Co. KG. 196 p.
15. Ehrgott, M. (2005), "Multicriteria Optimization". *Springer, Heidelberg*. Vol. XIII. 323 p. DOI: <https://doi.org/10.1007/3-540-27659-9>
16. Engau A., Sigler D. (2020), "Pareto solution sin multi criteria optimization underrun certainty". *European Journal of Operational Research*. No.281 (2). P. 357–368. DOI: 10.1016/j.ejor.2019.08.040
17. Zhou W., Zhu W., Chen Y., Chen J. (2022), "Dynamic changes and multi-dimensional evolution of portfolio optimization". *Economic Research-Ekonomika Istraživanja*. Vol. 35(11):1-26. P. 1431–1456. DOI:10.1080/1331677X.2021.1968308
18. Bakurova, A. V., Ropalo, H. M., Tereshchenko, E. V. (2021), "Analysis of the Effectiveness of the Successive Concessions Method to Solve the Problem of Diversification". MoMLeT+DS 2021: 3rd International Work shop on Modern Machine Learning Technologies and Data Science. P. 231–242. URL: <https://ceur-ws.org/Vol-2917/paper21.pdf>
19. Mathworks, "MATLAB for Artificial Intelligence". URL: <https://www.mathworks.com/campaigns/products>.

Received 17.11.2023

Відомості про авторів / About the Authors

Бакурова Анна Володимирівна – доктор економічних наук, професор, НУ «Запорізька політехніка», професор кафедри системного аналізу та обчислювальної математики, Запоріжжя, Україна; e-mail: abaka111060@gmail.com; ORCID ID: <https://orcid.org/0000-0001-6986-3769>

Савранська Алла Володимирівна – кандидат фізико-математичних наук, доцент, НУ «Запорізька політехніка», доцент кафедри системного аналізу та обчислювальної математики, Запоріжжя, Україна; e-mail: savranskaya-alla@ukr.net; ORCID ID: <https://orcid.org/0000-0003-0193-8722>

Терещенко Еліна Валентинівна – кандидат фізико-математичних наук, доцент, НУ «Запорізька політехніка», доцент кафедри системного аналізу та обчислювальної математики, Запоріжжя, Україна; e-mail: elina_vt@ukr.net; ORCID ID: <https://orcid.org/0000-0001-6207-8071>

Ширококорад Дмитро Вікторович – кандидат фізико-математичних наук, доцент, НУ «Запорізька політехніка», доцент кафедри системного аналізу та обчислювальної математики, Запоріжжя, Україна; e-mail: hoveringphoenix@gmail.com; ORCID ID: <https://orcid.org/0000-0002-2784-4081>

Шевчук Марк Валерійович – НУ «Запорізька політехніка», аспірант кафедри системного аналізу та обчислювальної математики, Запоріжжя, Україна; e-mail: shevchuk.marko@gmail.com; ORCID ID: <https://orcid.org/0000-0001-6245-1331>

Bakurova Anna – Doctor of Sciences (Economics), Professor, National University "Zaporizhzhia Polytechnic", Professor at the Department of Systems Analysis and Computational Mathematics, Zaporizhzhia, Ukraine.

Savranska Alla – PhD (Physics and Mathematics), Associate Professor, National University "Zaporizhzhia Polytechnic", Associate Professor at the Department of Systems Analysis and Computational Mathematics, Zaporizhzhia, Ukraine.

Tereschenko Elina – PhD (Physics and Mathematics), Associate Professor, National University "Zaporizhzhia Polytechnic", Associate Professor at the Department of Systems Analysis and Computational Mathematics, Zaporizhzhia, Ukraine.

Shyrokorad Dmytro – PhD (Physics and Mathematics), Associate Professor, National University "Zaporizhzhia Polytechnic", Associate Professor at the Department of Systems Analysis and Computational Mathematics, Zaporizhzhia, Ukraine.

Shevchuk Mark – National University "Zaporizhzhia Polytechnic", Postgraduate Student at the Department of Systems Analysis and Computational Mathematics, Zaporizhzhia, Ukraine.

АНАЛІЗ ЗАДАЧІ ВИБОРУ КРИТЕРІЇВ У МОДЕЛЯХ ДИВЕРСИФІКАЦІЇ

Цифровізація економіки знижує вартість ведення бізнесу завдяки автоматизації відповідних процесів, але будь-яка трансформація генерує нові ризики, нестійкість економіки. Економічна нестабільність призводить до падіння рівня життя та, як наслідок, негативно впливає на діяльність торговельних підприємств. Особливо чутливими до будь-яких змін є середній та малий бізнес. Зниження попиту на більшість товарів повсякденного вживання болісно позначається на діяльності торговельних підприємств малого та середнього бізнесу, призводить до появи нових ризиків. Ці ризики істотно впливають на зниження прибутковості підприємств. Тому важливим для кожного підприємства є завдання диверсифікації діяльності, що передбачає розширення товарного асортименту, переорієнтацію ринків збуту та оптимальний розподіл товарів між підрозділами одного підприємства. **Предметом дослідження** статті є багатокритеріальні моделі диверсифікованого портфеля, що мінімізують ризики, які виникають в епоху цифрової економіки в управлінні мережами торговельних підприємств. Для формалізації задачі запропоновано п'ять моделей, що різняться векторними цільовими функціями як за кількістю, так і за якістю обраних критеріїв. **Метою роботи** є аналіз проблеми вибору критеріїв у відповідних багатокритеріальних, або векторних, задачах диверсифікації. У статті досліджуються переваги введення до складу критеріїв класичної двокритеріальної моделі портфельної теорії критерію максимізації ентропії, що визначає ступінь різноманітності складу портфеля. Застосовується комплексне поєднання методів класичної теорії портфеля та багатокритеріальної оптимізації. **Результатами дослідження** є порівняння трьох методів розв'язування таких задач: згортка критеріїв, послідовні поступки та комп'ютерне моделювання множини Парето. **Висновки:** здобуті результати будуть корисними для автоматизації управління ризиками торговельних мереж. Практична цінність роботи полягає в тому, що досягнуті результати на реальних даних для мережі продемонстрували можливість застосування розробленого інструменту для автоматичного розподілу ресурсів у вигляді паретооптимальних портфелів із метою мінімізації ризиків.

Ключові слова: багатокритеріальна задача; задача оптимального портфеля; згортка критеріїв; метод послідовних поступок; множина Парето; ентропія.

Бібліографічні описи / Bibliographic descriptions

Бакурова А. В., Савранська А. В., Терещенко Е. В., Ширококорад Д. В., Шевчук М. В. Аналіз задачі вибору критеріїв у моделях диверсифікації. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 5–15. DOI: <https://doi.org/10.30837/ITSSI.2023.26.005>

Bakurova, A., Savranska, A., Tereschenko, E., Shyrokorad, D., Shevchuk, M. (2023), "Analysis of the criteria selection problem in diversification models", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 5–15. DOI: <https://doi.org/10.30837/ITSSI.2023.26.005>

O. BARKOVSKA, A. HAVRASHENKO

ANALYSIS OF THE INFLUENCE OF SELECTED AUDIO PRE-PROCESSING STAGES ON ACCURACY OF SPEAKER LANGUAGE RECOGNITION

The **subject matter** of the study is the analysis of the influence of pre-processing stages of the audio on the accuracy of speaker language recognition. The importance of audio pre-processing has grown significantly in recent years due to its key role in a variety of applications such as data reduction, filtering, and denoising. Taking into account the growing demand for accuracy and efficiency of audio information classification methods, evaluation and comparison of different audio pre-processing methods becomes an important part of determining optimal solutions. The **goal** of this study is to select the best sequence of stages of pre-processing audio data for use in further training of a neural network for various ways of converting signals into features, namely, spectrograms and mel-cepstral characteristic coefficients. In order to achieve the goal, the following **tasks** were solved: analysis of ways of transforming the signal into certain characteristics and analysis of mathematical models for performing an analysis of the audio series by selected characteristics were carried out. After that, a generalized model of real-time translation of the speaker's speech was developed and the experiment was planned depending on the selected stages of pre-processing of the audio. To conclude, the neural network was trained and tested for the planned experiments. The following **methods** were used: mel-cepstral characteristic coefficients, spectrogram, time mask, frequency mask, normalization. The following **results** were obtained: depending on the selected stages of pre-processing of voice information and various ways of converting the signal into certain features, it is possible to achieve speech recognition accuracy up to 93%. The practical significance of this work is to increase the accuracy of further transcription of audio information and translation of the formed text into the chosen language, including artificial languages. **Conclusions:** In the course of the work, the best sequence of stages of pre-processing audio data was selected for use in further training of the neural network for different ways to convert signals into features. Mel-cepstral characteristic coefficients are better suited for solving our problem. Since the neural network strongly depends on its structure, the results may change with the increase in the volume of input data and the number of languages. But at this stage, it was decided to use only mel-cepstral characteristic coefficients with normalization.

Keywords: mel-cepstral characteristic coefficients; spectrogram; time mask; frequency mask; normalization; neural network; voice; audio series; speech.

Introduction

Voice is the result of the vibration of the vocal folds (vocal cords) in the human larynx. These vibrations generate sound, which is then modified by various parts of the mouth and nasal cavities, as well as the lips and tongue. The physical basis of voice includes aspects such as frequency, amplitude, waveform, and other parameters [1, 2]. Here are some of the main criteria of the voice and how they are represented in a spectral diagram.

Frequency determines the pitch of the voice and is measured in hertz (Hz). Figure 1 shows the frequencies increasing along the vertical axis.

Amplitude defines the loudness of a sound and is measured in decibels (dB). The legend on the right (Fig. 1) shows that color intensity increases with density.

Dynamics indicates variations in the volume of the speech, such as pauses and quiet zones. Pauses in speech are displayed with minimal dynamics. In the figure, we can see a pause in the first 0.5 seconds, and from 4.5 to 6 seconds.

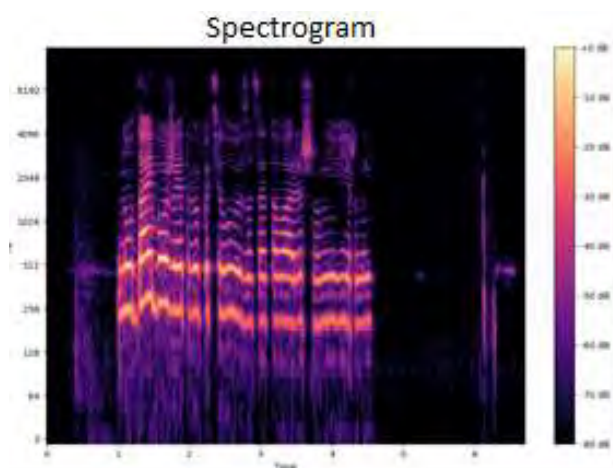


Fig. 1. Example of a spectrogram

Timbre defines the unique characteristics of a voice that distinguish it from other voices. Different timbres are represented by different shapes on the spectrogram.

Intonation determines the melody and rhythm of speech. Changes in intonation can be detected in the shape and location of intense sectors on the spectrogram.

Each language can have its own peculiarities in the spectral composition of the voice, due to the phonetic features and nature of the sounds that make up it. Voice spectral diagrams can be used to analyze these differences and to study various language features.

The relevance of this study lies in the fact that in today's world, where globalization and international communication are becoming a necessity, speaker recognition systems with subsequent translation into the chosen language are becoming extremely important. These technologies are revolutionizing the way we communicate and interact in different cultural

and linguistic environments. Thanks to this, speaker recognition systems can become an integral part of our digital lives, helping to make the world more connected and diverse. The practical significance of this work is to increase the accuracy of further transcription of audio information and translation of the generated text into the chosen language.

Materials and methods

A generalized traditional voice signal processing pipeline is shown in Figure 2.

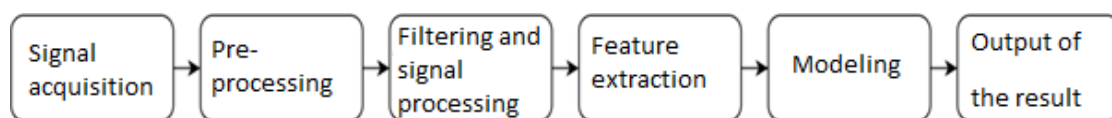


Fig. 2. Traditional voice signal processing pipeline

Each of these stages has certain tasks, the quality of which will affect the next step. The purpose and tasks of each stage are described below.

- signal acquisition is recording or selecting an audio file to be processed;
- pre-processing involves converting the signal from analog to digital format for further processing;
- filtering and signal processing usually involves the use of various filters to extract useful information, such as reducing background noise or enhancing harsh sounds;
- feature extraction is necessary to reduce the amount of input data, which speeds up the processing time of incoming audio files without losing processing accuracy;
- modeling in the context of voice signal processing means creating mathematical models or algorithms that can learn and recognize certain properties of voice signals;
- the result is displayed in a form suitable for the user, with the possibility of further processing of the received data.

In this paper, the research is focused on the feature extraction and modeling stages, so let's consider them in more detail. Among the existing possible ways to convert a signal into certain features, we can distinguish:

- mel-cepstral characteristic coefficients
- linear coding with prediction;
- spectrogram.

The shallow mel-cepstral characteristic coefficients are one of the frequently used features used in a variety of applications, especially in voice signal processing, such as speaker recognition, voice recognition, and gender identification. The mel-cepstral characteristic coefficients can be calculated by performing five sequential processes (Figure 3), namely, framing the signal, calculating the power spectrum, applying a bank of chalk filters to the obtained power spectra, calculating the logarithms of all filter banks, and finally applying the discrete cosine transform [3, 4, 5].



Fig. 3. Scheme for determining fine-frequency capstral coefficients

Speech signal analysis aims to find more informative, compact, and relevant knowledge than the raw speech signal data itself. Vocal tract features (also called segmental, spectral, or system features) [6] are one of the well-known representations of speech analysis. Among the most well-known applications that use the mel-cepstral characteristic coefficients of the

speech signal are emotion recognition in speech, speech and dialect recognition, and speech recognition, as shown in the following subsections.

Linear Predictive Coding (LPC) is a method of analyzing and modeling an audio signal. Linear prediction analysis can provide a set of speech signal model parameters that accurately represent the spectral

amplitude of the speech signal, taking into account a variety of articulatory and phonetic features. LPC can mimic the human articulatory system very well and, therefore, has some advantages in extracting speech feature parameters [7, 8].

The obtained LPC coefficients can serve as an acoustic signature of the voice, which allows them to be used for speaker recognition and other tasks in the field of speech processing.

One of the main advantages of LPC is its ability to reduce the data dimensionality in a well-controlled manner, which makes it effective for transmitting and storing speech information. Another important property of LPC is the ability to reconstruct a signal based on

a limited number of parameters, which makes it effective in resource-constrained systems such as real-time speech recognition systems or mobile applications.

A spectrogram is a visual representation of the time-frequency analysis of sound. It allows you to analyze changes in sound intensity as a function of frequency and time. A spectrogram is usually created in one of two ways: approximated as a set of filters derived from a series of bandpass filters (this was the only way before the advent of modern digital signal processing methods), or calculated from a time signal using the windowed or fast Fourier transform. These two methods actually produce different quadratic frequency-time distributions, but are equivalent under certain conditions.

Table 1. Comparison of feature extraction methods

	Mel-cepstral characteristic coefficients	Linear prediction coding	Spectrogram
Description	Mel-cepstral characteristic coefficients are coefficients obtained from the taken cepstral coefficients of the sound spectrum calculated on the basis of the chalk scale (chalk scale).	LPC models the voice signal as a linear combination of previous samples.	A spectrogram visualizes a time series as a function of time and frequency. It shows which frequencies are present in the signal at a particular moment in time.
Advantages	Effective for speech recognition, takes into account the peculiarities of auditory perception.	It reproduces formants (resonant peaks) in speech well, which is important for speech recognition.	It shows the time and frequency characteristics of a signal well, and helps to separate different sound sources.
Disadvantages	Requires a lot of calculations to obtain the cepstral coefficients and may be limited in higher frequencies.	It can be sensitive to noise, and its effectiveness decreases with non-stationary signals.	It is not always effective for speech recognition due to the lack of distinct speech characteristics.

Spectrograms are widely used in audio analysis, music production, speech processing, medical research, and many other fields to visualize information about the time and frequency character of sound signals [9, 10].

The result of the comparative analysis of feature extraction methods is shown in Table 1.

In further work, we will consider in detail the fine-frequency cepstral coefficients and spectrograms, because these methods generate compact representations, which facilitates their processing and analysis.

In addition, they help to extract important acoustic features, such as formants in speech.

A variety of mathematical models are used for voice recognition, among which the most common are neural networks (deep and shallow), support vector method (SVM), random forest, decision trees, and hidden Markov models.

The result of the comparative analysis of mathematical models for audio data analysis is shown in Table 2.

Table 2. Comparative analysis of mathematical models for audio data analysis

Model	Advantages	Disadvantages
Neural networks	Ability to learn complex dependencies, great flexibility in working with different types of data.	High computational resource requirements, need for a large amount of data for training.
Support vector method	Effective in high-dimensional spaces, works well with small data sets.	May be sensitive to noise, limited in cases where data is not linearly resolved.
Decision tree	Easy to interpret, does not require significant preliminary data analysis.	May be prone to overfitting, does not always work well with complex dependencies in data.
Random forest	Reduces the tendency to overlearn, works well with large amounts of data.	Can be computationally expensive, loss of interpretability compared to a separate decision tree.
Hidden Markov models	Good for modeling time series and event sequences.	Assumes independence between states, limited in modeling complex dependencies [15].

These mathematical models are used in various voice technologies, including speech recognition, speaker identification, emotion analysis, and others. The choice of a particular model may depend on the specific task, data availability, processing volume, and other factors. In this work, modeling is performed using neural networks because they can learn complex relationships and have great flexibility in working with different types of data.

Training a neural network is impossible without a prepared training and test set. The dataset was selected based on the results of the analysis of existing audio corpora.

For the test data, we chose the corpus described in [16]. The Common Voice corpus is a multilingual collection of transcribed speech intended for research and development of speech technologies. Common Voice is designed for the purposes of automatic speech recognition, but can be useful in other areas (e.g., language detection). To achieve scale and sustainability, Common Voice uses crowdsourcing

for both data collection and data validation. The latest release includes 29 languages, and as of November 2019, a total of 38 languages are being collected. More than 50,000 people have participated so far, resulting in 2,500 hours of audio. To the best of our knowledge, this is the largest publicly available audio corpus for speech recognition in terms of both hours and languages.

Another significant feature of Common Voice is its constant updating and expansion with fresh and representative data to improve and refine speech recognition models.

The reasons for choosing this dataset are shown in Table 3. The main reason for choosing this particular corpus was the large number of different languages and the large number of speakers in each set. In addition, this dataset is public under the CC0 license. Its use will allow us to simplify the model because there is no noise and will help train the model on different accents of different languages.

Table 3. Advantages of the chosen corpus

Selected corpus for training an artificial voice command analyzer	Advantages	Consequences of the advantages
Common Voice	A large number of languages	Possibility of scaling
	Absence of noise	No need to eliminate noise that could reduce accuracy
	Different audio authors	No false dependencies are created due to person/gender
	Public license	Can be used in this project.

To build the neural network, we used the TensorFlow library, which can help a researcher train neural networks to detect and decipher patterns and correlations, similar to the learning and understanding used by humans[17].

The goal is to analyze the impact of audio data preprocessing methods on the accuracy of speaker speech detection for use in further neural network training for different ways of converting signals into features, namely spectrograms and fine-frequency cepstral coefficients.

To achieve this goal, the following tasks should be solved

- analysis of methods for converting signals into features;
- review of mathematical models for performing audio order transformation by selected features;
- development of a generalized model for real-time translation of a speaker's speech;
- planning the experiment depending on the selected stages of audio pre-processing;
- raining and testing of the neural network for the planned experiments;
- comparative analysis of the training results.

Solution of the stated problem.

Discussion of the results

In this paper, we propose a generalized model of real-time translation of a speaker's speech into one of the selected languages in Figure 4. The model consists of stages that sequentially determine the speaker's language, convert the speech to text, and translate the prepared text into one of the selected languages, including artificial languages [18, 19, 20].

These mathematical models are used in various voice technologies, including speech recognition, speaker identification, emotion analysis, and others. The choice of a particular model may depend on the specific task, data availability, processing volume, and other factors. In this work, modeling is performed using neural networks because they can learn complex relationships and have great flexibility in working with different types of data.

This study examines the impact of preprocessing stages on speech recognition accuracy.

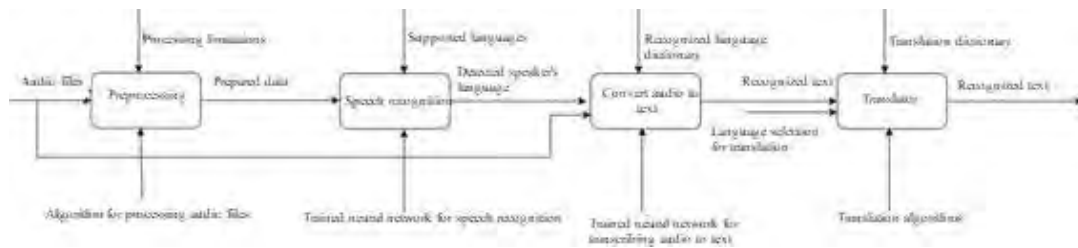


Fig. 4. Generalized model for real-time translation of a speaker's speech

The experiments were planned depending on the selected audio preprocessing stages – normalization, frequency mask and time mask for two main ways of transforming the signal into certain features – for fine-frequency cepstral coefficients and for spectrograms.

The time-domain severity transform coefficients and the spectrogram represent different ways of representing an audio signal in the form of feature vectors, and each of these methods can determine different aspects of the audio signal.

For each case, a separate neural network was created, and, taking into account the peculiarities of the results of each type of processing, the structure of the network was changed - the number of network layers, cutoff coefficients, number of epochs, etc.

A total of 8 experiments were generated for two different ways of converting the signal into

features – for low-frequency cepstral coefficients and for spectrograms.

The neural network was trained on datasets containing the same number of audio tracks for different languages:

1. Ukrainian;
2. English;
3. French;
4. German.

The system accepts randomly selected files from the dataset as input. The dataset was divided into training and test samples in the following ratio: 80% of the input data (out of 1000 audio files) was included in the training set, 20% is the test data on which we will test our neural network (200 test files).

The neural network itself consists of several levels and is shown in Figure 5.

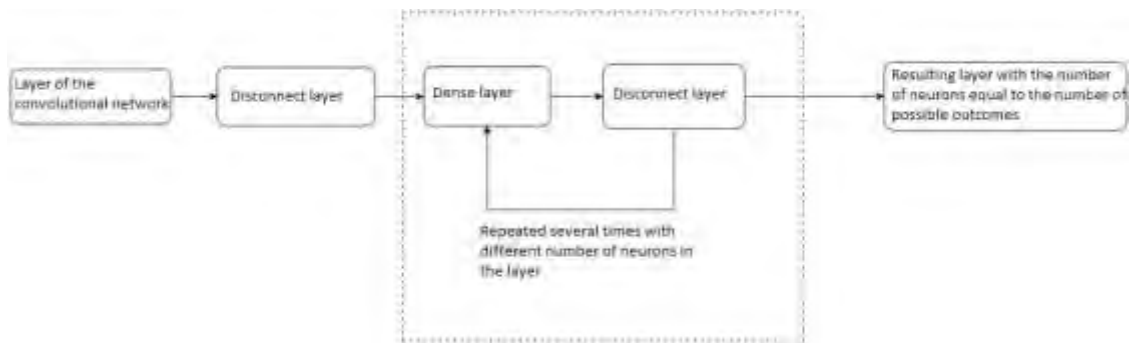


Fig. 5. Generalized structure of the neural network

The first layer uses a Convolutional Neural Network in the TensorFlow library. After this layer, we will have levels of convolutional representation, which can then be aligned and passed to the fully connected layer for further analysis and classification.

Next are several Dense layers with different numbers of neurons on them. With each subsequent layer, the number of neurons decreases. This is due to the dimensionality reduction technique. Reducing the number of neurons helps to control the number of parameters in the model, which is important to avoid overfitting, especially when data is limited.

In addition, to reduce overfitting, each Dense layer is followed by a random disconnect layer. This is one of the regularization techniques that helps to avoid overfitting the model. In each training epoch, a certain portion of the layer weights is randomly selected and set to zero at that stage. This helps to generalize the model and increase its resistance to overfitting.

The last layer is the result layer, in which the number of neurons is equal to the number of classes of possible outcomes.

An assessment of the impact of audio preprocessing methods on the accuracy (probability of

correct recognition) of speaker speech recognition in accordance with the experiments conducted using fine-frequency cepstral coefficients is shown in Table 4.

Table 4. *Experimental results using mel-cepstral characteristic coefficients*

Normalization	Time mask	Frequency mask	Accuracy of speaker language detection, %
–	–	–	90
+	–	–	93
–	+	–	24
+	+	–	24
–	–	+	24
+	–	+	25
–	+	+	22
+	+	+	23

As can be seen from the results, the time mask and frequency mask remove useful information and cannot be used with mel-cepstral characteristic coefficients.

An assessment of the impact of audio preprocessing methods on the accuracy (probability of correct recognition) of speaker speech recognition in accordance with the experiments conducted using spectrograms is given in Table 5.

Table 5. *Experimental results when using a spectrogram*

Normalization	Time mask	Frequency mask	Accuracy of speaker language detection, %
–	–	–	61
+	–	–	65
–	+	–	63
+	+	–	67
–	–	+	63
+	–	+	68
–	+	+	65
+	+	+	73

Based on the results of the spectrogram, we can conclude that masking helps to increase accuracy, but this method is still worse than mel-mean characteristic coefficients in this task, since mel-mean characteristic coefficients also take into account the features of the audio signal that are important for speech or sound recognition, and include high-frequency and low-frequency coefficients, differentiation, and other characteristics. A spectrogram, on the other hand, simply represents the distribution of signal energy at different frequencies over time. It may include less information about specific aspects of the audio signal that are important for certain tasks, such as speech recognition.

A comparison of the results presented in the tables is shown in Figure 6.

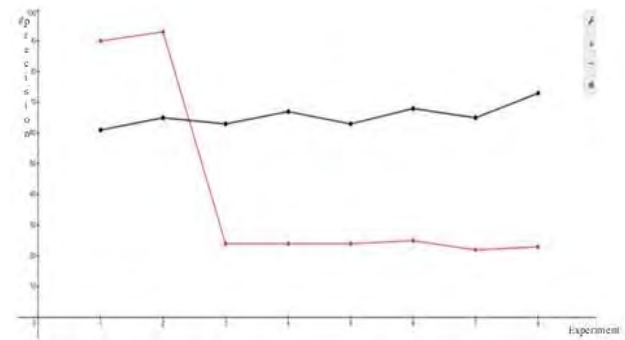


Fig. 6. The resulting graph of speaker language detection accuracy for each experiment. Red – mel-cepstral characteristic coefficients, black – spectrogram

Conclusions

This paper analyzes the impact of audio data preprocessing methods on the accuracy of speaker language detection for use in further neural network training for different methods of converting signals into features, namely spectrograms and mel-cepstral characteristic coefficients. Depending on the chosen methods of pre-processing voice information and different ways of converting the signal into features, we managed to achieve 93% accuracy in determining the speaker's language. The practical significance of this work is to increase the accuracy of further transcription of audio information and translation of the generated text into the chosen language, including artificial languages.

Based on the results of the study, a neural network model using mel-cepstral characteristic coefficients with normalization at the stage of input data preprocessing was chosen for further experiments on speaker speech recognition. This decision is based on several important factors that confirm the advantages of mel-cepstral characteristic coefficients in this context:

1. mel-cepstral characteristic coefficients allow to reduce the amount of input data. This allows to increase the speed of both processing and training of the neural network.
2. They effectively recognize speech and take into account the peculiarities of auditory perception, which interacts well with neural networks.
3. Experiments have shown that mel-cepstral characteristic coefficients have high classification accuracy for different sizes of training samples. This demonstrates their ability to provide information for effective analysis and generalization of patterns in the data.

4. mel-cepstral characteristic coefficients showed a stable result regardless of the input languages of the training samples. This may be an important aspect for further research, as it allows scaling up the study for more classes of input data.

All the justifications confirm that further speaker recognition and translation into the chosen language, including artificial languages, will use mel-cepstral characteristic coefficients as a method of extracting features of the input audio sequence, since the accuracy of speaker recognition using spectrograms, as shown experimentally in this paper, is 22% worse. In addition,

it was concluded that the use of time or frequency masks with mel-cepstral characteristic coefficients reduces the recognition probability to the level of guessing. Therefore, these feature extraction methods should be abandoned.

So, taking into account all the data and analysis results, we can conclude that mel-cepstral characteristic coefficients with preliminary data normalization is a reasonable pickup line for speaker speech detection. Its high accuracy and reduced input data size make this method the best choice for further research and implementation in practical applications.

References

1. Zhang, Z.(2016), "Mechanics of human voice production and control". *The journal of the acoustical society of america* Vol.140.4. P. 2614–2635. DOI: <https://doi.org/10.1121/1.4964509>
2. Garellek, M.(2022), "Theoretical achievements of phonetics in the 21st century: Phonetics of voice quality". *Journal of Phonetics* Vol.94(24). DOI: <https://doi.org/10.1016/j.wocn.2022.101155>
3. Abdul, Z. K., Al-Talabani A. K.(2022), "Mel Frequency Cepstral Coefficient and its Applications: A Review", *IEEE Access*, Vol. 10, P. 122136–122158. DOI: <https://doi.org/10.1109/ACCESS.2022.3223444>
4. Ayvaz, U.(2022), "Automatic speaker recognition using mel-frequency cepstral coefficients through machine learning". *CMC-Computers Materials & Continua*. Vol.71(3), P. 5511–5521. DOI: <https://doi.org/10.32604/cmc.2022.023278>
5. Shalbbya, A. (2020), "Mel frequency cepstral coefficient: a review". *ICIDSSD*, P. 1–10. DOI: <https://doi.org/10.4108/eai.27-2-2020.2303173>
6. Ramakrishnan, S. (2012), "Recognition of emotion from speech: A review". *Speech Enhancement, Modeling and Recognition Algorithms and Applications. Rijeka, Croatia: InTech*, P. 121–136. DOI: <https://doi.org/10.5772/39246>
7. Wang, L.(2022), "A Machine Learning Assessment System for Spoken English Based on Linear Predictive Coding". *Mobile Information Systems*, Vol. 2022 (5). P. 1–12. DOI: <https://doi.org/10.1155/2022/6131572>
8. Darling, D., Hinduja, J.(2022), "Feature Extraction in Speech Recognition using Linear Predictive Coding: An Overview". *i-Manager's Journal on Digital Signal Processing* Vol. 10.2. 16 p. DOI: <https://doi.org/10.26634/jdp.10.2.19289>
9. Lonce, W. (2017), "Audio spectrogram representations for processing with convolutional neural networks". *arXiv preprint arXiv*. P. 37–41. DOI: <https://doi.org/10.48550/arXiv.1706.09559>
10. Gong, Y., Chung, Y., Glass, J. (2021), "Audio spectrogram transformer". arXiv preprint arXiv:Version 3. available at: <https://arxiv.org/abs/2104.01778>
11. Qiuqiang, K. (2020), "Large-scale pretrained audio neural networks for audio pattern recognition". *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 28 (2020). P. 2880–2894. DOI: <https://doi.org/10.48550/arXiv.1912.10211>
12. Sandhya, P.(2020), "An Analysis of the Impact of Spectral Contrast Feature in Speech Emotion Recognition". *Third International Conference on Advances in Electronics, Computers and Communications (ICAIECC)*. *IEEE*, Vol. 9. No. 2. P. 87–95. DOI: <https://doi.org/10.3991/ijes.v9i2.22983>
13. Charbuty, B., Abdulazeez, A. (2021), "Classification based on decision tree algorithm for machine learning". *Journal of Applied Science and Technology Trends*, Vol. 2.01 (2021). P. 20–28. DOI: <https://doi.org/10.38094/jastt20165>
14. Breiman, L. (2001), "Random forests". *Machine learning* Vol. 45 (2001). P. 5–32. DOI: <http://dx.doi.org/10.1023/A:1010933404324>
15. Deshmukh, A.(2020), "Comparison of hidden markov model and recurrent neural network in automatic speech recognition", *European Journal of Engineering and Technology Research*, Vol. 5.8 (2020). P. 958–965. DOI: <https://doi.org/10.1051/itmconf/20235401016>
16. Ardila, R., Branson, M., Davis, K., Kohler, M., Meyer, J., Henretty, M., Morais, R., Saunders, L., Tyers, F., Weber, G. (2020), "Common Voice: A Massively-Multilingual Speech Corpus". *Proceedings of the Twelfth Language Resources and Evaluation Conference*, P. 4218–4222. DOI: <https://doi.org/10.48550/arXiv.1912.06670>
17. Goldsborough, P. (2016), "A tour of tensorflow". arXiv preprint arXiv:1610.01178 (2016). available at: <https://arxiv.org/abs/1610.01178>
18. Havrashenko, A., Barkovska, A. (2023), "Analysis of word search algorithms in the dictionaries of machine translation systems for artificial languages". *Computer systems and information technologies*. No 2. P. 17–24. DOI: <https://doi.org/10.31891/csit-2023-2-2>

19. Havrashenko, A., Barkovska, O.(2023), "Analysis of text augmentation algorithms in artificial language machine translation systems". *Advanced Information Systems*. No.7(1). P. 47–53. DOI: <https://doi.org/10.20998/2522-9052.2023.1.08>

20. Barkovska, O., Havrashenko, A., Kholiev, V., Sevostianova, O.(2021), "Automatic text translation system for artificial languages", *Computer systems and information technologies*. No. 3. P. 21–30. DOI: <https://doi.org/10.31891/CSIT-2021-5-3>

Received 30.11.2023

Відомості про авторів / About the Authors

Барковська Оlesia Юріївна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри електронних обчислювальних машин, Харків, Україна; e-mail: olesia.barkovska@nure.ua; ORCID ID: <https://orcid.org/0000-0001-7496-4353>

Гаврашенко Антон Олегович – Харківський національний університет радіоелектроніки, аспірант кафедри електронних обчислювальних машин, Харків, Україна; e-mail: anton.havrashenko@nure.ua; ORCID ID: <https://orcid.org/0000-0002-8802-0529>

Barkovska Olesia – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Electronic Computers, Kharkiv, Ukraine.

Havrashenko Anton – Kharkiv National University of Radio Electronics, PhD student at the Department of Electronic Computers, Kharkiv, Ukraine.

НЕЙРОМЕРЕЖНА МОДЕЛЬ У ЗАДАЧАХ ОБРОБЛЕННЯ ТА АНАЛІЗУ АУДІОФАЙЛІВ

Предметом дослідження є аналіз впливу етапів попереднього оброблення аудіоряду на точність визначення мови спікера. Значущість такого оброблення помітно зросла в останні роки завдяки її ключовій ролі в різноманітних застосуваннях, зокрема: зменшення обсягу інформації, фільтрація та шумопригнічення. Унаслідок збільшення попиту на рішення задач класифікації аудіоінформації оцінювання та порівняння різних методів оброблення аудіоряду стають важливими для визначення точності та ефективності отриманого рішення. **Мета роботи** – аналіз впливу методів попереднього оброблення аудіоінформації на точність визначення мови спікера для використання в подальшому навчанні нейромережі для різних способів перетворення сигналів в ознаки, а саме спектрограми та мелчастотні кепстральні коефіцієнти. Для досягнення поставленої мети були визначені такі **завдання**: проаналізувати способи перетворення сигналу в ознаки та аналіз математичних моделей для виконання аналізу аудіоряду за обраними ознаками; розробити узагальнену модель перекладу мови спікера в реальному часі та спланувати експеримент залежно від обраних етапів попереднього оброблення аудіоряду; змодельовати експеримент способом навчання та тестування згортової нейронної мережі. Використані такі **методи**: мелчастотний кепстральний аналіз, спектральний аналіз, математичні методи штучного інтелекту. **Досягнуті результати**: залежно від обраних методів попереднього оброблення голосової інформації та різних способів перетворення сигналу в ознаки вдалося досягти 93% точності визначення мови спікера. Практичною значущістю цієї роботи є збільшення точності подальшого транскрибування аудіоінформації та перекладу сформованого тексту обраною мовою, зокрема штучними мовами. **Висновки**. У процесі роботи було обрано найкращу послідовність етапів попереднього оброблення аудіоінформації з метою використання в подальшому навчанні нейромережі для різних способів перетворення сигналів в ознаки. Для вирішення окресленої задачі краще підходять мелчастотні кепстральні коефіцієнти. Оскільки точність нейромережі залежить від її структури, то зі збільшенням обсягів вхідної інформації та кількості мов результати можуть змінюватися. Але на певному етапі було прийнято рішення використовувати лише мелчастотні кепстральні коефіцієнти з нормалізацією на етапі попереднього підготовки вхідної інформації.

Ключові слова: мелчастотні кепстральні коефіцієнти; спектрограма; часова маска; частотна маска; нормалізація; нейромережа; голос; аудіоряд; мова.

Бібліографічні описи / Bibliographic descriptions

Барковська О. Ю., Гаврашенко А. О. Нейромережна модель у задачах оброблення та аналізу аудіофайлів. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 16–23. DOI: <https://doi.org/10.30837/ITSSI.2023.26.016>

Barkovska, O., Havrashenko, A. (2023), "Analysis of the influence of selected audio pre-processing stages on accuracy of speaker language recognition", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 16–23. DOI: <https://doi.org/10.30837/ITSSI.2023.26.016>

O. GAVRILENKO, M. MIAHKYI

FORECASTING THE CRYPTOCURRENCY EXCHANGE RATE BASED ON THE RANKING OF EXPERT OPINIONS

To date, most existing cryptocurrency exchanges do not have in their arsenal tools that would allow them to verify and investigate the information disseminated on social networks regarding a particular cryptocurrency. This makes it possible to conduct a relevant research with the subsequent development of a tool that, if used correctly, will provide users with advisory advice on further actions in relation to the cryptocurrency under study in the system. Based on this advice, interested parties will be able to adjust their decisions regarding further financial steps. The basis of most recommender systems is always the need to identify some influencing factors, which are later given certain weights to facilitate and simplify the formulation of further advice for users. In this paper, we study the influence of celebrity publications on the formation of prices for a particular cryptocurrency at a certain point in time. The importance and existence of this influence was previously proven by statistical methods. **The purpose of the study** is to develop an algorithm for studying the level of influence of posts of each of the selected group of experts in social networks on the cryptocurrency rate. **The object of the study** is the forecast of cryptocurrency rates. The input data used were the list of experts whose level of influence will be studied, the time interval of the study, the number of posts made by each of the experts in question over the specified period of time, and the actual cryptocurrency rates for the relevant period. The experts were well-known personalities who are either knowledgeable in the field of finance in general and cryptocurrencies in particular, or whose activities are somehow related to a particular cryptocurrency. **Research methods.** Experts are ranked based on the full probability and Bayesian formulas. Forecasting of cryptocurrency rates in a selected period of time is carried out using the algorithm for forecasting cryptocurrency rates based on expert posts on social networks (ATAPSN). To control the accuracy of forecasts, the relative average error is calculated. Recommendations for financial transactions with cryptocurrencies are formed by entering the critical value of the exchange rate and calculating the arithmetic mean of cryptocurrency exchange rates for a specified period of time. **Results.** As a result of the research, an algorithm has been developed that allows taking into account the impact of the posts of each of the selected ranked group of experts on changes in the rates of a particular cryptocurrency. On the basis of the obtained forecasts, the paper presents a methodology for forming recommendations for financial transactions with them.

Keywords: cryptocurrency exchange rate; forecasting algorithm; social media posts; ranking of a group of experts; information technology of intellectual analysis.

1. Introduction

At the moment, there are a huge number of publications and various information on various cryptocurrencies, which leads to a constant growth in the popularity of this topic among all age groups. This is due to the relative ease of entering cryptocurrency exchanges for further interaction with cryptocurrencies. Users of most cryptocurrency exchanges are provided with a fairly wide range of financial transactions and other opportunities. The simplest and therefore the most popular are buying and selling cryptocurrencies. These operations require minimal skills from the user, but they also do not have high liquidity. That is, the exchange does not provide users with conditions for rapid enrichment through the use of basic skills. However, developers and owners of cryptocurrencies are interested in the widest possible distribution of their own products, so exchanges, in turn, try to encourage people to use more "advanced" technologies and

opportunities provided by the system. Certain tools allow you to analyze the state of the market in relation to a selected cryptocurrency and thus formulate a theory about its future course. These information manipulations, once disseminated on social media, encourage many users to take appropriate action.

Thus, the relevance of this study is due to the growth of an already large amount of information about cryptocurrencies, which is aimed specifically at encouraging people to take appropriate actions in relation to certain cryptocurrencies.

However, most of the existing cryptocurrency exchanges currently do not have tools in their arsenal that would allow them to verify and investigate the information disseminated on social media regarding a particular cryptocurrency. This allows us to conduct a relevant study with the subsequent development of a corresponding tool that, if used correctly, will provide users with advisory advice on further actions regarding the cryptocurrency under study in the system. Based on

this advice, interested parties will be able to adjust their decisions regarding further financial steps.

Most recommender systems are always based on the need to identify certain factors of influence, which are later given certain weights to facilitate and simplify the formulation of further advice for users.

In our case, it is the influence of publications of famous people on the formation of prices for a certain cryptocurrency at a certain point in time that is being studied. The importance and existence of this influence has been proven by the authors of this paper in previous studies.

2. Literature review and problem statement

In general, the collection and processing of data from publications on cryptocurrencies from social networks and other online information platforms is a rather significant task, as they can provide a lot of useful and comprehensive information that will be necessary to build a high-quality mathematical model for further forecasting.

Paper [1] discusses the process of computer detection and categorization of opinions expressed in a piece of text in order to determine whether the writer's attitude toward a particular topic, product, etc. is positive, negative, or neutral. In the present study, a detailed study was conducted: sentiment analysis and its cause-and-effect relationship. Also, with the help of sentiment analysis, a generalized event was determined on its basis and taking into account the time. The results of the analysis of the cause-and-effect relationship can be used not only to determine the causes and effects, but also to further predict user sentiment. The main part of the publication is an overview of the combination of these approaches, which are combined into a single model that allows you to determine the mood during future events, as well as create a time forecast for the length of the interval between certain events. The average relative error was used to assess the accuracy.

To search for publications that meet certain requirements, such as the number, a single text format, and others, you need to select a specific social network. Twitter (currently X) meets the requirements well, and work [2] describes in detail the special linguistic analysis and statistics in this social network. The main purpose of the study, the authors noted, was to identify criminal elements in the United States by modeling topics of discussion and then incorporating them into a crime prediction model. A thorough analysis of

the impact of publications in social networks on the potential for certain criminal acts to occur in the future was conducted.

Paper [3] provides a comprehensive reference for researchers and practitioners, covering all areas that contribute to the construction and analysis of social networks.

Paper [4] is quite relevant today due to the difficult epidemiological situation in the world. It analyzed microblogs on Twitter and proposed several methods for identifying messages. It was determined that over ten weeks, out of more than five hundred thousand reports, the best model achieved a correlation of 0.78 using the CDC statistical method.

Also, one should not miss online blogs, where many people express their own opinions and visions of certain problems. In [5], a study was conducted to identify hate groups. The proposed approach is semi-automatic and consists of four modules, namely: blog spider, information retrieval, network analysis, and visualization. The study was conducted on the Xanga blog site. The results of the analysis were to identify some interesting demographic and topological characteristics in hate groups and to identify at least two large communities in addition to smaller ones. The proposed approach is also appropriate for studying hate groups and other related communities on blogs.

For business and the financial market, the process of analyzing large amounts of data and understanding the needs of the majority of people is very important, as it directly affects the income of the company and individuals. The research conducted in [6] was aimed at identifying the dominant factors that lead to currency crises. Also, this publication is intended to identify and characterize currency crises, as well as to predict the potential occurrence of the latter at an early stage. This will save managers some time to improve their crisis management policies and adjust their responses.

Paper [7] investigated the dynamics of linear and nonlinear, serial dependencies in financial time series within a moving window. In particular, attention is focused on identifying episodes of statistically significant two- and three-point correlation in the returns of several leading exchange rates, which may offer some potential for their predictability. The moving window approach was used to capture the dynamics of correlation for different window lengths and to analyze the distribution of periods with statistically significant correlations. It was found that for sufficiently large window lengths, these distributions correspond well to a power law.

The predictability itself is measured by the hit rate, i.e., the level of consistency between the actual return features and their predictions obtained using a simple correlation-based predictor.

It should be noted that all of the above studies are general in nature and do not provide results of forecasting currency rates, including cryptocurrencies. Accordingly, the factors that influence them were not investigated.

In [8], the authors study the main macroeconomic indicators of influence on the US dollar exchange rate in Ukraine: purchase/sale of cash currency, purchase/sale of non-cash currency, balance of purchase/sale of cash and non-cash currency, current year inflation, nominal and real GDP, purchase/sale by bank customers, transactions between banks, gross and net international reserves, unemployment rate, discount (interest) rate, balance of foreign exchange interventions, and volume of transactions of nominal value. The main economic components of exchange rate formation were identified using the principal components method. Using the statistical models ARIMA, Exponential Smoothing and SSA, the values of the selected factors of influence are predicted. The values of exchange rates are forecasted using regression models built by Fast Tree, Fast Forest, Fast Tree Tweedie and Gam algorithms, and the obtained values are tested for accuracy. This work did not forecast cryptocurrency rates and did not study the impact of such a factor as publications in social networks.

Paper [9] analyzes the methods, areas of application, and approaches to analyzing publications and forecasting events based on the collected data, and presents a model for assessing the impact of publications on changes in the cryptocurrency rate, taking into account the posts of only one expert. The author justifies the relevance of the topic and describes the possibilities of appropriate application of the results of the work. The main stages of working with event forecasting data are identified, namely: data pre-processing, further analysis and forecasting. This paper did not study the level of influence of publications of a group of experts in social networks on the cryptocurrency rate.

Paper [10] presents an algorithm for assessing the impact of the publications of a "main expert" on the cryptocurrency rate. This paper describes the process of determining the most influential ("main") expert and obtaining forecasts based on his or her posts on the rate of the selected cryptocurrency. This paper did not investigate the level of influence of all selected experts' publications on social media on the cryptocurrency rate. However, this algorithm is a transitional stage

to algorithms for predicting the cryptocurrency rate taking into account the posts of each of the selected group of experts.

As part of the research presented in papers [8–10], appropriate information systems were created to implement the above tasks of data mining.

The above analysis shows that the influence of certain factors on the cryptocurrency rate, in particular the influence of posts by famous people in social networks, is still not sufficiently developed and requires further research.

3. Aim and objectives of the study

The aim of the study is to develop an algorithm for studying the level of influence of posts of each of the selected group of experts in social networks on the cryptocurrency rate.

This algorithm will potentially provide an opportunity to increase the reliability of the forecast regarding the rate of the selected cryptocurrency for the purpose of further formulating recommendations.

To achieve this aim, the following tasks were set:

- to select a group of experts;
- to rank the selected group of experts according to the level of influence on the rate of the specified cryptocurrency;
- to obtain a forecast for the rate of a certain cryptocurrency, taking into account the posts in a certain social network of the selected group of experts based on the previously conducted ranking of experts;
- to formulate recommendations for financial transactions with cryptocurrencies based on the obtained forecasts.

4. Materials and methods of the study

The object of the study is the forecast of cryptocurrency rates.

The information required to analyze the level of influence of social media posts on cryptocurrency rates is a list of experts whose level of influence will be studied, the time interval of the study, the number of posts made by each of the experts in question during the specified period of time, as well as the actual cryptocurrency rates for the relevant period.

The experts were well-known personalities who are either knowledgeable in the field of finance in general and cryptocurrencies in particular, or whose activities are somehow related to a particular cryptocurrency.

Approximate samples of the datasets are shown in Table 1 and Table 2.

Table 1. The rate of the selected cryptocurrency for the specified period of time

A moment in time	Rate value
...	...
t_1	x_1
t_2	x_2
...	...
t_s	x_s
...	...

Table 2. Number of posts by selected experts for the specified period of time

Expert	Number of posts	Number of posts related to cryptocurrency
Expert 1	m_1	k_1
Expert 2	m_2	k_2
...		
Expert n	m_n	k_n

Table 1 shows the rates of the selected cryptocurrency x_1, x_2, \dots, x_s in the selected time period t_1, t_2, \dots, t_s , which can be taken from the Binance crypto exchange website [11].

In Table 2, m_1, m_2, \dots, m_n are the frequencies of expert posts, k_1, k_2, \dots, k_n are the frequencies of expert posts related to a particular cryptocurrency.

The data generated in this way is the input for this study. In the future, it is necessary to:

- analyze the number of posts by selected experts on the social network;
- to rank the group of experts;
- to obtain a forecast of the cryptocurrency rate, taking into account the posts of the selected group of experts after their ranking;
- formulate recommendations for financial transactions in relation to the selected cryptocurrency.

The use of a mathematical apparatus based on the full probability and Bayesian formulas makes it possible to use this information to rank the selected group of experts, depending on the probability with which each expert will post on a social network during the period under consideration [10].

Forecasting the rates of the selected cryptocurrency is carried out using the ATAPSN algorithm [9] (an algorithm based on posts in social networks). To control the accuracy of forecasts, the average relative error is calculated.

Recommendations for financial transactions with cryptocurrencies are formed by entering a critical rate value and calculating the arithmetic mean of cryptocurrency rates for a specified period of time.

5. An algorithm for predicting cryptocurrency rates based on posts of a group of experts in social networks

5.1. Selection of a group of experts

Problem statement:

From the set of users of a social network, we select a subset $A = (a_1, a_2, \dots, a_n)$ of users who satisfy the following requirements [12]

- the user must be a public figure;
- each user must be active in the selected social network and have a significant number of subscribers;
- users have different professional activities;
- the main professional activity of the users is somehow related to the use of cryptocurrency;
- each pair of users a_i and a_j , $i, j = 1, 2, \dots, n$, $i \neq j$, does not maintain communication in the selected social network (they are not friends and do not respond to each other's posts);
- each user has a sufficient qualification level in the financial sector.

Let's call these users experts. In the future, we will take their posts in the selected social network into account when predicting the rate of a certain cryptocurrency.

Let's assume that over a certain period of time, experts have made m posts in a social network, and k of them are related to a certain cryptocurrency. We consider the context of the posts to be arbitrary. Expert a_1 published m_1 posts during the specified period of time, of which k_1 posts are related to a certain cryptocurrency, expert a_2 published m_2 posts, of which k_2 posts are related to a certain cryptocurrency, etc. The last expert a_n has published m_n posts, of which k_n posts are related to a certain cryptocurrency. Then:

$$\begin{aligned} m_1 + m_2 + \dots + m_n &= m. \\ k_1 + k_2 + \dots + k_n &= k, \end{aligned} \quad (1)$$

where m_1, m_2, \dots, m_n – frequency of expert posts;

k_1, k_2, \dots, k_n – the frequency of expert posts related to a particular cryptocurrency.

It is necessary to calculate the frequency of posts of all selected experts for an optional time interval [13].

Justification:

Such a choice of experts is due to the need to form a set of such experts who will be independent of each other both in the space of the chosen social network and in the professional space.

**5.2. Ranking of a group of experts
by the level of influence on the rate
of the selected cryptocurrency**

Problem statement:

Based on the list of experts: $A = (a_1, a_2, \dots, a_n)$ obtained in paragraph 5.1. and taking into account the frequencies of their posts in the selected social network for the specified period of time – m_1, m_2, \dots, m_n (also k_1, k_2, \dots, k_n), given in Table 2, it is necessary to rank the experts in terms of their influence on the rate of the selected cryptocurrency.

The formulated problem can be easily interpreted as a classical probabilistic problem: for a certain period of time, m posts were written. It is known that n experts published posts during this period, where m_1, m_2, \dots, m_n are the frequencies of expert posts, k_1, k_2, \dots, k_n are the frequencies of expert posts related to the selected cryptocurrency. Event A means that in an arbitrary period of time one of the experts wrote a post related to the selected cryptocurrency. It is necessary to determine which expert is more likely to have made this post [14].

Justification:

Event A – at any moment of time t from the interval $[0; T]$ (in the table 1 $t_1 = 0$, $t_s = T$) a post related to the chosen cryptocurrency was written.

Hypothesis H_1 – the post was written by an expert 1,

hypothesis H_2 – the post was written by an expert 2,

...

hypothesis H_n – the post was written by an expert n .

Let us assume that events H_i and H_j , $i, j = 1, 2, \dots, n$, $i \neq j$ are pairwise independent. These assumptions can be made on the basis of the list of requirements that experts must meet (see Section 5.1).

According to the full probability formula:

$$P(A) = \sum_{i=1}^n P(H_i)P(A/H_i) \quad (2)$$

where

$$P(H_i) = m_i/m, \quad P(A/H_i) = k_i/m_i, \quad (3)$$

m_i – number of publications made by the i -th expert;

k_i – the number of publications made by the i -th expert related to the selected cryptocurrency;

m – total number of publications for the period $[0; T]$;

$P(H_i)$ – the probability that the post was published by the i -th expert;

$P(A/H_i)$ – is the probability that a post related to the selected cryptocurrency was written at any time, provided that this post was written by the i -th expert.

Then, using the Bayesian formula: for each expert, we calculate the probability that the i -th expert created the post, if it is known that it was definitely written by one of the experts during the period under consideration.

$$P(H_i/A) = \frac{P(H_i)P(A/H_i)}{P(A)}, \quad (4)$$

where $P(H_i)$ – the probability that the post was published by the i -th expert;

$P(A/H_i)$ – is the probability that at any time t from the interval $[0; T]$ a post related to the selected cryptocurrency was written, provided that the publication was written by the i -th expert, $P(H_i/A)$ – is the probability that the post was written by the i -th expert, provided that we know that at any time t a post related to the selected cryptocurrency was created.

The obtained a posteriori probabilities are arranged in descending order. This means that the expert with the highest probability $P(H_i/A)$ will have the greatest impact on the rate of the selected cryptocurrency in the interval $[0; T]$. Next, the expert with the next highest probability $P(H_i/A)$ is selected, and so on. As a result, we get a ranked list of experts according to the impact of their posts on social media on the cryptocurrency rate $A^* = (a_1^*, a_2^*, \dots, a_n^*)$.

It should be noted that such a ranking also allows you to assess the impact of a group of experts on the rate of a particular cryptocurrency over the forecast period $[T; T + \Delta t]$, since the impact of publications made during the time period $[0; T]$ also extends to a certain period of time $[T; T + \Delta t]$.

In addition to the above, the following ranking criteria can also be used [15]:

- the number of subscribers is directly proportional to the value of the influence of this expert on a certain audience;
- activity – frequency and regularity of publications on the topic of cryptocurrencies;
- forecast history – the percentage of successful forecasts;
- social capital – cooperation with financial organizations;
- regional influence – influence in a specific geographical region;
- thematic socialization – focus on specific cryptocurrencies or blockchain technologies;
- time factor – the relevance of information and the speed of its dissemination.

It should be noted that the use of each of these alternative criteria for ranking experts requires the use of a unique mathematical apparatus and is planned for further research.

5.3. Obtaining a forecast of the cryptocurrency rate, taking into account the posts in the social network of the selected group of experts based on the ranking of experts

Problem statement:

Let there be some training set $X = (x_1, x_2, \dots, x_s)$, where s – is the sample size, which consists of the actual rate of a certain cryptocurrency for the time period $[0; T]$ (the sample is based on the time series from Table 1). It is necessary to obtain forecasts of the rate of the selected cryptocurrency for the next period of time $[T; T + \Delta t]$, taking into account the ranking of experts carried out in paragraph 5.2.

Justification:

The proposed algorithm is based on the following principle: each selected expert has a certain influence on the cryptocurrency rate. In other words, in order to obtain a forecast with the desired accuracy, it is necessary to gradually take into account the posts of not only the "main" (most influential) expert, but also others [10].

Thus, taking into account the posts of each expert from the most influential to the least influential is an iterative process consisting of the following steps:

Step 0. Set a threshold for the mean relative prediction error, as well as the values of T and Δt .

Stage 1

Step 1.1. Using the ATAPSN [9], using the training sample X , obtain forecasts of cryptocurrency rates for the period $[T; T + \Delta t]$, taking into account the posts of the expert who is the first in the ranked list.

Step 1.2. Form a sample X from the time series shown in Table 1 for the time period $[T; T + \Delta t]$.

Step 1.3. From the forecasts obtained in step 1.1 for the time period $[T; T + \Delta t]$, form a sample Y .

Step 1.4. To check the accuracy of the forecast, calculate the average relative forecasting error:

$$MAPE_l = \frac{1}{s} \sum_{l=1}^s \frac{|x_l - y_l|}{x_l} 100\%, \quad (5)$$

where x_l – elements of the sample X ,

y_l – elements of the sample Y ,

$l = \overline{1; s}$, s – the volume of samples X and Y [16].

Step 1.5. If the $MAPE_l$ value does not exceed the specified threshold, the algorithm is complete. If the $MAPE_l$ value exceeds the specified threshold, then proceed to the next step.

Stage 2

Step 2.1. Using the ATAPSN [9], using the training sample X and the forecasts from step 1.1, obtain forecasts of cryptocurrency rates for the period $[T + \Delta t; T + 2\Delta t]$, taking into account the posts of the expert who is second in the ranked list.

Step 2.2. Form a sample X from Table 1 for the time period $[T + \Delta t; T + 2\Delta t]$.

Step 2.3. From the forecasts obtained in step 2.1 for the time period $[T + \Delta t; T + 2\Delta t]$, form a sample Y .

Step 2.4. To verify the accuracy of the forecast, calculate the $MAPE_l$ using formula (5).

Step 2.5. If the $MAPE_l$ value does not exceed the specified threshold, the algorithm is complete. If the $MAPE_l$ value exceeds the specified threshold, then proceed to the next step.

...

Stage n

Step n.1. Using the ATAPSN [9], using the training sample X and the forecasts from step $n-1.1$, obtain forecasts of cryptocurrency rates for the period $[T + (n-1)\Delta t; T + n\Delta t]$, taking into account the posts of the expert who is next in the ranked list.

Step n.2. Form a sample X from Table 1 for the time period $[T + (n-1)\Delta t; T + n\Delta t]$.

Step n.3. From the forecasts obtained in step n.1 for the time period $[T + (n-1)\Delta t; T + n\Delta t]$, form the sample Y .

Step n.4. To check the accuracy of the forecast, calculate the $MAPE_i$ using formula (5).

Thus, the maximum possible number of steps of the algorithm after receiving forecasts of cryptocurrency rates for the period $[T + (n-1)\Delta t; T + n\Delta t]$, taking into account the posts of the expert who is the last in the ranked list and calculating the corresponding $MAPE_i$, is n .

It should also be noted that at each stage, the sample X is formed at the corresponding time interval based on the time series from Table 1, and the sample Y is built from the forecasts of the cryptocurrency rate at the same time interval at the same time points as the corresponding sample X . This makes it possible to justify the legitimacy of the transition from time series to statistical samples and thus correctly calculate all MAPEs.

5.4. Formation of recommendations for financial transactions regarding cryptocurrencies based on the obtained forecasts

Problem statement:

Based on the forecast of the rates of the selected cryptocurrency obtained in clause 5.3, formulate recommendations for possible further financial transactions in relation to this cryptocurrency.

Justification:

Obtaining sufficiently accurate forecasts of the rates of the selected cryptocurrency (see paragraph 5.3) allows you to track the trends of their change over the period $[T + (n-1)\Delta t; T + n\Delta t]$. Based on this, it is possible to make recommendations on the expediency of buying or selling cryptocurrencies in order to make a profit.

It should be noted that when making recommendations, it is necessary to take into account the needs of users. To do this, it is necessary to set a critical value

$$y_{cr} = \begin{cases} y_{am} \geq y_{cr} \Rightarrow \text{cryptocurrency should be sold;} \\ y_{am} < y_{cr} \Rightarrow \text{cryptocurrency should be bought,} \end{cases} \quad (6)$$

where y_{am} – is the arithmetic mean of the sample Y , calculated by the formula:

$$y_{am} = \frac{\sum_{l=1}^s y_l}{s}, \quad (7)$$

y_l – sample elements, $l = \overline{1; s}$, s – sample volume.

The value of y_{cr} can be set at the discretion of the user or, based on the opinion of an expert, according to the situation on the cryptocurrency market.

6. Conclusions

This paper presents a modification of the ATAPSN algorithm [9] and expands the list of requirements for selecting experts [12].

The proposed algorithm makes it possible to take into account the level of influence of each expert from the selected group. At the same time, the use of ATAPSN [9] at each stage of the algorithm can significantly simplify the process of obtaining forecasts.

This approach makes it possible to increase the accuracy of forecasts of the rates of the selected cryptocurrency by taking into account the contributions of each of the selected experts, in contrast to the algorithms presented in [9] and [10], in which either a single expert was selected or the "main" expert was determined from the group of experts. That is, in both cases, the potential influence of other experts was neglected.

The proposed algorithm is iterative. At each stage, the posts of one expert were taken into account in the order of the ranked list. That is, at each stage, the previous forecast is refined in case of an unsatisfactory error.

The number of stages of the algorithm, depending on the value of the $MAPE_i$, can vary from 1 to n . This depends on the number of experts whose posts need to be taken into account to achieve the required forecast accuracy.

This approach calculates the a posteriori probabilities that a post related to the selected cryptocurrency was written by a particular expert during the forecasting interval. They were used to rank the experts.

It should be noted that for different time intervals, different expert ranking results can be obtained.

To use this approach, it is recommended to consider small time intervals (up to a week), each of which can be used to rank experts more accurately. This should potentially increase the accuracy of forecasts of the selected cryptocurrency rates in the specified time interval.

Using the frequency of posts in social media as a parameter for determining the influence of experts allows us to apply the classical apparatus

of probability theory, which guarantees the correctness of the results obtained.

The obtained forecasts are used to generate recommendations for buying or selling the selected cryptocurrency, depending on the user's needs and the situation on the cryptocurrency market.

The disadvantages include the fact that the accuracy of the forecast may be negatively affected by an unsuccessfully chosen time interval for which the forecast is made, since determining the duration of the impact of a particular expert's forecast is beyond the scope of the task at hand.

Also, the quality of the forecast depends on the selected group of experts, as a poorly

selected expert group can negatively affect the algorithm's performance.

These features indicate the need for constant monitoring of both cryptocurrency rates and expert posts on social media.

In order to further improve the accuracy of cryptocurrency rate forecasts and the relevance of recommendations for buying or selling it, it is planned to improve the obtained algorithm by using alternative criteria for ranking experts [15], developing alternative algorithms and incorporating them into the general information technology for determining the impact of social media posts on cryptocurrency rates.

References

1. Preethi, P., Uma, V., Kumar. A. (2015), "Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction", *Procedia Computer Science*. №48. P. 84–89. DOI: <https://doi.org/10.1016/j.procs.2015.04.154>
2. Gerber, Matthew S. (2014), "Predicting crime using Twitter and kernel density estimation". *Decision Support Systems*. №61. P. 115–125. DOI: <https://doi.org/10.1016/j.dss.2014.02.003>
3. Alhadjj, R., Rokne, J., (eds). (2018), "Encyclopedia of Social Network Analysis and Mining", *Analysis and Mining*. Springer, New York, NY. P. 2699. DOI: <https://doi.org/10.1007/978-1-4614-7163-9>
4. Culotta, A., Huberman, B. "Towards detecting influenza epidemics by analyzing Twitter messages", *Proceedings of the First Workshop on Social Media Analytics*. ACM. P. 115–122. 2010. URL: http://snap.stanford.edu/soma2010/papers/soma2010_16.pdf (accessed: 13.10.2023).
5. Chau, M., Xu, J. (2007), "Mining communities and their relationships in blogs: A study of online hate groups", *Int. J. Hum.-Comput. Stud.* 65(1). P. 57–70. DOI: <https://doi.org/10.1016/j.ijhcs.2006.08.009>
6. Karahoca, D., Karahoca, A., Yavuz, Ö. (2013), "An early warning system approach for the identification of currency crises with data mining techniques", *Neural Comput & Applic.* Volume 23. P. 2471–2479. DOI: <https://doi.org/10.1007/s00521-012-1206-9>
7. Žukovič, M. (2012), "Dynamics of episodic transient correlations in currency exchange rate returns and their predictability". *centr.eur.j.phys.* Volume 10. P. 615–624. DOI: <https://doi.org/10.2478/s11534-011-0120-6>
8. Gavrilenko, O., Novakivska, K., Shumeiko, O. (2022), "Select the most influx economic factors for forecasting the US dollar exchange rate", *Bulletin of NTU*. No. 54. P. 26–35. DOI: <https://doi.org/10.33744/2308-6645-2022-4-54-026-035>.
9. Gavrylenko, O., Miahkyi, M., Zhurakovskiy, Y. (2022), "The task of analyzing publications to build a forecast for changes in cryptocurrency rates", *Adaptive automatic control systems*. Volume 2. No. 41. P. 90–99. DOI: <https://doi.org/10.20535/1560-8956.41.2022.271349>
10. Bidyuk, P., Gavrilenko, O., Myagkyi, M. (2023), "The algorithm for predicting the cryptocurrency rate taking into account the influence of posts of a group of famous people in social networks", *System research and information technologies: an international scientific and technical journal*. № 2. P. 22–34. DOI: <https://doi.org/10.20535/SRIT.2308-8893.2023.2.02>
11. Binance cryptocurrency exchange. URL: <https://www.binance.com/en/strategy/spot/grid/DOGEUSDT>. (accessed: 13.10.2023).
12. Gavrilenko, O., Myagky, M. (2023), "Information system for analyzing the influx of publications of experts on the course of cryptocurrency exchanges based on the rich-agent approach", *IV International Conference "Software Security Engineering and Advanced Information Technologies" (SoftTech-2023)*. P. 67–70. URL: <https://ela.kpi.ua/handle/123456789/59988> (accessed: 13.10.2023).
13. Kartashov, M. "Internationality, processes, statistics, Publishing and printing center". Kiev University, Kiev, 2008. 504 p. URL: https://probability.knu.ua/userfiles/kmv/VPS_Pv.pdf (accessed: 13.10.2023).
14. Denisyuk, V., Bobkov, V., Pogrebetska, T., Repeta, V. "Higher mathematics. Part 4. The theory of probabilities and mathematical statistics", Edition "NAU-druk", Kiev, 2009. 256 p. URL: <https://studfile.net/preview/5797767/> (accessed: 13.10.2023).
15. Gavrilenko O., Myagkyi M. "Study of the influence of the expert group's posts on the course of cryptocurrencies", *XIX International Scientific Conference named after academician Mykhailo Kravchuk Abstracts of reports*.2023. P. 162–164. URL: <https://matan.kpi.ua/media/2023/kravchuk-conf-2023/kravchuk2023-abstracts.pdf> (accessed: 13.10.2023).
16. Time series. URL: <https://kstat.pnu.edu.ua/wp-content/uploads/sites/63/2018/04/%D0%A7%D0%B0%D1%81%D0%BE%D0%B2%D1%96-%D1%80%D1%8F%D0%B4%D0%B8.pdf> (accessed: 13.10.2023).

Відомості про авторів / About the Authors

Гавриленко Олена Валеріївна – кандидат фізико-математичних наук, доцент, Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", доцент кафедри інформаційних систем та технологій, Київ, Україна; e-mail: gelela1980@gmail.com; ORCID ID: <https://orcid.org/0000-0003-0413-6274>

Мякий Михайло Юрійович – Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", аспірант кафедри інформаційних систем та технологій, Київ, Україна; e-mail: mishamyagkiy3@gmail.com; ORCID ID: <https://orcid.org/0000-0002-8038-8839>

Gavrilenko Olena – PhD (Physics and Mathematics Sciences), Associate Professor, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Associate Professor at the Department of Information Systems and Technologies, Kyiv, Ukraine.

Miahkyi Mykhailo – National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Postgraduate Student at the Department of Information Systems and Technologies, Kyiv, Ukraine.

ПРОГНОЗУВАННЯ КУРСУ КРИПТОВАЛЮТИ НА ОСНОВІ РАНЖУВАННЯ ЕКСПЕРТНИХ ОЦІНОК

На сьогодні більшість криптовалютних бірж не мають інструментів, що дали б змогу перевірити й дослідити інформацію, поширену в соціальних мережах щодо тієї чи іншої криптовалюти. Це допомагає організувати відповідне дослідження з подальшим розробленням необхідного інструменту, що за умови коректного застосування надаватиме користувачам рекомендаційні поради про подальші дії щодо досліджуваної в системі криптовалюти. Ґрунтуючись на цих порадах, зацікавлені особи зможуть коригувати свої рішення щодо подальших фінансових кроків. В основі загальної більшості рекомендаційних систем завжди лежить необхідність установаження деяких факторів впливу, яким пізніше надаються певні вагові коефіцієнти для сприяння та спрощення формування подальших порад для користувачів. У цій роботі розглядається саме вплив публікацій відомих осіб на формування цін за певну криптовалюту в конкретний момент часу. Важливість і наявність зазначеного впливу доведено раніше з допомогою статистичних методів. **Мета дослідження** – розроблення алгоритму для вивчення рівня впливу дописів кожного експерта з обраної групи в соціальних мережах на курс криптовалюти. **Об'єктом роботи** є прогноз курсу криптовалют. Як вхідні показники застосовувалися перелік експертів, рівень впливу яких вивчатиметься; інтервал часу дослідження; кількість дописів, що зробив кожен із розглянутих експертів за вказаний період часу; реальні курси криптовалют за відповідний період. Експерти обиралися з-поміж відомих особистостей, які обізнані в галузі фінансів загалом і криптовалют зокрема, або діяльність яких так чи так пов'язана з певною криптовалютою. **Методи дослідження.** Ранжування експертів проводиться на основі формул повної ймовірності та Баєса. Курси криптовалют за обраний період часу прогноуються за допомогою певного алгоритму, зважаючи на дописи експертів у соціальних мережах (ATAPSN). Для контролю точності прогнозів обчислюється відносна середня похибка. Рекомендації щодо фінансових операцій із криптовалютою формуються способом уведення критичного значення курсу та обчислення середнього арифметичного курсів криптовалют за вказаний період часу. **Здобуті результати.** Унаслідок досліджень було розроблено алгоритм, що дає змогу брати до уваги вплив дописів кожного експерта з обраної ранжованої групи на зміни курсів певної криптовалюти. На основі отриманих прогнозів у роботі подана методика формування рекомендацій щодо фінансових операцій з ними.

Ключові слова: курс криптовалюти; алгоритм прогнозування; дописи в соціальних мережах; ранжування групи експертів; інформаційна технологія інтелектуального аналізу.

Бібліографічні описи / Bibliographic descriptions

Гавриленко О. В., Мякий М. Ю. Прогнозування курсу криптовалюти на основі ранжування експертних оцінок. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 24–32. DOI: <https://doi.org/10.30837/ITSSI.2023.26.024>

Gavrilenko, O., Miahkyi, M. (2023), "Forecasting the cryptocurrency exchange rate based on the ranking of expert opinions", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 24–32. DOI: <https://doi.org/10.30837/ITSSI.2023.26.024>

M. GRINCHENKO, M. ROHOVYI

A MODEL FOR IDENTIFYING PROJECT SPRINT TASKS BASED ON THEIR DESCRIPTION

The subject of research in this article is the identification of project sprint tasks. **The purpose of the article** is to find approaches to reducing the risks of not fulfilling sprint tasks. The article solves the following **tasks**: analyzing research on the classification and visualization of project tasks, developing an algorithm that can automatically classify text descriptions of sprint tasks, collecting and preparing a training sample of text descriptions of sprint tasks for training and testing the classification model, applying natural language processing methods to improve classification and ensure the accuracy of the results, validating the model on real data to assess the efficiency and accuracy of classification, and analyzing the results. The following **methods** have been used: machine learning methods for classification, text vectorization methods, methods for classifying text descriptions, natural language processing methods, methods for semantic analysis of task description text, methods for processing expert opinions. The following **results** were obtained: a comprehensive approach to using machine learning algorithms, including the collection and processing of textual descriptions of tasks, for classification and involvement of expert opinions to improve the quality of task perception by the project team. Text expressions were classified based on the Bayesian classifier and neural classifiers. A visual representation of the data was implemented. Semantic analysis of the text of the description and title of the tasks was performed. Data markup was obtained to classify the quality of the wording, which was performed by a team of experts. To measure the reliability of the obtained expert assessments, we calculated Cohen's kappa coefficient for each pair of markers. According to the experimental results, the accuracy of the Bayesian classifier is 70%. For the classifier based on deep learning, a neural network for binary classification based on the transformer architecture was selected. The neural network was trained using the Python programming language and deep learning frameworks. The result is a classifier that gives an accuracy score of 83% on a test dataset, which is a good result for a small dataset and data with conflicting labels. **Conclusions**: the analysis of textual data confirms that the existing data in the tracking system is incomplete and contains abbreviations, conventions, and slang. The results show that the assessment of the quality of the wording is determined by the level of expert knowledge of the specifics and context of the project, while increasing the number of experts has almost no effect on the result. In further research, it is recommended to test the hypothesis that the effectiveness of the classifier depends on the specific project and the use of unsupervised learning methods for the task of identifying the quality of formulations.

Keywords: project; task description; project task management system; model; classifier; vector representation.

Introduction

Nowadays, flexible or adaptive management methodologies are widely used to implement IT projects. The use of these methodologies makes it possible to respond quickly to changes and risks that arise during the development of a software product. The effectiveness of the project team is influenced by a large number of factors [1], among which an inaccurate description of the project tasks or their misunderstanding by developers is significant. Thus, identifying ambiguities in sprint task descriptions often becomes critical to successful software development. This allows you to reformulate unclear tasks in a timely manner and avoid misunderstandings between developers.

Tracking systems (also known as task management systems or project management systems) play a key role in IT project management by ensuring that work

tasks, resources, deadlines, and communications are effectively organized and tracked. Tracking systems allow you to create and assign tasks to team members. Each task can contain a description, deadlines, attached files, and other important information. Data collection and analysis allows you to create reports and analytical tools to evaluate team performance, identify trends, and make strategic decisions. Thus, project task management systems store information about program tasks and comments that can be useful for predicting the success of tasks. However, at the moment, there are practically no studies on this topic and no results that would allow us to analyze how accurately and clearly the tasks are set.

Assessing the quality and unambiguity of task descriptions is an important part of software development project management. High-quality and accurate task statements allow the project manager and team to better understand the scope of work, resources required, and assess the risks for each task or requirement. It also helps

in making decisions about priorities, resource allocation, and balanced project planning.

Text classification plays an important role in the field of natural language processing (NLP) and is becoming a key component in the development and improvement of project tasks. The rapid development of this field allows us to improve the results in project activities. This, in turn, helps to avoid possible misunderstandings, errors, and delays in project implementation. The text classifier allows you to assess the quality of task formulation, determining their clarity and conciseness. The use of classification helps to create a single standard for interpreting textual descriptions of tasks, which facilitates common understanding across the team. In general, the introduction of text classification into the project management process can improve the quality of project tasks, ensuring their more efficient implementation and avoiding possible difficulties in the development process.

Analysis of last achievements and publications

Fixing tasks and describing them in the training system is an important element for managing and controlling project development processes. This practice helps to avoid misunderstandings, set priorities, and determine the progress of each task during the sprint. Brief task notes in natural language allow team members to quickly grasp the essence of tasks. They can include keywords and phrases that identify the main aspects of the tasks.

The project team spends considerable time manually categorizing incoming tasks. Unfortunately, this process is difficult, and team members do not have clear guidelines for best practices based on historical data. As the amount of data grows, it becomes critical to implement automated task classification. Typically, sprint task descriptions are unstructured, making them difficult to process using natural language. Typos and abbreviations in the text of tasks are also complicating factors. Classification of sprint tasks is a critical process that ensures their proper addressing to the appropriate performers.

This paper analyzes studies on the classification and visualization of project tasks. The research can be divided into two groups.

The first group studies the quality problems of describing functional and non-functional requirements. Paper [2] proposes an approach that can be used to identify patterns that meet well-known standards with

a measure of F1 of 0.90. Also, this approach allows detecting common syntactic features for non-standard patterns in more than 73.5% of cases. The evaluation showed that these results are reliable regardless of the volume and duration of the processed requirements.

The study of the second group [3] focuses on the differences between vectorization methods that transform requirements written in natural language into numerical vectors for further classification. Through an empirical experiment using 5 open datasets, it was found that advanced vectorization methods significantly improve classification performance. It was also found that using pre-prepared data for vectorization is sometimes the most effective method.

First, let's take a closer look at the studies that belong to the first group.

In [4], the researchers defined the concept of "Requirement Smells" as an indicator of the quality of the requirements description for quick analysis during their formation. The researchers also presented an approach to identifying requirements, which was empirically evaluated in the study of many cases. The paper contains conclusions that relate them to the available evidence on the detection of natural language quality defects in requirements artifacts, and analyzes the impact and limitations of this approach and its evaluation.

Researchers in [5] tried to redefine approaches to measuring standard IEEE metrics (defines the software maintenance process) to improvise a quality assessment of SRS (software requirements specification). The authors recommended an SRS template and a strategy for assessing SRS quality metrics. In contrast to modern methods, this approach brought out the metrics of the IEEE SRS quality assessment standards as more modular for collecting requirements.

In [6], the authors proposed a semi-supervised text categorization approach for the automatic identification and classification of non-functional requirements (NFRs). The goal of this approach is to integrate into a recommender system to help analysts and software developers in the architectural design process. A small number of requirements identified during the discovery process allows learning an initial classifier for NFRs that can consistently identify further requirements for the problem statement in an iterative process.

Paper [7] presents an approach to reducing privacy risks in agile software development by automatically detecting information related to the privacy of user history, which is an important notation in agile requirements engineering (RE). Experimental results

have shown that deep learning algorithms allow obtaining better predictions than those achieved using conventional (shallow) machine learning methods. The use of Transfer Learning can significantly improve the accuracy of forecasts by up to 10%.

The authors of [16] presented an ontology-based approach to supporting requirement tracking. The developed ontology contains theoretical knowledge for developing requirements in an agile environment, which allows a group of developers to effectively manage the development of requirements for a software product. This facilitates rapid decision-making on revising project artifacts related to changes and allows for a more accurate determination of the scope of these changes.

In the second group of works on the classification and visualization of project tasks, researchers focus on the quality problems of describing user stories in a project.

In [10], two experiments were conducted to investigate the factors that could potentially affect the derivation of static conceptual models from specifications written in the form of user stories and use cases. The analysis of the results showed that the notation of requirements has a limited impact on the quality of the resulting conceptual model in terms of both reliability and completeness. Systematic grammatical analysis leads to conceptual models with a high level of completeness and validity in both experiments.

Using a linguistic classification of ambiguity problems, the authors of [11] offer a new perspective on current knowledge about understanding, avoiding, and resolving ambiguity in user stories. The possibilities of increasing the effectiveness of the user story technique in supporting requirements engineering (RE) activities in application system development (ASD) projects are investigated.

The approach proposed by the researchers [13] consists of a set of rules that receive scattered information in user stories and organize it in Lexicon symbols. The Lexicon provides a consolidated and organized structure to alleviate the problem of confusing information that causes a lack of tracking between different sprints. This approach has advantages for tracking requirements in agile methodologies, which is supported by previous evaluation results.

Study [15] defines a classification of developers' cognitive representation styles and cognitive interaction patterns in agile RE. The cognitive abilities of developers based on their statements when separating and defining user stories were investigated. The conclusions showed that developers tend to focus on the technology

of cognitive style of representation, even in RE, and have cognitive difficulties in activating and detailing user stories.

In [17], the authors investigate the effectiveness of using word embedding to classify tasks in the IT help desk. Experimental results indicate that the traditional Total Frequency Inverse Document Frequency (TFIDF) method, combined with the Support Vector Machines (SVM) method, provides competitive results and is sometimes more effective than static methods for building vector representations of words, such as word2vec. The use of TFIDF-SVM allows achieving good results at a low computational cost.

The authors of [9] propose an instrumental approach that combines information visualization with two natural language processing techniques: conceptual model extraction and semantic similarity. Experiments have shown that detecting terminological ambiguities requires considerable time, even with the use of technical means, and that it is difficult to determine whether a synonym can interfere with the correct direction of software system development.

Paper [16] presents a visual analysis method called DeepNLPVis, which is used to understand and analyze the model's behavior in text classification tasks and study the reasons for successful and unsuccessful cases. Studies have shown that task descriptions and comments can be used to predict the success of a task with a high level of accuracy.

In [18], the authors examine the usefulness of textual task descriptions and comments in predicting whether problems will be solved successfully. Experiments have shown that problem descriptions and comments can be used to predict problem solving success with more than 85% accuracy, and that predictions of problem solving success change over time.

In [19], the authors propose a method for classifying tasks by clustering and visualizing the characteristics of each category. This method allows experts and stakeholders to better understand the project and features of the target software system by reviewing the keywords of the task categories.

The analysis of the conducted research suggests that there are approaches to describing tasks and user stories, but the issue of qualitative description of project sprint tasks remains less studied. The need for a high-quality textual description of sprint tasks is especially relevant, since an incorrectly formulated or unclear description can lead to misunderstandings and errors in the execution of tasks.

Unlike assessing the quality of requirements and user stories, assessing the quality of sprint tasks requires taking into account the specifics of this context. At the stage of creating the sprint backlog, it is important to determine how clearly and unambiguously each task is formulated. This will help avoid misunderstandings and incorrect task performance by project team members.

The process of identifying and classifying sprint tasks into "clear" and "unclear" is critical at the stage of backlog formation. This will allow you to identify tasks that may be a source of misunderstanding and risk of project delays due to poor textual description. Classifying tasks will provide the project team with the opportunity to devote special attention and resources to clarifying and improving the description of these tasks, thereby reducing the likelihood of errors and misunderstandings.

Considering the quality and unambiguity of the textual description of sprint tasks as a factor influencing task performance is a key element of effective project management. This helps ensure that tasks are properly understood and effectively completed by team members.

Aim and tasks of the study

In the tracking system, tasks are formulated in such a way as to convey the essence of the work to be done as clearly and understandably as possible. It is important to ensure that each team member understands the task in the same way, to avoid ambiguities and misunderstandings. Experience shows that project executors save information about the task in its title, create descriptions, and add comments. Recording all this information in the tracking system provides the team with a convenient tool for collaboration, simplifies communication, and helps ensure that all project participants understand the tasks accurately and completely.

Thus, the purpose of the study is to find approaches to reduce the risks of failure to complete the tasks of an IT project sprint by automating the classification of text descriptions. To achieve this goal, the following tasks need to be solved:

1. Developing an algorithm that can automatically classify text descriptions of sprint tasks into "clear" and "unclear" according to quality and unambiguity criteria.

2. Collect and prepare a training set of textual descriptions of sprint tasks for training and testing the classification model.

3. Application of advanced natural language processing techniques to improve classification and ensure the accuracy of the results.

4. Validate the model on real data to assess the efficiency and accuracy of the classification.

5. Analyzing the results, identifying the factors that affect the quality of task formulation, and drawing conclusions on the application of the developed model.

Materials and methods

To achieve this goal and develop a model for identifying tasks with poor formulation in an IT project sprint, we propose an integrated approach that includes collecting and processing textual descriptions of tasks, using machine learning algorithms for classification, and involving expert opinions to improve the quality of task perception by the project team.

This study uses a task tracking system as the source of data on sprint tasks. The dataset includes 1000 tasks that were exported from the Jira project management system. Based on data from a real project, this dataset contains information about various task parameters such as name, description, unique key, performer, time spent, time estimate, task status, comments, and others. The language of the received textual descriptions of tasks for wording analysis is English, which is typical for software development projects. The main characteristics for identifying the task wording are the name and description. It should be noted that the description is not a mandatory characteristic of the task, i.e., the task formulation can consist of only the name.

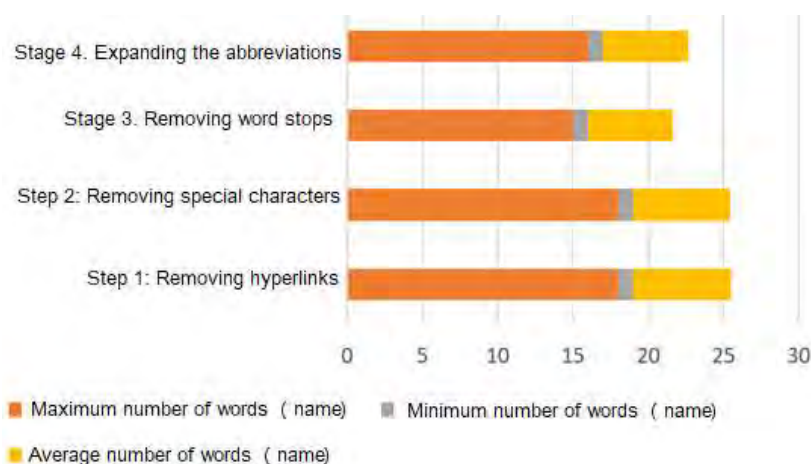
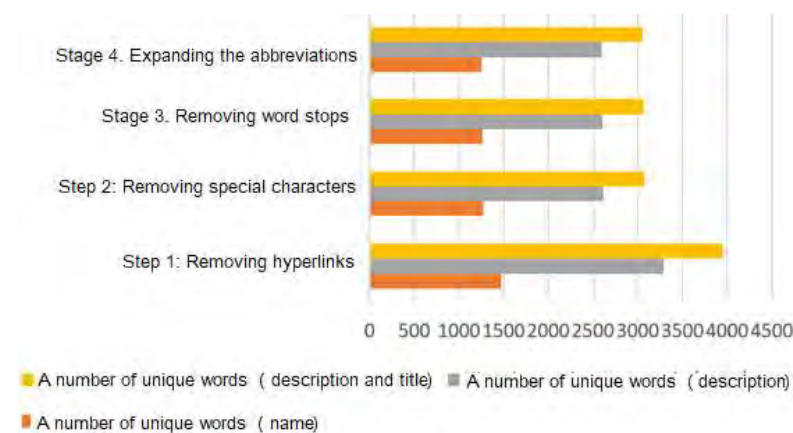
In the process of data pre-processing, a number of cleaning procedures must be performed. In addition to standard textual data parsing, links, code fragments, and uninformative symbols such as currency signs and likes are removed from the dataset. We also replace abbreviations with their full meanings.

The analysis of the dataset reveals the following characteristics: the average number of words in the title is 6.4, and the average number of words in the description is 37.67. The total number of tasks in the set is 1000, of which the number of tasks with descriptions is 461.

The number of unique words in the title and description is 5080. Conjunctions and articles turned out to be the most popular words, and among those that carry an independent semantic load, the most popular are: "add", "image", "inspection", "defect", "object". These words are relevant to this project.

Table 1. Statistics of the dataset at different stages of cleaning

Stage	Delete hyperlinks	Delete special characters	Delete word stops	Extension of abbreviations
Number of unique words (title)	1470	1270	1260	1251
Number of unique words (description)	3293	2610	2601	2596
Number of unique words (description + title)	3943	3069	3061	3051
Maximum number of words (title)	18	18	15	16
Average number of words (title)	6.47	6.42	5.62	5.68

**Fig. 1.** Statistics of the dataset at different stages of cleaning by the total number of words**Fig. 2.** Statistics of the dataset at different stages of cleaning by the number of unique words

Next, at the second stage of the study, the textual descriptions of the sprint tasks are converted into vector representations for further use in the classification model, which involves obtaining a vector representation of each task. This process starts with converting each task into a set of tokens using a tokenizer. A tokenizer, also known as a lexical analyzer, divides text into small units called tokens [20]. Tokens can be words, symbols, phrases, or even sentences, depending on the type of task and the processing language. Tokenization helps to further represent the text in a format suitable for further analysis and processing by computer algorithms.

There are a variety of approaches to obtaining vector representations of text fragments. These vectors can be generated by using unsupervised learning algorithms or retrained neural networks. Choosing a method of text vectorization is an important research task. The following methods can be used to obtain vector representations:

- Word2Vec – a method of vector representation of words [21];

- TF-IDF (Term Frequency-Inverse Document Frequency) is a statistical method that is widely used in the field of information retrieval and text data analysis [22];

– GloVe (Global Vectors for Word Representation) is a type of word representation using statistical properties of words in large text corpora [23];

– FASTTEXT is a type of vector word representation method based on neural networks and has several unique features that make it popular and effective in natural language processing [24].

At the third stage, tasks are marked up using expert annotations, and two classes are formed. However, key research questions arise: what number of experts is sufficient to ensure an independent and reliable assessment of the quality of task formulation? Can experts' knowledge of the project context influence the results of their expert evaluation? An important aspect is also to determine methods for assessing the consistency of expert opinions and establishing criteria for matching certain classes for further use in training the classification model.

The fourth stage involves training the classifier, where Bayesian classifier and neural classifiers can be effective for classifying text fragments. Determining the optimal classification model that meets the characteristics of the data of a particular project is the task of an experimental study. It is important to note that the classification results will be affected not only by the chosen classification method but also by the approach to text vectorization.

The final stage involves validating the model on the available data and selecting the optimal classifier for further use in other projects. One of the key criteria for evaluating the model is the F-measure. The F-measure is a metric that combines precision and recall into a single numerical score, allowing you to assess the balance between these two considerations. The F-measure provides a trade-off between precision and completeness, which is especially important when dealing with unsorted classes or unbalanced datasets. The overall goal is to achieve the best possible balance between these two aspects so that the model has high accuracy and, at the same time, avoids omissions (missed detections) of a particular class.

This approach to building a classifier allows us to systematize sprint tasks into categories according to the quality of their formulations. In addition, this approach takes into account the project context and expert opinions, which makes it flexible and adaptive to the specifics of a particular project and team characteristics, which is critical for further improving project management processes.

Study results and their discussion

The study was conducted on real data generated from the project management system to automate the identification of poor-quality sprint task formulations that affect the developers' understanding of them and, as a result, lead to project risks. The features of the data collected for the study are semi-structured, incomplete, textual, and lack of project context. Therefore, the data set requires special processing and cleaning.

As mentioned above, for classification tasks, we propose to use textual descriptions of tasks contained in the tracking system. Thus, each task is represented by a title and a more detailed description. In our experiments, we used either only the title or the title concatenated with the description.

In the process of processing the text of the tasks, all titles go through several stages:

- replacing all known abbreviations with their meanings;
- filtering all irrelevant characters (if any);
- normalizing all elements of the task text.

Each task is converted into a set of tokens using a tokenizer that is used by the model and understood by it. For all experiments, we use the Python programming language and deep learning frameworks, including Transformers. Next, we need to obtain a vector representation of the task names. At the initial stage, we used the TF-IDF method, the corpus of which was all the tasks available in the dataset of a real project.

The dataset of 1000 tasks was first analyzed into two classes (positive and negative) according to the expert's opinion on the quality of the wording. Then, using dimensionality reduction tools (PCA and t-SNE), visual representations of the data were obtained (Fig. 3, Fig. 4).

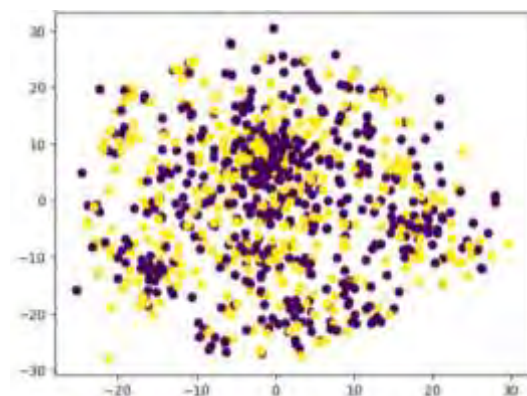


Fig. 3. Visual representation of data using t-SNE in two-dimensional space (yellow – positively annotated tasks, purple – negatively annotated tasks)

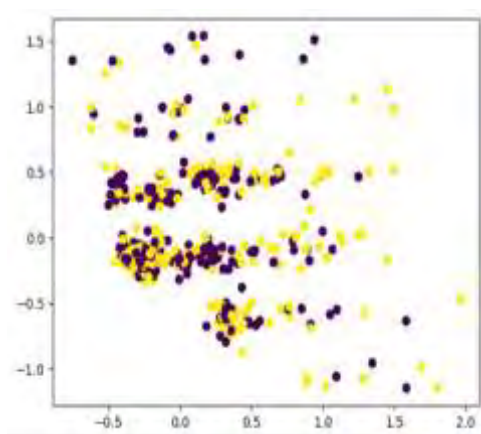


Fig. 4. Visual representation of data using PCA in two-dimensional space (yellow – positively annotated tasks, purple – negatively annotated tasks)

The semantic analysis of the text of the task description and title revealed that the known methods cannot be used due to the lack of context in the description. This is due to the fact that when forming the name and description of a task, the team has its own internal context, which is not transferred to the text of the task. This, in turn, is one of the main reasons for comprehension problems, which affects the timing and quality of the project.

Thus, it has been established that the project context is outside the textual descriptions of tasks, which affects their understanding by new team members, and also limits the use of natural language processing methods to analyze the quality of wording and identify factors that affect the understanding of the content of tasks.

At the next stage of the experiments, additional experts were involved in annotating the dataset. A team of four experts independently labeled the dataset for binary classification. Positive tasks include tasks with high quality, which are understandable to the marker and the time estimate for them is most likely to be adequate. Tasks that are incomprehensible to the expert receive negative marks. In Fig. 5 shows the distribution of positive and negative marks for each of the markers (experts).

It should be noted that when marking up data, as the expert's time and the number of tasks processed increase, the markers become more confident about the project context and a higher percentage of tasks receive positive marks. It is concluded that the more an expert works within the project, the more accurate the estimates become, that is, the clarity of the tasks whose descriptions are recorded in the tracking system significantly depends not only on the wording

of their textual description, but also on understanding the content of the project itself. It is assumed that only experts from the middle of the project can provide adequate markup for the data to classify the quality of task formulations.

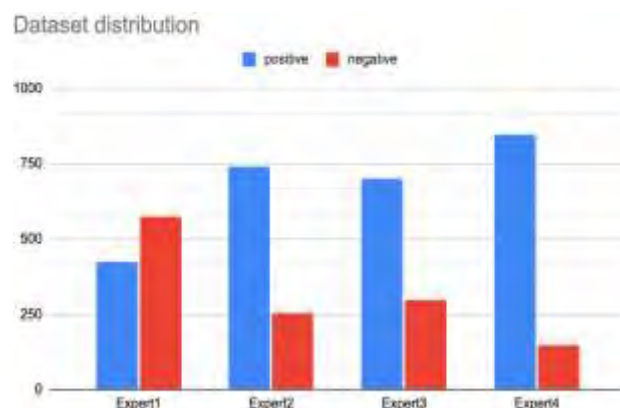


Fig. 5. Distribution of expert markup marks

Thus, we obtained four labeled data sets and analyzed the correlation of the estimates provided by different experts. As can be seen from Fig. 6, the consistency of experts' opinions is low: at best, the markers overlap in only 76% of the tasks, and sometimes only in 43%.

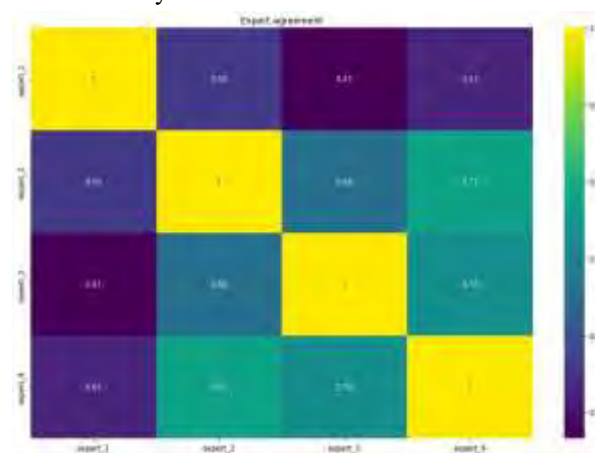


Fig. 6. The level of consistency of the four experts' opinions

Cohen's kappa coefficient is a statistic used to measure inter-rater reliability (as well as intra-rater reliability) for qualitative (categorical) items [25]. It is believed to be a more reliable measure than simply calculating the percentage of agreement, as kappa also takes into account the possibility of random agreement.

Cohen's kappa measures the agreement between two raters, each of whom classifies N items

into C mutually exclusive categories. The definition of κ is as follows:

$$\kappa = \frac{p_0 - p_e}{1 - p_e} - 1 - \frac{1 - p_0}{1 - p_e} \quad (1)$$

where p_0 – is the relative observed agreement among raters, and p_e – the hypothetical probability of random agreement, using the observed data to calculate the probability that each observer will randomly see each category.

If the raters are in full agreement, $\kappa = 1$.

If there is no agreement among the raters, other than what would be expected by chance (as indicated by p_e), $\kappa = 0$. It is possible that the statistics will be negative, which may occur by chance if there is no relationship between the estimates of two appraisers, or it may reflect a real tendency of appraisers to give different estimates.

For k categories, N observations to be classified, and n_{ki} is the number of times that evaluator i predicted category k :

$$p_e = \frac{1}{N^2} \sum n_{k1} n_{k2} \quad (2)$$

In machine learning and statistics, to evaluate binary classifications, the Kappa Cohen formula can be written as follows:

$$\kappa = \frac{2(TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)} \quad (3)$$

Where TP – true positives;

FP – false positives;

TN – true negatives;

FN – false negatives.

Based on the expert ratings, we calculated Cohen's kappa coefficient for each pair of markers. The results of calculating Cohen's kappa coefficient for the experiment are shown in Fig. 7.

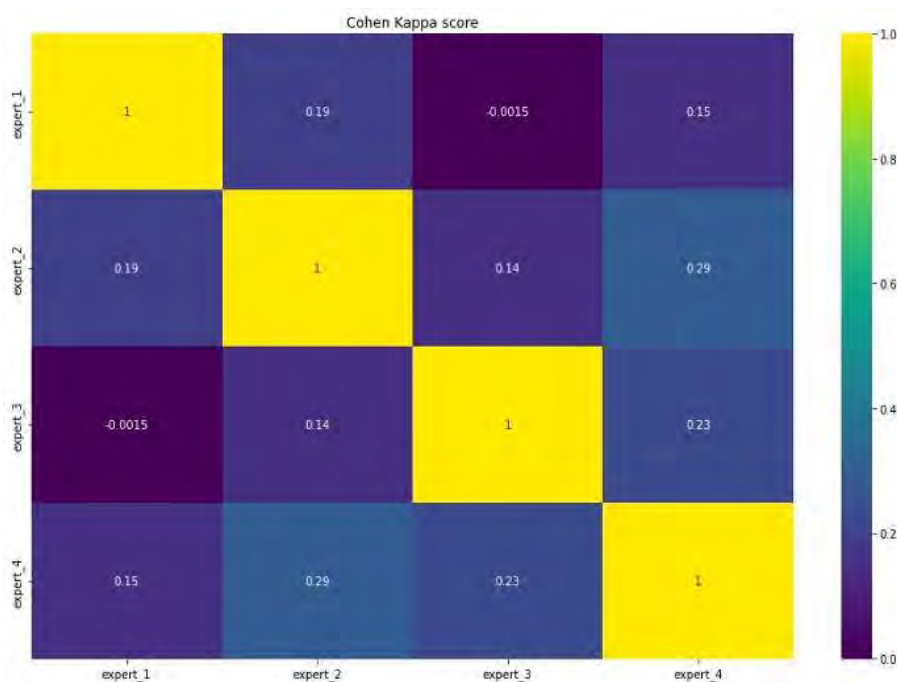


Fig. 7. The Cohen Kappa

Thus, the assumption that task understanding depends on the availability of information about the project, involvement in the development team, etc. was confirmed.

Next, to form the dataset for classification, we selected only those tasks where at least 3 experts gave the same answer about the quality of the task. Thus, we obtained 623 positive and 187 negative task evaluations.

The analysis of modern approaches to automatic text data processing has shown that Bayesian classifier and neural classifiers are well suited for classifying text expressions. A series of experiments were conducted.

Since one of the experiments is training a neural network based on a vectorized representation of the text and, according to the literature, the most representative are the vectors created by artificial neural networks, and given the small amount of data and resources,

the trained BERT (bert-base-uncased) model from the Python package for transformer class models was chosen for the experiment. This model was trained on a large dataset, which was tokenized using an appropriate tokenizer, which includes 28996 natural language tokens, 5 of which are the following tokens:

- [PAD] – to add to the sequence to get the required size;
- [UNK] – to replace unknown characters (in this case, this token was checked and was not present in the dataset);
- [CLS] – token to indicate the token in which the class should be located in the original sequence;
- [SEP] – a separator token;
- [MASK] – a masking token used in training with Masked Language Modelling. This training methodology is often used in pre-training natural language models to obtain representative vectors of input text sequences.

Transformers were first introduced in [26] as an encoder-decoder architecture, i.e., the input data is first collapsed to obtain a representation and then unfolded to perform a specific task. The main idea is to use a self-attention mechanism, which allows the model to effectively interact with different parts of the input text simultaneously. The model consists of several layers, each of which contains attention mechanisms and notation (normalization, activation).

The formula for the attention mechanism in Transformers is as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (4)$$

where Q , K , V – query, key and value matrices, respectively;

d_k – the dimension of the key vector.

Also, it was found that using only one level of attention is less efficient than using the attention mechanism on different linear projections of the query, key, and learned values, then collecting them all together and adding another linear projection to the original representation size. This operation can be described as follows:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_n)W^0, \quad (5)$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$.

One of the key innovations of the BERT deep learning model, which uses an encoder model that does not perform autoregressive prediction, is two-way contextual encoding, which allows the model to use information from all directions in the text. During

training, two types of input representations are used: segmental and positional embeddings.

The formulas for positional and segmental representations in the BERT neural network look like this:

$$Positional = Embedding(w_i) + PositionalEncoding(i), \quad (6)$$

$$Segments = Embedding(s_j), \quad (7)$$

where w_i – token on the position i ;

s_j – segment identifier of the sentence j ;

$PositionalEncoding(i)$ – is the vector of the position designation i in the input sentence.

There are many options for modeling the position, such as uniformly increasing functions or periodic functions.

The Bayes classifier was chosen as the first classification method. It is based on Bayes' theorem, which determines the probability of an event based on the probabilities of previous events. In the case of text classification, the Bayesian classifier uses the probabilities that a certain word or phrase belongs to a certain class.

The basic idea of a Bayesian classifier is to determine the probabilities for each class, given the occurrence of specific words or phrases in the text. Mathematically, this is expressed as follows:

$$P(C_x|X) = \frac{P(X|C_x) \times P(C_x)}{P(X)}, \quad (8)$$

where $P(C_x|X)$ – is the probability that document X belongs to the class C_k ;

$P(X|C_x)$ – is the probability of document X given the class C_k ;

$P(C_k)$ – a priori probability of the class C_k ;

$P(X)$ – total probability of the document X .

To simplify the calculations, additional assumptions are often used, for example, the independence of word occurrences, which allows expressing the probability of a document given a class as the product of the probabilities of individual words:

$$P(X|C_k) = P(w_1|C_k) \times P(w_2|C_k) \times \dots \times P(w_n|C_k), \quad (9)$$

where w_i – a single word from a document.

For the Bayesian classifier, vector representations are obtained from the trained BERT neural network. These representations are then used as independent datasets, unrelated to the text, on which the classifier

is trained. According to the experimental results, the Bayesian classifier achieved an accuracy of 70%. For the deep learning-based classifier, we chose a neural network for binary classification based on the BERT transformer architecture from the above package. The same model was chosen as a training model. For evaluations, the F1 criterion is used – a metric

for measuring the accuracy of models in classification tasks and is a harmonic mean between precision and recall. The result is a classifier that gives an accuracy score of approximately 83% on the test dataset, which is a good result for a small dataset and data with conflicting labels.

The training schedule is shown in Fig. 8.

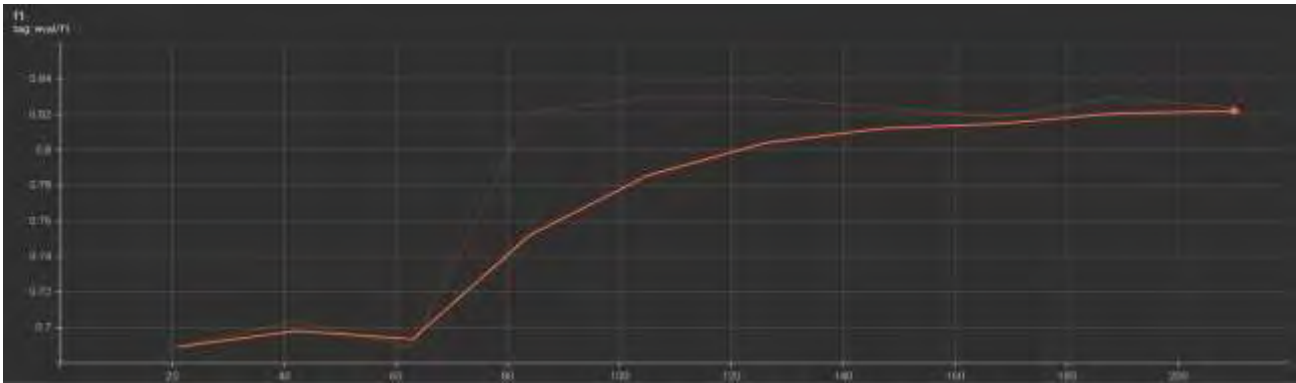


Fig. 8. Graph of F1 metric dependence on the training step

Thus, the experiments have shown that developing a classification model to identify the quality of textual descriptions of sprint tasks requires the involvement of experts who know the project context, determining the vectorization approach, and choosing a classifier model that depends on the available data collected from the project tracking system.

Conclusion and perspectives of further development

Thus, based on the results of the study, the following conclusions can be drawn.

This study focuses on solving the problem of identifying the quality of the textual description of sprint tasks in the project management system, which, in turn, determines an important aspect of trying to reduce the risks of project failure. In solving this problem, the developed and implemented sprint task classifier is used, which automatically identifies poor-quality task formulations. It was found that this can act as a catalyst for improving the formulated tasks and adding detailed descriptions, which contributes to the effective work of the project team as a whole.

The analysis of textual data confirms that the existing data in the tracking system is incomplete and contains abbreviations, symbols, and slang. This undoubtedly makes it difficult to understand.

It should be noted that the team understands the project context and is able to perceive the information provided. The results indicate that the quality of the wording is determined by the level of the expert's awareness of the specifics and context of the project, while increasing the number of experts has almost no effect on the result.

Adding formulation assessments during the sprint retrospective to train the classifier model and involve project team members as experts appears to be advisable. This will facilitate data collection for regular model training and improve the consistency of ratings assigned by different experts.

Despite the experiments that did not reveal the superiority of a particular classifier, it is recommended to use several classifiers, compare their results by F-measure, and take into account the choice of vectorizer to select the best one. Experiments on real data from a software development project for a Bayesian classifier and a classifier based on the Transformer architecture showed an accuracy of 0.7 and 0.83, respectively, which is quite acceptable given the training data.

In further research, it is planned to test the hypothesis that the classifier's effectiveness depends on a particular project and the use of unsupervised learning methods for the task of identifying the quality of formulations.

References

1. Rohovyi, M., Grinchenko, M. (2023), "Project team management model under risk conditions". *Vestn. Khar'k. politekhn. in ta. Ser.: Strategichne upravlinnya, upravlinnya portfelyamy, programamy ta proektamy* [Bulletin of the Kharkov Polytechnic Institute. Series: Strategic Management, Portfolio Management, Programs and Projects], Kharkov: NTU "KhPI", No. 1 (7), P. 3–11. DOI: <https://doi.org/10.20998/2413-3000.2023.7.1>
2. Sonbol, R., Rebdawi, G., Ghneim, N. (2022), "Learning software requirements syntax: An unsupervised approach to recognize templates, *Knowledge-Based Systems*, Vol. 248, 108933 p. <https://doi.org/10.1016/j.knosys.2022.108933>
3. Leelaprute, P., Amasaki, S. (2022), "A comparative study on vectorization methods for non-functional requirements classification", *Information and Software Technology*, Vol. 150, 106991 p. <https://doi.org/10.1016/j.infsof.2022.106991>
4. Femmer, H., Fernández, D., Wagner, S., Eder, S. (2017), "Rapid quality assurance with Requirements Smells", *Journal of Systems and Software*, Vol. 123, P. 190–213. <https://doi.org/10.1016/j.jss.2016.02.047>
5. Ramesh, M.R.R., Reddy, C.S. (2021), "Metrics for software requirements specification quality quantification", *Computers & Electrical Engineering*, Vol. 96, Part A, 107445 P. 3–11. <https://doi.org/10.1016/j.compeleceng.2021.107445>
6. Casamayor, A., Godoy, D., Campo, M. (2010), "Identification of non-functional requirements in textual specifications: A semi-supervised learning approach", *Information and Software Technology*, Vol. 52, Issue 4, P. 436–445. <https://doi.org/10.1016/j.infsof.2009.10.010>
7. Casillo, F., Deufemia, V., Gravino, C. (2022), "Detecting privacy requirements from User Stories with NLP transfer learning models", *Information and Software Technology*, Vol. 146, P. 106853. <https://doi.org/10.1016/j.infsof.2022.106853>
8. Dalpiaz, F., et al. (2019), "Detecting terminological ambiguity in user stories: Tool and experimentation", *Information and Software Technology*, Vol. 110, P. 3–16. <https://doi.org/10.1016/j.infsof.2018.12.007>
9. Dalpiaz, F., Gieske, P., Sturm, A. (2021), "On deriving conceptual models from user requirements: An empirical study", *Information and Software Technology*, Vol. 131, 106484 P. 1–13. <https://doi.org/10.1016/j.infsof.2020.106484>
10. Amna, A.R., Poels, G. (2022), "Ambiguity in user stories: A systematic literature review", *Information and Software Technology*, Vol. 145, P. 1–12. <https://doi.org/10.1016/j.infsof.2022.106824>
11. Urbietta, M., et al. (2020), "The impact of using a domain language for an agile requirements management", *Information and Software Technology*, Vol. 145, P. 1–16. <https://doi.org/10.1016/j.infsof.2020.106375>
12. Jia, J., et al. (2019), "Understanding software developers' cognition in agile requirements engineering", *Science of Computer Programming*, Vol. 178, P. 1–19. <https://doi.org/10.1016/j.scico.2019.03.005>
13. Murtazina, M., Avdeenko, T. (2019), "An Ontology-based Approach to Support for Requirements Traceability in Agile Development", *Procedia Computer Science*, Vol. 150, P. 628–635. <https://doi.org/10.1016/j.procs.2019.02.044>
14. Y. Wahba, Y., Madhavji, N., Steinbacher, J. (2020), "A Hybrid Machine Learning Model for Efficient Classification of IT Support Tickets in The Presence of Class Overlap", *Proceedings of the 32nd Annual International Conference on Computer Science and Software Engineering*, P. 151-156. DOI: 10.1109/ICIT58465.2023.10143149
15. Ramírez-Mora, S., Oktaba, H., Gómez-Adorno, H. (2020), "Descriptions of issues and comments for predicting issue success in software projects", *Journal of Systems and Software*, Vol. 168, P. 1–19. <https://doi.org/10.1016/j.jss.2020.110663>
16. Li, Z., A "Unified Understanding of Deep NLP Models for Text Classification", available at: <https://arxiv.org/abs/2206.09355> (last accessed 08.11.2023).
17. Ishizuka, R., et al. (2022), "Categorization and Visualization of Issue Tickets to Support Understanding of Implemented Features in Software Development Projects", *Applied Sciences*. № 12(7):3222. <https://doi.org/10.3390/app12073222>
18. Devlin, J., et al. (2019), "BERT: Pre-training of deep bidirectional transformers for language understanding", *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, P. 4171–4186. DOI: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)
19. Chawla, P., Hazarika, S., Shen, H.-W. (2020), "Token-wise sentiment decomposition for convnet: Visualizing a sentiment classifier", *Visual Informatics*, Vol. 4 Issue 2, P. 132–141. <https://doi.org/10.1016/j.visinf.2020.04.006>
20. Bird, S., Klein, E., Loper, E. "Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Beijing", 2009. 504 p. available at: <https://tjzhifei.github.io/resources/NLTK.pdf> (last accessed 08.11.2023).
21. "Word2vec", available at: <https://www.tensorflow.org/text/tutorials/word2vec> (last accessed 08.11.2023).
22. "TF-IDF (Term Frequency-Inverse Document Frequency)", available at: <https://www.learnatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/> (last accessed 08.11.2023).
23. Pennington, J., Socher, R., Manning, C. (2014), "GloVe: Global Vectors for Word Representation", *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, P.1532–1543. [http://dx.doi.org/10.3115/v1/D14-1162](https://dx.doi.org/10.3115/v1/D14-1162)
24. "Open-source FastText", available at: <https://fasttext.cc/> (last accessed 08.11.2023)
25. McHugh, Mary L. (2012), "Interrater reliability: the kappa statistic", *Biochemia Medica*, Vol. 22 Issue 3, P. 276–282. <https://doi.org/10.11613/BM.2012.031>

26. Ashish Vaswani, Noam Shazeer, Niki Parmar et al., (2017), "Attention Is All You Need", *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*, P.1–15. DOI: <https://doi.org/10.48550/arXiv.1706.03762>

Received 04.12.2023

Відомості про авторів / About the Authors

Гринченко Марина Анатоліївна – кандидат технічних наук, доцент, Національний технічний університет "Харківський політехнічний інститут", завідувачка кафедри стратегічного управління, Харків, Україна; e-mail: marinagrunchenko@gmail.com; ORCID ID: <https://orcid.org/0000-0002-8383-2675>

Роговий Микита Антонович – Національний технічний університет "Харківський політехнічний інститут", аспірант кафедри стратегічного управління, Харків, Україна; ORCID ID: <https://orcid.org/0000-0002-7902-3592>; e-mail: nikrogovoy@gmail.com

Grinchenko Marina – PhD (Engineering Sciences), Associate Professor, National Technical University "Kharkiv Polytechnic Institute", Head at the Department of Strategic Management, Kharkiv, Ukraine.

Rohovyi Mykyta – National Technical University "Kharkiv Polytechnic Institute", PhD student at the Department of Strategic Management, Kharkiv, Ukraine.

МОДЕЛЬ ІДЕНТИФІКАЦІЇ ЗАВДАНЬ СПРИНТУ ПРОЄКТУ НА ОСНОВІ ЇХ ОПИСУ

Предметом дослідження є ідентифікація завдань спринту проєкту. **Мета статті** – пошук підходів до зниження ризиків невиконання завдань спринту. У роботі вирішуються такі **завдання**: аналіз досліджень щодо класифікації та візуалізації завдань проєкту; розроблення алгоритму, який здатний автоматично класифікувати текстові описи завдань спринту; збір і підготовка навчальної вибірки текстових описів завдань спринту для навчання та тестування моделі класифікації; застосування методів оброблення природної мови для вдосконалення класифікації та забезпечення точності результатів, проведення валідації моделі на реальних показниках для оцінювання ефективності й точності класифікації; проведення аналізу результатів. Використовуються такі **методи**: машинне навчання для класифікації, векторизація текстів, класифікація текстових описів, оброблення природної мови, семантичний аналіз тексту опису завдань та оброблення експертних оцінок. **Досягнуті результати**. Запропоновано комплексний підхід використання алгоритмів машинного навчання, що передбачає збір та оброблення текстових описів завдань, для класифікації та залучення експертних оцінок з метою вдосконалення якості сприйняття завдань командою проєкту. Проведено класифікацію текстових висловів на основі класифікатора Баеса та нейронних класифікаторів. Реалізовано візуальну репрезентацію даних. Проведено семантичний аналіз тексту опису та назви завдання. Отримано розмітку даних для класифікації якості формулювань, яка була виконана командою експертів. Для вимірювання надійності отриманих оцінок експертів розраховано коефіцієнт каппа Коена для кожної пари розмітників. За результатами експериментів для класифікатора Баеса отримано точність 70%. Для класифікатора на основі глибокого навчання обрано нейронну мережу для бінарної класифікації на основі архітектури *transformer*. Проведено навчання нейронної мережі за допомогою мови програмування *Python* і фреймворків для глибокого навчання. Унаслідок отримано класифікатор, що на тестовому наборі оцінює з точністю 83%, що є гарним результатом для малого набору даних і даних із суперечливими мітками. **Висновки**. Аналіз текстової інформації підтверджує, що наявні в системі трекінгу дані не повні та містять скорочення, умовні позначки та сленг. Здобуті результати свідчать про те, що оцінка якості формулювань визначається рівнем обізнаності експерта щодо особливостей і контексту проєкту, водночас збільшення кількості експертів майже не впливає на результат. У подальших дослідженнях рекомендується перевірити гіпотезу про залежність ефективності класифікатора від конкретного проєкту та використання методів навчання без учителя для завдання ідентифікації якості формулювань.

Ключові слова: проєкт; опис завдань; система управління завданнями проєктів; модель; класифікатор; векторна репрезентація.

Бібліографічні описи / Bibliographic descriptions

Гринченко М. А., Роговий М. А. Модель ідентифікації завдань спринту проєкту на основі їх опису. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 33–44. DOI: <https://doi.org/10.30837/ITSSI.2023.26.033>

Grinchenko, M., Rohovyi, M. (2023), "A model for identifying project sprint tasks based on their description", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 33–44. DOI: <https://doi.org/10.30837/ITSSI.2023.26.033>

УДК 004.82

DOI: <https://doi.org/10.30837/ITSSI.2023.26.045>

В. Ляшик, І. ШУБІН

МЕТОД ЛОГІЧНИХ МЕРЕЖ ДЛЯ МОДЕЛЮВАННЯ СИСТЕМ АДАПТИВНОГО ТЕСТУВАННЯ ЗНАТЬ

Предметом дослідження є розроблення математичного та алгоритмічного забезпечення інтелектуального інструментарію аналізу наборів тестових завдань і моделювання процесу інтерпретації якості наборів тестових завдань, що дає змогу об'єктивно та всебічно проводити безперервний контроль знань суб'єктів навчання за умови впровадження концепції віртуального розподіленого навчання (перепідготовки). **Мета роботи** – підвищення ефективності контролю знань суб'єктів освітнього процесу в умовах дистанційної форми навчання з допомогою застосування адаптивних методів комп'ютерного тестування на основі моделей логічних мереж і алгебри скінченних предикатів. У статті вирішуються такі **завдання**: формування моделі тестування в розподіленому віртуальному навчальному середовищі та моделі оцінювання валідності за змістом наборів тестових завдань. Використовуються такі **методи**: методи логічних мереж і алгебраїчного програмування на основі алгебри скінченних предикатів і предикатних операцій, інтелектуальний аналіз інформації. Здобуто такі **результати**: сформульовано принципи інтелектуального аналізу в процесі розроблення моделі універсальної логічної мережі та її застосування до актуальних завдань штучного інтелекту в галузі неформалізованого оброблення інформації, а саме в побудові систем тестування знань для розподіленого віртуального навчання. **Висновки**. Удосконалено алгоритми оптимального багатоступінчастого адаптивного тестування знань як частини моделей розподіленого віртуального навчання та методів аналізу успішності суб'єктів навчання. Використано кон'юнктивну декомпозицію з бінарними предикатами та досягнуто мету дослідження, оскільки таким способом будь-який багатомісний предикат можна подати логічною мережею, що моделює процес тестування знань. Описано модель суб'єкта навчання.

Ключові слова: програмна інженерія; бази знань; алгебра скінченних предикатів; логічні мережі; логічні правила; надійність тестів; використання знань; модель суб'єкта навчання.

1. Вступ

Наразі серед проблем дистанційної освіти [1] особливе місце посідає проблема автоматизації контролю знань. Питання комп'ютерного контролю становлять значний інтерес для викладачів вищих навчальних закладів і розробників програмних засобів реалізації такого контролю. Водночас проблеми комп'ютерного контролю знань недостатньо широко висвітлені теоретично, і зацікавленість ними здебільшого реалізується з допомогою створення чергової програми комп'ютерного контролю із заздалегідь розробленим набором завдань. Засоби адаптивного тестування мало досліджені й опрацьовані недостатньо прозоро.

Контроль знань або тестування – процес, що проводиться з метою визначення рівня знань суб'єкта навчання [2]. Це найбільш стандартизований і об'єктивний метод контролю та оцінювання знань, умінь і навичок суб'єкта навчання, який позбавлений таких традиційних недоліків інших методів контролю знань, як неоднорідність вимог, суб'єктивність екзаменаторів, невизначеність системи оцінювання тощо. Рівні знань зазвичай дискретизуються.

За умови такого підходу тестування може розглядатися як деякий діагностичний процес, а стани, що характеризують оцінки знань суб'єкта навчання, – як діагностичні стани. Тести є ефективним засобом перевірки якості знань, набутих студентами, і оперативного контролю етапів навчання [3]. Інформаційні освітні ресурси (ІОР), що містять тестові матеріали, можна поділити на дві категорії:

– орієнтовані на проходження студентами тестів у письмовій формі з подальшою перевіркою вручну викладачем (як варіант – сканування результатів тестування для їх подальшої автоматизованої перевірки);

– системи комп'ютерного тестування (СКС) з відповідним наповненням тестовими матеріалами.

Переваги другої категорії тестових ІОР очевидні. Вони звільняють викладача від рутинної роботи під час іспитів і проміжного оцінювання знань у традиційному навчальному процесі, а в навчанні з використанням дистанційних технологій стають основним засобом контролю; уможливають автоматизацію оброблення результатів, об'єктивність контролю та швидкість перевірки якості підготовки значної кількості суб'єктів навчання із широкого кола

питань. Це дає змогу визначити розділи, що є найбільшою складністю у вивченні, та в режимі адаптації коригувати процес навчання залежно від результатів тестування. СКС виконують навчальну функцію та допомагають упровадити методи індивідуалізації процесу набуття знань суб'єктами навчання [4].

Функції контролю знань виконують водночас навчально-виховне та розвивальне завдання. Тестування є важливим елементом не тільки контролю знань, але й навчання. Навчальний складник тестування полягає в тому, що користувачеві після проходження тесту надаються посилання на розділи навчального матеріалу, з яких він відповів неправильно. Для досягнення зазначених результатів необхідно швидкими темпами розбудовувати дистанційну форму навчання, упровадження якої в Україні передбачене Національною програмою інформатизації.

Функція контролю [5] полягає у виявленні стану знань і вмінь студентів, рівня їхнього розвитку, у вивченні ступеня засвоєння певних компетенцій, навичок раціональної навчальної діяльності. За допомогою контролю визначається вихідний рівень для подальшого набуття знань, умінь і навичок, вивчається глибина та обсяг їх засвоєння. Порівнюється планований рівень із дійсними результатами, встановлюється ефективність використовуваних викладачем методів, форм і засобів навчання.

Навчальна функція [6] контролю знань полягає в удосконаленні знань і вмінь та їх систематизації. У процесі перевірки студенти повторюють і закріплюють вивчений матеріал. Вони не лише відтворюють раніше вивчене, але й застосовують знання та вміння в новій ситуації. Перевірка допомагає виокремити основне в досліджуваному матеріалі, зробити виявлення наявних знань і вмінь більш зрозумілим і точним. Контроль сприяє також узагальненню та систематизації знань.

Сутність діагностичної функції контролю [7] полягає в отриманні інформації про помилки й прогалини в знаннях і вміннях суб'єкта навчання, а також про причини, що їх породжують. За умови численних помилок викладач аналізує труднощі, з якими стикаються студенти у вивченні навчального матеріалу. За результатами діагностичних перевірок обирається найбільш інтенсивна методика навчання, а також уточнюються напрями подальшого вдосконалення змісту, методів і засобів освітнього процесу.

Прогностична функція [8] допомагає отримати випереджальну інформацію про навчально-виховний процес. Після перевірки знань з'являються підстави для прогнозування певного етапу навчального процесу: чи достатньо сформовані конкретні знання, уміння й навички для засвоєння наступної порції навчального матеріалу (розділу, теми). Результати прогнозування використовують для створення моделі подальшої поведінки суб'єкта навчання, що припускається помилок конкретного типу або має певні прогалини в системі пізнавальної діяльності. Завдяки прогнозуванню викладач робить правильні висновки для подальшого планування навчальної роботи.

Мета роботи – підвищення ефективності контролю знань суб'єктів освітнього процесу в умовах дистанційної форми навчання з допомогою застосування адаптивних методів комп'ютерного тестування на основі моделей логічних мереж і алгебри скінченних предикатів.

Метою роботи є розроблення математичного та алгоритмічного забезпечення інтелектуального інструментарію аналізу наборів тестових завдань і моделювання процесу інтерпретації якості наборів тестових завдань. Це дає змогу об'єктивно та всебічно проводити безперервний контроль знань студентів (суб'єктів навчання) в умовах упровадження концепції віртуального розподіленого навчання на основі створення методів автоматизації оцінювання якості освітніх тестів та інтерпретації результатів, опису математичного апарату, що допомагає якнайкраще описувати складні інтелектуальні процеси й методи оцінювання комплектів тестових завдань під час перевірки рівня знань суб'єктів навчання.

2. Постановка завдання

Під адаптивним тестовим контролем розуміють комп'ютеризовану систему науково обґрунтованої перевірки та оцінювання результатів навчання, що має високу ефективність завдяки оптимізації процедур генерації, подання та оцінювання результатів виконання адаптивних тестів [9].

Ефективність контрольних-оцінних процедур підвищується за умови використання багатокрокової стратегії відбору та подання завдань, основаної на алгоритмах із повною контекстною залежністю, у яких черговий крок відбувається тільки внаслідок оцінювання результатів попереднього кроку

(або кроків). Після виконання чергового завдання щоразу виникає потреба в ухваленні рішення про вибір рівня складності наступного завдання залежно від того, правильною чи неправильною була попередня відповідь [10].

Алгоритми підбору та подання завдань будуються за принципом зворотного зв'язку, коли в разі правильної відповіді, наданої студентом, чергове завдання обирається підвищеної складності, а неправильна відповідь спричиняє подання наступного, менш складного, завдання, порівняно з тим, на яке суб'єкт навчання дав неправильну відповідь. Також можливі додаткові запитання з тем, які студент знає не дуже добре, для більш точного з'ясування рівня знань у межах певних навчальних модулів. Отже, адаптивна модель нагадує викладача на іспиті: якщо суб'єкт навчання відповідає на запитання впевнено й правильно, викладач досить швидко ставить йому позитивну оцінку. Якщо той, хто проходить іспит, починає давати неправильні відповіді, то викладач ставить йому додаткові або навідні запитання того самого рівня складності або з тої самої теми. І, нарешті, якщо студент із самого початку відповідає погано, оцінку викладач теж ставить досить швидко, але негативну.

Для застосування тестової перевірки знань у системах віртуального навчання природною є вимога мінімального обсягу тестових завдань, на які має відповісти суб'єкт навчання. Однак для статистично повної оцінки сформованих компетенцій студента необхідно, щоб кількість його відповідей була якомога більшою. Останнє є суперечливим і, зважаючи на те, що розроблені тести мають бути адекватним відбиттям предметної галузі, сама процедура адаптивного тестування (зокрема розроблення тестів, алгоритм тестування, алгоритм оцінювання знань) має надійно контролюватися в процесі створення й тестуватися як продукт.

Вибір такого алгоритму тестування, структури тестів, які б задовольняли викладеним вимогам і характеристикам, не є тривіальним завданням і надалі визначається як завдання дослідження.

Алгебра предикатів є зручним і ефективним способом формального запису інформації найбільш загального вигляду. Модель, побудована засобами алгебри предикатів та предикатних операцій, є системою предикатних рівнянь, для ефективного рішення якої слугує логічна мережа, що уможлиблює як моделювання і верифікацію інтелектуальних

моделей, так і безпосередню реалізацію як інтелектуального ядра.

Запропоновано два поняття: "логіт рівня знань" і "логіт рівня труднощів завдання" [11, 12]. Перше визначено як натуральний логарифм відношення частки правильних відповідей випробуваного на всі завдання тесту до частки неправильних відповідей, а друге поняття – як натуральний логарифм іншого відношення: частки неправильних відповідей на завдання тесту до частки правильних відповідей на ті самі завдання за множиною учасників тестування. Єдина логарифмічна шкала дає змогу встановити необхідну відповідність між рівнем підготовки й труднощами завдання й, крім того, скоригувати результати тестування за умови тестів різної складності [13].

Адаптивне тестування визначається як "сукупність процесів генерації, подання та оцінювання результатів виконання адаптивних тестів, що забезпечує приріст ефективності вимірів порівняно з традиційним тестуванням завдяки оптимізації добору характеристик завдань, їх кількості, послідовності й швидкості подання щодо особливостей підготовки учасників тестування" [14, 15].

В умовах адаптивного тестування (під час проходження тесту або набору тестів) будується модель суб'єкта навчання, що застосовується для генерації або вибору наступних завдань тестування залежно від рівня підготовки суб'єкта навчання. У комплексних системах отримана модель також може використовуватися в процесі навчання. У наш час адаптивне тестування реалізується здебільшого у вигляді алгоритмів комп'ютерного тестування. Адаптивне тестування має задовольняти такі вимоги:

- можливість регулювати пропорції підготовлених легких, середніх і важких завдань залежно від кількості правильних відповідей;
- можливість регулювати пропорцій запропонованих різних тематичних розділів навчальної програми в тесті;
- можливість регулювати рівні складності пропонованих тестів, зважаючи на семантичну компетенцію суб'єкта навчання;
- долучення адаптивного механізму переведення на більш високий рівень тестування на тому самому рівні запропонованих завдань;
- кожне завдання більш складного рівня оцінюється вищими балами.

Вибір алгоритмів тестування на сьогодні фактично обмежений формами подання тестових завдань і алгоритмами оцінювання результатів тестування. Досягнення більш високих результатів і підвищення мотивації навчання в остаточному підсумку є основною метою перевірки знань [16–18]. Сам процес тестування, обумовлений певним алгоритмом, має бути максимально формалізованим і водночас гнучким, для того щоб адекватно оцінювати знання суб'єктів навчання. Крім того, можливість регулювати пропорції пропонованих легких, середніх і складних завдань залежно від кількості правильних відповідей є нетривіальною вимогою. Це пов'язано з тим, що зрештою в оцінюванні знань і вмінь студента так чи інакше використовуються статистичні методи для валідної апроксимації функції успіху.

Вирішення таких завдань передбачає формування моделі тестування в розподіленому віртуальному навчальному середовищі та моделі оцінювання валідності за змістом наборів тестових завдань.

3. Опис проведених теоретичних досліджень

Самі набори тестових завдань, що розробляються, мають бути адекватним відбиттям предметної галузі, а сама процедура адаптивного тестування (зокрема розроблені тести, алгоритми тестування, алгоритми оцінювання знань) має надійно контролюватися під час розроблення й тестуватися як продукт.

Отже, вибір такого алгоритму тестування, структури тестів, які б задовольняли викладеним вимогам і характеристикам не є тривіальним завданням і надалі визначається як завдання дослідження [19, 20].

Моделі логічних предикатних мереж і логічного програмування, використані в роботі для конструювання систем багатоступінчастого тестування, ґрунтуються на критерії максимуму інформації, що дозволяє брати до уваги різноманітні обмеження для різних тестових структур. Описи цих відношень за допомогою скінченних предикатів уможливають долучення таких обмежень у системи східчастого адаптивного тестування. Одним із видів багатоступінчастого підходу, описаного логічними мережами, є пірамідальне тестування (рис. 1).

На поданому рисунку жирною лінією показаний маршрут одного суб'єкта навчання (траєкторія відповідей на запитання).



Рис. 1. Пірамідальна модель логічної мережі, що описує процес адаптивного тестування

Якщо зміна порядку подання тестового завдання відбувається на кожному кроці тестування (постійна адаптація), тоді траєкторія відповідей на запитання прийме вигляд, що зображений на рис. 2.

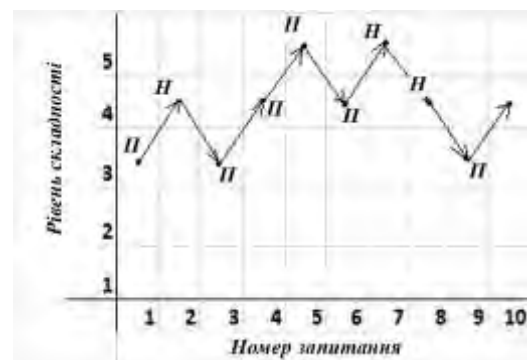


Рис. 2. Схема проходження завдань за умови постійної адаптації (H – неправильна відповідь; П – правильна відповідь)

Якщо ухвалення рішення про зміну порядку проходження завдань здійснюється після аналізу результатів звітів випробуваного на спеціальному блоку завдань (блокова адаптація), то схема проходження завдань має такий вигляд (див. рис. 3).

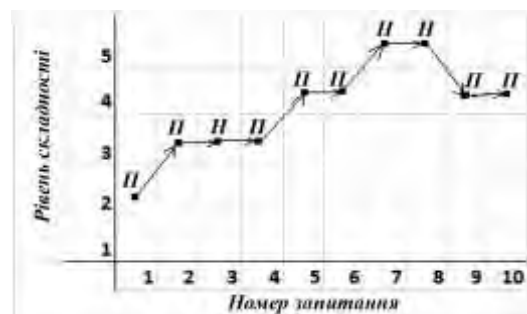


Рис. 3. Схема проходження завдань за умови блокової адаптації (H – неправильна відповідь; П – правильна відповідь)

На протипагу дискретним завданням, множини завдань встановлюють так звані блоки, тобто зв'язані завдання, об'єднані загальною тематикою (або за іншою ознакою). До того ж ці зв'язки розглядаються як неподільні одиниці. Блоки запитань об'єднані таким чином, що припускають фіксований порядок подання тестових завдань [17, 18].

Стохастичний вибір тем лекційного курсу зумовлений тим, що вони, як правило, не є рівнозначними. Інакше кажучи, теми можуть відрізнятися за кількістю розглянутих питань, за рівнем складності теми загалом і, як наслідок, конкретних запитань, за важливістю теми та за іншими ознаками, суб'єктивно обумовленими викладачем.

Запитання (у межах наведеного вище сценарію) обирається після вибору теми. Вибір теми здійснюється стохастично з обліком наведених вище обставин. Доцільним у цьому разі є використання вагового коефіцієнта K_{wt} теми таким чином, щоб:

$$\sum_{i=1}^t K_{w_i}(i) = 1,$$

де $T = (1...t)$ визначає i -ту тему із загальної кількості t тем.

Тоді вибір дозволеної теми здійснюється в такий спосіб.

Крок 1.

Для множини дозволених тем із використанням конгруентного генератора псевдовипадкових чисел генерується випадкове $A_t^{(i)} = \{0..1\}$ число, відповідне до рівномірного розподілу функції ймовірності.

Крок 2.

Визначається для кожної теми добуток $K_{w_t}^{(i)} \times A_t^{(i)}$.

Крок 3.

Обирається тема (з якої потім буде обрано запитання) за ознакою найбільшого значення добутку $K_{w_t}^{(i)} \times A_t^{(i)}$.

Аналогічно можуть обиратися запитання для кожної теми.

Вибір додаткового запитання або їх ланцюжка потребує іншого підходу. Так само, як і для тем, і основних запитань існують переваги до їх вибору (відсутність переваг може розглядатися як їх рівність). Однак, якщо під час складання реального іспиту вибір запитань здійснюється практично миттєво й визначається рівномірною функцією

розподілу, то для додаткових запитань існують такі правила:

– вони ставляться по одному, тобто ситуація, коли ставиться понад одне запитання одночасно, не можлива;

– вибір додаткового запитання здійснюється на розсуд викладача, який має з тих або інших причин певні переваги в цьому виборі.

Значені переваги полягають у частоті запитань, що ставляться екзаменатором, за одиницю часу. Це приводить до визначення подібного роду вибору як потоку подій, а процедури вибору – як моделі марківських мереж.

Отже, для вибору додаткового запитання метод має такі кроки.

Крок 1.

Для кожного додаткового запитання, що може бути як одне, так і кореневе в ланцюзі (дереві) додаткових запитань, за допомогою конгруентного генератора псевдовипадкових чисел визначається число $A_t^{(i)} = \{0..1\}$.

Крок 2.

Відповідно до заданого для кожного додаткового запитання $\lambda_{(i)}^{(i)}$ визначається число $b_{(i)}^{(i)}$:

$$b_{(i)}^{(i)} = -\frac{1}{\lambda_{(i)}^{(i)}} \ln(A_t^{(i)}).$$

Крок 3.

Обирається те додаткове запитання, для якого значення $b_{(i)}^{(i)}$ є найменшим.

Значення $b_{(i)}^{(i)}$ за умови обґрунтованого $\lambda_{(i)}^{(i)}$ відповідає часу, після якого буде поставлено додаткове запитання.

Наведений алгоритм із ваговими коефіцієнтами для тем і основних запитань є доречним, а результат у двох випадках практично однаковий, незважаючи на різницю в шкалах оцінки (нормальну й логарифмічну). Однак у першому випадку (для тем і основних запитань) фактично маємо справу з об'єднаною моделлю формування білетів традиційного іспиту. У другому – з моделлю додаткового опитування.

Під час тестування та верифікації тесту як продукту зазначений роздільний підхід дозволить оцінити різницю витрат часу за умови автоматизованого тестування і, з іншого боку, може бути використаний для прогнозування традиційного

іспиту як частина моделі, що визначає час, витрачений на вибір додаткових запитань.

У послідовному підході вибір завдань і реалізація обмежень є пов'язаними. Хоча зазначений підхід припускає оптимальну адаптацію, послідовна реалізація обмежень не є ідеальною. Алгоритми з такими особливостями мають схильність до вибору завдань із найбільшою кількістю зв'язків з іншими темами на початку тестування. Однак вибір деяких запитань може виявитися не оптимальним у подальшому проходженні тесту. У цьому разі це призведе до того, що результат оцінювання знань буде менш адекватним, ніж за умови оптимального адаптивного тестування, та/або до неможливості завершення тесту без порушення обмежень.

В експертних методах є опція експертного оцінювання всього тестового матеріалу перед процедурою тестування. Явне кодування суттєвих для завдань атрибутів не завжди є необхідним [19, 20]. Також зазначені підходи не мають повної адаптації у виборі завдання. Ці методи дозволяють часткову адаптацію з ієрархічними переходами. Угрупування блоків тестових завдань і створення багатоступінчастої тестової системи таким чином, щоб будь-яка частина системи задовольняла всі обмеження, призводить до значної комбінаторної проблеми, яка зі свого боку може спричинити виникнення неоптимальної системи з ієрархічними переходами та/або до порушення обмежень. Однак, як було зазначено, формальні методи для створення багатоступінчастих систем тестування існують.

Використання формальних методів для розроблення багатоступінчастих систем тестування дає змогу отримати явне кодування всіх важливих для завдання атрибутів.

Аналіз має здійснюватися за такими критеріями:

- інформативність – застосування принципу максимуму інформації для кінцевої кількості заданих тестових завдань;

- у процесі тестування здійснюється оцінювання знань і вмінь студента методами математичної статистики (наприклад j -та ймовірність правильної відповіді на завдання i -го рівня складності); крім цього, статистичне оброблення дозволяє визначити момент завершення тестування, що відбувається, коли функція розподілу j -х імовірностей правильних відповідей на завдання i -го рівня складності змінюється несуттєво;

- імовірність помилок – це ймовірність того, що тестування проводитиметься неадекватно знанням студента або відбудеться порушення обмежень системи тестування;

- складність проєктування;

- адекватність оцінювання знань суб'єкта тестування;

- рівень адаптації, що пов'язаний із критерієм інформативності, – що вищий рівень адаптації, то вища інформативність системи;

- експертна оцінка – результативний висновок, що дає змогу робити висновки про ті або інші властивості тестових завдань;

- формалізація процедур створення тестових завдань і обмежень, що припускає застосування рутинного алгоритму для досягнення заданої мети за певного набору початкових умов.

Аналіз наявних методів комп'ютерного адаптивного тестування приводить до дуже важливої дилеми. Алгоритм із оптимальними властивостями мав би обирати завдання в такій послідовності, щоб досягати оптимальної адаптації та одночасно брати до уваги всі обмеження. Це потрібно для того, щоб запобігти порушень деяких із них або не отримати неоптимальну адаптацію під час подальшого тестування.

Можливі такі вирішення цієї дилеми:

- реалізація алгоритму з можливістю повернення назад, щоб поліпшити наступні результати;

- реалізація лінійного алгоритму, який би зважав на наслідки прийнятих у майбутньому рішень. За умови адаптивного тестування відстеження у зворотному порядку неможливе, алгоритм застосовується в масштабі реального часу й попередній вибір не може бути скасований. Отже, залишається тільки один варіант – використовувати алгоритм, який щоразу обирає нове завдання. Це є новим класом алгоритмів.

4. Опис методу адаптивного тестування та застосування логічних мереж

Для визначення базового алгоритму необхідно навести сценарій роботи системи. У його основі лежить парадигма приймання іспиту викладачем у студента як модель адаптивного тестування. Такий вибір сценарію роботи системи зумовлений тим, що, по-перше, ця процедура історично склалася дуже давно й добре формалізована, по-друге,

у проєктуванні тестів їх розробникам необхідно спиратися на загальноприйняті, відомі та використовувані ним методи з мінімальною модифікацією. Сам алгоритм адаптивного тестування має супроводжуватися на кожному кроці (під час переходу від одного завдання до іншого) максимальною інформативністю (надавати максимум інформації про відповіді студента на кожне запропоноване запитання). Водночас необхідно зважати на суб'єктивні властивості учасника тестування, що можуть спричинити нерозуміння очевидно поставленого запитання або завдання.

Подання предиката P у вигляді кон'юнкції предикатів P_1, P_2, \dots, P_n називається його кон'юнктивною композицією. Розкладання предиката P у кон'юнкцію тих самих предикатів P_1, P_2, \dots, P_n називається його кон'юнктивною декомпозицією. Важливим висновком декомпозиції є так звана бінарна декомпозиція предиката P , яка характеризується тим, що кожний предикат у системі $\{P_1, P_2, \dots, P_n\}$ має в точності два істотні аргументи.

Тепер можна формально визначити логічну мережу для оцінювання складності наборів тестових завдань і встановлення шляху надання комплектів тестів для оцінювання знань суб'єктів навчання – це подання результату бінарної кон'юнктивної декомпозиції багатомісного предиката.

Предикати саме і є моделями наборів тестових завдань. Саме в прийнятному формальному визначенні логічна мережа виявилася простим і універсальним засобом подання структури будь-якого об'єкта. Це вагомий довід на користь того, щоб ототожнити формальне визначення логічної мережі з її змістовим визначенням.

Додаткові запитання є невід'ємною частиною тестів за певних умов, однак потрібно брати до уваги таке:

- не всі додаткові запитання є тотожними як з погляду складності, так і з погляду повноти відповідності основному запитанню;
- ланцюжки додаткових запитань є логічно пов'язаними послідовностями;
- запитання (як основні, так і додаткові) ставляться послідовно, тобто неможливо поставити два або більше запитань одночасно;
- частота подання додаткових запитань різна.

За базову модель системи адаптивного тестування обрано подання у вигляді логічних мереж

і математичні моделі алгебри скінченних предикатів, що має такі властивості:

- основною одиницею є запитання певної складності, у якого можуть бути ланцюги додаткових запитань;
- вибір додаткового запитання визначається з огляду на ймовірності появи кожного додаткового запитання як потоку найпростіших подій із шляхом логічного висновку на вибір неправильної відповіді;
- система тестів є замкненою в розумінні логічних мереж, тобто якщо запитання (основне або додаткове) є станом S_i , що має предикат виконання тестового завдання $P(S_i)$, то логічна сума предикатів тестових відповідей та перебування системи в стані S тотожно визначається як

$$P = \vee P(S_i),$$

де S_i – визначений стан – тестове завдання як процес оцінювання знань і вмінь студента; воно має виконуватися в разі вдалої відповіді на одне із запитань та/або додаткового запитання (або їх ланцюга), тобто вилучення одного із запитань не має призвести до присвоєння стану нульового значення.

Остання вимога дозволяє обходити запитання нескінченно, тобто завершення тесту можливе в таких ситуаціях:

- вичерпані всі запитання в банку тестових завдань;
- досягнуто кінець тесту;
- рівень знань оцінений з достатньою точністю;
- рівень знань суб'єкта навчання визначений як недостатній для досягнення критерію проходження тесту;
- суб'єкт навчання демонструє свою неспроможність, відповідаючи на запитання тесту.

Інформація береться з відповідей студента в процесі розв'язання завдання або з його передісторії. Комплексна модель адаптивної комп'ютерної системи навчання й тестування (КСНТ) будується на основі моделей учасників навчального процесу, подання знань предметної галузі, системи тестування й керування процесом навчання.

Комплексна модель адаптивної систем навчання й тестування має містити:

- семантичну модель навчального курсу;
- множину моделей суб'єктів навчання;
- модель організації процесу навчання й тестування знань тих, кого навчають.

– предметну галузь навчальної дисципліни, подану моделлю знань навчальних матеріалів.

Моделлю знань про структуру навчального матеріалу адаптивної системи навчання й тестування знань обрана логічна мережа, що містить відомості про поняття предметної галузі навчальної дисципліни, про їхні взаємозв'язки, складність окремих понять. Під методичним матеріалом будемо розуміти сукупність інформації, що містить знання про цілі, форми, методи й засоби навчання та розв'язання навчально-виховних завдань.



Рис. 4. Взаємодія моделей і процесів адаптивної комп'ютеризованої системи тестування

Модель системи адаптивного керування навчальним процесом і тестування знань студентів призначена для автоматичного визначення порядку подання їм матеріалів навчальної дисципліни залежно від результатів тестування на початку вхідного рівня знань із навчальної дисципліни й потім з уже вивчених тем і розділів.

Подання процесу тестування у вигляді предикатного опису логічних мереж не є вичерпним. Як зазначалося раніше, за умови і правильної, і неправильної відповіді на запитання має ухвалюватися одне з таких рішень:

- перехід до наступного основного запитання з вибором рівня його складності;
- перехід до додаткових запитань (до їх дерев), у цьому разі потрібно відкинути вже поставлені додаткові запитання (як окремі, так і їх дерева);
- повернення до основного запитання, якщо отримана відповідь на додаткове (або додаткові) запитання;
- завершення тестування.

Завдання як суб'єкт у розглянутому процесі тестування передбачає виконання правил, заданих екзаменатором. Це відповідає й традиційному

Модель суб'єкта навчання вирішує завдання опису попередніх знань, знань про навчальний курс та індивідуальних особистісних особливостей. Вона містить досить повну інформацію про суб'єкта навчання: рівень знань, умінь і навичок, здатність до навчання й виконання завдань (чи вміє студент використовувати отриману інформацію), особистісні характеристики (тип, орієнтація) та інші параметри, що можуть бути як кількісними, так і якісними.

Модель адаптивної системи тестування наведено на рис. 4. Вона містить декілька рівнів.

прийманні іспиту, тобто залежно від відповіді на поставлені запитання екзаменатор ухвалює зазначені вище рішення. Водночас він бере до уваги як фрагментарно розділені відповіді (оцінка за кожне запитання й висновок середнього), так і за певними логічними правилами весь ланцюжок відповідей суб'єкта тестування.

Загальним підходом у прийнятих методах, окрім загальних правил для всіх завдань, є компонування завдань із різних частин (рисуноків, таблиць, мультимедійного контенту), зумовлених як стимули. Це дає змогу заощаджувати ресурси з розміщення стимулів і виділяти програму оброблення відповідей, загальні правила в різних частинах логічної мережі, що розробляється.

Такий підхід не дозволяє достатньо індивідуалізувати тестові запитання. Насамперед це пов'язано з тим, що кожне запитання поєднує як безпосереднє завдання, так і рішення, пов'язане з виконанням цього завдання, з відповідями на додаткові запитання.

Для забезпечення гнучкості в прийнятті рішень, простоти створення запитань і логічних правил, що визначають прийняття рішень для конкретного

запитання, доцільним є об'єднання запитання й процедури ухвалення пов'язаного з ним рішення. Такий підхід спрощує як процедуру тестування, так і саму систему тестування з погляду задоволення вимог мінімальної складності застосовуваних алгоритмів.

Прийняття загальних для всього процесу тестування рішень потребує загальних підходів в одному сеансі тестування. Ці підходи визначаються:

- застосуванням загального методу (або методів), що визначає той крок у процесі тестування, коли додаткова інформація про знання суб'єкта тестування буде надлишковою;
- процедурою початку роботи системи тестування (вибір першого запитання) і стратегією переходу від одного запитання до наступного;
- наданням докладних результатів тестування, як у природній, так і обробленій тими або іншими методами формі.

Це вимагає використання протоколу тестування й застосування:

- алгоритмів логічного операційного та статистичного аналізу результатів тестування з погляду надмірності або недостатності інформації;
- алгоритмів, що визначають рівень підготовки суб'єкта тестування;
- алгоритмів, що забезпечують стохастичні переходи по мережах тестових завдань.

В окремій сесії модель суб'єкта навчання визначається такими параметрами:

- протоколом опитування;
- результатами оцінювання знань.

Отже, зазначений підхід формує парадигму тестування, що є природною для викладача, має аналог у класичному розумінні іспиту й визначається як модель суб'єкта навчання. Необхідно наголосити, що безпосереднє використання подання логічних мереж не забезпечує розв'язання поставленого завдання, тому що є ціла низка обмежень логічного характеру, що визначають як перехід зі стану в стан, вилучення станів, так і обмеження кількості переходів (тестування не може проводитися нескінченно й запитання не мають повторюватися).

Процедуру додаткових запитань пропонується виконувати двома способами. Перший спосіб полягає в тому, що до закінчення множини додаткових запитань (одного основного або єдиного запитання) не здійснюється жодних дій, крім фіксування відповідей з наступної передачі протоколу

щодо основного запитання, де ухвалюється одне з таких рішень:

- зарахувати відповідь на основне запитання як правильну з можливим коригуванням складності;
- вважати запитання незадовільним і перейти до групи запитань більш низького рівня;
- повторно поставити додаткове запитання із залученням додаткових запитань, що залишилися;
- повторно поставити основне запитання без залучення додаткових.

Другий спосіб полягає в тому, що функції прийняття рішень про правильності відповіді та/або перехід до інших основних, додаткових запитань або про завершення тестування передається додатковому запитанню.

Останнє є загальною формою, тобто псевдоінтерактивна процедура може бути зведена до маршової шляхом вилучення логічного аналізу й прийняття рішень (логічний перехід у розумінні теорії логічних мереж).

Наявність логічного аналізу й ухвалення рішення в тестовому завданні моделює правило поведінки екзаменатора. Прийнятий ситуативний підхід на підставі правил для конкретного запитання про поведінку (передача керування іншому запитанню) у межах окремо взятого завдання.

Додаткові запитання можуть бути обрані з основних із будь-яким рівнем складності та/або створені окремо (для конкретного тесту). Як і основним запитанням, їм має бути привласнений рівень складності. Отже, можна узагальнити запитання в наборі тестових завдань як структуру (клас) даних і визначити його поведінку. Необхідно, щоб запитання мали:

- ознаку, що визначає зв'язки (такі запитання, у які з нього можна перейти), обумовлені логічними висновками;
- ознаку, що визначає зв'язки (ті запитання, у які з нього можна перейти звичайним для логічних мереж способом);
- ознака "зупинки" запитання в логічних мережах, тобто ситуація, коли запитання вже було поставлено;
- ознака, що визначає, чи є запитання унікальним для певної мережі;
- ознака, що визначає, яке або які наступні запитання можуть бути викликані цим запитанням, тобто яким запитанням може бути передане керування.

Наведемо приклад створення логічної мережі адаптивного тестування.

Запитання мають відповідати таким положенням:

- забезпечити приймання керування від попереднього запитання з отриманням поточного протоколу опитування;

- забезпечити видачу тестового завдання і варіантів відповідей до них;

- забезпечити початок відліку власного часу в момент видачі завдання;

- підтримувати логічну функцію, що визначає зниження рівня складності, у разі виходу за межі часу, встановленого для відповіді на конкретне запитання та/або прийняти (за певних умов) для цього запитання рішення про правильні / неправильні відповіді;

- передати протокол опитування наступному запитанню і контролерові мережі запитань;

- передати ознаки, що визначають стан запитання контролерові мережі запитань;

- перевірити роботу запитання (передбачається передача керування контролерові мережі та/або наступному запитанню).

На рис. 5 зображений фрагмент побудованої мережі запитань (основне запитання з додатковими).

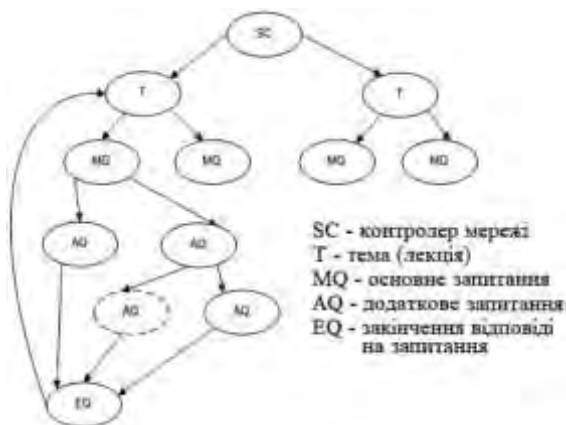


Рис. 5. Фрагмент мережі запитань

Використання логічних мереж на основі алгебри скінченних предикатів та предикатних операцій припускає наявність моделі системи як абстракції, що має сукупність станів $S = \{S_1, \dots, S_n\}$. Ці стани є взаємовиключними й перехід з одного стану в інший здійснюється різними для мережі шляхами через появу події, що є загалом предикатом подій. Необхідно, щоб кожний стан системи S_i мав як мінімум один вхід і один вихід, тобто так звані

"глухі" або кінцеві стани відсутні. Фактично будь-який стан системи може бути обраний як початковий.

Останній метод перевіряє предикат, чи сформувалася множина такої ознаки, як складність запитань для всієї групи суб'єктів навчання й одночасно чи сформувалася множина предикатів розподілу відповідей залежно від їх складності для конкретного суб'єкта тестування. Також збільшення складності алгоритму тестування не спричинить суттєве збільшення витрат ресурсів.

Висновки

Ідея адаптивного тестування, основана на блоках запитань, має безпосереднє відношення до одного із найбільш поширених форматів багатоступінчастого тестування, за умови якого суб'єкт навчання проходить послідовність тестів, рухаючись у бік більш складних запитань, даючи успішні відповіді, або до більш простих, якщо його відповіді не правильні. У цьому разі перехід від запитання до запитання відбувається за певними правилами, що описані алгебро-логічними рівняннями. Автоматизація в практиці тестування також уможливує статистичне оцінювання знань на кожному кроці тестування.

Виявлено низку обмежень логічного характеру, що визначають як перехід зі стану в стан, вилучення станів, так і обмеження кількості переходів (тестування не може проводитися нескінченно й запитання не мають повторюватися). Однак у разі незначної різниці між зазначеними передбачуваними й дійсними властивостями групи студентів застосування зазначеної методики має позитивний результат. Із цього можна зробити висновок, що впровадження запропонованого методу є ефективним для будь-яких обсягів та інших властивостей, що визначають тестування як процес.

Використання кон'юнктивної декомпозиції з бінарними предикатами дає змогу досягти мети дослідження, оскільки таким способом будь-який багатомісний предикат $P(x_1, x_2, \dots, x_m)$ можна подати логічною мережею. Долучення в кон'юнкцію, крім бінарних, унарних предикатів жодної користі не приносить, тому робити так немає необхідності. Можна будувати мережі, використовуючи кон'юнктивну декомпозицію предиката P на предикати P_1, P_2, \dots, P_m більшого числа їх істотних аргументів, але з інженерного погляду це нераціонально, оскільки

складність такої мережі (та її вартість і швидкодія) стрімко зростає зі збільшенням бінарності предикатів P_1, P_2, \dots, P_m . Отже, логічно приходимо до мереж з бінарними предикатами. Збільшення складності алгоритму тестування не призведе до суттєвих витрат ресурсів ПК.

Очікуваний результат щодо впровадження розроблених моделей логічних мереж моделювання наборів тестових завдань відповідно до параметричної моделі тестування досягнутий за умови апроксимації предикатної функції успіху, що припускає визначення уточнених значень дискримінаторів завдань способом розв'язання завдання оптимізації із використанням відомих алгоритмів. Це пов'язане насамперед з тим, що застосований метод опису тестових завдань є універсальними, а вирішення проблеми потребує

додаткового дослідження, метою якого є часткове заміщення цих алгоритмів операційними методами, що використовують як властивість поставленого завдання тестування, так і іншу апріорну інформацію.

Різниця в рівнях складності основних і додаткових запитань і запропонований зв'язок між основними запитаннями й гілками додаткових запитань дає змогу в процесі тестування мінімізувати кількість необхідних відповідей суб'єкта навчання для визначення рівня його знань і водночас суттєво поліпшити адаптаційні властивості тестування.

Застосування методів побудови логічних мереж для реалізації етапу перевірки знань у системах адаптивного навчання є ефективним для будь-яких обсягів та інших властивостей, що визначають тестування як процес.

Список літератури

1. Шубін І., Пітюкова М. Логічні мережі та їх використання для вирішення морфологічних завдань. Матеріали III Міжнародної конференції "Інноваційні технології в науці та освіті". Амстердам, Нідерланди, 2019. С. 402–405. URL: <http://dspace.opu.ua/jspui/bitstream/123456789/10382/1/%D0%86%D0%BD%D0%BD%D0%BE%D0%B2-%D0%A2%D0%B5%D1%85%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D1%96%D1%97-2019-%D0%90%D0%BC%D1%81%D1%82%D0%B5%D1%80%D0%B4%D0%B0%D0%BC-%D0%9B%D0%BE%D0%B7%D0%B0%D0%BD%D0%BE%D0%B2%D0%B0-%D0%9F%D1%94%D1%82%D1%83%D1%88%D0%BA%D0%BE%D0%B2%D0%B0-%D0%94%D1%80%D1%83%D0%BA%D0%A1%D1%82%D0%B0%D1%82%D1%82%D1%8F.pdf>
2. Backer, P., Siemens G. Educational data mining and learning analytics. The Cambridge handbook of the learning sciences, 2019. 274 p. DOI:10.1017/CBO9781139519526.016
3. Fourier J. Un modele d'indexation relationnel pour les graphes conceptuels fondee sur une interpretation logique, Phd thesis Universitee. Grenoble, 1998. 302 p. URL: https://www.academia.edu/2686445/Un_mod%C3%A8le_dindexation_relationnel_pour_les_graphes_conceptuels_fond%C3%A9_sur_une_interpr%C3%A9tation_logique
4. Gruzdo I., Kyrychenko I., Tereshchenko G., Shanidze O. Analysis of Models Usability Methods Used on Design Stage to Increase Site Optimization, Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems. (COLINS), Volume III: Intelligent Systems Workshop, 2023. In CEUR Workshop Proceedings, Vol. 3403, P. 387-4093. URL: <https://ceur-ws.org/Vol-3403/paper31.pdf>
5. Shubin I. Development of conjunctive decomposition tools. CEUR Workshop Proceedings, 2021. P. 890–900. URL: <https://ceur-ws.org/Vol-2870/>
6. Karataiev O., Sitnikov D., Sharonova N. A Method for Investigating Links between Discrete Data Features in Knowledge Bases in the Form of Predicate Equations, CEUR Workshop Proceedings. 2023. P. 224–235. URL: <https://ceur-ws.org/Vol-3387/paper17.pdf>
7. Тест як інструмент педагогічного моніторингу. URL: <http://opentest.com.ua/test-kak-instrument-izmereniya-urovnyaznaniy> (дата звернення: 20.11.2023.)
8. MyTestXPro – Система програм для створення та проведення комп'ютерного тестування, збору та аналізу їх результату, URL: <http://mytest.net> (дата звернення: 28.11.2023.)
9. Комп'ютерна програма тестування OpenTEST2. URL: <http://opentest.com.ua/kompyuternaya-programma-testirovaniyaznaniy-opentest-2> (дата звернення: 21.11.2023.)
10. Конструктор тестів Keepsoft. URL: <http://www.keepsoft.ru/simulator.htm> (дата звернення: 20.11.2023.)
11. Безкоштовна програма для тестування знань та онлайн-підготовки. URL: <http://xtls.org.ua/test.html> (дата звернення: 20.11.2023.)
12. Sharonova N. et al. Issues of Fact-based Information Analysis. International Conference on Computational Linguistics and Intelligent Systems. 2018. 178 p. URL: http://web.kpi.kharkov.ua/iks/wp-content/uploads/sites/113/2021/10/preface_colins_volume2_2018.pdf
13. Williams P. E-learning: what the literature tells us about distance education. *An overview. Aslib Proceedings*. Vol. 57. 2005. P 109–122. DOI: <https://doi.org/10.1108/00012530510589083>
14. Omran P. G., Wang K., Wang Z. An Embedding-based Approach to Rule Learning in Knowledge Graphs, *IEEE Transactions on Knowledge and Data Engineering*. Vol. 33(4). 2021. P. 1348–1359. DOI: 10.1109/TKDE.2019.2941685

15. Pellissier-Tanon T., Weikum G., Suchanek F. F. YAGO 4: A Reasonable Knowledge Base, *17th International Conference, ESWC 2020*, Heraklion, Crete, Greece, May 31–June 4. 2020, P. 583–596. DOI:10.1007/978-3-030-49461-2_34
16. Kyrychenko I., Malikin D. Research of Methods for Practical Educational Tasks Generation Based on Various Difficulty Levels, *6th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2022)*, May 12–13, 2022, Gliwice, Poland. CEUR Workshop Proceedings 3171, Volume I: Main, 2022. P. 1030–1042. URL: <https://ceur-ws.org/Vol-3171/paper74.pdf>
17. Omran P. G., Wang Z., Wang K. Scalable rule learning via learning representation, *Twenty-Seventh International Joint Conference on Artificial Intelligence. IJCAI-18*. 2018. P. 2149–2155. DOI:10.24963/ijcai.2018/297
18. Svato M., Schockaert S., Davis J. STRiKE: Rule-Driven Relational Learning Using Stratified k-Entailment, in: *ECAI, 2020*. URL: <https://ida.fel.cvut.cz/~kuzelka/pubs/ecai2020.pdf>
19. Sharonova N., Gruzdo I., Tereshchenko G. Generalized Semantic Analysis Algorithm of Natural Language Texts for Various Functional Style Types. *6th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2022)*, May 12–13, 2022, Gliwice, Poland. CEUR Workshop Proceedings 3171, Volume I: Main, P. 16–26. URL: <https://ceur-ws.org/Vol-3171/paper4.pdf>
20. Barkovska, O. Research into Speech-to-text Transformation Module in the Proposed Model of a Speaker's Automatic Speech Annotation. *Innovative Technologies and Scientific Solutions for Industries*. № 4 (22). 2022. P. 5–13. DOI: <https://doi.org/10.30837/ITSSI.2022.22.005>

References

1. Shubin, I., Pitiukova, M. "Logichni mezhy ta jih vykorystannja dl'a vyrishennia morfologichnyh zavdan. Materials of the 3rd International Conference Innovative Technologies in Science and Education. Amsterdam, the Netherlands", 2019. P. 402–405. available at: <http://dspace.opu.ua/jspui/bitstream/123456789/10382/1/%D0%86%D0%BD%D0%BD%D0%BE%D0%B2-%D0%A2%D0%B5%D1%85%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D1%96%D1%97-2019-%D0%90%D0%BC%D1%81%D1%82%D0%B5%D1%80%D0%B4%D0%B0%D0%BC-%D0%9B%D0%BE%D0%B7%D0%B0%D0%BD%D0%BE%D0%B2%D0%B0-%D0%9F%D1%94%D1%82%D1%83%D1%88%D0%BA%D0%BE%D0%B2%D0%B0-%D0%94%D1%80%D1%83%D0%BA%D0%A1%D1%82%D0%B0%D1%82%D1%82%D1%8F.pdf>
2. Backer, P., Siemens, G. (2019), "Educational data mining and learning analytics". *The Cambridge handbook of the learning sciences*, 274 p. DOI:10.1017/CBO9781139519526.016
3. Fourier, J. (1998), "Un modele d'indexation relationnel pour les graphes conceptuels fondee sur une interpretation logique", *Phd thesis Universitee. Grenoble*, 302 p. available at: https://www.academia.edu/2686445/Un_mod%C3%A8le_dindexation_relationnel_pour_les_graphes_conceptuels_fond%C3%A9_sur_une_interpr%C3%A9tation_logique
4. Gruzdo, I., Kyrychenko, I., Tereshchenko, G., Shanidze, O. "Analysis of Models Usability Methods Used on Design Stage to Increase Site Optimization", *Proceedings of the 7th International Conference on Computational Linguistics and Intelligent Systems. (COLINS)*, Volume III: Intelligent Systems Workshop, 2023. In CEUR Workshop Proceedings, Vol. 3403, P. 387–4093. available at: <https://ceur-ws.org/Vol-3403/paper31.pdf>
5. Shubin, I. "Development of conjunctive decomposition tools". CEUR Workshop Proceedings, 2021. P. 890–900. available at: <https://ceur-ws.org/Vol-2870/>
6. Karataiev, O., Sitnikov, D., Sharonova, N. "A Method for Investigating Links between Discrete Data Features in Knowledge Bases in the Form of Predicate Equations", CEUR Workshop Proceedings, 2023, P. 224–235. available at: <https://ceur-ws.org/Vol-3387/paper17.pdf>
7. "The test as a tool of pedagogical monitoring" ["Test yak instrument pedahohichnoho monitorynhu"] available at: <http://opentest.com.ua/test-kak-instrument-izmereniya-urovnya-znaniy> (last accessed 20.11.2023.)
8. "MyTestXPro – System of programs for creating and conducting computer testing, collecting and analyzing their results". ["Systema prohram dlia stvorennia ta provedennia kompiuternoho testuvannia, zboru ta analizu yikh rezultativ"] available at: <http://mytest.net> (last accessed 28.11.2023.)
9. "Computer testing program OpenTEST2" ["Kompiuterna prohrama testuvannia OpenTEST2"]. available at: <http://opentest.com.ua/kompyuternaya-programma-testirovaniya-znaniy-opentest-2> (last accessed 21.11.2023.)
10. "Test designer Keepsoft". ["Konstruktor testiv Keepsoft"]. available at: <http://www.keepsoft.ru/simulator.htm> (last accessed 20.11.2023.)
11. "Free program for knowledge testing and online preparation". ["Bezkoshtovna prohrama dlia testuvannia znan ta onlain pidhotovky"]. available at: <http://xtls.org.ua/test.html> (last accessed 20.11.2023.)
12. Sharonova, N. et al. "Issues of Fact-based Information Analysis". *International Conference on Computational Linguistics and Intelligent Systems*. 2018. 178 p. available at: http://web.kpi.kharkov.ua/iks/wp-content/uploads/sites/113/2021/10/preface_colins_volume2_2018.pdf
13. Williams, P. (2005), "E-learning: what the literature tells us about distance education". *An overview. Aslib Proceedings*. Vol. 57. P 109–122. DOI: <https://doi.org/10.1108/00012530510589083>
14. Omran, P. G., Wang, K., Wang, Z. (2021), "An Embedding-based Approach to Rule Learning in Knowledge Graphs", *IEEE Transactions on Knowledge and Data Engineering*. Vol. 33(4). P. 1348–1359. DOI: 10.1109/TKDE.2019.2941685

15. Pellissier-Tanon, T., Weikum, G., Suchanek, F. (2020), "F. YAGO 4: A Reasonable Knowledge Base", *17th International Conference, ESWC 2020*, Heraklion, Crete, Greece, May 31–June 4. P. 583-596. DOI:10.1007/978-3-030-49461-2_34
16. Kyrychenko, I., Malikin, D. "Research of Methods for Practical Educational Tasks Generation Based on Various Difficulty Levels" 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2022), May 12–13, 2022, Gliwice, Poland. CEUR Workshop Proceedings 3171, Volume I: Main, 2022. P. 1030–1042. available at: <https://ceur-ws.org/Vol-3171/paper74.pdf>
17. Omran, P. G., Wang, Z., Wang, K. (2018), "Scalable rule learning via learning representation", *Twenty-Seventh International Joint Conference on Artificial Intelligence. IJCAI-18*. P. 2149–2155. DOI: 10.24963/ijcai.2018/297
18. Svato, M., Schockaert, S., Davis, J. "STRiKE: Rule-Driven Relational Learning Using Stratified k-Entailment", in: ECAI, 2020. available at: <https://ida.fel.cvut.cz/~kuzelka/pubs/ecai2020.pdf>
19. Sharonova, N., Gruzdo, I., Tereshchenko, G. "Generalized Semantic Analysis Algorithm of Natural Language Texts for Various Functional Style Types". 6th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2022), May 12–13, 2022, Gliwice, Poland. CEUR Workshop Proceedings 3171, Volume I: Main, 2022. P. 16–26. available at: <https://ceur-ws.org/Vol-3171/paper4.pdf>
20. Barkovska, O. (2022), "Research into Speech-to-text Transformation Module in the Proposed Model of a Speaker's Automatic Speech Annotation". *Innovative Technologies and Scientific Solutions for Industries*. № 4 (22). P. 5–13. DOI: <https://doi.org/10.30837/ITSSI.2022.22.005>

Надійшла 07.12.2023

Відомості про авторів / About the Authors

Ляшик Володимир Андрійович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна; e-mail: volodymyr.liashyk@nure.ua; ORCID ID: <https://orcid.org/0000-0001-7326-0813>

Шубін Ігор Юрійович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри програмної інженерії, Харків, Україна; e-mail: igor.shubin@nure.ua; ORCID ID: <https://orcid.org/0000-0002-1073-023X>

Liashyk Volodymyr – Kharkiv National University of Radio Electronics, Postgraduate, Kharkiv, Ukraine.

Shubin Ihor – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Professor at the Department of Software, Kharkiv, Ukraine.

METHOD OF LOGIC NETWORKS FOR MODELING SYSTEMS OF ADAPTIVE KNOWLEDGE TESTING

The subject of the research is the development of mathematical and algorithmic support of the intellectual toolkit for the analysis of sets of test tasks and the modeling of the process of interpreting the quality of sets of test tasks, which allows for objective and comprehensive continuous control of the knowledge of subjects of training, subject to the implementation of the concept of virtual distributed training (retraining). **The purpose** of the research is to improve the effectiveness of monitoring the knowledge of subjects of education in the distance form of education through the use of adaptive computer testing methods based on models of logical networks and the algebra of finite predicates. The following **tasks** are solved in the article: the formation of a testing model in a distributed virtual learning environment and a model of validity assessment based on the content of sets of test tasks. The following **methods** are used: methods of logical networks and algebraic programming based on the algebra of finite predicates and predicate operations, intellectual analysis of information. The following **results** were obtained: the principles of intellectual analysis were formulated in the development of a model of a universal logical network and its application to actual tasks of artificial intelligence in the field of informal information processing, namely, in the construction of knowledge testing systems for distributed virtual learning. **Conclusions.** Algorithms for optimal multi-stage adaptive testing of knowledge as part of distributed virtual learning models and methods for analyzing the success of training subjects have been improved. The use of conjunctive decomposition with binary predicates achieves the goal of the research, because in this way any multi-place predicate can be represented by a logical network simulating the process of knowledge testing, the model of the subject of learning is described.

Keywords: software engineering; knowledge bases; algebra of finite predicates; logical networks; logical rules; reliability of tests; use of knowledge; model of the learning subject.

Бібліографічні описи / Bibliographic descriptions

Ляшик В. А., Шубін І. Ю. Метод логічних мереж для моделювання систем адаптивного тестування знань. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 45–57. DOI: <https://doi.org/10.30837/ITSSI.2023.26.045>

Liashyk, V., Shubin, I. (2023), "The Method Of Logical Networks For Modeling Adaptive Knowledge Testing Systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 45–57. DOI: <https://doi.org/10.30837/ITSSI.2023.26.045>

A. POROSHENKO, A. KOVALENKO

AUDIO SIGNAL TRANSMISSION METHOD IN NETWORK-BASED AUDIO ANALYTICS SYSTEM

The **subject matter** of the article is audio signal transmission method in network-based audio analytics system. The creation of a network-based audio analytics system leads to the emergence of new classes of load sources that transmit packetized sound data. Therefore, without constructing adequate mathematical models, it is impossible to build a well-functioning network-based audio analytics system. A fundamental question in traffic theory is the question of load source models. The development of an method for transmitting audio signals in a network-based audio analytics system becomes necessary. Based on this, the **goal** of the work is to create methods an method for transmitting audio signals in a network-based audio analytics system to ensure efficiency and accuracy in audio analytics. The following **tasks** were solved in the article: the formation of a model for the system's load sources, investigation of connection and traffic management, implementation of control and traffic monitoring functions in the network, research of methods to ensure the quality of audio signal transmission and the development of a method of transmitting an audio signal by virtual routes switching. To achieve these goals, the following **methods** are used: mathematical signal processing, data compression algorithms, optimization of network protocols, and the use of high-speed network connections. The obtained **results** include modeling of the system's load sources, examination of connection and traffic management, investigation of methods to ensure the quality of audio signal transmission and a method of transmitting an audio signal by virtual routes switching was proposed. In **conclusion**, the possibilities of using simulation modeling of nodes in the network-based audio analytics system are highly limited. This is explained by the fact that the acceptable level of information loss in data centers is very low. The use of the developed method enables effective control and processing of sound information in real-time. This method can find broad applications in various fields, including security, healthcare, management systems, and other industries where the analysis of audio signals is a crucial element.

Keywords: audio analytics; audio signal transmission method; traffic management; virtual routes; delay.

Introduction

In the era of continuous development of audio technology and network systems, the importance of audio signal transmission in a networked audio analytics system is becoming indisputable. The method of audio signal transmission in a networked audio analytics system plays an important role in the development and implementation of modern technologies aimed at improving the safety, convenience, and efficiency of various areas of our lives.

Networked audio analytics systems are used to monitor and detect events in public places, offices, transportation, etc. The effectiveness of these systems depends on the accuracy and speed of audio signal transmission, which allows for real-time response to danger or violations. In the field of building and infrastructure management, network systems are used that analyze audio signals to detect changes in the environment. This allows you to automatically respond to changes in the use of premises and energy efficiency. By transmitting audio signals in a networked system, it is possible to implement interactive control and interaction systems, such as voice assistants or voice recognition systems, which make the use of technology more

convenient and accessible. In the medical field, the transmission of audio signals in a networked system is used to record and analyze sounds, which can be useful in diagnosing various conditions or monitoring patients.

The ability of the transmission method to operate in real time is of great importance. Delay in the transmission of audio signals can negatively affect the system's response time, which is critical in security or instant control scenarios. Therefore, the audio transmission method in a networked audio analytics system becomes a key tool for improving various aspects of everyday life, providing the ability to use audio information with maximum efficiency and safety.

Analysis of recent research and publications

The efficiency of an audio analytics system depends on the accuracy of audio signal recognition. Papers [1–4] consider modern methods for processing various audio signals, including human speech. The analysis of the dataset for training the audio signal recognition model is presented in [5]. The task of evaluating the accuracy of an audio signal processing method is presented in [6]. The transmission method must ensure stable transmission

of high-quality audio data so that the system can correctly identify and analyze audio information.

Delay and real-time capability are critical parameters for the effectiveness of the audio transmission method in a networked audio analytics system, especially in demanding applications such as security, medicine, and other areas where rapid response to events is important. Additionally, the transmission method should be optimized to minimize the bandwidth consumed, especially in the face of large amounts of data being transmitted. This makes the system more efficient and cost-effective. These issues are discussed in [7–9].

The transmission method must demonstrate resistance to errors and noise that may occur during the transmission of audio signals over the network. Ensuring reliable transmission even in the presence of various interferences is a key aspect of efficiency. The issues of audio signal compression and audio codecs are considered in [10–12].

The ability of the transmission method to adapt to an increase in the amount of data and the scale of the system is an important factor in the context of the development and expansion of applications of the networked audio analytics system. The issues of scalability of complex networks are discussed in [13, 14] and the results of these works can be applied to the construction of networked audio analytics systems.

Due to the increase in the number of connected devices and the volume of audio data exchange, it is important to develop methods to protect against potential attacks and ensure the confidentiality of transmitted data. Security issues are discussed in [15, 16].

Identifying parts of a common problem

In the digital era, the transmission of audio signals in a networked audio analytics system is proving to be a key aspect for the development and improvement of audio technology. Methods of transmitting audio data over a network are becoming increasingly important in ensuring not only the quality of sound reproduction, but also in analyzing, processing and understanding large amounts of audio information. The creation of a network audio analytics system leads to the emergence of new classes of load sources that transmit packetized audio data. Therefore, without building sufficiently adequate mathematical models, it is impossible to build a well-functioning network audio analytics system. A fundamental issue in the theory of telecommunication traffic is the issue of load source models.

Thus, it becomes necessary to develop a method for audio signal transmission in a networked audio analytics system. Based on this, the purpose of this paper is to develop a method for transmitting audio signals in a networked audio analytics system by switching virtual routes.

Modeling of system load sources

The main feature of the load sources of network audio analytics systems is that they are characterized by different information transmission rates. The introduction of new models of load sources requires new descriptions for the load itself. When describing the load models that arise in networked audio analytics systems, a three-level load processing model is used, as shown in Fig. 1.

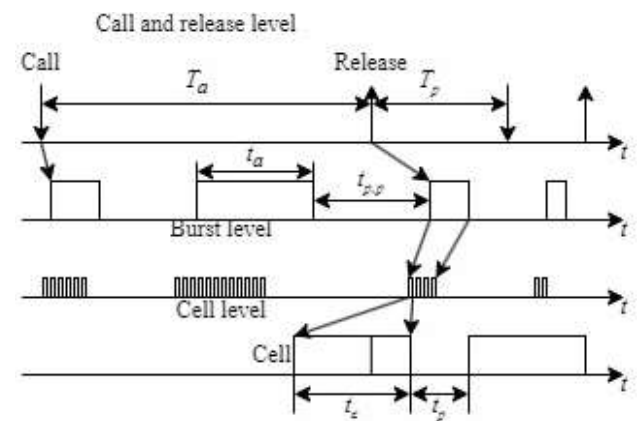


Fig. 1. Three-level load model:

- T_a – time of the user's active state;
- T_p – time of the user's passive state;
- t_a – time of the user's active information transmission;
- t_p – duration of the pause in information transmission;
- $t_{p,p}$ – duration of the pause between cell transmissions;
- t_c – time of cell transmission.

The first level – the Call Level – corresponds to the duration of a communication session. At this level, it is necessary to take into account the activity of individual load sources, to know the intensity of calls from them and the intensity of their service.

The second level (Burst Level) determines the nature of the bitstream at the time of user service and corresponds to the moments of activity and passivity. The distribution of pause durations between bit streams determines the burstiness of the message.

The third level is the Cell Level, which is the level of transmission of cells transmitted to the transport network. Transmitted cells from all load sources form

the traffic of the audio analytics network system. This level directly determines the capacity of buffer drives, the probability of cell loss, and the time they wait in the queue.

A networked audio analytics system is a system of high structural complexity, for which analytical models have been developed relatively recently. Simulation modeling is widely used to verify analytical models of user load and quality of service (GoS). In networked systems, audio analysts estimate GoS not only by the probability of loss of information cells, but also by the probability of their delay, and even by the value of delay jitter.

Call admission rules and the service strategy itself have a significant impact on these service quality characteristics. A large number of studies have been devoted to solving these three fundamental problems when creating networked audio analytics systems [17, 18].

A certain flexibility is required from the traffic management system because it is necessary to maintain the specified QoS and GoS levels. At the same time, the control system is required, on the one hand, to be relatively simple to implement at the ATM level and, on the other hand, to be robust, i.e., to ensure high efficiency of resource management when the traffic situation on the network changes. When organizing access to the network system, audio analysts distinguish two aspects, namely the rule for call access in the network (Call Admission) and the strategy for servicing a call when the call has already been accepted for service (Source Policing).

The first aspect is related to the study of the impact of priorities on the quality of service, and the second aspect is related to the uniformity of information cells in the network. The rules for admitting a call to the network are based on knowledge of the link state, i.e. the number of calls of different types accepted for service. The decision to accept the next call is made if the probability of loss of information cells is within acceptable limits.

To satisfy these aspects, the network audio analytics system needs to implement two functions of managing and controlling traffic in the network: user access control (Connection Admission Control) and Usage Parameter Control (UPC).

Managing connections in a networked audio analytics system

The Connection Admission Control (CAC) algorithm defines the actions performed by the network that allow

a connection to be established. With the CAC algorithm, a connection is established only if the network can support this connection with the required QoS and does not degrade the quality of service of already established connections. The algorithm allows you to block the admission of new connections to the network if the accumulation of cells in the queues increases. The ACS algorithm performs actions both at the initial stage of call service and at the stage of renegotiation of call service conditions. The ACS algorithm decides whether to provide a virtual channel connection (VC) or a virtual path connection (VP) to the consumer. This algorithm strikes a balance between efficiency and QoS.

To decide whether to accept or reject a connection, the ACS algorithm must determine the acceptable traffic volume limits and the required QoS class (values of cell delays, cell delay jitter, cell losses and bursts).

At the stage of implementing the CAC algorithm, the so-called Traffic contract is determined. This contract requires compliance with a set of service parameters, including QoS parameters. According to the ITU recommendations, the contract is monitored by checking it for compliance with the accepted service rule and network QoS. The contract identifies the user with whom it is possible to change its further service. The set of parameters describing the traffic is called the traffic descriptor of the connection being established. The traffic descriptor is regulated by the ITU and is associated with the establishment of a single BP connection or a connection BC. According to the ITU recommendations: the traffic descriptor defines the traffic parameters, the compliance of the traffic parameters with the validation rule, and the network QoS.

Thus, the CAC algorithm represents a set of actions that must be performed at the connection establishment stage in order to accept or reject user service. A connection request is accepted for service only if there are sufficient resources to realize a connection throughout the network with the required GoS. At the same time, it is a mandatory requirement to maintain the accepted quality of service of already established connections.

Traffic management in a networked audio analytics system

User-generated load policing and Usage/Network Parameter Control (UPC/NPC) are performed at the User-Network Interface (UNI) and Network-Node Interface (NNI) level. This control is a set of actions

that must be performed by the network to monitor and manage the traffic generated by ATM connections. The main purpose of managing the parameters of the traffic being served is to monitor each established connection for its stricter compliance with the traffic contract. An ideal UPC/NPC algorithm should be able to detect any unacceptable traffic situations, respond quickly to deteriorating traffic parameters, and have a simple implementation.

Among the methods proposed to meet these requirements, considerable attention is paid to the method that has received the conventional name of "Leaky bucket". Another method widely used in data transmission networks is the so-called Sliding Window Protocol.

Two approaches are used to build a control procedure: statistical and based on the use of some a priori rules (the so-called operational approach). The disadvantage of the statistical approach is the need for long-term observation and, as a result, a slow management response. Therefore, in practice, operational approaches are used.

ITU Recommendation 1.371 presents two equivalent versions of the operational algorithm – the so-called Generic Cell Rate Algorithm (GCRA). The first version is called the Virtual Scheduling Algorithm (VSA), and the second version is called the Leaky Bucket Algorithm (LBA).

Both versions of GCRA, for any sequence of cell arrival moments ($t_a \geq 1$), determine the cells that can be transmitted along the link or their transmission will be delayed.

The VSA virtual scheduling algorithm uses two parameters: I – increment; L – some limit of its increment and a set of intermediate variables $GCRA(I, L)$. The theoretically predicted moment of TAT cell appearance is calculated under the assumption that the cells are uniformly distributed in time, and the distance between two consecutive cells is equal to I at the moment of source activity. If the actual arrival time exceeds $t = TAT - L$, where L is some acceptable value, then the cell is allowed, otherwise the cell is delayed.

A diagram of the virtual scheduling algorithm is shown in Fig. 2.

The LBA algorithm (Fig. 3) shows how the burst flow of cells fills bucket B1. If bucket B1 is not overflowing, then cells enter the transmission medium at a guaranteed rate. If bucket B1 is full, the excess flow goes to bucket B2. If bucket B2 is not overflowing, then from B2 the flow of cells enters the transmission

medium at an arbitrary rate, but less than the guaranteed rate. If B2 is full, the excess flow is discarded. The described flow control scheme corresponds to the algorithm shown in Fig. 4.

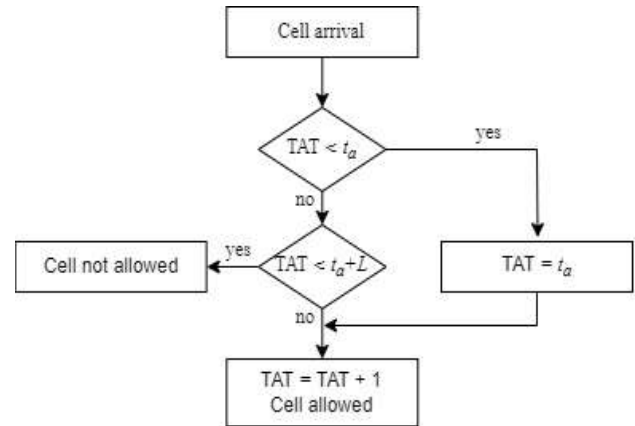


Fig. 2. Virtual scheduling algorithm

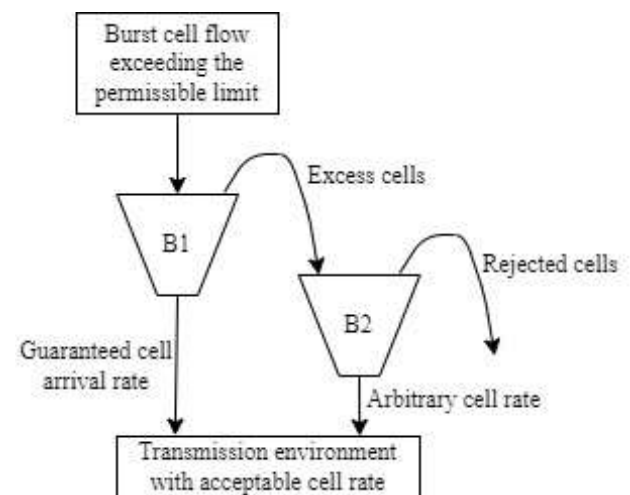


Fig. 3. Physical interpretation of the "leaky bucket" method

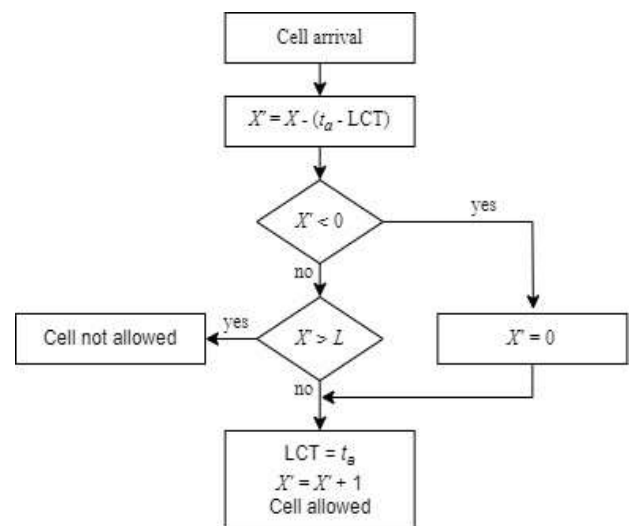


Fig. 4. "Leaky bucket" bitstream control algorithm

It should be noted that unlike the sliding window mechanism, the leaky bucket mechanism operates with individual cells. The flow of cells is copied to the so-called pseudo-queue ("leaky bucket"). The queue is serviced at a certain speed (the rate of the bucket). When the buffer is full, the cells of the outgoing stream are delayed.

Each established connection is assigned a counter, the content of which increases by one when an information cell arrives and decreases accordingly if the current bit rate is acceptable in the transmission link. The counter has a certain threshold L . If the counter is equal to L , the access of the cells to the network is terminated, and they are transferred to the buffer drive. Thus, this method of service controls two parameters: the current information transfer rate (the number of cells transmitted per unit of time) and the threshold L (bucket capacity).

Estimates of the probability of cell loss and waiting time in the buffer drive during statistical multiplexing, taking into account the above factors, are important system characteristics.

Ensuring transmission reliability

Semantic transparency (ST) of a network is defined as the ability of a network to transport messages from a source to a destination with an acceptable number of errors. With a given transmission system, network transparency is ensured by encoding the message, repeating the message at the receiver's request, and a combination of these techniques. The most important characteristic that defines the ST is the possibility of distortion of bits of information – the Bit Error Rate (BER). BER is defined as the total number of erroneous bits divided by the total number of transmitted bits.

For different transmission systems, the BER measurement is practically reduced to a set of statistics. Since the probability of bit distortion is different in the transmission systems and the transmission medium, it takes different time to obtain representative statistics. In modern fiber-optic communication lines, the value of $BER = 10^{-9}$. As the BER probability increases, the number of retransmissions of damaged packets increases. The increase in load can be estimated by the following formula, which takes into account the use of the sliding window control protocol:

$$R(n) = \frac{W}{2} \times \frac{1 - (1 - BER)^{nL}}{(1 - BER)^{nL}}, \quad (1)$$

where W – window size;

n – number of transmission links;

L – packet length, bits.

Figure 5 shows the dependence of traffic growth on the change in BER on the transmission link. The graph is plotted for the case of five transmission links ($n = 5$). In this figure, the traffic is evaluated relative to the original traffic. For example, if $R(BER) = 1$, it means that the number of transmitted packets doubles.

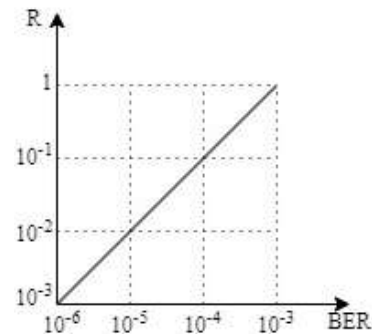


Fig. 5. Dependence of load increase on the probability of BER error on the transmission link

At a low BER value, link-to-link control is not very effective, because the probability of errors is extremely low. In this case, end-to-end control is the most effective.

Figures 6 and 7 illustrate the case of link control and end-to-end control. They show the dependence of the load increase on the change in the BER error probability. For end-to-end control, the case of five links is considered. Fig. 6 corresponds to the case when the packet length is 53 octets, Fig. 7 corresponds to the case when the packet length is 1000 octets.

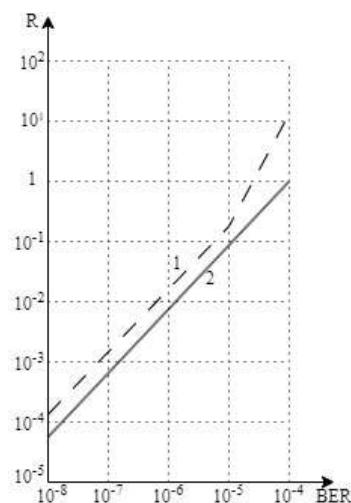


Fig. 6. Dependence of the change in link utilization on the change in the error probability on the link at a packet length of 53 octets: 1 – end-to-end control, five links; 2 – control on one transmission link

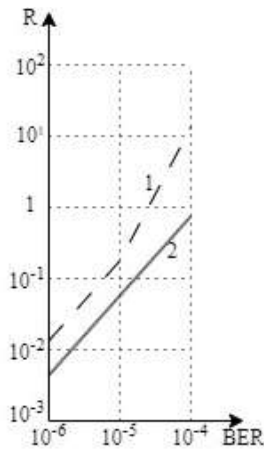


Fig. 7. Dependence of the change in link utilization on the change in the error probability on the link at a packet length of 1000 octets:
1 – end-to-end control, five transmission links;
2 – control on one transmission link

The lower the BER value in the channel, the more efficient the end-to-end control is compared to the link control. Fig. 8 shows the dependence of the efficiency of end-to-end control compared to link control for a standard cell of length $53 \times 8 = 424$. The effectiveness will be evaluated by the ratio of the loads of the path with n-link control and one-link control:

$$C = R(n)/R(1). \quad (2)$$

It is assumed that there are five transmission links between users ($n = 5$), i.e.:

$$C = R(n)/R(5). \quad (3)$$

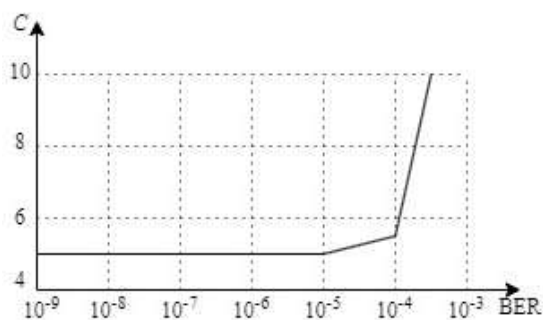


Fig. 8. Dependence of efficiency change on error probability

From the graph shown in Fig. 8, it can be determined that up to a value $BER \leq 10^{-5}$, "end-to-end" control is practically indistinguishable from link control. Indeed, if $BER \leq 10^{-9}$ then $C = 5,0$, and if $BER \leq 10^{-5}$ then $C = 5,04$.

In packet-oriented networks, a different assessment is used – PER (Packets Error Rate). The PER value is determined for a certain normalized time and

significantly depends on the properties of the transmission medium. It is calculated as the number of erroneously received packets to the total number of transmitted packets.

Buffer overflow can also lead to an increase in the load on the link. However, in practice, the probability of cell loss is extremely low ($P_{cell} \leq 10^{-9}$). The capacity of buffer drives that ensure a low probability of cell loss can be seen in Fig. 9, which was calculated using the model $M/D/1/L$ ($L \neq \infty$). The curve is plotted at the load intensity applied to the link $a = 0,8$ Erl.

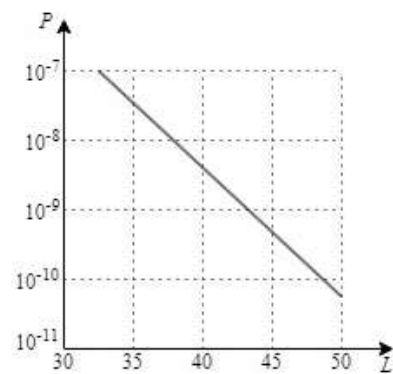


Fig. 9. Dependence of the probability of cell loss on the change in the buffer storage capacity at $a = 0,8$ Erl

Compliance with time requirements

The time transparency of the network is measured by the message delay and the end-to-end jitter of the message delay. Delay significantly affects the quality of communication for users who transmit real-time audio or video. Long delays for audio messages are manifested in the form of an echo effect. The amount of signal delay during end-to-end transmission is normalized.

The delay time consists of two components, namely, the delay caused by the transmission medium and the delay caused by the processing of messages (packets). In digital switching systems, the amount of delay is limited by the value of $T_{csk} \leq 450$ μ s.

The delay time caused by message processing can be estimated as a first approximation as the probability that the actual waiting time will exceed the permissible time t_a and can be defined as follows:

$$P(> t_a) = 1 - (1 - \lambda) \left[\sum_{k=0}^{[t]} \frac{(\lambda(k-t))^k}{k!} e^{\lambda(t-k)} \right]. \quad (6)$$

The form of the function $P(> t_a)$ for $\lambda = 0,2, 0,4$ and $0,8$ as a function of waiting time is shown in Fig. 10.

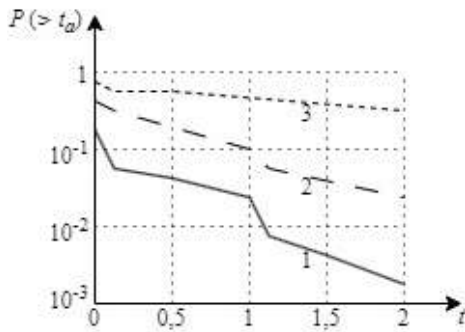


Fig. 10. Change of probability $P(> t_a)$ for three load values: 0,2 Erl; 0,4 Erl; 0,8 Erl

Dependence of the average time spent in the queue T_c of the delayed cells from the change in load a is shown in Fig. 11.

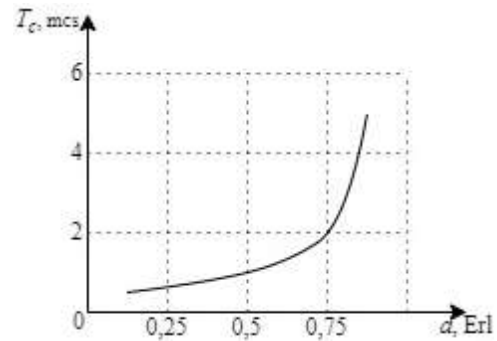


Fig. 11. Dependence of the average delay time on changes in load intensity

The nature of connection delays for different transmission speeds on the transmission link (1 and 10 Gbps) can be seen in Table 1. The data in Table 1 are obtained with the probability of packet loss and the intensity of the incoming load $a = 0,8$ Erl.

Table 1. The nature of connection delays at different bit rates

Delays, μ s	Transmission speed on the link, Gbit/s					
	1			10		
	Packet size, bytes					
	256	512	1024	256	512	1024
Transmissions	4000	4000	4000	4000	4000	4000
Fixed on the switching node	16	32	64	4	8	16
In the queue and during depacketization	100	200	400	10	20	40
During packetization	500	1000	2000	100	200	300
In a synchronous network	900	900	900	900	900	900

Method of audio signal transmission by switching virtual routes

The method of audio signal transmission in an audio analytics network system by switching virtual routes is based on the principle of fast packet switching. This method is designed to optimize the transmission of audio signals in a network audio analytics system. The main goal is to ensure the efficiency and accuracy of real-time audio analytics. When creating this method, mathematical models were developed to adequately reflect new classes of load sources that transmit packetized audio data and analysis and improvement of connection and traffic management methods in the network system for resource utilization were performed. As a result, mechanisms for managing user access to the network and managing resource utilization parameters were implemented, and an efficient switching method based on the use of virtual routes for routing audio packets was created.

The principle of fast packet switching is shown in Fig. 12. A cell arriving at a switching node (SN)

contains an information field and a header. At the switching node, using a special routing process, the cell is assigned bits of an additional address representing the address of the route of the cell through the switching system. As the cell passes through the cascades of the switching system, the bits of the additional address are used to select a route. The output cell, as before, contains only the information field and the header.

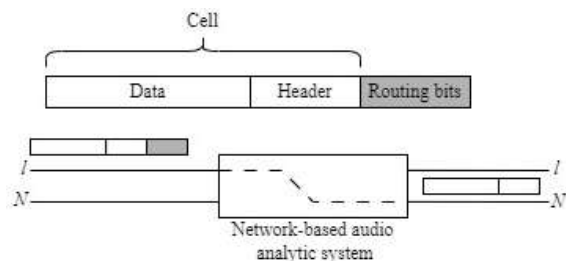


Fig. 12. The principle of fast packet switching

To connect two terminals to each other, you need to switch not only virtual channels, but also virtual paths. A VP switch is a virtual path switch, a VP/VC switch is a virtual path switch, and a virtual channel switch.

A VP/VC switch consists of a virtual path switch (VP switch) and a virtual channel switch (VC switch). The switching scheme of virtual paths and channels is shown in Fig. 13.

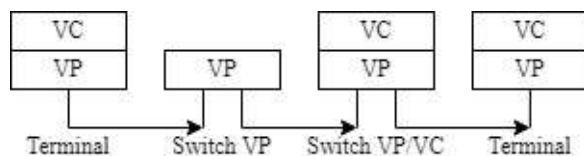


Fig. 13. Switching scheme for virtual paths and channels on the network

Assigning a route address and routing a cell by a switching system requires certain management resources. Fig. 14 shows the conditional dependence of control costs on the transmission rate for different types of switching. As follows from Fig. 14, the control costs of the FPS are significantly higher than in the case of channel switching and multichannel switching, but lower than in the case of packet switching in virtual channel mode and datagram mode.

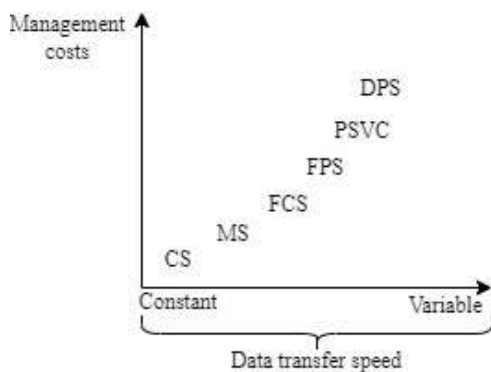


Fig. 14. Dependence of management costs on the transmission rate for different switching modes:

CS – channel switching; MS – multichannel switching;
FCS – fast channel switching; FPS – fast packet switching;
PSVC – packet switching with virtual channels;
DPS – datagram packet switching

The possibilities of using simulation modeling of the nodes of the audio analytics network system are very limited. This is due to the fact that the acceptable level of loss of information cells is very low. When constructing probabilistic models of audio analytics network systems, various approximations are used: Markov processes, combinatorial approaches, methods based on decomposing the system into subsystems with a simpler structure. In some cases, the Wiener-Hopf approximation is used to estimate the delays of information centers, which sometimes allows

obtaining fairly simple formulas when the incoming flow of requests is a Markov modulated Poisson flow.

The quality of service of information cells in a networked audio analytics system is significantly affected by routing. The so-called FPS mode is implemented at the AC of the audio analytics network system. This switching mode has both the properties of the channel switching mode and the packet switching mode.

When creating networked audio analytics systems, the search for networks with an optimal structure is continuously underway, as well as the development and research of methods for calculating their probabilistic and time characteristics. Intensive research is currently underway to create new classes of networks. In particular, networks with so-called "variable" routing are of interest. In such networks, an information cell can pass through only a part of the system's cascades, not all of them. To create such opportunities, a special switching element with a capacity of 2×4 . Special classes of networks are being studied that allow for the selection of a shared buffer. Due to the rapid improvement of the technology, it is interesting to study networks built with the use of high-capacity switches: 8×8 , 16×16 , 32×32 inputs and outputs.

An important problem when building networked audio analytics systems is the problem of network management. When creating a sufficiently branched network audio analytics system, the problem of distributing the total transmission rate along the path between individual network nodes arises. Network management also affects the admission of calls to the network and routing. These issues are of great importance and are still poorly understood.

Conclusions

The main requirements for the efficiency of the audio transmission method in a networked audio analytics system include minimizing latency to meet high standards of interaction and security. Research in the field of optimizing audio data processing algorithms and the use of fast network connections are identified as important factors in achieving these goals.

Consideration of scalability and error tolerance requirements is also important for the successful implementation of the audio transmission method at various levels. A balanced approach to the development and improvement of this technology allows to take into account the diverse requirements of different areas and opens the way for innovative solutions.

Thus, the following results were obtained in this article:

1. Modeling of the system load sources was performed.
2. Connection and traffic management in a networked audio analytics system is considered.
3. Methods for ensuring the quality of audio signal transmission are investigated.
4. A method of audio signal transmission by switching virtual routes is proposed.

Future research in the field of audio signal transmission methods in a networked audio analytics system may focus on the development of data compression technologies, the use of artificial intelligence to improve recognition accuracy, and the study of integration with other technologies, which will help to increase the functionality and versatility of the method.

References

1. Kholiev, V., Barkovska, O. (2023), "Comparative analysis of neural network models for the problem of speaker recognition", *Innovative Technologies and Scientific Solutions for Industries*, Vol. 24, No. 2, P. 172–178. DOI: 10.30837/ITSSI.2023.24.172
2. Barkovska, O., Kholiev, V., Lytvynenko, V. (2022), "Study of noise reduction methods in the sound sequence when solving the speech-to-text problem", *Advanced Information Systems*, Vol. 6, No. 1, P. 48–54. DOI: 10.20998/2522-9052.2022.1.08
3. Mykhailichenko, I., Ivashchenko, H., Barkovska O., Liashenko O. (2022), "Application of Deep Neural Network for Real-Time Voice Command Recognition", *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek)*, P. 1–4. DOI: 10.1109/KhPIWeek57572.2022.9916473
4. Kovalenko, A., Poroshenko, A. (2022), "ANALYSIS OF THE SOUND EVENT DETECTION METHODS AND SYSTEMS", *Advanced Information Systems*, Vol. 6, No. 1, P. 65–69. DOI: 10.20998/2522-9052.2022.1.11
5. Kholiev, V., Barkovska, O. (2023), "Analysis of the of training and test data distribution for audio series classification", *Інформаційно-керуючі системи на залізничному транспорті*, Vol. 28, No. 1, P. 38–43. DOI: 10.18664/iksz.v28i1.276343
6. Barkovska, O. (2022), "Performance study of the text analysis module in the proposed model of automatic speaker's speech annotation", *Computer systems and information technologies*, Vol. 4, P. 13–19. DOI: 10.31891/csit-2022-4-2
7. Poroshenko, A., Kovalenko, A., Sedlaček, P. (2022), "Organization of Audio Analytics Systems Topologies", *2022 IEEE 9th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*, P. 517–521, DOI: 10.1109/PICST57299.2022.10238687
8. Poroshenko, A., Kovalenko, A. (2023), "Optimization of a basic network in audio analytics systems", *Advanced Information Systems*, Vol. 7, No. 1, P. 23–28. DOI: 10.20998/2522-9052.2023.1.04
9. Kuchuk, N., Kovalenko, A., Kuchuk, H., Levashenko, V., Zaitseva, E. (2022), "Mathematical Methods of Reliability Analysis of the Network Structures: Securing QoS on Hyperconverged Networks for Traffic Anomalies". *Future Intent-Based Networking. Lecture Notes in Electrical Engineering*, Vol. 831. Springer, Cham. P.223-241. DOI: 10.1007/978-3-030-92435-5_13
10. Ghido, F., Tabus, I. (2013), "Sparse Modeling for Lossless Audio Compression", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 1, P. 14–28. DOI: 10.1109/TASL.2012.2211014
11. Huang, H., Shu, H., Yu, R. (2014), "Lossless audio compression in the new IEEE Standard for Advanced Audio Coding", *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, P. 6934–6938. DOI: 10.1109/ICASSP.2014.6854944
12. Zeghidour, N., Luebs, A., Omran, A., Skoglund, J., Tagliasacchi, M. (2022), "SoundStream: An End-to-End Neural Audio Codec". *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 30, P. 495–507. DOI: 10.1109/TASLP.2021.3129994
13. Petrosky, E. E., Michaels, A. J., Ridge, D. B. (2019), "Network Scalability Comparison of IEEE 802.15.4 and Receiver-Assigned CDMA", *IEEE Internet of Things Journal*. Vol. 6, No. 4, P. 6060–6069. DOI: 10.1109/JIOT.2018.2884455
14. Meng, X., Pappas, V., Zhang, L. (2010), "Improving the Scalability of Data Center Networks with Traffic-aware Virtual Machine Placement", *2010 Proceedings IEEE INFOCOM*, P. 1–9. DOI: 10.1109/INFCOM.2010.5461930
15. Amarudin, Ferdiana, R., Widyawan (2020), "A Systematic Literature Review of Intrusion Detection System for Network Security: Research Trends, Datasets and Methods", *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)*, P. 1–6. DOI: 10.1109/ICICoS51170.2020.9299068
16. Do, E. H., Gadepally, V. N. (2020), "Classifying Anomalies for Network Security", *ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, P. 2907–2911. DOI: 10.1109/ICASSP40776.2020.9053419
17. Szymanski, A., Lason, A., Rzasa J., Jajszczyk, A. (2007), "Grade-of-service-based routing in optical networks [Quality-of-Service-Based Routing Algorithms for Heterogeneous Networks]", *IEEE Communications Magazine*, Vol. 45, No. 2, P. 82–87. DOI: 10.1109/MCOM.2007.313400

18. Liu, C. H., Gkelias, A., Leung, K. K. (2008), "Connection admission control and grade of service for QoS routing in mesh networks", *2008 IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, P. 1–5. DOI: 10.1109/PIMRC.2008.4699895

Received 30.11.2023

Відомості про авторів / About the Authors

Порошенко Антон Ігорович – Харківський національний університет радіоелектроніки, аспірант кафедри електронних обчислювальних машин, Харків, Україна; e-mail: anton.poroshenko@nure.ua; ORCID ID: <https://orcid.org/my-orcid?orcid=0000-0001-7266-4269>

Коваленко Андрій Анатолійович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри електронних обчислювальних машин, Харків, Україна; e-mail: andriy.kovalenko@nure.ua; ORCID ID: <https://orcid.org/my-orcid?orcid=0000-0002-2817-9036>

Poroshenko Anton – Kharkiv National University of Radio Electronics, Postgraduate Student at the Department of Computing Machines, Kharkiv, Ukraine.

Kovalenko Andriy – Doctor of Technical Sciences, Phd (Computer engineering), Kharkiv National University of Radio Electronics, Professor at the Department of Computing Machines, Kharkiv, Ukraine.

МЕТОД ПЕРЕДАЧІ АУДІОСИГНАЛУ У МЕРЕЖНІЙ СИСТЕМІ АУДІОАНАЛІТИКИ

Предметом дослідження в статті є метод передачі аудіосигналу в мережній системі аудіоаналітики. Створення зазначеної системи приводить до появи нових класів джерел навантаження, що передають пакетизовані звукові дані. Тому без побудови достатньо адекватних математичних моделей не можна створити мережну систему аудіоаналітики, яка ефективно працює. Фундаментальним у теорії телетрафіку є питання моделей джерел навантаження. Отже, стає необхідним розроблення методу передачі аудіосигналу в мережній системі аудіоаналітики. З огляду на сказане **метою статті** є створення методу передачі аудіосигналу в мережній системі аудіоаналітики з метою забезпечення ефективності та точності аудіоаналітики. У статті вирішуються такі **завдання**: формування моделі джерел навантаження системи, дослідження управління з'єднаннями і трафіком у мережній системі аудіоаналітики, реалізація функцій управління та контролю трафіком у мережі (управління доступом користувача до мережі та управління параметрами використання), дослідження методів забезпечення якості передачі аудіосигналу та розроблення методу передачі аудіосигналу з допомогою комутації віртуальних маршрутів. Для досягнення поставлених завдань використовуються такі **методи**: математичне оброблення сигналів, алгоритми компресії даних, оптимізація мережних протоколів і використання швидких мережних з'єднань. Досягнуто таких **результатів**: змодельовано джерела навантаження системи, розглянуто управління з'єднаннями і трафіком у мережній системі аудіоаналітики, досліджено методи забезпечення якості передачі аудіосигналу та запропоновано метод передачі аудіосигналу з допомогою комутації віртуальних маршрутів. **Висновки**. Можливості використання імітаційного моделювання вузлів мережної системи аудіоаналітики дуже обмежені. Це пояснюється тим, що допустимий рівень втрат інформаційних осередків дуже низький. Упровадження розробленого методу передачі аудіосигналу в мережній системі аудіоаналітики дає змогу досягти ефективного контролю та оброблення звукової інформації в режимі реального часу. Цей метод може широко застосовуватися у сферах безпеки, медицини, систем управління та інших галузях, де аналіз аудіосигналів є важливим складником.

Ключові слова: аудіоаналітика; метод передачі аудіосигналу; управління трафіком; віртуальні маршрути; затримка.

Бібліографічні описи / Bibliographic descriptions

Порошенко А.І., Коваленко А.А. Метод передачі аудіосигналу у мережній системі аудіоаналітики. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 58–67. DOI: <https://doi.org/10.30837/ITSSI.2023.26.058>

Poroshenko, A., Kovalenko, A. (2023), "Audio signal transmission method in network-based audio analytics system", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 58–67. DOI: <https://doi.org/10.30837/ITSSI.2023.26.058>

Н. РОМАНКІВ, Д. СИТНІКОВ

АНАЛІЗ І ВИБІР МЕТОДІВ КАСТОМІЗАЦІЇ SAAS-РІШЕНЬ, ПОБУДОВАНИХ ЗА ДОПОМОГОЮ ТЕХНОЛОГІЙ CLOUD-NATIVE

Предметом дослідження є методи кастомізації SaaS-рішень. **Мета статті** – визначити цілісну стратегію кастомізації SaaS-рішень, розроблених за допомогою технологій *cloud-native*. **Завдання:** проаналізувати сучасні підходи щодо архітектури SaaS-застосунків; виявити основні методи кастомізації для сучасних SaaS-застосунків; дослідити та встановити спосіб кастомізації інтерфейсу SaaS-застосунків; на підставі проведеного дослідження визначити цілісну стратегію кастомізації SaaS-застосунків. Упроваджуються такі **методи:** аналіз і синтез – для вивчення технологій, що використовуються для побудови SaaS-застосунків; абстрагування та узагальнення – для визначення загальної архітектури SaaS-застосунку; синтез вебтехнологій – для вибору методів кастомізації SaaS-застосунків та побудови алгоритму вибору методу кастомізації. Досягнуті **результати:** досліджено сучасну архітектуру SaaS-застосунків; обрано методи кастомізації SaaS-застосунків, а також визначено цілісну стратегію щодо кастомізації SaaS-застосунків; запропоновано метод модифікації інтерфейсу SaaS-застосунків, що дає змогу виконувати кастомізацію вебінтерфейсу SaaS-застосунку незалежно від конкретного фреймворку *front-end*. **Висновки:** сучасна архітектура SaaS-застосунків є комплексною та основана на мікросервісній архітектурі, підході *multi-tenant*, хмарних технологіях і веббраузері; методи кастомізації SaaS-застосунків мають розроблятися незалежно від певних технологічних стеків, щоб задовільнити потреби більшості або всіх користувачів SaaS-застосунку в кастомізації; підхід *API-first* є фундаментальним у побудові кастомізованого SaaS, оскільки він є основою для створення будь-якої автоматизації чи пов'язаної бізнес-логіки; *webhooks* є найкращою опцією для реалізації підписки на події, що відбуваються в SaaS-застосунку, незалежно від технологічного стеку; визначена стратегія кастомізації SaaS-застосунків дає змогу робити кастомізацію незалежно від певного технологічного стеку й, окрім реалізації бізнес-логіки, також покриває модифікацію інтерфейсу.

Ключові слова: SaaS; кастомізація; web; webhooks, Iframe, API-first.

Вступ

Сучасні технології та інтернет дали змогу моделі SaaS [1] (програмне забезпечення як послуга) з'явитися та домінувати у світі. Чимало багатомільярдних компаній були побудовані на SaaS, зокрема *Netflix, Amazon, Facebook, Airbnb*. А згідно з 360 *Research Reports* компанія SaaS у глобальному ринку оцінювалася в 1 777 098 доларів 2021 р., а також продовжуватиме зростати й досягне 1 777 188 доларів до 2027 р. Побудова сучасного SaaS-застосунку зазнає чимало технічних проблем, які необхідно вирішити, особливо в сучасному світі з широким упровадженням доступних хмарних технологій і нових технологій *cloud-native* [2], що продовжують з'являтися.

Окрім того, дуже високої популярності набуває *MACH Alliance* [3] й використання архітектури *composable ecommerce* [4], що допомагає бізнесу будувати систему як конструктор, який збирається із різних SaaS-систем і в якому окремі системи відповідають за різні технічні можливості електронної комерції, таких як пошук, персоналізація, керування замовленнями тощо.

Хоча такий підхід дає змогу бізнесу бути гнучким, а також прискорює розроблення IT-системи, водночас він обмежує розробників системи, оскільки сторонні сервіси є *black box* і зазвичай позбавляють можливості модифікувати їх системи. Важливо, щоб *Enterprise*-системи мали доступ до налаштувань, але часто вимоги до налаштувань від клієнта виходять за межі того, що вендор програмного забезпечення може передбачити наперед. В епоху використання локальних систем клієнти робили глибокі налаштування поза передбаченням вендора, безпосередньо змінюючи його вихідний код, а потім будуючи та експлуатуючи вендор у власному дата-центрі. Коли програмне забезпечення перейшло до хмарного багатокористувацького SaaS, клієнтам уже не можна безпосередньо змінювати вихідний код вендора, оскільки той самий екземпляр коду використовується декількома клієнтами одночасно в процесі виконання. Отже, постає питання розроблення та проектування методів кастомізації SaaS-систем, які давали б змогу інженерам і користувачам систем модифікувати її під свої потреби, без впливу на інших користувачів SaaS-системи.

Аналіз останніх досліджень і публікацій

Аналіз джерел довів, що кастомізації *SaaS*-систем приділялась увага в багатьох наукових працях, і це також підтверджує актуальність окресленої проблеми. Так, Ху Сонг, Франк Чаувел та інші провели ґрунтовне дослідження щодо підтримки кастомізації *SaaS* [5]. Автори вивчають очікування консультантів, які спеціалізуються на кастомізації корпоративних програмних систем. Кастомізація традиційно ґрунтується на припущенні, що програмне забезпечення встановлено у власному дата-центрі клієнта, тому клієнт має повний контроль над системою та може вільно її змінювати. Це припущення вже не правильне, оскільки корпоративне програмне забезпечення переходить до багатокористувацького *Software as a Service*, бо воно працює в хмарі та контролюється вендором *SaaS*. Отже, дехто вважає, що налаштування *SaaS* є занадто складним і від нього необхідно відмовитися. Водночас практика в галузі *SaaS* визнає перевагу конфігурації над налаштуванням. Два вендори програмного забезпечення, які взяли участь у цьому дослідженні, також розпочали з підтримки конфігурації на своїх рішеннях *SaaS*. Однак виявляється, що чимало їхніх партнерів (користувачів *SaaS*-системи) мають інший погляд, і тому ці вендори постійно отримують запитання щодо підтримки налаштувань на своїх продуктах *SaaS*. Також у роботі [5] подається теорія, побудована на висновках, що кастомізація залишається важливим для *SaaS*, однак багатокористувацькість (*multi-tenancy*) змінює відповідальність зацікавлених сторін. Партнери (користувачі *SaaS*) тепер не можуть самостійно змінити продукт, щоб досягти побажань своїх клієнтів, але мають покладатися на вендорів *SaaS* для розроблення та впровадження змін і нових функцій системи. Зрештою партнерам необхідно відмовитися від старого методу вільного розроблення з довільними модифікаціями, розсіяними в продукті. Замість цього вони мають прийняти новий формат розроблення з більшою кількістю обмежень та участю вендорів і приділити значну увагу бізнесу. Ця трансформація конфліктує з традиційними інтересами партнерів, а саме: мінімізацією витрат для клієнтів, безшовною інтеграцією власного коду з основним продуктом і забезпеченням гнучкості. Тому критичним фактором успіху для екосистеми налаштувань є те, що вендори мають пропонувати

не лише висококастомізований *SaaS*, але й ефективні засоби підтримки, які будуть відповідати новим видам налаштувань і водночас компенсувати ці основні інтереси. Необхідність у кастомізації виникає через розрив між клієнтами та вендорами: останні зосереджуються на загальному корпоративному програмному забезпеченні, але не мають знань і ресурсів, щоб зрозуміти бізнес кожного окремого клієнта. Утім конфігурація не може замінити кастомізацію, оскільки вона ігнорує цей розрив, намагаючись зрозуміти, що вендори здатні передбачити та реалізувати всі ймовірні функції.

Також Ральф Мітзнер, Андреас Матзгер та інші в роботі [6], наводять такий аргумент: щоб використовувати економію масштабу, вендорам *SaaS* необхідно успішно привертати значну кількість користувачів для своїх *SaaS*-застосунків. Очевидно, що функціональність і якість, які індивідуальні клієнти очікують від програмного застосування, можуть відрізнитися. Як наслідок, постачальникам *SaaS* необхідно брати до уваги варіативні вимоги широкого спектра потенційних клієнтів. Це означає, що *SaaS*-застосунки мають дозволяти конфігурацію та налаштування для кожного користувача.

Окрім цього, Веі Сун та інші в роботі [7] дійшли висновку, що конфігурація та кастомізація є критичними аспектами для вендорів *SaaS* у процесі розроблення своїх систем. Хоча найважливіше – це добре розробити функції *SaaS*-застосунку, щоб задовольняти якнайбільше вимог у цільовому сегменті клієнтів і галузі застосування; рівень можливостей, наданих для конфігурації та кастомізації, дає ключові конкурентні переваги на ринку. У багатьох ситуаціях, якщо дуже важко розробити *SaaS*-застосунок як стандартну пропозицію для більшості клієнтів, висока здатність до конфігурації та кастомізації стане ключовим фактором успіху.

У студіях [8, 9] автори досліджували використання мікросервісів для кастомізації *SaaS*-систем. Такий метод кастомізації пропонується для застосування, коли необхідно зробити глибоку кастомізацію.

Підхід, згаданий у праці [8], занадто детально прив'язаний до стеку технологій *.Net* та розроблений на припущенні, що вендор *SaaS* надасть доступ до коду системи, який здебільшого є неможливим, оскільки для компаній, що продають *SaaS*, код програмної системи – це основний актив бізнесу, який приносить прибуток.

Водночас підхід, розроблений у роботі [9], пропонує сучасний *cloud-native* та динамічний підхід до використання мікросервісів. У такому підході кожна кастомізація реалізується за допомогою мікросервісу. Робота демонструє використання синхронного виклику, окремої бази даних *NoSQL* та середовища на основі *Docker* для реалізації такої архітектури. Але цей метод не покриває підхід до модифікації інтерфейсу *SaaS*-застосунку й побудований на особливому фреймворку *MiSC-Cloud*, що робить такий підхід майже непридатним до застосування.

Еспен Нордлі в роботі [10] розглядає можливості глибокої модифікації завдяки подіям. Архітектура застосунку також залежить від специфічного фреймворку *MiSC-Cloud* й побудована на припущенні, що вендор застосовуватиме специфічний *event bus* (шина подій) й надасть доступ до шини для мікросервісів, які мають виконувати кастомізацію способом підписки на події та їх модифікацію.

Визначення не вирішених раніше частин загальної проблеми. Мета роботи, завдання

Отже, аналіз літературних джерел довів, що в роботах, присвячених кастомізації *SaaS*-систем, описуються дуже специфічні методи модифікації, основані на конкретному наборі технологій, що є помилковим і не може масштабуватися на будь-який *SaaS*. Також немає цілісної стратегії щодо кастомізації *SaaS*-застосунків, яка передбачала б сучасні й різнобічні методи для модифікації. Окрему увагу необхідно приділити кастомізації інтерфейсу *SaaS*, оскільки це важливий аспект для модифікації *SaaS* під потреби окремих клієнтів компанії, який був упущений у згаданих роботах.

Мета статті – визначення цілісної стратегії кастомізації *SaaS*-систем. Для досягнення поставленої мети потрібно виконати такі **завдання**: проаналізувати сучасні підходи щодо архітектури *SaaS*-застосунків; визначити основні методи кастомізації для сучасних *SaaS*-застосунків; дослідити та обрати спосіб модифікації інтерфейсу *SaaS*-застосунків; на підставі проведеного дослідження визначити цілісну стратегію кастомізації *SaaS*-застосунків.

Матеріали й методи

Методологічною основою роботи є такі **методи** наукового дослідження: аналіз і синтез

застосовуються для вивчення технологій, що впроваджуються з метою побудови *SaaS*-застосунку; абстрагування та узагальнення – для визначення загальної архітектури *SaaS*-застосунку; синтез вебтехнологій – для вибору методів кастомізації *SaaS*-застосунків і побудови алгоритму вибору методу кастомізації.

Результати досліджень та їх обговорення

Сучасна архітектура *SaaS*-застосунків є складною та комплексною, що підтверджується статтями [11, 12], оскільки впровадження сучасних *SaaS*-застосунків супроводжується такими труднощами:

- забезпечення високого рівня безпеки для збереження конфіденційності та цілісності даних користувачів;
- забезпечення ефективного масштабування, щоб застосунок міг витримати зростання обсягу користувачів і бізнесу;
- реалізація зручних інтеграцій з іншими системами та послугами для покращення функціональності;
- розроблення механізмів відмовостійкості для забезпечення неперервності роботи застосунку в умовах можливих збоїв;
- ефективне керування витратами та оптимізація використання ресурсів для забезпечення ефективності бізнес-моделі;
- забезпечення швидкого розроблення та впровадження нового функціоналу для відповіді на мінливі потреби ринку.

Отже, розглянемо детально підходи, що застосовуються в сучасній архітектурі *SaaS*-застосунків.

Більшість успішних *SaaS*-застосунків, наприклад *Netflix*, *Twitter*, *Slack*, використовують мікросервісну архітектуру, що є достатньо ефективною. Гнучкість і масштабованість є ключовими перевагами мікросервісної архітектури в контексті *SaaS*. Кожен мікросервіс може функціонувати незалежно, що дає змогу розробникам ефективно впроваджувати новий функціонал та масштабувати окремі частини системи відповідно до потреб користувачів. Незалежність та ізоляція мікросервісів стає важливою особливістю для *SaaS*. Це дає змогу забезпечити високий рівень ізоляції між різними функціональними частинами системи, що важливо для забезпечення безпеки та конфіденційності даних користувачів. Можливість швидко розгорнути

та оновлювати окремі мікросервіси без впливу на інші компоненти системи є важливим аспектом для реалізації гнучких методологій розроблення та впровадження нового функціоналу в *SaaS*-застосунках. Однак попри всі переваги мікросервісна архітектура стикається з викликами. Складність управління багатьма мікросервісами та їх координація може виявитися непростим завданням, особливо в розробленні та експлуатації великих *SaaS*-застосунків. Важливо враховувати збільшені витрати на інфраструктуру через підтримку багатьох мікросервісів, особливо за умови великої кількості користувачів та операцій. Складність тестування та відлагодження може спричинити розподілений характер мікросервісів, що потребує ретельного планування та впровадження відповідних інструментів для забезпечення якості застосунку.

Окрім цього, сучасні *SaaS*-застосунки використовують архітектуру *multi-tenant* [13, 14]. Зазначена архітектура є ключовою для реалізації *SaaS*-застосунків, оскільки вона дає змогу ефективно обслуговувати багато користувачів на одній і тій самій інфраструктурі та програмному забезпеченні. У *multi-tenant* інфраструктурні ресурси використовуються спільно, але водночас дані користувача ізольовані. Такий підхід дозволяє економити ресурси, оскільки різні користувачі застосовують спільну інфраструктуру та обслуговуються однією версією програмного забезпечення. Для постачальників *SaaS* це означає, що вони можуть забезпечити більш ефективне використання ресурсів і зменшити загальні витрати на утримання системи та збільшити прибутки. Також, оскільки в такій архітектурі використовується один екземпляр програмного забезпечення, це полегшує управління та оновлення системи. Постачальники *SaaS* можуть ефективно впроваджувати нові версії та функціонал, не припиняючи роботу інших користувачів.

Під час створення методів модифікації необхідно брати до уваги те, що одна й та сама інфраструктура *SaaS*-застосунку використовується одночасно безліччю користувачів, а тому модифікації на рівні інфраструктури *SaaS*-застосунку для досягнення кастомізації є неможливими або занадто складними для реалізації. А отже, будь-яка модифікація має відбуватися поза рівнем інфраструктури *SaaS*-застосунку.

Також у розробленні *SaaS*-застосунків використовують технології *cloud-native*, оскільки впровадження хмарних сервісів (зокрема *AWS*, *Azure*,

Google Cloud) має чисельні переваги для розроблення та експлуатації *SaaS*-застосунків. Розглянемо деякі з них більш детально.

– Еластичність. Хмарні платформи дають змогу легко змінювати обсяг ресурсів залежно від навантаження. Це допомагає збільшувати чи зменшувати потужність обчислення та інші ресурси відповідно до потреб.

– Масштабованість. Хмарні сервіси дають змогу ефективно масштабувати застосунки горизонтально або вертикально, щоб забезпечити високу доступність і продуктивність.

– Гнучкість. Хмарні рішення допомагають легко адаптувати конфігурацію та параметри інфраструктури під потреби конкретного *SaaS*-застосунку.

– Автоматизація. Засоби автоматизації в хмарних сервісах спрощують розгортання, моніторинг, інтеграцію та керування застосунками, що полегшує розроблення та експлуатацію.

– Хмарні технології дають змогу користувачам отримати доступ до *SaaS*-застосунків із будь-якого місця світу, де є інтернет. Це робить робочий процес більш гнучким і мобільним.

– Відсутність потреби в інвестуванні у власну апаратуру та інфраструктуру допомагає знизити витрати на створення та утримання інфраструктури.

– Хмарні послуги дають змогу швидко впроваджувати нові версії застосунків і вносити зміни в реальному часі.

Варто приділити особливу увагу розподілу ринку хмарних сервісів між основними провайдерами зазначених послуг. Інформація про розподіл [15] подається в табл. 1. Аналізуючи показники з розподілу ринку між різними провайдерами, можемо дійти висновку, що методи кастомізації мають бути *cloud-agnostic*, тобто не прив'язуватися до певної публічної хмари. Або якщо й розробляти специфічний для публічної хмари метод модифікації, то це мають бути найбільші гравці на ринку, а саме *AWS*, *GCP*, *Azure*, щоб покрити найбільшу долю ринку із найменшими витратами на розроблення та підтримку таких методів.

Тепер, коли ми подали ключові елементи архітектури *back-end* *SaaS*-застосунків, необхідно також розглянути архітектуру *front-end*.

SaaS-модель стала можлива завдяки активному розвитку вебтехнологій, що дають змогу завантажити користувачам застосунок безпосередньо у веббраузер без необхідності встановлювати будь-що на свій персональний комп'ютер чи мобільний пристрій.

Отже, традиційний *SaaS*-застосунок оснований саме на вебтехнології для реалізації *front-end*, тобто клієнтської частини застосунку.

Таблиця 1. Розподіл ринку між провайдерами хмарних послуг

Провайдер хмарних послуг	Відсоток ринку станом на 2023 р.
AWS	32
Azure	22
GCP	11
Alibaba Cloud	4
IBM Cloud	3
SalesForce	3
Oracle	2
Tencent Cloud	2

Застосування браузера для доступу *SaaS* має численні переваги, що сприяють зручності та доступності для користувачів. Перелічимо деякі ключові причини.

– Користувачі можуть отримати доступ до *SaaS*-застосунків у будь-якому місці, де є інтернет, використовуючи будь-який пристрій з веббраузером. Це робить використання застосунків більш зручним і мобільним.

– Браузерний доступ до *SaaS* робить застосунки незалежними від конкретної операційної системи. Це означає, що користувачі можуть застосовувати *SaaS* як на комп'ютерах, так і на мобільних пристроях, незалежно від конкретного середовища.

– Користувачі мають миттєвий доступ до оновлень і нових функцій, оскільки оновлення відбуваються на рівні серверів *SaaS*, і не потрібно чекати на встановлення нових версій на локальних пристроях.

– Постачальники *SaaS* можуть здійснювати оновлення та покращення централізовано на своїх серверах, замість того, щоб вимагати в користувачів регулярно встановлювати нові версії застосунків.

– Використання браузера дає змогу уникнути потреби в установленні та оновленні спеціалізованого клієнтського програмного забезпечення. Користувачам просто потрібно відкрити браузер та увійти в обліковий запис *SaaS*.

JavaScript є основною мовою для розроблення клієнтської частини веб-, а отже, *SaaS*-застосунків. Вона дає змогу розробникам створювати динамічні та інтерактивні інтерфейси, що покращують користувацький досвід. Хоча всі або принаймні більшість вебзастосунків використовують *JavaScript*

для побудови клієнтської частини, у сучасній веброботі домінують фреймворки *React*, *Angular*, *Vue* тощо. Популярність фреймворків *front-end* подається в табл. 2 [16].

Таблиця 2. Розподіл ринку між фреймворками *front-end*

Провайдер хмарних послуг	Відсоток ринку станом на 2023 р.
React	37
GSAP	14
Vue.js	11
Styled-components	8
Backbone.js	7
Require.js	6
Emotion	6
AngularJS	4
Handlebars	4
Stimulus	4

Тобто під час розроблення методів модифікації інтерфейсу необхідно зважати, що користувачі *SaaS*-застосунків в ідеалі мають давати змогу використовувати будь-який фреймворк для кастомізації вебчастини, щоб задовільнити потребу більшості користувачів *SaaS*-застосунків, без необхідності наймати розробника із знанням однієї технології, тільки заради кастомізації застосунку.

Отже високорівнева архітектура *SaaS*-застосунку має такі ключові елементи (рис. 1):

- мікросервіси;
- архітектура *multi-tenant*;
- *cloud-native*;
- браузер як клієнтська частина.

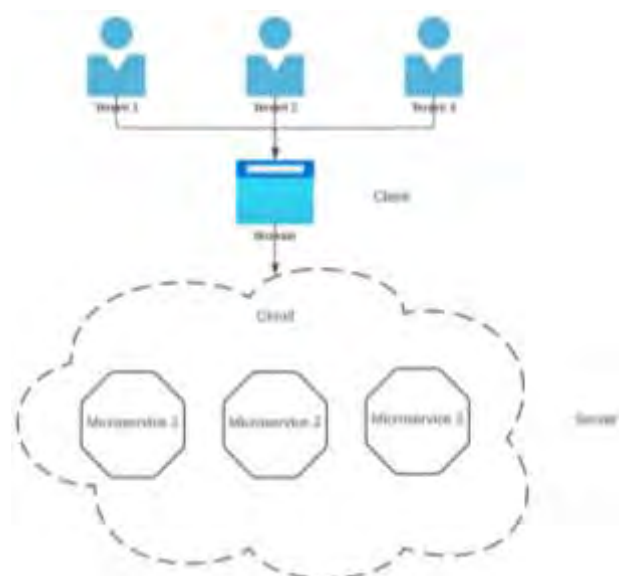


Рис. 1. Високорівнева архітектура *SaaS*-застосунку

Для вибору методів модифікації *SaaS*-застосунків і визначення цілісної стратегії кастомізації спершу визначимо потреби для модифікації *SaaS*-застосунків. Такі потреби можуть виникати з різноманітних причин, пов'язаних з особливими вимогами користувачів та бізнес-потребами. Перелічимо деякі основні потреби:

- кастомізація може бути необхідною для того, щоб адаптувати *SaaS*-застосунок до конкретних бізнес-процесів і потреб компанії. Це може передбачати налаштування робочих потоків, структури даних та автоматизації завдань;

- компанії можуть мати власні системи, інструменти чи сервіси, з якими *SaaS*-застосунок має інтегруватися;

- підприємства можуть потребувати оптимізації робочих процесів. Це передбачає кастомізацію застосунку для полегшення виконання завдань та підвищення продуктивності;

- користувачі можуть мати різні вподобання та потреби. Модифікація інтерфейсу, функціональності або налаштувань дає змогу користувачам адаптувати застосунок до своїх власних потреб і стандартів.

Фундаментальним підходом у побудові кастомізованого *SaaS* вважаємо використання підходу *API-first* у розробленні *SaaS*-застосунку. Підхід *API-first* передбачає, що дизайн і розроблення *API* (інтерфейсу програмування застосунків) визначається та впроваджується перед створенням інших частин застосунку. Цей підхід має декілька переваг, особливо для *SaaS*-продуктів, але в контексті кастомізації *SaaS*-застосунків для нас важливо, що зосередження на розробленні *API* сприяє зручній внутрішній інтеграції між різними частинами системи. Одночасно це також полегшує забезпечення зовнішньої інтеграції для інших сервісів і розробників.

Наявність *API* в *SaaS*-застосунку дає змогу його клієнтам використовувати *API* для кастомізації такими способами:

- користувачі можуть застосовувати *API* для інтеграції *SaaS*-застосунку зі своїми власними даними чи іншими системами;

- *API* може використовуватися для внесення власної бізнес-логіки чи автоматизації в робочі потоки *SaaS*-застосунку;

- за допомогою *API* можна створювати власні звіти та аналітичні інструменти, використовуючи дані, що надає *SaaS*-застосунок.

Для розроблення публічного *API* необхідно брати до уваги, що *API* має бути незалежним

від мови програмування чи технологічного стеку, яким послуговуються клієнти, щоб гарантувати можливість інтеграції з *API* для кожного користувача *SaaS*-застосунку. *REST API* [17] є найкращим вибором для реалізації публічного *API*, оскільки:

- *REST (Representational State Transfer)* – простий і легкий для розуміння стандарт. Він використовує стандартні *HTTP*-методи (*GET*, *POST*, *PUT*, *DELETE*) та оснований на стандарті ресурсів (ресурси, що можуть бути подані та змінені за допомогою *URI*);

- *REST* – стандарт для взаємодії в мережі Інтернет, його підтримка вбудована у більшість сучасних технологій та бібліотек. Це робить його легко інтегрованим і доступним для різних платформ і мов програмування;

- *REST API* природно інтегрується з архітектурою вебзастосунків. Користувачі можуть легко взаємодіяти з *REST API* з допомогою стандартних веббраузерів, що робить його доступним для великої аудиторії.

Окрім *API*, також необхідно вирішити проблему відстежування та реагування на події в *SaaS*-застосунку, оскільки це базова потреба в разі реалізації будь-якої бізнес-логіки в застосунку. Зважаючи, що використання інфраструктурних змін не можливе з причин, описаних вище, необхідно знайти підхід, який би дав змогу отримувати події із *SaaS*-системи незалежно від певної публічної хмари, брокера подій чи інфраструктури.

Проаналізувавши можливі опції, дійшли висновку, що найкращим варіантом для реалізації підписки на події в *SaaS*-застосунку є *webhook* [18] – механізм, який дає змогу одній системі (часто вебсерверу) автоматично надсилати дані або повідомлення іншій системі за умови виникнення якоїсь події. Коли в системі, що підтримує вебхуки, відбувається певна подія, вона надсилає *HTTP*-запит (зазвичай *POST*-запит) іншій системі за попередньо визначеним *URL*. Основні властивості вебхуків:

- допомагають системам взаємодіяти асинхронно. Замість того щоб постійно запитувати іншу систему про оновлення, система отримує повідомлення тільки тоді, коли відбувається певна подія;

- їх використовують для реагування на конкретні події в системі. Наприклад, вебхук може бути застосований для повідомлення про створення нового користувача, завершення оплати чи оновлення даних;

- спрощують інтеграцію між різними системами, оскільки вони передають дані безпосередньо

від одного сервера до іншого в разі виникнення певної події;

- дають змогу отримувати інформацію майже в реальному часі. Це особливо важливо для систем, де потрібно негайно реагувати на події;

- допомагають реалізувати будь-якою мовою програмування чи на стеку технологій як для сервера, так і клієнта.

Реалізація надійного й масштабованого механізму вебхуків є питанням, що необхідно детально розглянути окремо. Зазначена реалізація має відповідати таким вимогам:

- процес підписання на вебхуки має бути доступним з допомогою *API* або *UI*. Також має бути можливість конфігурувати кожну окрему підписку на те, які події ми хочемо отримати;

- доставка подій до вебхука має бути надійною, і необхідна стратегія для повторення спроб доставки, якщо сервіс вебхука був тимчасово не досяжний;

- інтерфейс *SaaS*-застосунку має давати змогу переглядати історію відправлень подій до вебхуків.

Як вже зазначалося вище, для кастомізації інтерфейсу необхідно визначити такий метод, який би дав змогу використовувати будь-який фреймворк для кастомізації. Серед наявних можливостей веббраузера є лише один інструмент, що задовольняє такі потреби. Це *Inline Frame*, або просто *iframe*, – HTML-тег, який допомагає вбудовувати інший документ *HTML* у поточний документ. Цей тег є частиною веброзмітки й використовується для створення вкладених вікон (фреймів), що можуть мати вміст іншого ресурсу або сторінки.

Тобто для реалізації кастомізації інтерфейсу пропонується застосовувати вбудовані вебсторінки. Отже, користувач, який хоче кастомізувати інтерфейс застосунку, може зробити це, застосовуючи будь-який фреймворк. Для цього необхідно буде облаштувати хостинг вебсторінки, який підключатиметься в *iframe*.

Цей підхід можна поєднувати із підходом *API-first*, оскільки вебсторінка, завантажена в *iframe*, має змогу використовувати публічне *API SaaS*-застосунку для виконання бізнес-функцій тощо. Але щоб такий підхід працював, вебсторінка, завантажена в *iframe*, має авторизуватися із *SaaS*-вебзастосунком та отримати необхідну інформацію про конфігурацію тощо. Для вирішення цієї проблеми пропонується використовувати стандартний *API* веббраузера, що уможливило спілкування основної вебсторінки з вбудованою. Вебсторінка, завантажена в *iframe*, здатна прослуховувати повідомлення,

відправлені основною сторінкою, використовуючи підписку на події в об'єкта *window*, і надсилати повідомлення основній сторінці, застосовуючи метод *postMessage*. Водночас аналогічні дії може виконувати основна сторінка щодо вбудованої в *iframe*. Так, під час завантаження сторінки можливе виконання ініціалізації та передачі конфігурації та необхідної авторизаційної інформації до вбудованої сторінки.

Отже, високорівнева архітектура, що зображує всі три розглянуті методи кастомізації *SaaS*-застосунків, подана на рис. 2. Високорівнева архітектура відображає такі елементи:

- *Client, Static Hosting for iframe* – хостинг, що містить статичні файли, необхідні для роботи вебсторінки, яку завантажують в *iframe* для кастомізації інтерфейсу;

- *Client Web Server* – сервіс *back-end*, що містить бізнес-логіку, яку хочуть реалізувати клієнти *SaaS*-застосунку. Цей вебсервер отримує вебхук повідомлення від *SaaS Webhook Delivery Servers* та реагує на події. Водночас цей вебсервер використовує *SaaS API* для реалізації бізнес-логіки;

- *SaaS Webhook Delivery Servers* – сервери *back-end SaaS*-застосунку, відповідальні за доставку повідомлень до вебхуків;

- *SaaS API* – сервери *back-end SaaS*-застосунку, відповідальні за реалізацію публічного *API SaaS*-застосунку.

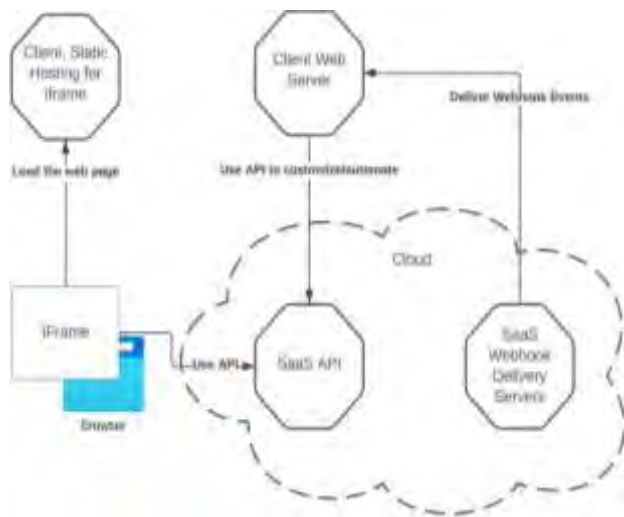


Рис. 2. Методи кастомізації *SaaS*-застосунків

Також для прийняття рішення щодо того, коли та який саме метод кастомізації з трьох обраних необхідно використовувати, було розроблено спеціальний алгоритм (рис. 3).

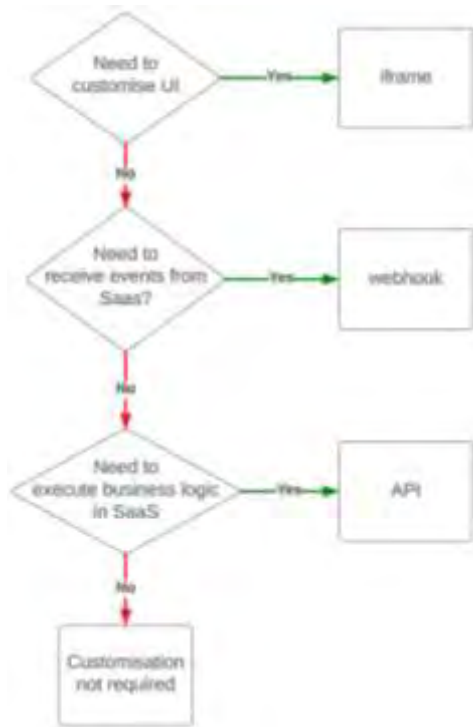


Рис. 3. Алгоритм вибору методів для кастомізації SaaS-застосунку

Висновки й перспективи подальшого розвитку

Під час дослідження розроблено цілісну концепцію кастомізації SaaS-застосунків. За результатами роботи можна зробити такі висновки:

- аналіз джерел довів, що проблема кастомізації SaaS-застосунків є актуальною;
- сучасна архітектура SaaS-застосунків є комплексною та будується на мікросервісній

архітектурі, підході *multi-tenant*, використанні хмарних технологій та веббраузера;

- методи кастомізації SaaS-застосунків мають розроблятися незалежно від певного технологічного стеку, щоб задовільнити потреби більшості чи всіх користувачів SaaS-застосунку в кастомізації;

- підхід *API-first* є фундаментальним у побудові кастомізованого SaaS, оскільки він фундаментальний для розроблення будь-якої автоматизації чи бізнес-логіки;

- вебхуки є найкращою опцією для реалізації підписки на події, що відбуваються в SaaS-застосунку незалежно від технологічного стеку;

- *Iframe* є найкращою опцією для реалізації кастомізації інтерфейсу вебзастосунку, оскільки уможливує використання будь-якого фреймворку для побудови інтерфейсу.

Для демонстрації всіх трьох опцій розроблено високорівневу діаграму, що зображує використання трьох методів кастомізації, а також створено алгоритм вибору методу кастомізації. Зазначена стратегія кастомізації SaaS-застосунків дає змогу виконувати кастомізацію незалежно від певного технологічного стеку й, окрім реалізації бізнес-логіки, також покриває модифікацію інтерфейсу завдяки використанню *iframe*.

Для подальшого розвитку поставленого завдання необхідно розглянути підходи, які б уможлилювали кастомізацію поведінки API SaaS-застосунку без зміни коду чи інфраструктури SaaS-застосунку, а також спроектувати надійний і масштабований механізм вебхуків.

Список літератури

1. Michael J. Kavis Architecting the Cloud Design Decisions for Cloud Computing Service Models (SaaS, PaaS, and IaaS): Wiley, 2014. 224 p. URL: <https://www.wiley.com/en-us/Architecting+the+Cloud%3A+Design+Decisions+for+Cloud+Computing+Service+Models+%28SaaS%2C+PaaS%2C+and+IaaS%29-p-9781118617618> (дата звернення: 27.11.2023)
2. Tomas Erl, Ricardo Puttini, Zaigham Mahmood. Cloud Computing, Concepts, Technology & Architecture: Pearson, 2013. 747 p. URL: <https://www.redalyc.org/pdf/6380/638067279007.pdf> (дата звернення: 27.11.2023)
3. MACH Allience. URL: https://en.wikipedia.org/wiki/MACH_Alliance (дата звернення: 27.11.2023)
4. Composable Commerce. URL: <https://www.elasticpath.com/composable-commerce> (дата звернення: 27.11.2023)
5. Hui Song, Franck Chauvel, Arnor Solberg, Bent Foyn, Tony Yates. How to support customization on SaaS: A Grounded Theory from Customisation Consultants. 2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C). 2017. P. 247–249. DOI: <https://doi.org/10.1109/ICSE-C.2017.136>
6. Ralph Mietzner, Andreas Metzger, Frank Leymann, Klaus Pohl. Variability modeling to support customization and deployment of multi-tenant-aware Software as a Service applications. 2009 ICSE Workshop on Principles of Engineering Service Oriented Systems. 2009. 814 p. DOI: <https://doi.org/10.1109/PESOS.2009.5068815>
7. Wei Sun, Xin Zhang, Chang Jie Guo, Pei Sun, Huis Su. Software as a Service: Configuration and Customization Perspectives. 2008 IEEE Congress on Services Part II (services-2 2008). 2008. 2946 p. DOI: <https://doi.org/10.1109/SERVICES-2.2008.29>

8. Franck Chauvel, Arnor Solberg. Using Intrusive Microservices to Enable Deep Customization of Multi-tenant SaaS. *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. 2018. 225 p. DOI: <https://doi.org/10.1109/QUATIC.2018.00015>
9. Hui Song, Phu H. Nguyen, Frank Chauvel, Jens Glattetre, Thomas Schjerpen Customizing Multi-Tenant SaaS by Microservices: A Reference Architecture". *2019 IEEE International Conference on Web Services (ICWS)*. 2019. 710 p. DOI: <https://doi.org/10.1109/ICWS.2019.00081>
10. Espen Tonnessen Nordli, Phu H. Nguyen, Franck Chauvel, Hui Song. Event-Based Customization of Multi-tenant SaaS Using Microservices. *2020 International Conference on Coordination Languages and Models*. 2020. P. 171–180. DOI: https://doi.org/10.1007/978-3-030-50029-0_11
11. What is SaaS Architecture? 10 Best Practices For Efficient Design. URL: <https://www.cloudzero.com/blog/saas-architecture/> (дата звернення: 29.11.2023)
12. SaaS application architecture best practices. URL: <https://acropolium.com/blog/saas-app-architecture-2022-overview/> (дата звернення: 29.11.2023)
13. Eng Lieh Ouh, Benjamin Kok Siew Gan. An Exploratory Study of Architectural Style and Effort Estimation for Multi-Tenant Microservices-Based Software as a Service (SaaS). *2023 IEEE 20th International Conference on Software Architecture Companion (ICSA-C)*. 2023. 103 p. DOI: <https://doi.org/10.1109/ICSA-C57050.2023.00043>
14. Wei-Tek Tsai, Peide Zhong. Multi-tenancy and Sub-tenancy Architecture in Software-as-a-Service (SaaS). *2014 IEEE 8th International Symposium on Service Oriented System Engineering*. 2014. 519 p. DOI: <https://doi.org/10.1109/SOSE.2014.20>
15. Worldwide market share of leading cloud infrastructure service providers. URL: <https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/> (дата звернення 29.11.2023)
16. JavaScript frameworks technologies market share. URL: <https://www.wappalyzer.com/technologies/javascript-frameworks/> (дата звернення 29.11.2023)
17. Leonard Richardson, Mike Amundsen, Sam Ruby. RESTful Web APIs: O'Reilly Media, Inc. 2013, 406 p.
18. Webhook. URL: <https://uk.wikipedia.org/wiki/Webhook> (дата звернення 29.11.2023)

References

1. Michael, J. "Architecting the Cloud Design Decisions for Cloud Computing Service Models" (SaaS, PaaS, and IaaS): Wiley, 2014. 224 p. available at: <https://www.wiley.com/en-us/Architecting+the+Cloud%3A+Design+Decisions+for+Cloud+Computing+Service+Models+%28SaaS%2C+PaaS%2C+and+IaaS%29-p-9781118617618> (last accessed: 27.11.2023)
2. Tomas, Erl. R., Puttini, Z. "Cloud Computing, Concepts, Technology & Architecture: Pearson", 2013. 747 p. available at: <https://www.redalyc.org/pdf/6380/638067279007.pdf> (last accessed: 27.11.2023)
3. "MACH Alliance", available at: https://en.wikipedia.org/wiki/MACH_Alliance (last accessed: 27.11.2023)
4. "Composable Commerce", available at: <https://www.elasticpath.com/composable-commerce> (last accessed: 27.11.2023)
5. Hui, Song, Franck, C., Arnor, S., Bent, F., Tony, Y. (2017), "How to support customization on SaaS: A Grounded Theory from Customisation Consultants". *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. P. 247–249. DOI: <https://doi.org/10.1109/ICSE-C.2017.136>
6. Ralph, M., Andreas, M, Frank, L., Klaus, P. (2009), "Variability modeling to support customization and deployment of multi-tenant-aware Software as a Service applications". *2009 ICSE Workshop on Principles of Engineering Service Oriented Systems*. 814 p. DOI: <https://doi.org/10.1109/PESOS.2009.5068815>
7. Wei, Sun; Xin, Zhang, Chang, Jie Guo, Pei, Sun, Huis, Su. (2008), "Software as a Service: Configuration and Customization Perspectives". *2008 IEEE Congress on Services Part II (services-2 2008)*. 2946 p. DOI: <https://doi.org/10.1109/SERVICES-2.2008.29>
8. Franck, C., Arnor, S. (2018), "Using Intrusive Microservices to Enable Deep Customization of Multi-tenant SaaS". *2018 11th International Conference on the Quality of Information and Communications Technology (QUATIC)*. 225 p. DOI: <https://doi.org/10.1109/QUATIC.2018.00015>
9. Hui, Song, Phu, H. Nguyen, Frank, C., Jens, G., Thomas, S. (2019), "Customizing Multi-Tenant SaaS by Microservices: A Reference Architecture". *2019 IEEE International Conference on Web Services (ICWS)*. 710 p. DOI: <https://doi.org/10.1109/ICWS.2019.00081>
10. Espen, T., Phu, H. Nguyen, Franck, C., Hui, Song (2020), "Event-Based Customization of Multi-tenant SaaS Using Microservices". *2020 International Conference on Coordination Languages and Models*. P. 171–180. DOI: https://doi.org/10.1007/978-3-030-50029-0_11
11. "What is SaaS Architecture? 10 Best Practices For Efficient Design", available at: <https://www.cloudzero.com/blog/saas-architecture/> (last accessed: 29.11.2023)
12. "SaaS application architecture best practices", available at: <https://acropolium.com/blog/saas-app-architecture-2022-overview/> (last accessed: 29.11.2023)

13. Eng, L., Benjamin, K. (2023), "An Exploratory Study of Architectural Style and Effort Estimation for Multi-Tenant Microservices-Based Software as a Service (SaaS)". *2023 IEEE 20th International Conference on Software Architecture Companion (ICSA-C)*. 103 p. DOI: <https://doi.org/10.1109/ICSA-C57050.2023.00043>
14. Wei-Tek, T., Peide, Z. (2014), "Multi-tenancy and Sub-tenancy Architecture in Software-as-a-Service (SaaS)". *2014 IEEE 8th International Symposium on Service Oriented System Engineering*. 519 p. DOI: <https://doi.org/10.1109/SOSE.2014.20>
15. "Worldwide market share of leading cloud infrastructure service providers", available at: <https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/> (last accessed: 29.11.2023)
16. "JavaScript frameworks technologies market share", available at: <https://www.wappalyzer.com/technologies/javascript-frameworks/> (last accessed: 29.11.2023)
17. Leonard Richardson, Mike Amundsen, Sam Ruby. *RESTful Web APIs*: O'Reilly Media, Inc. 2013, 406 p.
18. "Webhook", available at: <https://uk.wikipedia.org/wiki/Webhook> (last accessed: 29.11.2023)

Надійшла 30.11.2023

Відомості про авторів / About the Authors

Романків Назарій Дмитрович – Харківський національний університет радіоелектроніки, студент кафедри системотехніки, Харків, Україна; e-mail: ayzrian@gmail.com; ORCID ID: <https://orcid.org/0009-0004-9893-6823>

Ситніков Дмитро Едуардович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри системотехніки, Харків, Україна; e-mail: dsytnikov@googlemail.com; ORCID ID: <https://orcid.org/0000-0003-1240-7900>

Romankiv Nazarii – Kharkiv National University of Radio Electronics, Student at the Department of System Engineering, Kharkiv, Ukraine.

Sytnikov Dmytro – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Professor at the Department of System Engineering, Kharkiv, Ukraine.

ANALYSIS AND SELECTION OF METHODS FOR CUSTOMIZING SaaS SOLUTIONS BUILT USING CLOUD-NATIVE TECHNOLOGIES

The subject of the study is the methods of customization of *SaaS* solutions. **The purpose of the article** is to determine a holistic strategy for customizing *SaaS* solutions developed using *cloud-native* technologies. **Objectives**: to analyze modern approaches to the architecture of *SaaS* applications; to identify the main methods of customization for modern *SaaS* applications; to investigate and establish a method for customizing the interface of *SaaS* applications; based on the study, to determine a comprehensive strategy for customizing *SaaS* applications. The following **methods** are implemented: analysis and synthesis – to study the technologies that are used to build *SaaS* applications; abstraction and generalization – to determine the overall architecture of the *SaaS* application; synthesis of web technologies – to select methods of customization of *SaaS* applications and build an algorithm for selecting a customization method. **Results** achieved: the modern architecture of *SaaS* applications is studied; methods of customization of *SaaS* applications are selected, and a holistic strategy for customization of *SaaS* applications is defined; a method for modifying the interface of *SaaS* applications is determined, which allows customization of the web interface of a *SaaS* application regardless of the specific *front-end* framework. **Conclusions**: modern architecture of *SaaS* applications is complex and based on microservice architecture, *multi-tenant* approach, cloud technologies and web browser; methods of customization of *SaaS* applications should be developed independently of certain technology stacks to meet the customization needs of most or all users of *SaaS* applications; the *API-first* approach is fundamental in building a customized *SaaS*, as it is the basis for creating any automation or related business logic; *webhooks* are the best option for implementing subscriptions to events occurring in a *SaaS* application, regardless of the technology stack; a defined strategy for customizing *SaaS* applications allows customization regardless of a particular technology stack and, in addition to implementing business logic, also covers interface modifications.

Keywords: *SaaS*; customization; *web*; *webhooks*, *Iframe*, *API-first*.

Бібліографічні описи / Bibliographic descriptions

Романків Н. Д., Ситніков Д. Е. Аналіз і вибір методів кастомізації *SaaS*-рішень, побудованих за допомогою технологій *cloud-native*. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 68–77. DOI: <https://doi.org/10.30837/ITSSI.2023.26.068>

Romankiv, N., Sytnikov, D. (2023), "Analysis and selection of methods for customizing *SaaS* solutions built using *Cloud-Native* technologies", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 68–77. DOI: <https://doi.org/10.30837/ITSSI.2023.26.068>

V. FEDORCHENKO, O. SHMATKO, I. MYKHAILENKO, V. TRETIAK, O. KOLOMIITSEV

INTEGRATING ANALYTICAL STATISTICAL MODELS, SEQUENTIAL PATTERN MINING, AND FUZZY SET THEORY FOR ADVANCED MOBILE APP RELIABILITY ASSESSMENT

The study presents a new method for evaluating the reliability of mobile applications using the Corcoran model. This model includes several aspects of reliability, including performance, reliability, availability, scalability, security, usability, and testability. The Corcoran model can be applied to evaluate mobile applications by analysing key reliability metrics. Using the model significantly improves the reliability assessment of applications compared to traditional methods, which are primarily focused on desktop and server configurations. The aim of the study is to offer a more optimised approach to evaluating the reliability of mobile applications. The paper examines the problems faced by mobile app developers. This study represents a new application of the Corcoran model in evaluating the reliability of mobile applications. This model is characterised by an emphasis on the use of quantitative statistics and the ability to provide an accurate estimate of the probability of failure without any inaccuracies, which distinguishes this model from other software reliability models. The paper suggests using a combination of analytical statistical models, data extraction methods such as sequential pattern analysis, and fuzzy set theory to implement the Corcoran model. The application of the methodology is demonstrated by studying software error reports and conducting a comprehensive statistical analysis of them. To improve the results of future research, the paper suggests making more extensive use of the Corcoran model in various mobile applications and environments. It is recommended to change the model to take into account the constantly changing characteristics of mobile applications and their increasing complexity. In addition, it is advisable to conduct additional research to improve the data mining methods used in the model and explore the possibility of integrating artificial intelligence for more advanced software reliability analysis. Applying the Corcoran model to the mobile app development process to evaluate reliability can significantly improve the quality of applications, leading to increased customer satisfaction and trust in mobile apps. This model can serve as a guide for developers and companies to evaluate and improve their applications, driving innovation and continuous improvement in the competitive mobile app sector.

Keywords: mobile application; software development; reliability assessment; the Corcoran model.

Introduction

The term "software reliability" refers to the degree to which the procedures of the software development life cycle (SDLC) can be managed and controlled to produce reliable programs. This metric will be used until the conditions for completing the testing procedure are met. In addition, software reliability helps to maintain and predict the correctness of the software [1]. Software reliability engineering was developed to evaluate and quantify software quality. It demonstrates the fault-free operation of a program [2, 3]. Software reliability models are constantly being improved by both researchers and practitioners.

The following factors make it difficult to assess and predict the stability of a mobile application. First, mobile environments differ from traditional desktop computers and servers. Secondly, new forms of deficiencies are generated by the introduction of functionality and characteristics specific to the mobile context, such as energy, network, incompatibility, modified and restricted graphical user interface (GUI), interruptions, and notifications [4]. Third, there is a wide variety of mobile

platforms and hardware capabilities. Fourth, due to consumer demand, the development of mobile applications has accelerated, and the functionality of mobile applications has become more complex [5]. And, of course, mobile devices break when an app is published. Software engineers rely primarily on problem reports submitted by users in addition to testing.

Researchers should spend more time analyzing software stability to determine its value for mobile apps. More accurate results and analyses can be obtained if software reliability testing takes into account the specifics of mobile applications.

Software engineers, enterprises, and research institutions are interested in being able to predict failures in mobile applications. Thus, we propose to evaluate the reliability of mobile applications based on bug reports and generate more accurate results.

Literature review

We identified several studies and systematic literature reviews (SLRs) related to software reliability [4, 5–8].

Several studies and literature reviews were found [9, 10–13] that focused on software dependency.

However, none of these studies specifically addressed the current state of mobile application reliability; rather, they all focused on traditional software. In order to determine what is the most up-to-date research in software reliability, Singhal [14] conducted a SLR analysis that included materials released before 2011. Ten years ago, when the widespread use of mobile applications was just beginning, this was the case. The study categorized 141 publications based on the research question, methodology (e.g., survey, theory), and environment (e.g., academic, industrial). Since the information available at the time was not sufficient to prove industrial validity, the study recommended additional industrial research. Due to the lack of standardized usage of words related to software reliability, the authors emphasized the importance of manual searching to find relevant material on the topic.

In their analysis of the literature from 1990 to 2010, Shahrokhni and Feldt [15] focused on software resilience, which is described as a reliability characteristic in various standards such as IEEE-STD 610.12-1990. For this study, the authors analyzed and classified 144 articles according to the following criteria: type of study (e.g., evaluation, experience report), contribution (e.g., tool, metrics), type of evaluation (e.g., academic, industrial), and development stage (e.g., requirements, design, and implementation). The lack of research on identifying and defining software sustainability requirements was the largest gap identified in the study.

Most studies have mainly focused on one component of reliability (invalid inputs), while others, such as unexpected events, timeouts, interruptions, and stressful execution settings, have been completely ignored. Febrero et al. [16] analyzed 503 articles from 2003 to 2014 as part of their modeling of SMS software reliability. Static and architectural reliability models, as well as software reliability growth models, Bayesian approaches, test-based methods, AI-based methods, and other types of reliability models were divided into five groups.

Finding that many studies do not meet the established quality requirements, the study identified a knowledge gap. To fill this gap, the same authors conducted a systematic literature review (SLR) on software reliability assessment using the ISO/IEC 25000 SQuaRE quality standard for 1991–2014.

According to the results of the latter study, insufficient attention has been paid to adjusting quality

and reliability standards to take into account the interests of multiple stakeholders.

They also noted that the complexity of existing reliability models does not allow them to be used in routine situations. Lack of agreement and different definitions of reliability have also hindered the development of useful models. The authors noted, for example, that "reliability" and "fault tolerance" are often used synonymously, despite their differences.

They were more focused on how reliability models apply reliability requirements (e.g., ISO/IEC-25000 SQuaRE), whereas our work explores the current state of reliability in mobile applications. In addition, we review research conducted over the past six years or so.

Alhazaa and Andrews [17] performed a state-of-the-art SMS in which they examined reliability growth patterns that take into account the development of software systems. They summarized the trends in terms of year of publication, location, and nature of the study (academic, industrial). The studies were categorized based on the proposed approach (analytical and curve fitting) and research style (empirical or non-empirical), as well as the scale of the solution (type and number of changes: one change point, multiple change points). In addition, they used the criteria of Ali et al. [18] to assess the reliability of empirical studies. They suggested that researchers look for higher quality empirical studies with closer cooperation with industry. In addition, these authors recommended further research on the following questions: how far into the future can these models look? When do professionals need to change the models or adjust their settings? All of these previous studies (including Alhazaa and Andrews) agree that the solutions lacked industry validation because they were mostly studied in an academic context without involving or collaborating with practitioners throughout the study.

The proposed model

Thus, in order to successfully fulfill one of the main tasks of this work – creating an integrated model for assessing software reliability – we need to develop an idea of which model of software reliability analysis is most suitable for our project and how the statistical data for this model will be obtained.

Of all the software reliability assessment models considered, the Corcoran model was chosen as the most suitable for use in this work. There are several reasons for this, but the most important is the absence of the need for additional work (e.g., introduction of artificial errors)

and the focus of this model on the use of quantitative statistical data about the project.

Corcoran's model is an example of an analytical statistical model of software reliability because it does not use test time parameters and only takes into account the result of N tests in which N_i errors of the i -th type are detected. The model uses variable failure probabilities for different types of errors.

Unlike other methods of this type, the Corcoran model estimates the probability of fault-free program execution at the time of evaluation [19].

- The model requires knowledge of the following indicators;
 - The model contains non-static failure probabilities for different sources of errors and, accordingly, different probabilities of their correction;
 - The model uses only such parameters as the result of N tests in which N_i errors of the i -th type are observed;
 - Detection of errors of the i -th type during N tests occurs with probability a_i .

The reliability level indicator R is calculated by the formula:

$$R = \frac{N_0}{N} + \sum_{i=1}^K \frac{Y_i \times (N_i - 1)}{N}, \quad (1)$$

where N_0 – is the number of failed (or unsuccessful) tests performed in a series of N tests;

K – known number of error types;

Y_i – probability of errors

$$Y_i = \begin{cases} a_i, & \text{if } N_i > 0, \\ 0, & \text{if } N_i = 0; \end{cases} \quad (2)$$

a_i – probability of detecting errors of the i -th type during testing.

In this model, the probability of occurrence of a certain event is estimated based on a priori information or statistics for the previous period of software operation.

The number of tests N_i for the Corcoran formula for an incomplete set of test reports is defined as:

$$N_i = \frac{R_i \times N_i \times 0.6}{R_i}, \quad (3)$$

where R_i – number of reports imported to the system;

$$LCSuff(S_{1..p}, T_{1..q}) = \begin{cases} LCSuff(S_{1..p-1}, T_{1..q-1}) + 1, & \text{if } S[p] = T[q] \\ 0, & \text{otherwise,} \end{cases}$$

where line S of length p ;

line T of length q ;

R_i – total number of reports on the Socorro server;

N_i – total number of product installations.

The algorithm of sequential pattern mining was chosen as a method of data mining. This choice was made under the influence of the availability of a large number of algorithms for solving similar problems, the ease of understanding the principles of these algorithms, and the high adaptability of this method to the required tasks. Sequential pattern mining is a data mining activity aimed at finding statistically significant patterns between data in which values are presented sequentially. As a rule, the values are considered discrete, which distinguishes this activity from data extraction from a time series. Sequential pattern mining is a special case of structured data mining [37]. In this paper, we will use algorithms to find the longest unified sequence. In computer science, the problem of finding the longest common sequence is to find the longest sequence (substring) or substrings that are common to two or more strings. For example, the longest common sequence of the strings "ABABC", "BABCA", and "ABCBA" is the string "ABC", which is three letters long. Other common sequences include "a", "AB", "B", "BA", and "C".

The problem of finding these sequences can be formulated as follows: given two lines S of length p and T of length q , you need to find the longest line that is common to S and T . Another interpretation of this problem is the problem of generalizing k -common sublines: given a set of lines $S = \{S_1, \dots, S_k\}$, where $|S_i| = n_i$, $|S_j| = n_j$, for every $2 \leq k \leq k$, you need to find the longest lines that occur inside at least K lines.

In this paper, we will continue to consider and use only the dynamic programming approach, since the length of lines in the subject area of this paper usually does not exceed 20 elements, but the simplicity of implementation and the visibility of the dynamic programming algorithm are much higher. To solve this problem using dynamic programming, you must first find the longest common suffix for all pairs of prefixes in the lines. The longest common suffix is calculated by the following formula:

$LCSuff(S_{1..p}, T_{1..q})$ – is the longest line that is common to S and T .

For example, for the strings "ABAB" and "BABA", the result of this algorithm is the following table 1.

For example, for the strings "ABAB" and "BABA", the result of this algorithm is the following table:

Table 1. Example of using the dynamic programming algorithm

		A	B	A	B
	0	0	0	0	0
B	0	0	<u>1</u>	0	1
A	0	<u>1</u>	0	<u>2</u>	0
B	0	0	<u>2</u>	0	<u>3</u>
A	0	1	0	<u>3</u>	0

The maximum of these common longest suffixes for possible prefixes must be the longest common sequence (subline) of lines S and T . These suffixes are underlined on the diagonals of Table 1. For this example, the longest common sequences are "BAB" and "ABA".

$$LCSubstr(S,T) = \max_{1 \leq i \leq m, 1 \leq j \leq n} LCStuff(S_{1..i}, T_{1..j}),$$

where $LCSubstr(S,T)$ – is the largest common sequence (subline) of lines S and T .

This formula can be extended for the case of comparing more than two lines by adding additional dimensions to the table, but this is not necessary in our case. To better establish patterns and relationships between error reports, it is also necessary to consider algorithms for measuring edit distance. The reason for

$$d_{ij} = \begin{cases} d_{i-1,j-1} & \text{for } a_j = b_i \\ \min \begin{cases} d_{i-1,j} + w_{del}(b_i) \\ d_{i,j-1} + w_{ins}(a_i) \\ d_{i-1,j-1} + w_{sub}(a_j, b_i) \end{cases} & \text{for } a_j \neq b_i \end{cases} \quad \text{for } 1 \leq i \leq m, 1 \leq j \leq n,$$

simple recursive method of calculating these formulas takes exponentially long. Therefore, as a rule, the calculations are performed using the Wagner and Fisher dynamic programming algorithm. After completing the Wagner–Fisher algorithm, the minimal sequence of editing operations can be read as the Return Path of the operations (starting with dmn) used during the dynamic programming algorithm.

Example from the practice

Below is an example of using this method to assess software reliability using the proposed approach.

The input data are data about 100 tests of the program. Out of 100 tests, 20 were successful

this is the discrepancy between data in related reports and the need to reduce noise between data samples.

In computer science, edit distance is a way to quantify the similarity of two strings (e.g., two words) to each other by counting the minimum number of operations required to transform one string into another. Editing distance is used in natural language processing tasks where automatic spelling correction can identify possible edits for a misspelled word by selecting those words from the dictionary that have a low editing distance to that word. In bioinformatics, this distance can be used to quantify the similarity of DNA sequences, which can be represented as strings of letters A , C , G , and T .

Different definitions of edit distance use different sets of operations on strings. For example, Levenshtein distance uses deletion, insertion, or replacement of characters in a string. Since it is the most common metric, it is the Levenshtein distance that is usually referred to as "edit distance". The most common algorithm for finding edit distance uses a standard set of Levenshtein operations and determines the distance between $a = a_1 \dots a_n$ and $b = b_1 \dots b_m$ as d_{mn} , which is recursively calculated by the following formulas:

$$d_{i0} = \sum_{k=1}^i w_{del}(b_k) \quad \text{for } 1 \leq i \leq m,$$

$$d_{0j} = \sum_{k=1}^j w_{ins}(a_k) \quad \text{for } 1 \leq j \leq n,$$

(without failures), and in other cases, the data shown in Table 2 was obtained.

When all the necessary data are calculated, the Corcoran model must be applied to find the probability of program failure at the time of evaluation.

Thus, this approach requires a tool to analyze data from similar projects or analyze available statistics from the current project to establish the a_i parameter. Such a tool would be data mining methods using sequential pattern mining. Information on the total number of installations, the number of worlds and error groups, and the probability of each error group occurring will be used to calculate the software reliability index.

Table 2. Example of using the Corcoran model (part 1)

Type of error	Probability of error a_i	Number of N_i errors that occur during testing	$Y_i *$	$(Y_i * (N_i - 1)) / N **$
1. calculation errors	0,09	5	0,09	0,0036
2. logical errors	0,26	25	0,26	0,0624
3. input/output errors	0,16	3	0,16	0,0032
4. data manipulation errors	0,18	0	0	0
5. communication errors	0,17	11	0,17	0,017
6. data definition errors	0,08	3	0,08	0,0016
7. database errors	0,06	4	0,06	0,0018

* – value is calculated by the formula (2)
** – value is calculated by the formula (1)

Table 2. Example of using the Corcoran model (part 2)

Initial data	
$N =$	100
$N_0 =$	20
$R =$	0,2896

Mozilla Firefox receives 2.5 million error messages every day. That is why analyzing and finding software errors (bugs) is a very difficult task. Although errors can appear in different system modules and components or on different pages of web applications, they can also be the result of a general program flaw (bug). That is why there is a need to analyze and automatically search for duplicate and related reports.

The whole process of grouping reports is as follows:

1) Reports are sent to the server where they are stored. If it works, the reports will be automatically imported from the Socorro server, an open source bug report server for Mozilla products.

2) The server automatically groups reports into categories according to the cause of the bug. Each category has at least one or more reports.

3) Developers (in our case, a user from the moderator group) assign the corresponding software defect record to the general report categories. One record can correspond to one or more categories, and one category can have zero or more defect records. Programmers (in our case, users of any group) can also be assigned to defect records to resolve an existing issue.

A typical bug report for the Fennec Android mobile browser (Firefox for Android) consists of two parts, shown in Figures 1–3.

As part of the Mozilla Crash Reports project, information from the Socorro server is processed and presented in the form of statistics. For example, you can view the number of reports per day, the number of product installations, or statistics on the number of errors and reports for different versions of the product at different times. But the most interesting thing is the ability to view automatically created groups of bug reports and assigned records of software defects.

The Mozilla algorithm is quite simple and sometimes inefficient. This algorithm compares only the error signature from the top form. This leads to the appearance of double groups of errors shown in Figures 4–6, which should actually be combined into one group.

This is the reason for considering the problem of grouping related reports. Since there is very little data to analyze this problem, this paper only considers data obtained from the Socorro server.

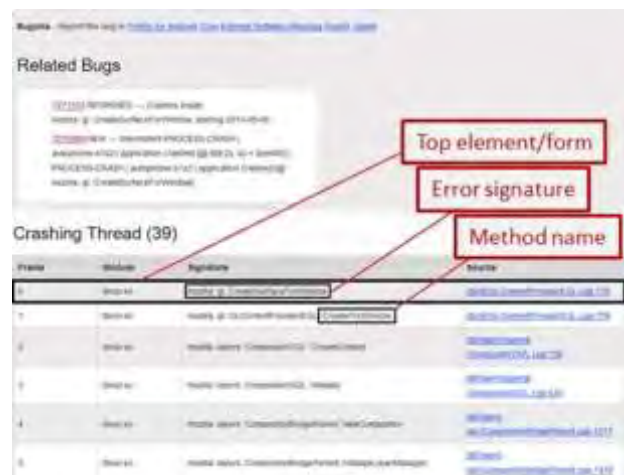


Fig. 1. Information from the defective flow

Details		Metadata	Modules	Raw Dump	Signature
Signature	mozilla::gl::CreateSurfaceForWindow More Reports Search				
UUID	7b376646-492e-4a56-8501-cddbc2160518				
Date Processed	2016-05-18T22:03:50.122684+00:00				
Uptime	7				
Last Crash	16 seconds before submission				
Install Age	14440 since version was first installed				
Install Time	2016-05-18 18:03:01				
Product	FennecAndroid				
Version	49.0a1				
Build ID	20160518030234				
Release Channel	nightly				
OS	Android				
OS Version	0.0.0 Linux 3.4.0-perf-gd1cced7 #1 SMP PREEMPT Thu Aug 27 17:06:58 2015 armv7l				
Build Architecture	arm				
Build Architecture Info	ARMV7 Qualcomm Krait features: swp,naif,thumb,fastmult,vfpv2,edsp,neon,vfpv3,fs,vfpv4,ldva,ldvt 4				
Crash Reason	SIGSEGV				
Crash Address	0x0				
User Comments					
App Notes	EGL: EGL+ GL Context? GL Context+ AdapterDescription: 'Model: SCL23, Product: SCL23_gp_kdi, Manufacturer: Sony, Hardware: qcom, OpenGL: Qualcomm -- Adreno (TM) 330 -- OpenGL ES 3.0 V866.0 ADU (CL#)' FF(D000-L10100-W00000000-T0100) Sony SCL23 EGL/SCL23_gp_kdi/SCL23:4.4.2/14.3.C.0.300/015_gpruser/release-keys				
Processor Notes	processor_proud-processor-l-9e74b745_17927; MozillaProcessorAlgorithm2015; stunk_classifier: reject - not a plugin hang				
EMCheckCompatibility	True				
Winsock LSP	None				
Adapter Vendor ID	Qualcomm				
Adapter Device ID	Adreno (TM) 330				
Android CPU ABI	armeabi-v7a				
Android Manufacturer	Sony				
Android Model	SCL23				
Android Version	19 (REL)				

Fig. 2. General information from the report

FennecAndroid 49.0a1 Crash Report [@ mozilla::gl::CreateSurfaceForWindow]

ID: 7b376646-492e-4a56-8501-cddbc2160518
Signature: mozilla::gl::CreateSurfaceForWindow

Fig. 3. Report ID and signature

Item	OS	Ver	Signature	Count	Min	Max	Min	Max	First Appearance	Signature	CrashTime
1	11.0%	4.0%	mozilla::gl::CreateSurfaceForWindow	181	0	0	0	0	2016-04-05	201110	2016-04-05
2	12.0%	4.0%	mozilla::gl::CreateSurfaceForWindow	174	0	0	0	0	2016-04-05	201110	2016-04-05
3	10.7%	3.4%	mozilla::gl::CreateSurfaceForWindow	98	0	0	0	0	2016-04-11	201110	2016-04-11
4	2.0%	0.7%	mozilla::gl::CreateSurfaceForWindow	16	0	0	0	0	2016-04-05	201110	2016-04-05
5	2.0%	0.4%	mozilla::gl::CreateSurfaceForWindow	16	0	0	0	0	2016-04-05	201110	2016-04-05
6	1.9%	0.6%	mozilla::gl::CreateSurfaceForWindow	14	0	0	0	0	2016-04-05	201110	2016-04-05

Fig. 4. Example of double error groups

Frame	Module	Signature	Symbol
0	0x00000000	0x00000000	
1	0x00000000	0x00000000	
2	0x00000000	0x00000000	
3	0x00000000	0x00000000	
4	0x00000000	0x00000000	
5	0x00000000	0x00000000	
6	0x00000000	0x00000000	
7	0x00000000	0x00000000	

Fig. 5. Report of the first group and the corresponding signature

Frame	Module	Signature	Symbol
0	0x00000000	0x00000000	
1	0x00000000	0x00000000	
2	0x00000000	0x00000000	
3	0x00000000	0x00000000	
4	0x00000000	0x00000000	
5	0x00000000	0x00000000	
6	0x00000000	0x00000000	
7	0x00000000	0x00000000	

Fig. 6. Report of the second group and the corresponding signature

Due to the imperfection or lack of tools for automatic categorization of bug reports in common bug trackers, in this article we use three rules to find related reports more accurately. These rules were built based on the analysis of the structure of reports from the Socorro server and look like this:

1) Comparison of error signatures

Examples of using this rule are the following cases:

- nsDiskCacheStreamIO::FlushBufferToFile()
Strstr | nsDiskCacheStreamIO::FlushBufferToFile()
- nsStyleContext::Release()
nsStyleContext::nsStyleContext

As you can see from the above examples, the comparison should not be performed carefully, letter by letter, taking into account the structure and special notation of the record. Please note that the signature of the highest form method with fully filled fields will be used as the signature.

To implement this rule, we will use the method of measuring the edit distance, namely the Levenshtein distance, using the Wagner–Fisher algorithm. To use this algorithm, the signature will be split into a sequence of components that will act as individual letters.

2) Uppercase comparison

This rule works on the same principle as the comparison of error signatures, but it compares not the signatures, but the file path specified in the upper forms. It's important to remember that in this article, the data from the highest form that has non-zero attributes in all its fields will be used to compare top forms and signatures.

3) Comparing frequent, closely spaced subsets of forms

This rule means that two reports are related if they have one or more of the same call stack paths or forms. For example, the reports "ABCDEF", "DEFA",

and "BDEF" have the longest common sequence – "DEF". In our case, instead of letters, we will use parts of the call stack.

To determine the length of a common element sufficient to establish a relationship, a threshold function will be used that takes the total length of the stack, the length of the common sequence, and its distance from the highest form. To determine the longest common sequences, a sequential pattern extraction algorithm will be used, namely the dynamic programming algorithm discussed above.

The previously mentioned mathematical methods and functions related to fuzzy sets will be used to evaluate the performance of these rules.

To calculate the degree of similarity between two reports, two fuzzy models were used: a model for analyzing the similarity of forms available in the reports and a model for analyzing the similarity of the reports themselves.

Conclusions

In summary, the Corcoran model offers a valuable and comprehensive approach to assessing the reliability of mobile applications, taking into account various dimensions such as performance, reliability, availability, scalability, security, usability, maintainability, and testability. By implementing this model, developers and organizations can gain valuable insights into the strengths and weaknesses of their applications, allowing them to make informed decisions and prioritize improvements.

Implementing the Corcoran model in the software development process can lead to higher quality mobile apps, increased end-user satisfaction, and increased

trust in the mobile app ecosystem. In addition, this model can serve as a benchmark for developers and organizations to compare their apps to industry standards and competitors, promoting innovation and continuous improvement of mobile apps.

In summary, the Corcoran model for assessing mobile application reliability represents a significant advancement in mobile application assessment, enabling organizations to better meet user needs and expectations while ensuring a high level of reliability in the increasingly competitive mobile application market.

In the future, it is planned to expand the use of the proposed approach based on the Corcoran model for various mobile applications and environments. In the future, it is proposed to modify the model to take into account the ever-changing characteristics of mobile applications and their growing complexity. In addition, it is desirable to conduct additional research to improve the data mining methods used in the proposed approach and to explore the possibility of integrating artificial intelligence for more advanced software reliability analysis.

References

1. Mangla, M., Sharma, N., Mohanty, S. N. (2021), "A sequential ensemble model for software fault prediction", *Innovations in Systems and Software Engineering*, P. 1–8. DOI: <https://doi.org/10.1007/s11334-021-00390-x>
2. Khuat, T. T., Le, M. H. (2019), "Ensemble learning for software fault prediction problem with imbalanced data", *International Journal of Electrical & Computer Engineering (2088–8708)*, Vol. 9, No 4. DOI: 10.11591/ijece.v9i4.pp3241-3246
3. Sales, A. M. A. et al. (2023), "Proposal of fault detection and diagnosis system architecture for residential air conditioners based on the Internet of Things", *2023 IEEE International Conference on Consumer Electronics (ICCE)*, P. 1–5. DOI: 10.1109/ICCE56470.2023.10043408
4. Joorabchi, M. E., Mesbah, A., Kruchten, P. (2013), "Real challenges in mobile app development", *2013 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, P. 15–24. DOI: 10.1109/ESEM.2013.9
5. Heitkötter, H., Hanschke, S., Majchrzak, T. A. (2012), "Evaluating cross-platform development approaches for mobile applications", *Web Information Systems and Technologies: 8th International Conference, WEBIST 2012*, Revised Selected Papers 8, P. 120–138. DOI: https://doi.org/10.1007/978-3-642-36608-6_8
6. Zhang, H., Babar, M. A. (2013), "Systematic reviews in software engineering: An empirical investigation", *Information and Software Technology*, Vol. 55, No 7, P. 1341–1354. DOI: <https://doi.org/10.1016/j.infsof.2012.09.008>
7. Garousi, V., Mäntylä, M. V. (2016), "A systematic literature review of literature reviews in software testing", *Information and Software Technology*, Vol. 80, P. 195–216. DOI: <https://doi.org/10.1016/j.infsof.2016.09.002>
8. Felizardo, K. R. et al. (2017), "Defining protocols of systematic literature reviews in software engineering: a survey", *43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, P. 202–209. DOI: 10.1109/SEAA.2017.17
9. Pachouly, J. et al. (2022), "A systematic literature review on software defect prediction using artificial intelligence: Datasets, Data Validation Methods, Approaches, and Tools", *Engineering Applications of Artificial Intelligence*, Vol. 111, 104773 p. DOI: <https://doi.org/10.1016/j.engappai.2022.104773>
10. Son, L. H., Pritam, N., Khari, M., Kumar, R., Phuong, P. T. M., & Thong, P. H. (2019), "Empirical study of software defect prediction: a systematic mapping", *Symmetry*, 11(2), 212 p. DOI: <https://doi.org/10.3390/sym11020212>
11. Li, Z., Jing, X. Y., Zhu, X. (2018), "Progress on approaches to software defect prediction", *Iet Software*, Vol. 12, No 3, P. 161–175. DOI: <https://doi.org/10.1049/iet-sen.2017.0148>
12. Zhou, T. et al. (2019), "Improving defect prediction with deep forest", *Information and Software Technology*, Vol. 114, P. 204–216. DOI: <https://doi.org/10.1016/j.infsof.2019.07.003>
13. Thota, M. K., Shajin, F. H., & Rajesh, P. (2020), "Survey on software defect prediction techniques", *International Journal of Applied Science and Engineering*, 17(4), P. 331–344. DOI: [https://doi.org/10.6703/IJASE.202012_17\(4\).331](https://doi.org/10.6703/IJASE.202012_17(4).331)
14. Singhal, S. et al. (2021), "Systematic literature review on test case selection and prioritization: A tertiary study", *Applied Sciences*, Vol. 11, No 24, P. 12121. DOI: <https://doi.org/10.3390/app112412121>
15. Shahroknii, A., Feldt R. (2013), "A systematic review of software robustness", *Information and Software Technology*, Vol. 55, No 1, P. 1–17. DOI: <https://doi.org/10.1016/j.infsof.2012.06.002>
16. Febrero, F., Calero, C., Moraga, M. Á. (2016), "Software reliability modeling based on ISO/IEC SQuaRE", *Information and Software Technology*, Vol. 70, P. 18–29. DOI: <https://doi.org/10.1016/j.infsof.2015.09.006>
17. Ali, S. et al. (2009), "A systematic review of the application and empirical investigation of search-based test case generation", *IEEE Transactions on Software Engineering*, Vol. 36, No 6, P. 742–762. DOI: 10.1109/TSE.2009.52
18. Rathi, G., Tiwari, U. K., Singh, N. (2022), "Software Reliability: Elements, Approaches and Challenges", *International Conference on Advances in Computing, Communication and Materials (ICACCM)*. P. 1–5. DOI: 10.1109/ICACCM56405.2022.10009422

Відомості про аспірів / About the Authors

Шматко Олександр Віталійович – PhD, доцент, Національний технічний університет "Харківський політехнічний інститут", доцент кафедри програмної інженерії та інтелектуальних технологій управління, Харків, Україна; e-mail: oleksandr.shmatko@khi.edu.ua; ORCID ID: <http://orcid.org/0000-0002-2426-900X>

Коломійцев Олексій Володимирович – доктор технічних наук, професор, Національний технічний університет "Харківський політехнічний інститут", професор кафедри комп'ютерної інженерії та програмування, Харків, Україна; e-mail: alexis_k@ukr.net; ORCID ID: <http://orcid.org/0000-0001-8228-8404>

Федорченко Володимир Миколайович – PhD, доцент, Харківський національний університет радіоелектроніки, доцент кафедри електронних обчислювальних машин, Харків, Україна; e-mail: volodymyr.fedorchenko@nure.ua; ORCID ID: <http://orcid.org/0000-0001-7359-1460>

Михайленко Ірина Володимирівна – PhD, доцент, Харківський національний автомобільно-дорожній університет, доцент кафедри вищої математики, Харків, Україна; e-mail: irinaamih@gmail.com; ORCID ID: <http://orcid.org/0000-0002-5961-3616>

Третяк Вячеслав Федорович – кандидат технічних наук, доцент, Харківський національний університет Повітряних Сил імені Івана Кожедуба, науковий співробітник наукового центру Повітряних Сил, Харків, Україна; e-mail: slava_tr@ukr.net; ORCID ID: <http://orcid.org/0000-0003-2599-8834>

Shmatko Oleksandr – PhD, Associate Professor, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department of Software Engineering and Intelligent Management Technologies, Kharkiv, Ukraine.

Kolomiitsev Olexsii – Doctor of Sciences (Engineering), Professor, National Technical University "Kharkiv Polytechnic Institute", Professor at the Department of Computer Engineering and Programming, Kharkiv, Ukraine.

Fedorchenko Volodymyr – PhD, Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Electronic Computers, Kharkiv, Ukraine.

Mykhailenko Iryna – PhD, Associate Professor, National Automobile and Road University, Associate Professor at the Department of Higher Mathematics, Kharkiv, Ukraine.

Tretiak Viacheslav – PhD, Associate Professor, Ivan Kozhedub Kharkiv National Air Force University, Senior Researcher, Kharkiv, Ukraine.

ІНТЕГРАЦІЯ АНАЛІТИЧНИХ СТАТИСТИЧНИХ МОДЕЛЕЙ, ПОСЛІДОВНОГО АНАЛІЗУ ЗАКОНОМІРНОСТЕЙ ТА ТЕОРІЇ НЕЧІТКИХ МНОЖИН ДЛЯ РОЗШИРЕНОГО ОЦІНЮВАННЯ НАДІЙНОСТІ МОБІЛЬНИХ ЗАСТОСУНКІВ

Дослідження є новим методом оцінки надійності мобільних додатків за допомогою моделі Коркорана. Ця модель включає в себе кілька аспектів надійності, включаючи продуктивність, надійність, доступність, масштабованість, безпеку, зручність використання і тестованість. Модель Коркорана може бути застосована для оцінки мобільних додатків шляхом аналізу основних показників надійності. Використання моделі значно поліпшує оцінку надійності застосунків в порівнянні з традиційними методами, які в першу чергу орієнтовані на конфігурації настільних комп'ютерів і серверів. Мета дослідження – запропонувати більш оптимізований підхід до оцінки надійності мобільних додатків. В роботі розглянуто проблеми з якими стикаються розробники мобільних застосунків. Це дослідження представляє нове застосування моделі Коркорана в області оцінки надійності мобільних додатків. Ця модель відрізняється акцентом на використання кількісної статистики та здатністю надавати точну оцінку ймовірності збою без будь-яких неточностей, що відрізняє цю модель від інших моделей надійності програмного забезпечення. В роботі пропонується використання комбінації аналітичних статистичних моделей, методів видобутку даних, таких як послідовний аналіз шаблонів, і теорію нечітких множин для реалізації моделі Коркорана. Застосування методології продемонстровано на прикладі дослідження звітів про помилки програмного забезпечення та проведення їх всебічного статистичного аналізу. Щоб покращити результати майбутніх досліджень, в роботі пропонується більш широко використовувати модель Коркорана у різних мобільних додатках та середовищах. Рекомендується змінити модель, щоб врахувати постійно мінливі характеристики мобільних додатків і їх зростаючу складність. Крім того, бажано провести додаткові дослідження для вдосконалення методів видобутку даних, що використовуються в моделі, та вивчити можливість інтеграції штучного інтелекту для більш просунутого аналізу надійності програмного забезпечення. Застосування моделі Коркорана у процесі розробки мобільних додатків для оцінки надійності може значно підвищити якість додатків, що призведе до підвищення рівня задоволеності клієнтів та довіри до мобільних додатків. Ця модель може слугувати орієнтиром для розробників та компаній при оцінці та вдосконаленні своїх додатків, сприяючи інноваціям та постійному вдосконаленню в конкурентному секторі мобільних додатків.

Ключові слова: мобільний застосунок; розробка програмного забезпечення; оцінювання надійності; модель Коркорана.

Бібліографічні описи / Bibliographic descriptions

Федорченко В. М., Шматко О. В., Михайленко І. В., Третяк В. Ф., Коломійцев О. В. Інтеграція аналітичних статистичних моделей, послідовного аналізу закономірностей та теорії нечітких множин для розширеного оцінювання надійності мобільних застосунків. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 78–86. DOI: <https://doi.org/10.30837/ITSSI.2023.26.078>

Fedorchenko, V., Shmatko, O., Mykhailenko, I., Tretiak, V., Kolomiitsev, O. (2023), "Integrating analytical statistical models, sequential pattern mining, and fuzzy set theory for advanced mobile app reliability assessment", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 78–86. DOI: <https://doi.org/10.30837/ITSSI.2023.26.078>

Д. ЧЕРНИШОВ, Д. СИТНИКОВ

БІНАРНА КЛАСИФІКАЦІЯ НА ОСНОВІ ПОЄДНАННЯ ТЕОРІЇ ПРИБЛИЗНИХ МНОЖИН І ДЕРЕВ РІШЕНЬ

Предмет дослідження – підвищення точності та ефективності алгоритмів класифікації на основі дерев рішень за допомогою інтеграції принципів теорії приблизних множин (*Rough Set*), математичного підходу до апроксимації множин.

Мета дослідження – розроблення гібридної моделі, яка об'єднує теорію приблизних множин з алгоритмами дерев рішень, тим самим вирішуючи вроджені обмеження цих алгоритмів у роботі з невизначеністю в даних. Ця інтеграція має суттєво покращити точність та ефективність бінарної класифікації на основі дерев рішень, роблячи їх більш стійкими до різних вхідних даних. **Завдання статті** передбачають глибоке вивчення можливих синергій між теорією приблизних множин та алгоритмами дерев рішень. З цієї метою комплексно досліджено інтеграцію теорії приблизних множин у межах алгоритмів дерев рішень. Це потребує розроблення моделі, що використовує принципи та алгебраїчні інструменти теорії приблизних множин для більш ефективного відбору ознак у системах, оснований на деревах рішень. Модель застосовує теорію приблизних множин для ефективної роботи з невизначеністю та вагомністю, що дає змогу удосконалювати та розширювати процеси відбору ознак у системах дерев рішень. Проведено серію експериментів на різних наборах даних для демонстрації ефективності та практичності цього підходу. Ці набори даних обрані для подання спектра складностей та невизначеностей із забезпеченням ретельного оцінювання можливостей моделі. **Методологія** використовує передові алгебраїчні інструменти теорії приблизних множин, зокрема формулювання алгебраїчних виразів та розроблення нових правил і технік, для спрощення та підвищення точності процесів класифікації даних за допомогою систем дерев рішень. Знахідки дослідження є важливими, оскільки свідчать про те, що інтеграція теорії приблизних множин у алгоритми дерев рішень може забезпечити більш точні та ефективні результати класифікації. Така гібридна модель демонструє значні переваги в роботі з інформацією із вбудованою невизначеністю, що є загальним викликом у багатьох додаткових сценаріях. Ефективність інтегрованого підходу продемонстровано його успішним застосуванням у сферах кредитного скорингу та кібербезпеки, що вказує на його потенціал як універсального інструмента в галузі видобутку даних і машинного навчання.

Висновки. З'ясовано, що інтеграція теорії приблизних множин може привести до більш точних та ефективних результатів класифікації. Покращуючи можливість дерев рішень, необхідно зважати на невизначеність і неточність інформації. Дослідження відкриває нові перспективи для надійного й ґрунтового аналізу та інтерпретації даних у різних галузях – від охорони здоров'я до сфери фінансів тощо. Інтеграція теорії приблизних множин і дерев рішень є важливим кроком у розвитку більш удосконалених, ефективних і точних інструментів класифікації в епоху великих обсягів інформації.

Ключові слова: класифікація дерев рішень; теорія приблизних множин; алгебраїчний підхід; машинне навчання.

Вступ

Класифікація за допомогою дерева рішень набула поширення як основна техніка в галузі видобутку даних і машинного навчання, що визнається своєю простою інтерпретацією та застосуванням у різних галузях, зокрема охорона здоров'я, фінансування, маркетинг і аналіз соціальних медіа. Поява значних обсягів інформації підвищила важливість ефективних методів класифікації, що можуть швидко й точно обробляти й категоризувати великі набори даних. Зокрема, дерева рішень відіграють ключову роль у завданнях, пов'язаних із класифікацією даних на окремі категорії або з виявленням закономірностей і тенденцій у складних наборах даних.

У галузі класифікації за допомогою дерев рішень основна **проблема** полягає в обмеженій здатності

алгоритму обробляти невизначеність і неточність інформації. Цей недолік особливо суттєвий у сценаріях, де точна й надійна класифікація важлива, наприклад, у медичній діагностиці, аналізі фінансового ризику та складних процесах прийняття рішень на основі даних. Традиційні моделі дерев рішень, хоча й ефективні в простих сценаріях, часто викликають труднощі в підтриманні високої точності під час зіткнення з неоднозначною або неповною інформацією. Дослідження спрямоване на вирішення цієї проблеми способом інтеграції теорії приблизних множин у фреймворк дерев рішень. Теорія приблизних множин зі своїми міцними математичними структурами для роботи з вагомністю та невизначеністю пропонує перспективне рішення для покращення здатності дерев рішень класифікувати дані точно за таких складних умов.

Успішна інтеграція теорії приблизних множин з алгоритмами дерев рішень спрямована на створення більш адаптивної, ефективної та точної моделі дерева рішень, здатної розв'язувати складні питання сучасного аналізу інформації. Для досягнення поставленої мети необхідно виконати завдання – детально дослідити потенційні синергії між теорією приблизних множин та алгоритмами дерев рішень.

Інтеграція теорії приблизних множин у алгоритми дерев рішень спрямована на покращення їх здатності ефективно обробляти й розуміти невизначені дані. Цей підхід особливо актуальний у галузях, де є важливими точність та інтерпретація класифікації, наприклад, у медичній діагностиці або оцінюванні фінансового ризику. Теорія приблизних множин надає міцний математичний фундамент для покращення результативності дерев рішень, особливо в сценаріях із неповною або неоднозначною інформацією.

У класифікації за допомогою дерев рішень застосовуються різні підходи для виявлення важливості ознак, починаючи від евристичних підходів на основі заздалегідь визначених правил і експертних знань, ефективних у простіших сценаріях, до більш класичних підходів, наприклад, алгоритмів дерев рішень, з традиційними критеріями відбору ознак, такими як ентропія. Проте кожен із цих методів має свої обмеження, особливо в умовах невизначеності даних, і прогалини, які має заповнити теорія приблизних множин.

Метою цього дослідження є розроблення гібридної моделі, що об'єднує теорію приблизних множин з алгоритмами дерев рішень, тим самим вирішуючи вроджені обмеження цих алгоритмів у роботі з невизначеністю даних. Щоб досягти окресленої мети, додатково планується провести всебічний порівняльний аналіз ефективності розробленої гібридної моделі та порівняти її з традиційними алгоритмами.

Це порівняльне оцінювання буде зосереджено на ключових метриках, таких як точність класифікації, ефективність оброблення даних і стійкість методологій до різних типів наборів даних. З допомогою цього дослідження сподіваємося визначити більш надійну й універсальну модель класифікації за допомогою дерева рішень, яка може ефективно обробляти складнощі сучасних даних і здійснити вагомий внесок у сфері видобутку інформації та машинного навчання.

Аналіз проблеми й наявних методів

Інтеграція теорії приблизних множин у класифікацію за допомогою дерев рішень є актуальним, проте складним завданням. Дерева рішень – це основний інструмент у галузі добування даних та машинного навчання, відомий своєю інтерпретованістю та застосовністю в різних сферах. Однак дерева рішень часто виявляються неефективними в роботі з нечіткою та невизначеною інформацією, що спричиняє зниження ефективності класифікації. Використання теорії приблизних множин, що ефективно долає нечіткість і невизначеність, має потенціал розв'язати ці проблеми, але така інтеграція не є простою та потребує всебічного аналізу як теоретичних, так і практичних аспектів.

Унаслідок ретельного огляду сучасної літератури можна говорити про значні досягнення як в алгоритмах дерев рішень, так і в теорії приблизних множин. Так, дослідження [1, 2] висвітлюють поліпшення в методологіях дерев рішень, наголошуючи на підвищенні їх адаптивності до різноманітних типів даних. Одночасно теорія приблизних множин усе частіше застосовується в галузях, де є невизначеність і неточність інформації, про що говориться в студіях [3, 4]. Однак злиття цих двох сфер наразі залишається малодослідженим.

У роботі [5] наголошується на важливості невизначеності даних у деревах рішень і прогалині, яку може заповнити теорія приблизних множин. Крім того, у дослідженнях [6, 7] обговорюються алгебраїчні підходи в теорії приблизних множин, що можуть надати систематичний спосіб для їх інтеграції в дерева рішень. Складність такої інтеграції розглядається у статті [8]. Її автори висвітлюють необхідність нових методологій, які могли б безперешкодно об'єднувати ці дві сфери.

Можливість такої інтеграції підтверджується останніми досягненнями в алгоритмічних підходах до дерев рішень [9, 10]. Ці досягнення створюють основу, до якої може бути долучена теорія приблизних множин. Також вивчення теорії приблизних множин [11, 12] має значний прогрес у алгебраїчних методах, натякаючи на шляхи покращення класифікації дерев рішень.

Практичні наслідки такої інтеграції є вагомими. У таких сферах, як охорона здоров'я та фінансування, де прийняття рішень є критичним і інформація часто містить елемент невизначеності, покращені

моделі дерев рішень можуть сприяти більш надійним і точним прогнозам, як це передбачається в дослідженнях, зокрема [13]. Однак ця інтеграція висуває нові виклики, особливо в збереженні простоти та інтерпретованості дерев рішень під час включення складності теорії приблизних множин.

Хоча в літературі йдеться про обнадійливий ефект у покращенні класифікації дерев рішень за допомогою теорії приблизних множин, у ній також наголошується на необхідності інноваційних підходів для подолання викликів, властивих такій інтеграції.

Це дослідження має на меті подолати цей розрив способом розроблення алгебраїчного підходу, що гармонізує надійність теорії приблизних множин із простотою алгоритмів дерев рішень. Постає необхідність створити гібридну модель, що поєднує теорію приблизних множин з алгоритмами дерев рішень, підвищуючи точність класифікації за умови невизначеності даних, а також зберігаючи простоту використання дерев рішень.

Вирішення завдання

У цьому дослідженні, коли розглядаємо граничну ділянку $B(X)$ для множини X , маємо на увазі різницю між верхньою та нижньою апроксимацією X . Математично це може бути виражено так:

$$B(X) = A^*(X) - A_*(X), \quad (1)$$

де $A^*(X)$ – верхня апроксимація X , що є множиною всіх елементів, які можуть бути в X на підставі доступної інформації (2). Вона містить усі елементи, що мають принаймні одну ознаку, спільну з членами X . Ця апроксимація зазвичай більш включаюча й охоплює елементи, які можуть потенційно належати розглянутій множині.

$$A^*(X) = \{x \in U, [x] \cap X \neq \emptyset\}, \quad (2)$$

де U – універсальна множина;

$[x]$ – клас еквівалентності (3).

$$[x] = \{y \in U, xRy\}. \quad (3)$$

З іншого боку, $A_*(X)$ – це нижня апроксимація X , що містить елементи, які точно належать X , на основі наявних даних. Вона має всі елементи, для яких кожна ознака відповідає ознакам в X (4). Ця апроксимація більш виключна і містить лише ті елементи, які обов'язково належать розглянутій множині.

$$A_*(X) = \{x \in U, [x] \subseteq X\}. \quad (4)$$

Після обчислення граничної ділянки наступним кроком є оцінювання важливості кожної ознаки. Це досягається способом систематичного вилучення ознак по одній і спостереження за результатом змін у граничній ділянці. Аналізуючи, як вилучення кожної ознаки впливає на граничну ділянку, можемо визначити відносну важливість кожної ознаки. Цей підхід дає змогу виявити критичні ознаки для визначення ясності або неоднозначності класифікації. Ознаки, що в процесі вилучення значно змінюють граничну ділянку, вважаються більш важливими, оскільки їх присутність істотно сприяє точності та однозначності класифікації. Навпаки, ознаки, вилучення яких призводить до мінімальних змін у граничній ділянці, вважаються менш критичними. Цей метод дає більш детальне розуміння важливості ознак поза традиційними метриками способом прямого зв'язку важливості ознаки з її впливом на динаміку границі класифікації.

$$V(P_i) = \frac{\Delta(BN_i)}{M(X)} \times 100\%, \quad (5)$$

де $M(X)$ – це кількість елементів у множині X .

У запропонованій гібридній моделі важливість ознаки (5) застосовуватиметься як критерій розбиття для побудови дерева рішень.

Матеріали й методи

Отримана гібридна модель була оцінена за допомогою метрики площі під кривою робочої характеристики приймача ($ROC AUC$). Цей метод оцінювання використовується для визначення точності класифікації моделі на наборі даних, який містить як позитивні, так і негативні приклади.

Оцінювання моделі проводитиметься за допомогою випадкової стратифікованої крос-валідації з 10 фолдами. Зазначений метод забезпечує більш точну оцінку ефективності моделі, ніж оцінка на одному наборі даних.

Стратифікована крос-валідація працює шляхом поділу набору даних на 10 рівних частин, або фолдів. Потім модель навчається на 9 фолдах і тестується на 10-му фолді. Цей процес повторюється 10 разів, і результати всіх 10 тестувань використовуються для оцінювання ефективності моделі.

За допомогою цього методу можна оцінити ефективність моделі на трьох різноманітних наборах

даних. Ці набори обиралися для подання різних рівнів складності та невизначеності, забезпечуючи всебічну оцінку можливостей моделі.

Набір даних *Titanic* [14], широко визнаний у сфері машинного навчання й науки про дані, дає ідеальний варіант для оцінювання покращеної моделі дерева рішень. Цей набір даних містить інформацію про пасажирів відомого корабля *Titanic*, зокрема такі відомості, як вік, стать, клас, вартість квитка та статус виживання. Складність і змінність цієї інформації роблять її ефективним вибором для оцінювання можливостей моделі у роботі з реальними, невизначеними й категоріальними даними.

Набір даних *Microsoft Malware Classification Challenge (BIG 2015)* [15] є бенчмарком у сфері кібербезпеки, особливо в класифікації шкідливих програм. Він містить значну кількість об'єктів, кожен з яких є різними типами шкідливих програм і має різні ознаки, такі як байткод, асемблерний код та інші характеристики файлів. Складність цього набору даних і критичне значення його застосування роблять його гарним кандидатом для оцінювання ефективності покращеної моделі дерева рішень в умовах високого ризику та технічної складності.

Набір даних *Home Credit Default Risk* [16], надає вичерпне середовище для оцінювання покращеної моделі дерева рішень у контексті оцінювання кредитного ризику. Цей набір даних особливо актуальний для фінансових установ, мета яких – точно передбачити здатність своїх клієнтів до повернення кредитів. Він має різноманітні ознаки, зокрема кредитна історія, деталі позики та демографічні відомості позичальника, що робить його цінним ресурсом для тестування класифікаційних здібностей моделі у фінансовому середовищі.

У цьому дослідженні всі три набори даних були однаково підготовлені для бінарної класифікації. Підготовка даних передбачала перетворення цільових змінних у бінарний формат. Також були відібрані та бінаризовані відповідні ознаки, здійснено оброблення відсутніх значень за допомогою відповідних технік та нормалізацію або стандартизацію числових ознак. Крім того, кожен набір даних був розподілений на навчальні й тестові набори із заходами, спрямованими на забезпечення балансу даних і подолання проблем нерівномірності класів. Цей стандартизований підхід до підготовки даних забезпечує послідовність між наборами даних і сприяє справедливому оцінюванню продуктивності моделей дерев рішень у різних завданнях бінарної класифікації.

Результати досліджень

У дослідженні проведено порівняльний аналіз продуктивності двох критеріїв для побудови дерев рішень: критерію приросту інформації та критерію зміни граничної ділянки. Для цього було використано три набори даних: *Titanic*, *Malware Classification* та *Credit Default Risk*.

Результати порівняння показали, що критерій зміни граничної ділянки в деяких випадках забезпечує більш високу точність, ніж критерій приросту інформації. Зокрема, для набору даних *Titanic* критерій зміни граничної ділянки забезпечує підвищення точності класифікації на 1,3 %. У табл. 1 подані значення оцінки якості двох критеріїв для кожного набору даних.

Таблиця 1. Оцінка якості критеріїв

	<i>Titanic</i>	<i>Malware Classification</i>	<i>Credit Default Risk</i>
Важливість ознаки відповідно до приросту інформації	0.828	0.5658	0.6155
Важливість ознаки відповідно до змін у граничній ділянці	0.841	0.5658	0.5787

Підвищення точності класифікації на наборі даних *Titanic* можна пояснити тим, що критерій зміни граничної ділянки дозволяє будувати дерева рішень, які краще відображають структуру даних. Зокрема, зазначений критерій бере до уваги не тільки кількість інформації, яка отримується в процесі розбиття на основі ознаки, але також і розмір граничної ділянки, що утворюється в цій ситуації. Це дає змогу обирати ознаки, які забезпечують найбільш чітке розбиття даних на два класи.

Також був проведений візуальний аналіз двох різних графіків дерев рішень, для датасету *Titanic* з параметром максимальної глибини, який був заданий значенням 4. Ці візуалізації є важливими для розуміння впливу інтеграції теорії приблизних множин у моделі дерева рішень. Були візуалізовані графіки як для критерію приросту інформації (рис. 1), так і для критерію різниці граничних ділянок (рис. 2). Така візуалізація дає змогу зрозуміти, як дерево рішень визначає пріоритети різних ознак на основі різних критеріїв.

Було створено три графіки кореляції (рис. 3) для аналізу взаємозв'язку змін у граничній ділянці

та ентропії для кожної ознаки в наборах даних. Результати свідчать про слабку кореляцію, яка вказує на те, що зміни в граничній ділянці, хоч і впливають на ентропію, але це не має систематичного характеру.

Ця слабка кореляція підкреслює складну природу відношень між коригуваннями граничної ділянки та загальним отриманням інформації в моделях дерева рішень.

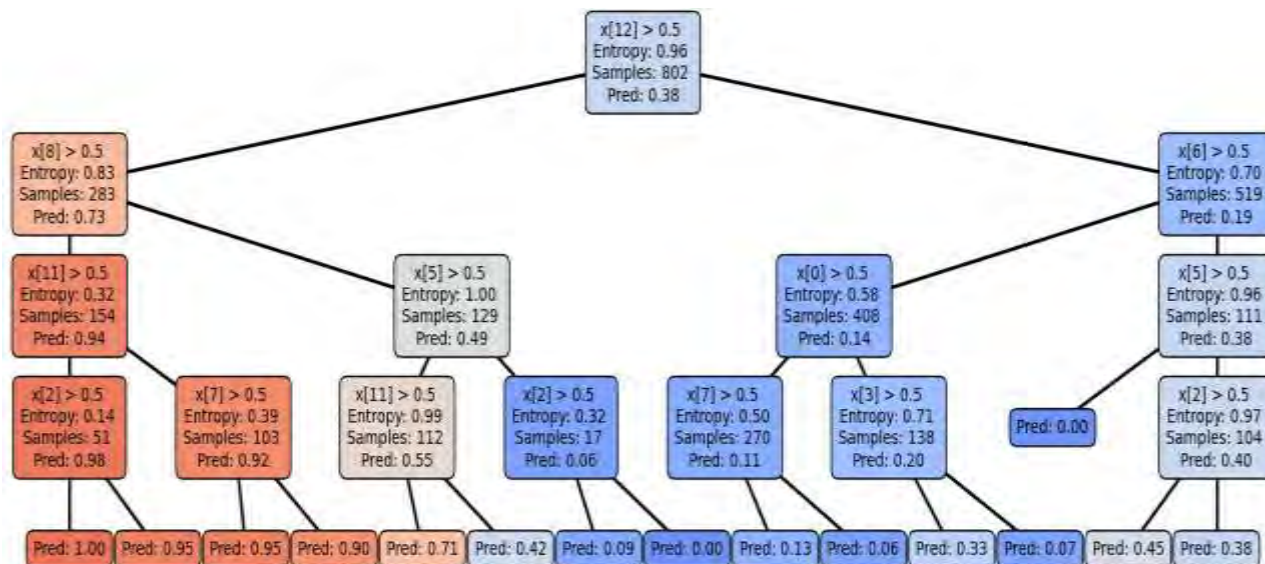


Рис. 1. Дерево рішень на базі критерію приросту інформації

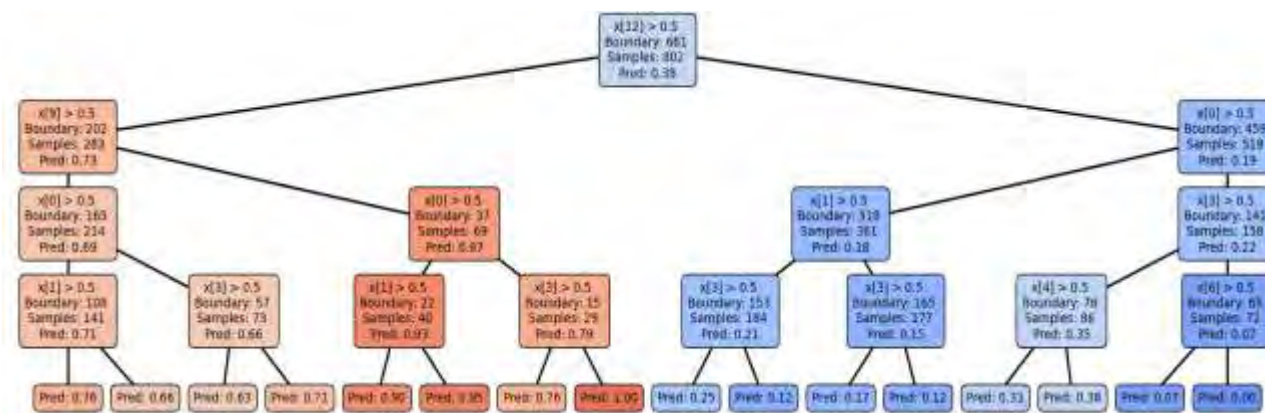


Рис. 2. Дерево на базі критерію різниці граничних ділянок

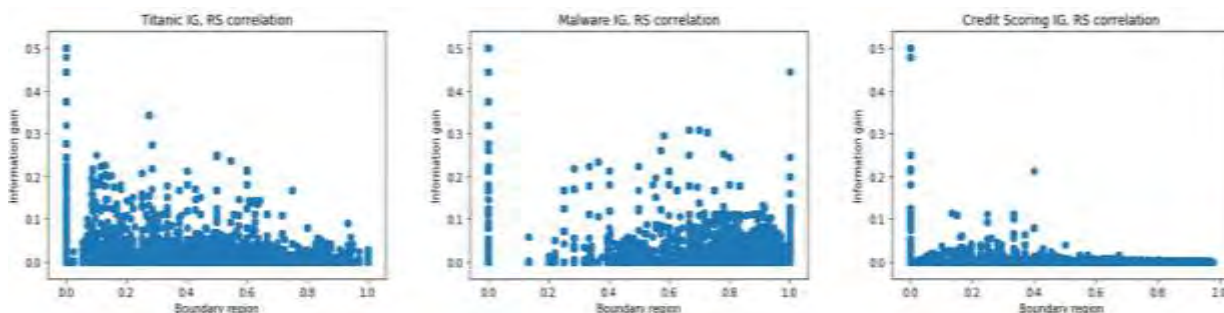


Рис. 3. Кореляція зміни граничної ділянки та приросту інформації

Підсумовуючи, можемо наголосити, що інтеграція теорії приблизних множин у класифікацію за

допомогою дерев рішень має перспективні результати в деяких наборах даних, особливо в покращенні

точності класифікації. Візуалізації надають більш глибокі уявлення про те, як ця інтеграція впливає на процес прийняття рішень моделей. Слабка кореляція між змінами в граничній ділянці та ентропією наголошує на недоліках і складнощах цієї інтеграції, указуючи на необхідність подальших досліджень та оптимізації для повного використання потенціалу теорії приблизних множин у нових моделях дерев рішень.

Висновки

На основі результатів цього дослідження можна прийти до висновку, що інтеграція теорії приблизних множин у алгоритми класифікації за допомогою дерев рішень значно підвищує їх продуктивність, особливо в умовах роботи з невизначеністю та вагомністю інформації. Кілька ключових факторів лежать в основі успіху цього гібридного підходу.

1. Об'єднана модель ефективно використовує переваги теорії приблизних множин у роботі з вагомністю ознак та невизначеністю даних. Ця інтеграція може сприяти значному покращенню точності класифікації, особливо в складних наборах даних, де традиційні дерева рішень зазвичай не досягають високої продуктивності.

2. Алгебраїчний підхід для теорії приблизних множин допомагає вдосконалити процес прийняття рішень у завданні класифікації. Це покращення дає змогу моделі виявляти та тлумачити витончені

зразки та взаємозв'язки в даних, які можуть бути пропущені традиційними алгоритмами дерев рішень.

3. Інтеграція зберігає фундаментальну простоту та можливість тлумачення дерев рішень, але й додаючи міцності теорії приблизних множин. Ця рівновага є важливою для практичного застосування, де одночасно значущими є як точність, так і зручність використання.

Порівняльний аналіз дослідження, що порівнює покращену модель дерева рішень із традиційними алгоритмами, розкриває переваги та потенціал цього гібридного підходу. Інтеграція не лише покращує точність класифікації, але й розширює застосовність дерев рішень на більш складні та невизначені сценарії оброблення даних.

Саме тому можна зробити висновок, що інтеграція теорії приблизних множин у класифікацію за допомогою дерев рішень є певним проривом. Здатність покращеної моделі вирішувати завдання невизначеності, а також її практичність роблять її дуже ефективним інструментом для завдань класифікації даних у різних галузях. Майбутні дослідження мають зосереджуватися на подальшому вдосконаленні та застосуванні цієї моделі, особливо в галузях, де невизначеність даних є значним викликом. Це дослідження є перспективним для створення більш надійних, ефективних і точних інструментів з метою прийняття рішень в епоху великих обсягів даних та складного оброблення інформації.

Список літератури

1. Costa V. G. and Pedreira C. E. Recent advances in decision trees: an updated survey. *Artificial Intelligence Review, Springer Science and Business Media LLC*. Vol. 56. No. 5. P. 4765–4800. 2022. DOI: 10.1007/s10462-022-10275-5
2. Hafeez M. A., Rashid M., Tariq H., Abideen Z. U., Alotaibi S. S., and Sinky M. H. Performance Improvement of Decision Tree: A Robust Classifier Using Tabu Search Algorithm. *Applied Sciences, MDPI AG*. Vol. 11. No. 15. 6728 p. 2021. DOI: 10.3390/app11156728
3. Wang Z., Zhang X., and Deng J. The uncertainty measures for covering rough set models. *Soft Computing, Springer Science and Business Media LLC*. Vol. 24. No. 16. P. 11909–11929. 2020. DOI: 10.1007/s00500-020-05098-x
4. Geetha M. A., Acharjya D. P., and Iyengar N. Ch. S. N. Algebraic properties and measures of uncertainty in rough set on two universal sets based on multi-granulation. *Proceedings of the 6th ACM India Computing Convention, ACM*. P. 1–8. 2013. DOI: 10.1145/2522548.2523168
5. Qian Y., Xu H., Liang J., Liu B., and Wang J. Fusing Monotonic Decision Trees. *IEEE Transactions on Knowledge and Data Engineering*. Vol. 27. No. 10. P. 2717–2728. 2015. DOI: 10.1109/TKDE.2015.2429133
6. Sitnikov D. and Ryabov O. An Algebraic Approach to Defining Rough Set Approximations and Generating Logic Rules. *Data Mining V, WIT Press*. 10 p. 2004. DOI: 10.2495/data040171
7. Sitnikov D., Titova O., Romanenko O., and Ryabov O. A method for finding minimal sets of features adequately describing discrete information objects. *Data Mining X, WIT Press*. 8 p. 2009. DOI: 10.2495/data090141
8. Wang D., Liu X., Jiang L., Zhang X., and Zhao Y. Rough Set Approach to Multivariate Decision Trees Inducing. *Journal of Computers, International Academy Publishing (IAP)*. Vol. 7. No. 4. P. 870–879. 2012. DOI: 10.4304/jcp.7.4.870-879

9. Blockeel H., Devos L., Fréney B., Nanfack G., and Nijssen S. Decision trees: from efficient prediction to responsible AI. *Frontiers in Artificial Intelligence, Frontiers Media SA*. Vol. 6. Jul. 26. 2023. DOI: 10.3389/frai.2023.1124553
10. Hu X., Rudin C., and Seltzer M. Optimal Sparse Decision Trees. *arXiv*. 2019. DOI: 10.48550/ARXIV.1904.12847
11. Chiaselotti G., Gentile T., and Infusino F. Decision systems in rough set theory: A set operatorial perspective. *Journal of Algebra and Its Applications, World Scientific Pub Co Pte Lt*. Vol. 18. No. 01. 2019. 1950004 p. DOI: 10.1142/s021949881950004x
12. Xu J., Qu K., Meng X., Sun Y., and Hou Q. Feature selection based on multiview entropy measures in multiperspective rough set. *International Journal of Intelligent Systems, Hindawi Limited*. Vol. 37. No. 10. 2022. P. 7200–7234. DOI: 10.1002/int.22878
13. Duan G., Ding D., Tian Y., and You X. An Improved Medical Decision Model Based on Decision Tree Algorithms. *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), Atlanta, GA, USA IEEE*. 2016. P. 151–156. DOI: 10.1109/BDCloud-SocialCom-SustainCom.2016.33
14. Cukierski W. Titanic – Machine Learning from Disaster. Kaggle. 2012. URL: <https://kaggle.com/competitions/titanic>
15. Ronen R., Radu M., Feuerstein C., Yom-Tov E., and Ahmadi M. Microsoft Malware Classification Challenge. *arXiv*. 2018. DOI: 10.48550/ARXIV.1802.10135
16. Montoya A., Odintsov K., and Kotek M. Home Credit Default Risk. Kaggle. 2018. URL: <https://kaggle.com/competitions/home-credit-default-risk>

References

1. Costa, V. G. and Pedreira, C. E. (2022), "Recent advances in decision trees: an updated survey". *Artificial Intelligence Review, Springer Science and Business Media LLC*. Vol. 56, No. 5. P. 4765–4800. DOI: 10.1007/s10462-022-10275-5
2. Hafeez, M. A., Rashid, M., Tariq, H., Abideen, Z. U., Alotaibi, S. S., and Sinky, M. H. (2021), "Performance Improvement of Decision Tree: A Robust Classifier Using Tabu Search Algorithm". *Applied Sciences, MDPI AG*. Vol. 11, No. 15. 6728 p. DOI: 10.3390/app11156728
3. Wang, Z., Zhang, X., and Deng, J. (2020), "The uncertainty measures for covering rough set models". *Soft Computing, Springer Science and Business Media LLC*. Vol. 24, No. 16. P. 11909–11929. DOI: 10.1007/s00500-020-05098-x
4. Geetha, M. A., Acharjya, D. P., and Iyengar, N. Ch. S. N. (2013), "Algebraic properties and measures of uncertainty in rough set on two universal sets based on multi-granulation". *Proceedings of the 6th ACM India Computing Convention, ACM*. P. 1–8. DOI: 10.1145/2522548.2523168
5. Qian, Y., Xu, H., Liang, J., Liu, B., and Wang, J. (2015), "Fusing Monotonic Decision Trees". *IEEE Transactions on Knowledge and Data Engineering*. Vol. 27, No. 10. P. 2717–2728. DOI: 10.1109/TKDE.2015.2429133
6. Sitnikov, D. and Ryabov, O. (2004), "An Algebraic Approach to Defining Rough Set Approximations and Generating Logic Rules". *Data Mining V, WIT Press*. 10 p. DOI: 10.2495/data040171
7. Sitnikov, D., Titova, O., Romanenko, O., and Ryabov, O. (2009), "A method for finding minimal sets of features adequately describing discrete information objects". *Data Mining X, WIT Press*. 8 p. DOI: 10.2495/data090141
8. Wang, D., Liu, X., Jiang, L., Zhang, X., and Zhao, Y. (2012), "Rough Set Approach to Multivariate Decision Trees Inducing". *Journal of Computers, International Academy Publishing (IAP)*. Vol. 7, No. 4. P. 870–879. DOI: 10.4304/jcp.7.4.870-879
9. Blockeel, H., Devos, L., Fréney, B., Nanfack, G., and Nijssen, S. (2023), "Decision trees: from efficient prediction to responsible AI". *Frontiers in Artificial Intelligence, Frontiers Media SA*. Vol. 6. Jul. 26. DOI: 10.3389/frai.2023.1124553
10. Hu, X., Rudin, C., and Seltzer, M. (2019), "Optimal Sparse Decision Trees". *arXiv*. DOI: 10.48550/ARXIV.1904.12847
11. Chiaselotti, G., Gentile, T., and Infusino, F. (2019), "Decision systems in rough set theory: A set operatorial perspective". *Journal of Algebra and Its Applications, World Scientific Pub Co Pte Lt*. Vol. 18, No. 01. 1950004 p. DOI: 10.1142/s021949881950004x
12. Xu, J., Qu, K., Meng, X., Sun, Y., and Hou, Q. (2022), "Feature selection based on multiview entropy measures in multiperspective rough set". *International Journal of Intelligent Systems, Hindawi Limited*. Vol. 37, No. 10. P. 7200–7234. DOI: 10.1002/int.22878
13. Duan, G., Ding, D., Tian, Y., and You, X. (2016), "An Improved Medical Decision Model Based on Decision Tree Algorithms". *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), IEEE*. P. 151–156. DOI: 10.1109/BDCloud-SocialCom-SustainCom.2016.33
14. Cukierski, W. (2012), "Titanic – Machine Learning from Disaster" Kaggle. available at: <https://kaggle.com/competitions/titanic>.
15. Ronen, R., Radu, M., Feuerstein, C., Yom-Tov, E., and Ahmadi, M. (2018), "Microsoft Malware Classification Challenge". *arXiv*. DOI: 10.48550/ARXIV.1802.10135

16. Montoya, A., Odintsov, K., and Kotek, M. (2018), "Home Credit Default Risk". Kaggle. available at: <https://kaggle.com/competitions/home-credit-default-risk>

Надійшла 30.11.2023

Відомості про авторів / About the Authors

Чернишов Дмитро Владиславович – Харківський національний університет радіоелектроніки, бакалавр комп'ютерних наук, Харків, Україна; e-mail: dmytro.chernyshov@nure.ua; ORCID ID: <https://orcid.org/0009-0003-2773-7467>

Ситніков Дмитро Едуардович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри системотехніки, Харків, Україна; e-mail: dmytro.sytnikov@nure.ua; ORCID ID: <https://orcid.org/0000-0003-1240-7900>

Chernyshov Dmytro – Kharkiv National University of Radio Electronics, Bachelor of Computer Science, Kharkiv, Ukraine.

Sytnikov Dmytro – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Professor at the Department of System Engineering, Kharkiv, Ukraine.

BINARY CLASSIFICATION BASED ON A COMBINATION OF ROUGH SET THEORY AND DECISION TREES

The subject of the study is to improve the accuracy and efficiency of classification algorithms using decision trees by integrating the principles of Rough Set theory, a mathematical approach to approximating sets. **The aim of the study** is to develop a hybrid model that integrates rough set theory with decision tree algorithms, thereby solving the inherent limitations of these algorithms in dealing with uncertainty in data. This integration should significantly improve the accuracy and efficiency of binary classification based on decision trees, making them more robust to different inputs. Research **objectives** include a deep study of possible synergies between approximate set theory and decision tree algorithms. For this purpose, we are conducting a comprehensive study of the integration of approximate set theory within decision tree algorithms. This includes the development of a model that utilizes the principles and algebraic tools of approximate set theory to more efficiently select features in decision tree-based systems. The model uses the theory of approximate sets to efficiently handle uncertainty and weighting, which allows for improved and extended feature selection processes in decision tree systems. A series of experiments are conducted on different datasets to demonstrate the effectiveness and practicality of this approach. These datasets are chosen to represent a range of complexities and uncertainties, providing a thorough and rigorous evaluation of the model's capabilities. The **methodology** uses advanced algebraic tools of approximate set theory, including the formulation of algebraic expressions and the development of new rules and techniques, to simplify and improve the accuracy of data classification processes using decision tree systems. The findings of the study are important because they show that integrating approximate set theory into decision tree algorithms can indeed provide more accurate and efficient classification results. Such a hybrid model demonstrates significant advantages in dealing with data with embedded uncertainty, which is a common challenge in many complementary scenarios. The versatility and effectiveness of the integrated approach is demonstrated by its successful application in the areas of credit scoring and cybersecurity, which emphasizes its potential as a versatile tool in data mining and machine learning. The **conclusions** show that integrating approximate set theory can lead to more accurate and efficient classification results. By improving the ability of decision trees to account for uncertainty and imprecision in data, the research opens up new possibilities for robust and sophisticated data analysis and interpretation in a variety of industries, from healthcare to finance and beyond. The integration of approximate set theory and decision trees is an important step in the development of more advanced, efficient, and accurate classification tools in the era of big data.

Keywords: decision tree classification; approximate set theory; algebraic approach; machine learning.

Бібліографічні описи / Bibliographic descriptions

Чернишов Д. В., Ситніков Д. Е. Бінарна класифікація на основі поєднання теорії приблизних множин і дерев рішень. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 87–94. DOI: <https://doi.org/10.30837/ITSSI.2023.26.087>

Chernyshov, D., Sytnikov, D. (2023), "Binary classification based on a combination of rough set theory and decision trees", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 87–94. DOI: <https://doi.org/10.30837/ITSSI.2023.26.087>

V. SHAPOVALOV, O. STRYZHAK

TRANSDISCIPLINARY INTEGRATION OF KNOWLEDGE IN THE ENVIRONMENT OF A VIRTUAL STEM CENTER

The **subject** of the article is the role of ontological systems in improving the processes of structuring and analyzing scientific content, especially in the context of STEM education. The **goal** of the work is to research and analyze the application of ontological systems in the context of managing educational materials in STEM education. The use of such approaches is focused on developing effective methods for structuring and presenting educational knowledge in the STEM education system, emphasizing the importance of integrating different scientific disciplines to optimize the educational process. In accordance with the purpose, the following tasks were set: to develop a methodology for creating and implementing ontological systems in STEM education and to develop the architecture of a virtual STEM center that would provide the implementation of the proposed principles. The research is based on the following **methods**: For the development of the architecture of the virtual STEM center, we used UML diagrams. We developed UML diagrams of roles and activities that illustrate the interaction of different users and systems, as well as demonstrate work processes and interactions in multi-agent systems. Special attention is paid to activity diagrams, which reflect the processing of user requests and the interaction of the stemua.science agent with other components of the STEM center. The following **results** were obtained: A modular system architecture of the virtual STEM center was developed and described using UML diagrams, which includes roles such as the STEM center administrator, editor, author, and user, as well as the administrator of the CIT "Polyhedron". The interaction of these roles with the virtual STEM center is described in detail, revealing the mechanisms of their interaction and joint work aimed at creating, filling, and editing content in the transdisciplinary STEM center. The process of optimizing work processes in the modular system of the virtual STEM center is also considered. Ways to fill and use the T-STEM center in an ontological form have been identified. The interaction of software entities of the T-STEM center in an ontological form has been analyzed. **Conclusions**: Based on the conducted research, it is concluded that the use of ontological systems in the context of managing educational materials in STEM education is an effective method for structuring and presenting scientific content, promoting the integration of various scientific disciplines, and optimizing the learning process. It is determined that ontological systems are an effective method for structuring and presenting scientific content, facilitating the integration of various scientific disciplines, and optimizing the learning process. The modular architecture of the system is found to facilitate efficient interaction among different roles and automate workflow processes. Integration with a multi-agent system allows for the use of external data sources and ensures interoperability with other systems. For the further development of the system, research is needed to enhance the efficiency of role interactions and workflow automation. Additionally, research on integrating the system with other STEM education systems is necessary.

Keywords: ontological systems; STEM-education; content structuring; data analysis; scientific materials; knowledge; educational process.

Introduction

Ontological systems are critical for structuring and analyzing scientific and educational content, in particular in STEM education (an approach to education based on the application of science, technology, engineering, and mathematics). They allow us to identify and systematize key concepts and their interrelationships, which contributes to a deeper understanding of subject areas.

Generally speaking, an ontology can be defined as a certain form of comprehensive and detailed formalization of descriptions of a certain area of knowledge that reflect images of the world picture using a conceptual scheme. Typically, such a scheme consists of a hierarchical structure of concepts (taxonomy) containing all relevant classes of objects,

their relationships and rules accepted in this area of knowledge. A conceptual scheme means a set of concepts and information about concepts (properties, relations, constraints, axioms and rules for using concepts that are necessary to describe problem-solving processes). The relationships between these concepts and the axioms that define the basic rules and principles are fundamental to creating a coherent ontological structure [1, 2]. In fact, an ontology is a conceptual and conceptual framework of all scientific theories without exception.

Ontologies, as a formal description of concepts and their relationships, are crucial for the exchange and reuse of knowledge, especially in education [3]. The creation of an ontological structure involves the use of axioms, which are basic rules and principles, to define these relationships [4].

This process can be facilitated by the systematic method, which involves the creation of ontologies in conceptual and logical-linguistic formats, where the former is intended for human understanding and the latter for machine processing [5].

Analysis of the problem and existing methods

Narratives and their ontological mapping STEM narratives, which include descriptions of various processes, facts, and their properties, can be represented through ontological models. This allows for greater semantic clarity and structural organization of scientific and educational texts [2]. Ontological models have been proposed as a means of structuring and representing scientific and educational information, which provides greater semantic clarity and organization [6]. These models can be used to develop deeply structured educational tasks, methods, textbooks, laboratory works, etc. They structurally reflect the semantics of academic disciplines, which is integrated in a transdisciplinary manner into interconnected content knowledge bases [7]. The problem of annotating scientific and educational texts with semantic information has been solved by using a modular ontology that allows building a reduced structure useful for automated annotation [8]. In addition, the development of ontological models for representing information accumulated by scientific and educational institutions has been investigated in order to facilitate the evaluation of their effectiveness [9].

In this context of scientific and educational activities, ontology can be considered as a certain explicit conceptualization of a logical theory, a certain calculus with certain rules. This theory allows to systematize the categories of reality as those expressed in the language of meanings of certain statements and utterances [10, 11], and which are in the content of narrative descriptions. Particular attention is paid to the presentation of their display formats. In the process of presenting a narrative in an ontological format, the contexts of all its concepts (ontology concepts) are hierarchically interconnected and distinguished by a strict order relationship. This determines the conditions for using the concepts of STEM disciplines in the formulation of educational and research tasks.

Hierarchically interconnected concepts of educational narratives form the basis of its taxonomic representation. Taxonomy in scientific and educational ontologies plays a key role, allowing to classify and systematize concepts according to their properties and

interrelationships of their contexts. This contributes to a better understanding of the hierarchical relationships between different elements of knowledge [2]. Taxonomy as a structural platform for the ontology of scientific and educational narratives is crucial for the classification and systematization of concepts in STEM disciplines, as it helps to understand the hierarchical relationships between different elements of knowledge [16]. This is especially important in the semantic analysis of narratives of scientific and technical subject areas that make up the subject of STEM education, where the taxonomy of relations can be used to identify semantic relationships and automatically build ontologies [16]. The use of formal ontology tools such as identity, entity, unity, and dependency can further enhance the properties of the taxonomy and improve human understanding [17]. In the context of web documents, the structure of narratives can be used to highlight relevant concepts and establish hierarchical relationships between them, contributing to the creation of a taxonomy [18].

The axiomatic principle of forming STEM narrative ontologies provides a scientific foundation for teaching STEM disciplines, based on the basic rules and conditions on which their scientific concepts and their interconnectedness are defined [2]. Axioms and the basic principles of their use are crucial in the development of ontological systems in STEM education, as they provide the basis for scientific concepts and their interaction [19]. These principles guide the development of ontologies, ensuring that they are reusable, application-independent, and easy to maintain [19]. They also play a role in the selection of concepts and hypotheses, the reconstruction of scientific theories, and the evaluation of scientific and educational outcomes of STEM users [20]. Top-level ontologies that describe domain-independent categories of reality are particularly important in this regard [21]. The development of ontological theories based on established work in various fields further contributes to the creation of fundamental ontologies [20].

As a tool for displaying taxonomies, oriented graphs without cycles are most often used, which is formed by the concepts of educational narratives that are hierarchically interconnected. This allows us to consider the contexts of taxonomy nodes as components of certain knowledge, and to display their entirety as a fragment of the world picture. However, when considering the whole variety of taxonomies, the researcher encounters the phenomenon of the hyper-property of reflection, which is expressed in the fact

that all taxonomies can be reflected on themselves. This reflexive reflection of a taxonomy on itself is verbal in nature. This can be deduced from the fact that the nodes of any taxonomy form specific statements that may have truth values [22].

Interactivity and consolidation in ontological systems. Ontological systems themselves are characterized by a high level of interactivity and the ability to consolidate with other information resources. This makes it possible to create more flexible and adaptive models for analyzing and studying scientific data. To do this, some researchers use logical-linguistic models and their components – concepts of a certain sequence of finite-length symbols (SFLS concepts), which form a certain set of names for marking the nodes of a taxonomic graph. The taxonomic graph itself is represented in the format of growing pyramidal networks (GPNs) [15].

This allows us to consider taxonomies as type-free constructions in the format of lambda expressions [23]. To prove the truth of taxonomies as a technological platform for ontologies of arbitrary narratives, the uniqueness of the SPNs to Bem trees is usually considered [23–25]. A constructive characteristic of the type-lessness of Bem trees and their uniqueness of the GPN is the formation on their basis of technological conditions for the implementation of taxonomies in order to ensure the implementation of scientific and educational narratives of different topics in the STEM environment and to identify the intercontextual coherence of the concepts that determine their content [1, 2, 15, 23–25].

The purpose of this article is to study and analyze the use of ontological systems in the context of learning materials management in STEM education. The use of such approaches is focused on the development of effective methods for structuring and presenting educational knowledge in the STEM educational system, emphasizing the importance of integrating various scientific disciplines to optimize the learning process. In accordance with the goal, the following **tasks** have been set: to develop a methodology for creating and implementing ontological systems in STEM education and to develop the architecture of a virtual STEM center that would ensure the implementation of the proposed principles.

Solving the problem. Materials and methods

In the course of the study, we developed the architecture of a virtual STEM center, which was used

later to develop a virtual STEM center template (stemua.science) and its other components. In this study, we developed an architecture and described it using UML (Unified Modeling Language) diagrams to describe the architectural and software components of a virtual STEM center for transdisciplinary knowledge integration in a STEM learning management system. The described UML diagrams perform different functions in the context of the article. UML role diagrams are used to illustrate and analyze the different roles or actors that interact with a system. They help to understand how different users or systems interact with the components being developed. A UML diagram of the agent interaction activity of a multiagent system shows the interaction between different agents in a multiagent system. It details the processes and relationships that occur between different agents and their impact on the overall functionality of the system. A UML process activity diagram is used to describe the flow of workflows or operations that occur in a system. It visualizes the steps of an algorithm or the interaction between system components.

The UML process activity diagram of the process of issuing user request results demonstrates how the virtual STEM center system processes user requests and returns results. It is useful for explaining the mechanisms of user interaction with the system and receiving answers. Interaction diagram of the stemua.science agent: shows how the stemua.science agent interacts with other components or agents of the virtual STEM center. It is important to reflect the role of this agent in the overall architecture and its impact on workflows.

Results

Transdisciplinary integration of management roles in the modular architecture of a virtual STEM center

Moving on from the conceptual description of ontological systems and their importance for structuring scientific content, especially in the context of STEM education. We will consider the architecture of a virtual STEM center, which provides for specific roles, such as STEM center administrator, editor, author and user, as well as the administrator of the CIT (Cognitive Information Technology) "Polyhedron". The discussion focuses on the interaction of these roles with the system, revealing the mechanisms of their interaction and collaboration aimed at creating, filling and editing content in a virtual STEM center, which is fundamental to ensuring an effective information flow and development of the educational process.

The architecture of the virtual STEM center includes roles such as STEM center administrator, editor, author, user, and administrator of the Polyhedron CIT. The architecture provides for the interaction of these roles in order to create a single field for them to fill and edit the educational resources of the virtual STEM center. It is advisable to consider the capabilities of each user and model their behavior in its environment.

The STEM center administrator determines its basic settings and provides access to users, determining their additional capabilities. The virtual STEM center settings include upgrading resources and ensuring data structuring in the system. The STEM center is based on the API (Application Programming Interface) client server architecture, which is operated by the STEM center administrator.

The main material is prepared by dispersed authors who have access to the virtual STEM center using their own automated workstations (offices), which provide the appropriate functionality that imitates the traditional functionality of text editors familiar to authors. It is assumed that teachers of schools and institutes, as well as methodologists, will play the role of authors. The author has the functionality to create and edit materials.

In order to ensure that the STEM Center template (stemua.science) provides high quality content, the role

of a publication editor is envisaged. The editor checks the quality and style of writing, completeness of sentences, and grammar of the texts uploaded to the STEM Center. Accordingly, the role of the administrator includes the functionality of validating and publishing materials previously submitted by authors for publication.

After the data is entered into the virtual STEM center, it is structured in accordance with the standard provided by the ARI exchange protocol and transferred to the STEM center based on the Polyhedron CIT, which provides for the automated generation of structured data sets in the form of ontological graphs that can provide transdisciplinary educational materials. The administrator of the Polyhedron CIT ensures that data is received according to a certain standard and that ontologies are generated automatically.

After completing the entire cycle, STEM Center users can use the content of the STEM Center. Users have the functionality of working with a search engine, using the functions of selecting and filtering materials, and further viewing educational information.

The described process is a complete cycle of creating and using information based on a virtual STEM center. Detailed scheme of roles of a virtual STEM center (Fig. 1).

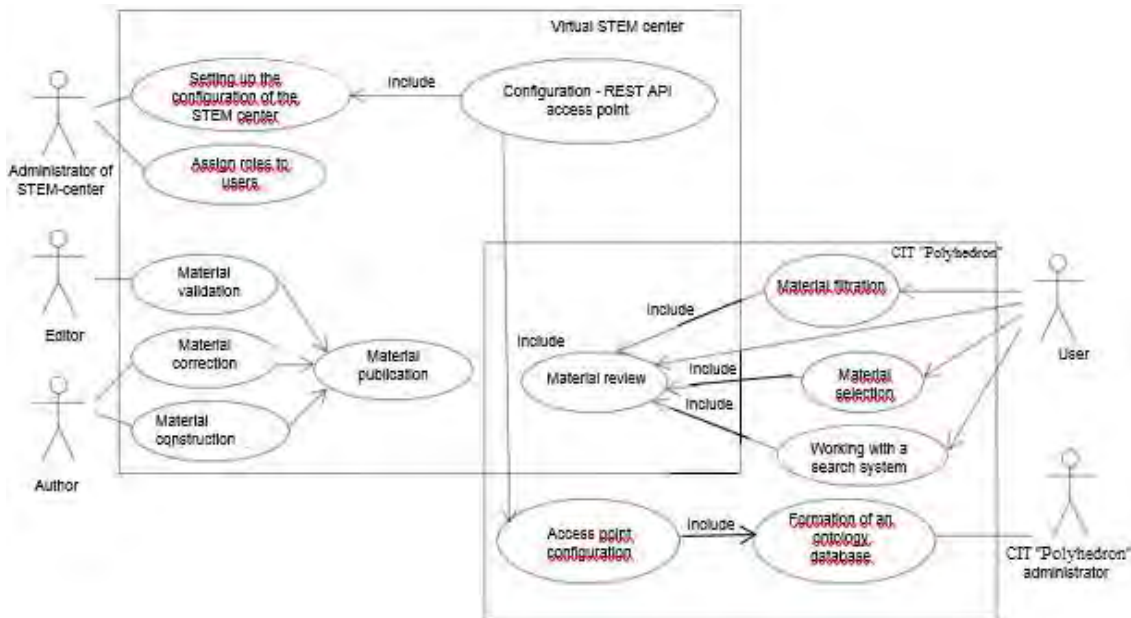


Fig. 1. Ways to use the virtual STEM center

Modeling user functional processes in a virtual STEM center using UML diagrams

To display behavioral models for different users performing different functional tasks in the virtual STEM

center, the processes are described using UML diagrams of the interaction activity of the agent of the multiagent system of the virtual STEM center and its subsequent automated transformation into a structured information

resource of the virtual STEM center (Fig. 2), the process of issuing the results of the request of users of the virtual

STEM center (Fig. 2), the process of working with the resource of the virtual STEM center (Fig. 4).

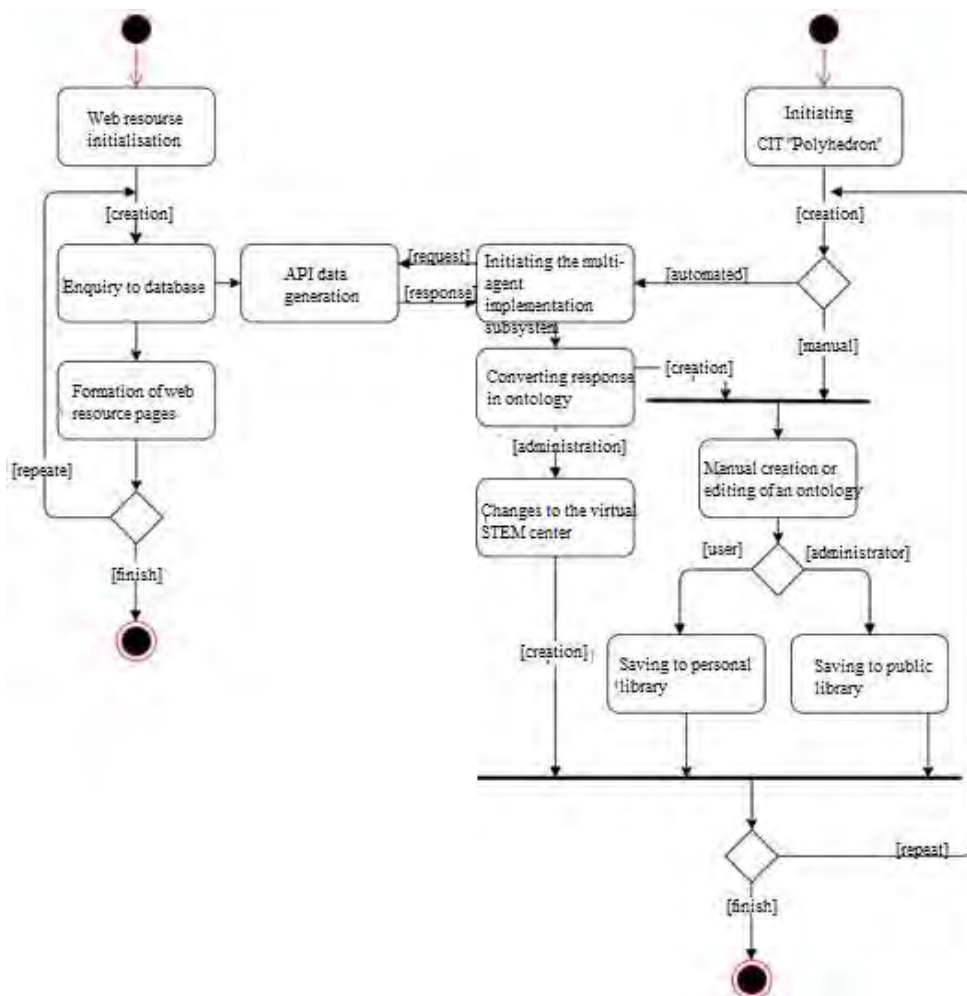


Fig. 2. UML diagram of the activity of interaction of the agent of the multi-agent system virtual STEM center and its further automated transformation into a structured information resource of the virtual STEM center

As can be seen from Fig. 2, to ensure the interaction of the agent of the multi-agent system of the virtual STEM center with the STEM center itself, a process is provided on the software side of the stemua. science agent, which consists of initializing a web resource, querying the database, depending on the type of query (determined by the user) to the database, the process can take place in two milestone flows – the formation of the page and the subsequent process of completing the work or creating a new resource (step querying the database); or the process of interaction between the stemua.science agent and the virtual one is initiated.

To simplify, systematize, and structure the STEM center data, the virtual STEM center uses an algorithmic mechanism for issuing the results of the virtual STEM center users' queries (Fig. 3), which involves the stage of initialization of the web resource and further milestone

variability, which involves the display of all materials in a list, filtering by subject or filtering by type of material at the choice of the virtual STEM center user; based on the user's request, a list of virtual STEM center objects is generated that matches the user's request. After that, the process can be repeated or completed.

Optimizing workflows in the modular system of a virtual STEM center

The process of working in a virtual STEM center is initiated by activating the Polyhedron information technology complex (ITC), where users review existing ontologies created in advance. Automated data processing is realized with the help of integrated modules of the Polyhedron ITC, which include Synopsis, Confor, Editor, Alternative, Search Engine, Linguistic Corpus, and Crypto.

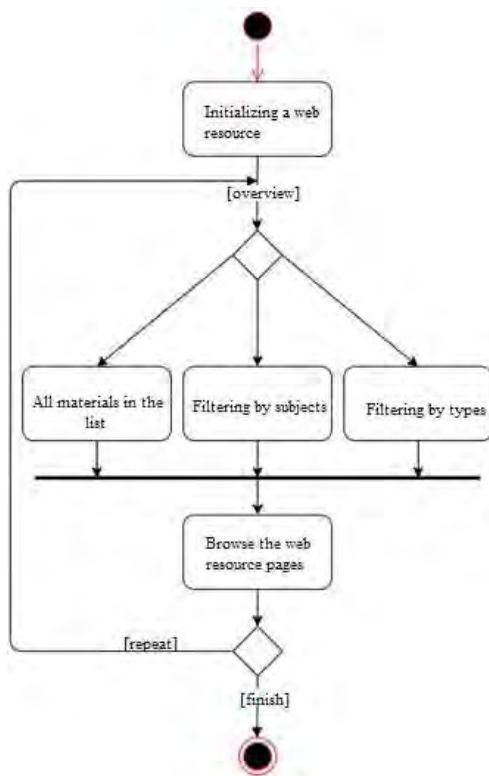


Fig. 3. UML diagram of the activity of the process of issuing the results of the request of users of the virtual STEM center

The data in the vertices of taxonomic graphs are processed in the Alternative module, which allows processing large data sets and selecting vertices with relevant semantic data and ranking ontological vertices by integral characteristics according to the user's request.

STEM-oriented systems are transdisciplinary and contain ordered disparate data sets that can be processed (filtered) according to the user's request to select information by the requested authors, attributes, subjects, and types, or to rank according to the user's request.

After processing by the Alternative module, sets of objects are formed and the user is given the opportunity to view them in a transdisciplinary view. Such processed data is subject to contextual linking of information resources and can be presented in an object representation, tabular representation, graphical representation, in the form of an ontological prism or in an ontology in the form of web resources. The user of the virtual STEM center can complete the work with the virtual STEM center or go through the algorithmic actions starting from the milestone of choosing options for forming sets of objects again. The UML diagram of the activity of the process of issuing the results of the request of users of the virtual STEM center is shown in Fig. 4.

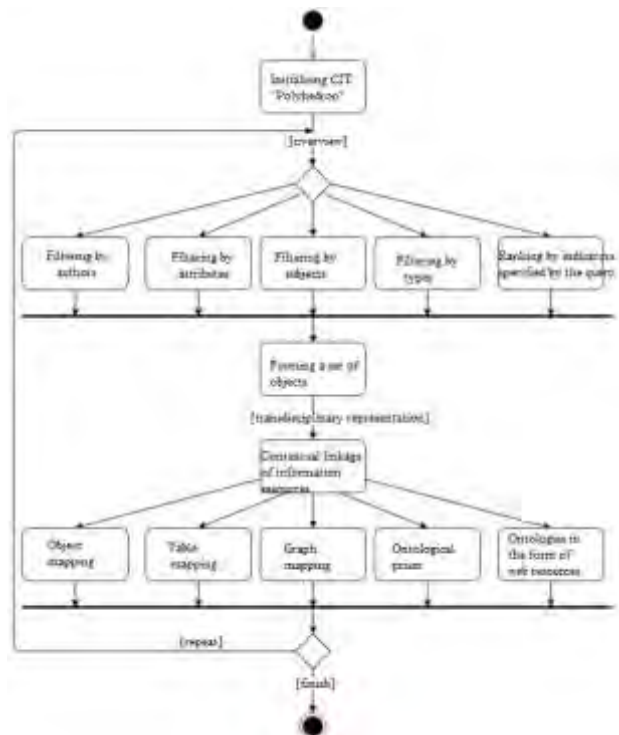


Fig. 4. UML diagram of the activity of the process of issuing the results of the request of users of the virtual STEM center

As can be seen from Fig. 5, there are several ways to fill and use the T-STEM center in ontological form. An expert using the program module "Outline" can generate basic information to fill the T-STEM center. With the help of a recursive reducer, such basic information is processed and arrays of methods and research papers are formed. The methods and research works formed are read by queries using the Polyhedron and Alternative APIs, which in turn form a data array for the ontological configuration analysis module. The ontological configuration analysis module generates content that can be displayed and processed in the ontological workbench and further processed in the testing module. The user of the testing module can use the material data to select them according to the query criteria.

In addition, the ontological configuration analysis module can receive information from the editor through ontological configurations. The information processed by the ontological configuration analysis module can be used for viewing in the viewer and, in particular, in the form of an ontological prism. The ontological data through the viewer can be processed in the Alternative software, which can also be accessed directly through the Alternative API.

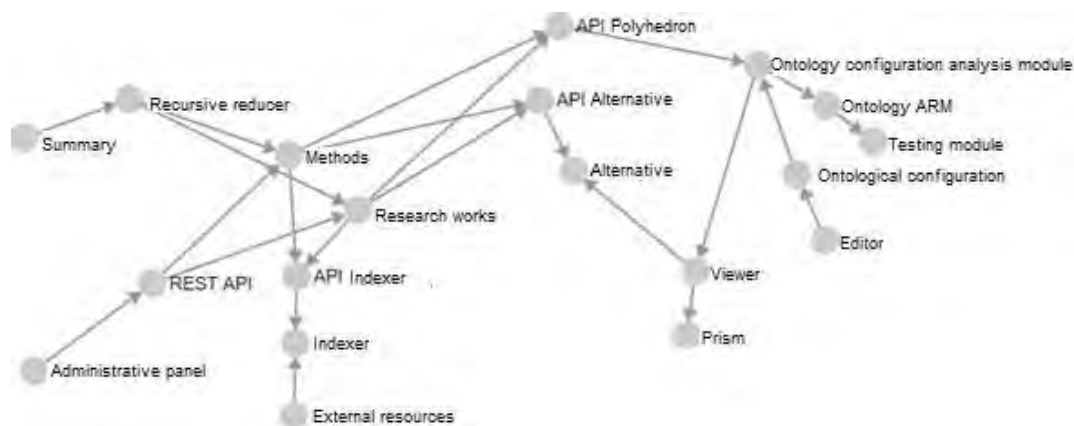


Fig. 5. General representation of the interaction of program entities of the T-STEM center in ontological form

Another way of filling is to use the administrative panel of the STEMUA templating tool and further use the REST API. The ontological entities of the T-STEM center are generated by analyzing natural language texts with the "Synopsis" module, processed by the "Indexer" module and using the REST API by converting data from the MySQL database into the XML format of ontologies.

In addition, the ontological T-STEM center can use data from external sources. To do this, the material from external resources is indexed by an indexer and an array of data on methods and research works is formed using the Indexer API.

The formation of data arrays is based on previous studies and provides for additional automation of a number of processes and simplification of data entry and formation of taxonomic arrays with data in the nodes. To process the input data and create relevant information output, the ITC "Polyhedron" software modules are used. For this purpose, an architectural solution has been developed, which consists in using the data of agents of a multiagent system. Let us consider the architectural solution on a practical example of using the external agent stemua.science and other external agents as sources of ontological data.

The formation of a display for the end user is created using the Polyhedron API or, if necessary, the Alternative API. The Polyhedron API transmits data to the ontological configuration analysis module, which, based on the analysis of the received data, loads the Ontology Viewer or the Ontology ARM.

The author of the publication creates content and enters it into the virtual STEM center, where the publication editor automatically generates, formats and processes the content. If the virtual STEM center has successfully processed the entered data, it generates a report on the content's compliance and notifies the

author. In response, the author generates a confirmation request for the content to be published. The publication editor transfers the processed data to the database, where the exchange process is initialized using data exchange protocols. The exchange process includes saving and structuring and receiving a response about the successful completion of the process. The database generates a request for data exchange from the multi-agent system to the controller for initializing the resource interoperability point. A data tracer passes between the database and the resource interoperability point initialization controller.

Using the data obtained from the stemua.science agent database, the resource interoperability point initialization controller provides the generation of a data array in the standard defined by the protocol, where the data is actually transferred, which involves the formation of a request from the ontological database creation controller and a response from the agent. After that, an exchange XML file is generated and the data is transferred to the multi-agent system of the CIT "Polyhedron".

The XML file is automatically transferred to the CIT "Polyhedron" administrator's office, where the administrator can perform a number of actions, including making a decision on saving. Further, the architecture of the virtual STEM center provides for the process that was previously proposed, which consists in recording in an ontological database. From the ontological database, it becomes possible to process the data with the tools of the CIT "Polyhedron", in particular, to receive a request from the user for interoperable interaction and calculation using semantic and integral characteristics. The processing takes place in the "Alternative" module.

In accordance with the proposed architecture, the user has the option of using the resource in the agent's database, using a taxonomized resource located in

the CIT "Polyhedron", or using data processed by the Alternative module. The diagram of interaction between

the stemua.science agent and the CIT "Polyhedron" is shown in Fig. 6.

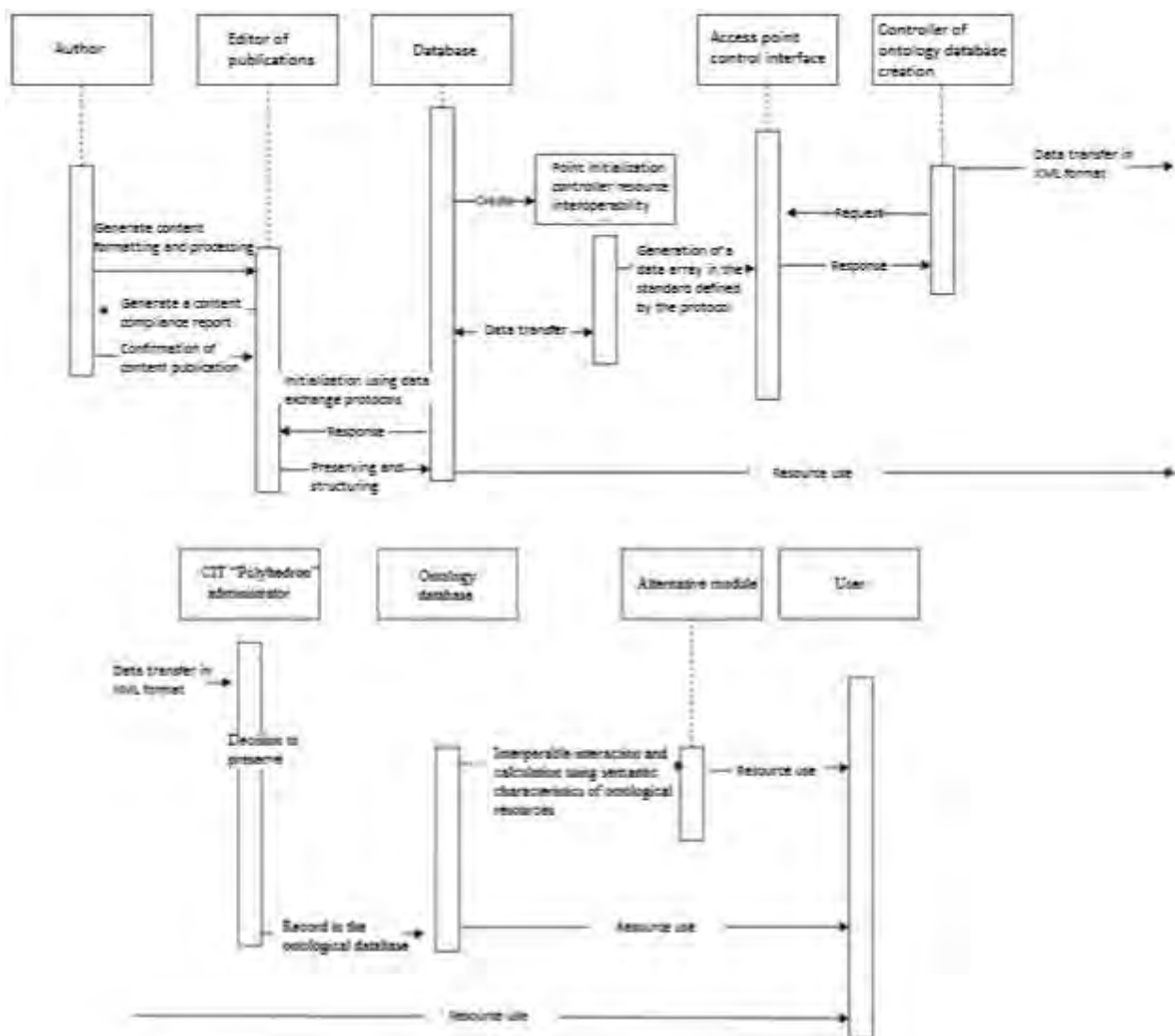


Fig. 6. Diagram of interaction between the stemua.science agent and the CIT "Polyhedron"

Structure and Functionality of Classes in the Modular System of the Virtual STEM Center

To ensure the integral functioning of the virtual STEM center, the architecture provides for the use of classes that can be divided into controller classes, data classes, and processing classes (Fig. 7).

The STEM center controller provides the use of agents of a multi-agent environment by generating request-responses from disparate web environments (for example, those using wordpress).

To generate structured information, the models of ontograph, node, property, alternative, criterion, and dessiontask were developed earlier.

The Ontograph class is responsible for displaying an ontological graph consisting of a set of graph nodes and allows you to access the parent and child nodes and elements of the selected node. Ontograph provides quick access to each node by its name or unique identifier. It uses xml/json to load and save the ontological taxonomy.

The Node class is used to generate information about the point itself and its attachments. This class ensures the correct storage and use of information about the name, properties, related nodes and vertices.

The Property class is used to store and display the properties of each vertex. It also contains tools for managing the properties of a node, including defining the property class, adding, editing, and deleting.

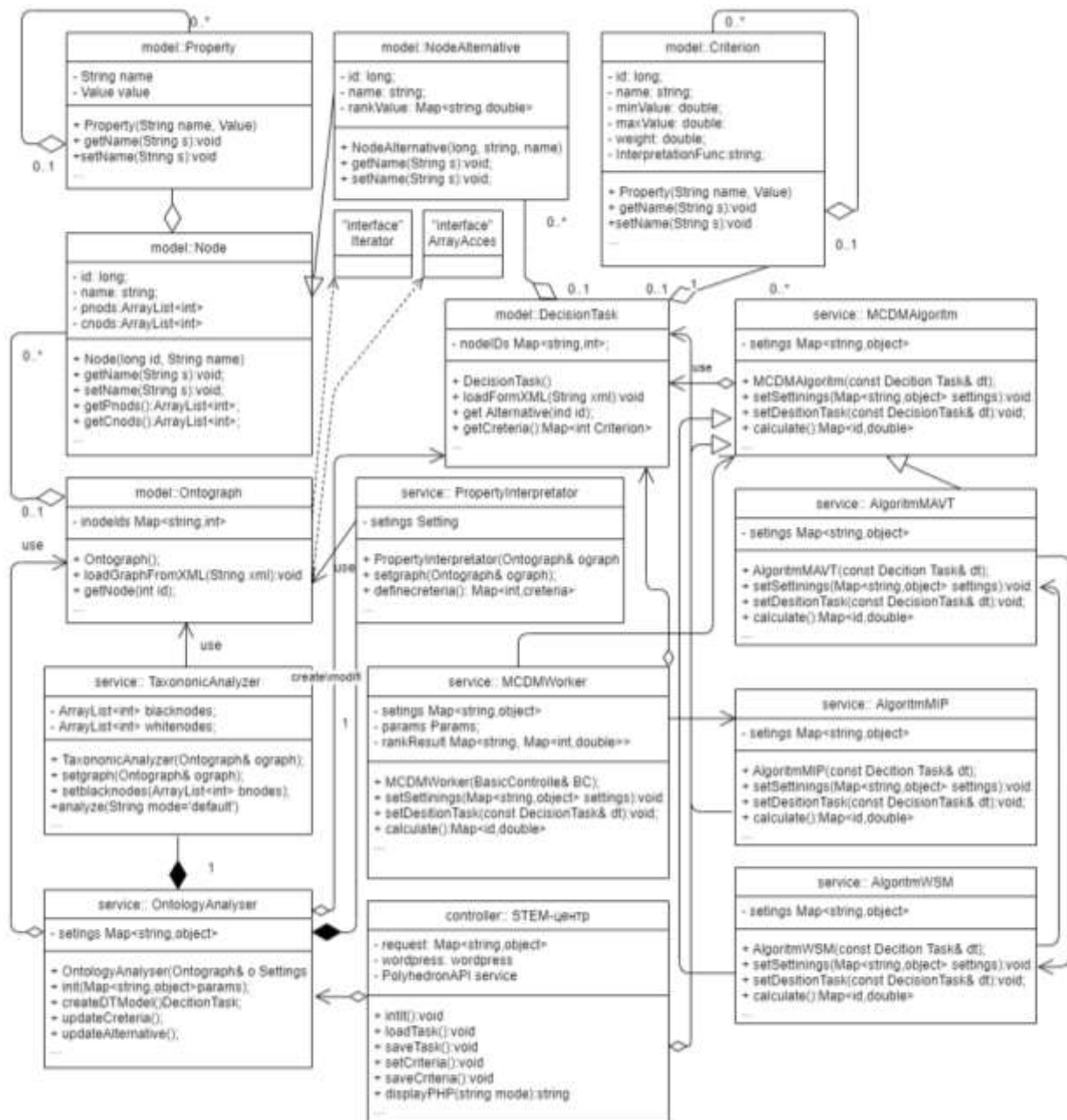


Fig. 7. Diagram of T-STEM center classes

The DecisionTask contains information about the ranking model, alternatives, and criteria, and allows you to use nodes for ranking tasks. It allows you to create weighting factors.

The NodeAlternative class is used to solve problems of alternatives. It is derived from the Node class and is compatible with Ontograph. It contains algorithms for processing properties and attributes.

The Criterion class is used to solve optimization problems and uses information about the properties of a point and forms criteria from them, including the name, range of possible values, and optimization vector.

Services are an important component of a virtual STEM center that use models to generate structured data sets from unstructured ones, in particular, agents of a multi-agent STEM center. These services include Ontology Analyzer, Ontology Analyzer, Taxonomic Analyzer, and Property Interpreter.

OntologyAnalyser generates ontologies according to the models. This class allows you to use ontological nodes and taxonomic structure to form a task of ranking alternatives based on an ontological model. The OntologyAnalyser analyzes the ontology and extracts information from it and generates data sets that are

further processed. It works in a complex and includes the TaxonomicAnalyzer and PropertyInterpreter classes.

The TaxonomicAnalyzer class analyzes the taxonomy and separates the nodes that contain relevant data and can be used as alternatives from those that do not contain relevant data and cannot be used as alternatives. It provides labeling of such vertices and passes the task formation for filtering vertices and alternatives.

The PropertyInterpreter class provides the formation of an array of input data of all vertices that can be used to process alternatives. It also provides the definition of common properties of alternative objects that form arrays of criteria for ranking and filtering.

Since the data in the virtual STEM center are formed by significant arrays of heterogeneous data, it is important to process them using the alternative tool for ease of use. Therefore, additional ranking classes are used, such as the solution of the formed problem of ranking alternatives (MCDMWorker), the class of the algorithm for solving the problem of ranking alternatives (MCDMAlgorithm), the class of the method of weighted sums of criteria (AlgorithmWSM), the class of using the ideal point method (AlgorithmMIP), the class of using the method of multi-attribute evaluation (AlgorithmMAVT).

The MCDMWorker class creates a problem of ranking alternatives to be solved based on its model, which is the mathematical core of the virtual STEM center. That is, this class aggregates the results obtained and forms the final result.

The abstract class MCDMAlgorithm provides data preparation for other classes. It uses methods for calculating criterion values by interpreting the criteria, ensures the generation of a single database by reducing the data of specific vertices (criterion values) to a common numerical scale. It also uses algorithms for using weighting coefficients for each criterion.

After the data and criteria required for processing are formed, specialized classes are used to perform calculations using various methods. For example, there are classes for solving the problem of ranking alternatives using the weighted sums of criteria method (WSM; AlgorithmWSM), using the method of ideal point (MIP; AlgorithmMIP), using the method of multi-attribute evaluation (MAVT; AlgorithmMAVT).

The main entities of a virtual STEM center are Author, Editor, Administrator, User, and Material. Each of these entities contains data that is necessary for the processes described in the behavior diagrams

of the virtual STEM center. The virtual STEM center contains information about the unique administrator identifier (AdminID), his/her name (Name) and permissions (Prirogatives) related to the administrator entity; the identifier (AuthorID), name (Name), permissions (Prerogatives) and materials (Materials) related to the author; EditorID, Name, Prerogatives, and ValidatedMaterials related to the editor; MaterialID, an array of properties, URL, TextData, Abstract, Published, and TimeStamp related to the materials. A diagram of the relationship between the T-STEM Center entities is shown in Fig. 8.

Conclusions.

Prospects for further development

This article addresses the urgent problem of systematizing large volumes of unstructured text in a global environment. The focus is on developing a model, methods, and tools for transdisciplinary representation and integration of this data. This research is important in several ways: The paper demonstrates the use of recursive reduction and other modern techniques to improve the efficiency of information retrieval, processing, classification, and selection. Implementation The transdisciplinary integration that promotes the creation of a modular system in STEM centers contributes to the development of innovative teaching and research methods. Implemented interactive modules and tools, such as "Synopsis", "Confor", "Editor", significantly improve user interaction with the content of the STEM environment. The recursive reduction method with a number of transformations was used to achieve the predicative form of ontologically specified descriptions, which facilitates deeper analysis and understanding of STEM educational content.

Using UML diagrams, we visualized the functional processes and interactions between different users and components of the virtual STEM center, providing a deeper understanding of its structure and functionality. The proposed virtual STEM center provides high flexibility and scalability, which allows it to adapt to the changing needs of users and expand as needed. The use of the category of transdisciplinarity helps to integrate various educational disciplines, facilitating the processes of synchronization of STEM education disciplines and intellectual exchange and cooperation in teaching the basics of science.

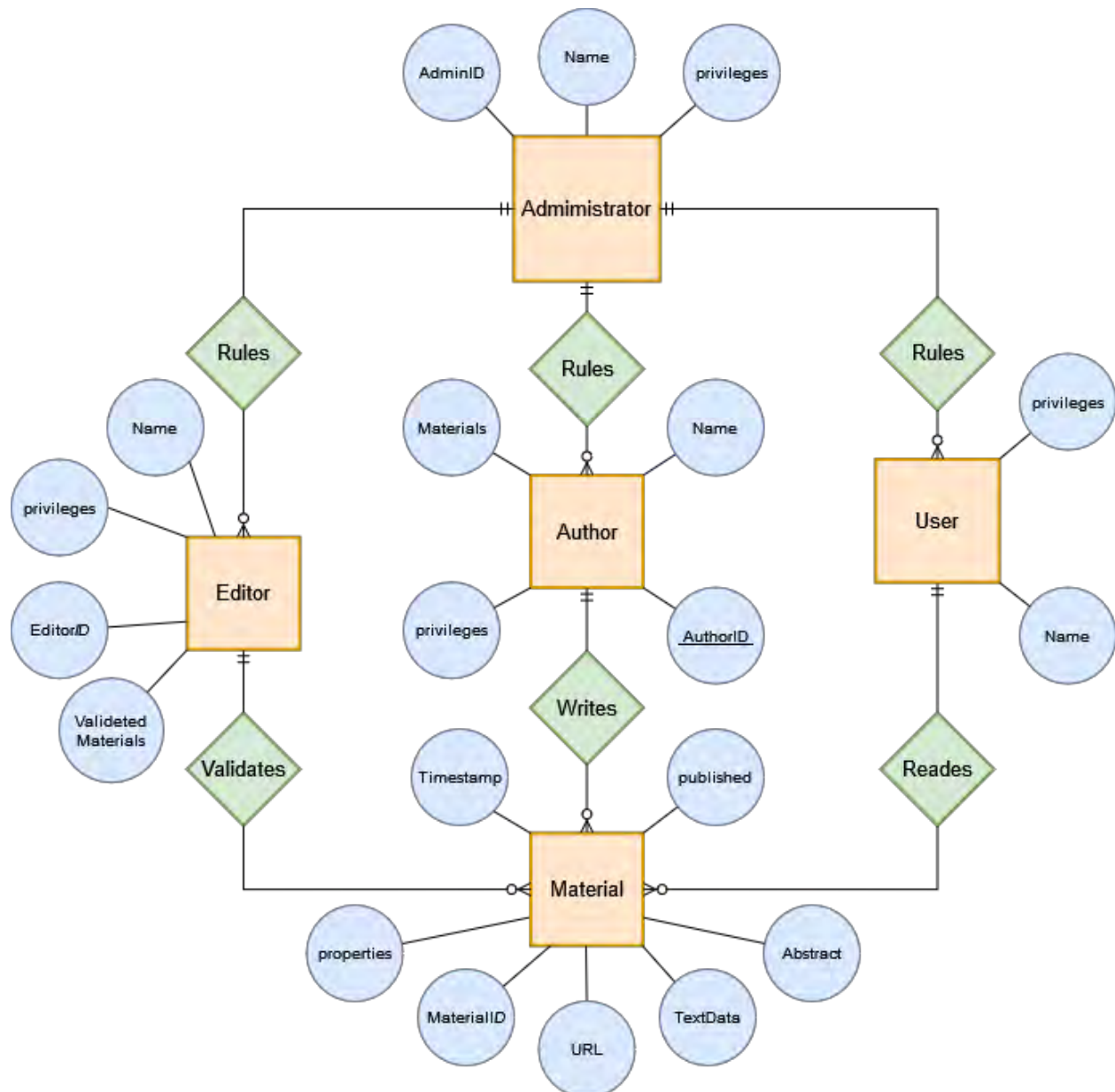


Fig. 8. Diagram of relationships between T-STEM center entities

In general, the study provides a significant contribution to the development of scientific approaches to managing large amounts of information, especially in the context of transdisciplinarity of educational narratives, becoming an important element in the field of STEM education and research and education activities.

Prospects for further development of the research presented in this article promise to be large-scale and diverse, as they cover a number of key areas of use and improvement of scientific and educational disciplines. The main directions for further development can be outlined as follows:

Given the success of designing ontology-driven transdisciplinary representation of educational

narratives as large arrays of unstructured information, the possibility of expanding their use in teaching disciplines is being considered. This may include public administration, industry, science, education, media, and other areas where effective information processing and structuring is key. The structured representation of information implemented in the virtual STEM center environment creates the basis for the implementation of decision support systems in the course of performing educational and research tasks. This can significantly increase the efficiency of the scientific and educational process.

The use of the developed virtual STEM center in education and science opens up new opportunities

Відомості про авторів / About the Authors

Шаповалов Віктор Борисович – Національний центр "Мала академія наук України", відділ створення навчально-тематичних систем знань, старший науковий співробітник, Київ, Україна; e-mail: svb@man.gov.ua; ORCID ID: <https://orcid.org/0000-0001-6315-649X>

Стрижак Олександр Євгенійович – доктор технічних наук, професор, Національний центр "Мала академія наук України", заступник директора з наукової роботи, Київ, Україна; e-mail: sac953@gmail.com; ORCID ID: <https://orcid.org/0000-0002-4954-3650>

Shapovalov Viktor – National Center "Junior Academy of Sciences of Ukraine", senior researcher, Kyiv, Ukraine.

Stryzhak Oleksandr – Doctor of Sciences (Engineering), Professor, Deputy Director for Scientific Work at the National Center "Junior Academy of Sciences of Ukraine", Kyiv, Ukraine.

ТРАНСДИСЦИПЛІНАРНА ІНТЕГРАЦІЯ ЗНАТЬ У СЕРЕДОВИЩІ ВІРТУАЛЬНОГО STEM-ЦЕНТРУ

Предмет дослідження – роль онтологічних систем у вдосконаленні процесів структурування та аналізу наукового контенту, особливо в контексті STEM-освіти. **Мета** роботи полягає у дослідженні та аналізі застосування онтологічних систем у контексті управління навчальними матеріалами в STEM-освіті. Використання таких підходів зосереджене на розробленні ефективних методів структурування та подання навчальних знань в освітній системі STEM, з огляду на важливість інтеграції різних наукових дисциплін для оптимізації навчального процесу. Відповідно до мети визначено такі **завдання**: розробити методологію створення та впровадження онтологічних систем у STEM-освіту та розробити архітектуру віртуального STEM-центру як платформи трансдисциплінарної інтеграції систем знань (T-STEM), що впроваджуються в STEM-освіті. Проведені дослідження ґрунтуються на таких **методах**: для розроблення архітектури віртуального STEM-центру застосовано UML-діаграми; створено UML-діаграми ролей та активностей, що ілюструють взаємодію різних користувачів і систем, а також демонструють робочі процеси та взаємодії у мультиагентних системах. Особлива увага приділяється діаграмам активностей, що відтворюють оброблення запитів користувачів і взаємодію агента *stemua.science* з іншими компонентами STEM-центру. Досягнуто таких **результатів**: розроблено та описано за допомогою UML-діаграм архітектуру модульної системи віртуального STEM-центру, що передбачає такі ролі, як адміністратор STEM-центру, редактор, автор і користувач, а також адміністратор КІТ "Поліедр". Детально описано взаємодію цих ролей з віртуальним STEM-центром та водночас розкрито механізми їх взаємодії та спільної роботи, спрямовані на створення, наповнення та редагування контенту в середовищі віртуального STEM-центру. Також розглянуто процес оптимізації робочих процесів у модульній системі віртуального STEM-центру. Визначено шляхи наповнення та використання T-STEM центру в онтологічному форматі. Проаналізовано взаємодію програмних сутностей T-STEM центру в онтологічному вигляді. **Висновки**. На основі проведеного дослідження зроблено висновок, що використання онтологічних систем у контексті управління навчальними матеріалами в STEM-освіті є ефективним методом структурування та подання наукового контенту, що сприяє інтеграції різних наукових дисциплін та оптимізації навчального процесу. Визначено, що онтологічні системи є ефективним методом структурування та подання наукового контенту, що сприяє інтеграції різних наукових дисциплін та оптимізації навчального процесу. Установлено, що модульна архітектура віртуального STEM-центру забезпечує ефективну взаємодію різних ролей та автоматизацію робочих процесів. Інтеграція з мультиагентною системою дає змогу використовувати зовнішні джерела інформації та забезпечує інтероперабельність з іншими системами. Для подальшого розвитку віртуального STEM-центру необхідно дослідити підвищення ефективності взаємодії різних ролей та автоматизації робочих процесів. Також планується вивчити інтеграцію віртуального STEM-центру з іншими системами STEM-освіти.

Ключові слова: онтологічна системи; STEM-освіта; структурування контенту; аналіз даних; наукові матеріали; знання; освітній процес.

Бібліографічні описи / Bibliographic descriptions

Шаповалов В. Б., Стрижак О. Е. Трансдисциплінарна інтеграція знань у середовищі віртуального STEM-центру. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 95–107. DOI: <https://doi.org/10.30837/ITSSI.2023.26.095>

Shapovalov, V., Stryzhak, O. (2023), "Transdisciplinary integration of knowledge in the environment of a virtual stem center", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 95–107. DOI: <https://doi.org/10.30837/ITSSI.2023.26.095>

Y. SHOVKOVYI, O. GRINYOVA, S. UDOVENKO, L. CHALA

AUTOMATIC SIGN LANGUAGE TRANSLATION SYSTEM USING NEURAL NETWORK TECHNOLOGIES AND 3D ANIMATION

Implementation of automatic sign language translation software in the process of social inclusion of people with hearing impairment is an important task. Social inclusion for people with hearing disabilities is an acute problem that must be solved in the context of the development of IT technologies and legislative initiatives that ensure the rights of people with disabilities and their equal opportunities. This substantiates the relevance of the research of assistive technologies, in the context of software tools, such as the process of social inclusion of people with severe hearing impairment in society. **The subject of research** is methods of automated sign language translation using intelligent technologies. **The purpose of the work** is the development and research of sign language automation methods to improve the quality of life of people with hearing impairments in accordance with the "Goals of Sustainable Development of Ukraine" (in the "Reduction of Inequality" part). **The main tasks** of the research are the development and testing of methods of converting sign language into text, converting text into sign language, as well as automating translation from one sign language to another sign language using modern intelligent technologies. Neural network modeling and 3D animation methods were used to solve these problems. The following **results** were obtained in the work: the main problems and tasks of social inclusion for people with hearing impairments were identified; a comparative analysis of modern methods and software platforms of automatic sign language translation was carried out; a system combining the SL-to-Text method is proposed and investigated; the Text-to-SL method using 3D animation to generate sign language concepts; the method of generating a 3D-animated gesture from video recordings; method of implementing the Sign Language1 to Sign Language2 technology. For gesture recognition, a convolutional neural network model is used, which is trained using imported and system-generated datasets of video gestures. The trained model has a high recognition accuracy (98.52%). The creation of a 3D model for displaying the gesture on the screen and its processing took place in the Unity 3D environment. The structure of the project, executive and auxiliary files used to build 3D animation for the generation of sign language concepts includes: event handler files; display results according to which they carry information about the position of the tracked points of the body; files that store the characteristics of materials that have been added to certain body mapping points. **Conclusions:** the proposed methods of automated translation have practical significance, which is confirmed by the demo versions of the software applications "Sign Language to Text" and "Text to Sign Language". A promising direction for continuing research on the topic of the work is the improvement of SL1-to-SL2 methods, the creation of open datasets of video gestures, the joining of scientists and developers to fill dictionaries with concepts of various sign languages.

Keywords: automation of sign speech; animated character; body position tracking; people with hearing impairments; sign language; neural networks; gesture recognition; ukrainian sign language; sign language translation; reduce inequality.

Introduction

Sign language is one of the oldest ways of communication for people with hearing impairments in all cultures, with its own rules and norms. However, despite the fact that sign language is no less important than verbal language, it has not received sufficient attention in research. This language consists of gestures, each of which is performed with the hands in combination with body position, facial expressions, shape or movement of the mouth and lips. The use of sign languages by people without hearing loss is secondary, but quite common, because there is often a need to communicate with people with hearing loss who use sign language. The use of sign language instead of voice communication can also be useful in situations where voice communication is impossible or difficult.

According to the World Health Organization (WHO), approximately 1.5 billion people (approximately 20% of the population) worldwide have hearing impairments, of which 430 million are completely deaf (approximately 5.7% of the population) [1]. More than 44 thousand people with hearing impairments are registered with the All-Ukrainian public organization Ukrainian Society of the Deaf. These citizens need attention and protection because of their hearing impairment, especially during the military aggression on the territory of Ukraine.

Today, many web developers are trying to make their sites more accessible to people with hearing impairments. For this purpose, in particular, WCAG web accessibility standards are used [2]. It should be noted that in most developed countries there is legislation that obliges organizations and companies to provide

accessibility to their resources for users with various disabilities, including people with hearing impairments. The use of special software tools allows this category of people to develop the practical skills and abilities necessary for education, communication with the outside world, assimilation of information, further employment, building a professional career and living in general. The Decree of the President of Ukraine "On the Sustainable Development Goals of Ukraine for the period up to 2030" (in the part "Reducing Inequality") recommends taking this aspect into account when determining the directions of scientific research [3].

Thus, social inclusion for people with hearing impairments is an acute problem that needs to be addressed in the context of the development of IT technologies and legislative initiatives that ensure the rights and equal opportunities of people with disabilities. This justifies the need and relevance of researching assistive technologies and software tools to facilitate the social inclusion of people with severe hearing impairments in society. Such impairments are partially compensated for by creating appropriate living conditions, which are achieved with the help of modern information technologies. The problem of using sign language is that this type of communication is extremely important for people, especially for those who have limitations in verbal communication, such as people with autism or hearing impairments. In addition, the growth of globalization and intercultural communication has made sign language even more relevant, as it can influence the perception of cultural differences and promote mutual understanding between people from different cultures and language groups. The study of sign language can open up new opportunities for those who have limitations in verbal communication and help to understand which gestures are acceptable for different cultures and situations.

Ukrainian Sign Language (USL) is a natural language system that is transmitted by visual and gestural means and has its own lexical and grammatical structure [4]. It has naturally developed and serves as a primary mode of communication for individuals using sign language residing in Ukraine, whether currently or in the past. In the 20th century, the USL began to develop actively and become more standardized. In 1963, a scientific laboratory of sign language was established at the Taras Shevchenko National University of Kyiv. As of January 1, 2022, there were about 40 preschools in Ukraine with deaf children under 6 years of age and 60 specialized general education schools for deaf

students between the ages of 6 and 18. From 2006 to 2018, the regulatory authority in Ukraine that was responsible for the study of linguistic features and the development of DMC published textbooks, manuals and scientific articles. The COVID-19 pandemic had a significant impact on the communication of people with hearing impairments around the world. As most people with hearing impairments depend on lip reading and visual communication, the introduction of restrictions and recommendations on social distancing and the use of masks became significant barriers to communication. Since February 24, 2022, the hostilities have particularly exacerbated the isolation and alienation of deaf Ukrainians and increased the threat to their lives. Ukrainians with hearing impairments faced a number of problems during the martial law: first, they do not have access to information about air alerts, as they cannot hear sirens or telephone signals; the second problem was the lack of access to the local administration's hotline, rescue service, remote consultations of medical and psychological workers, volunteers, etc.

Raising awareness and understanding among the general public of the problems faced by people with hearing impairments should help reduce discrimination and improve attitudes towards this category of people. To solve these problems, experts and specialists in the field of sign language interpretation, linguistics, artificial intelligence, and software development should be involved to comprehensively address the problems described above. It is necessary to offer accessible software applications for easy use in everyday life, as this will help to ensure the rights of people with hearing impairments and their participation in society. Therefore, the study of models and methods of automated sign language translation (namely, translation of sign language into text or sounds and vice versa, as well as translation from one sign language to another sign language) is an urgent task.

Comparative analysis of existing sign language interpretation systems and technologies.

Purpose and objectives of the study

Modern technologies that are used to communicate between people with hearing impairments can be classified according to the following features:

- a) the possibility to choose a sign language (SL): one or more;
- b) the type of translation:
 - Speech → Text;

- SL → Text;
- SL → Text → Speech;
- Text → SL;
- Speech → Text → SL;
- Text → Speech;
- Text1 → Text2;
- Speech1 → Speech2;
- SL1 → SL2;

c) the object of translation: a person or a program.

Speech → Text is an important type of communication for people with hearing impairments. An example of this is the use of subtitles, which can be useful for a variety of reasons: to help people with hearing loss who cannot clearly hear or understand the language used in a video or movie; to improve language comprehension (particularly in foreign language learning); to improve listening (subtitles can help people improve their listening skills by allowing them to listen more closely and understand content); to allow content to be translated into other languages, allowing people from different countries and cultures to understand the content.

Speech → SL is also a common type of communication from a hearing person to a person with a hearing impairment. This method describes the work of sign language interpreters, whose main tasks include: real-time translation of direct speech for hearing impaired people who interact with the world using SL; preparation for important events (conferences, meetings and other events where hearing impaired people may be present); creating subtitles for videos and movies to make them more accessible individuals with hearing loss; preparation of educational materials for hearing impaired people to provide them access to education. Sign language interpretation systems are used in various industries, including education, medicine, government agencies, etc. In particular, they use services such as Google Meet, Zoom, Skype Translator, WebEx, Microsoft teams, as well as applications for mobile devices (Ava, iTranslate, Sorenson Buzz). Some of these services provide free access to sign language interpretation, while others may be paid. A common mobile application from Microsoft is the Seeing AI app, which uses machine learning and artificial intelligence to help people with visual impairments. The program is available for iOS and Android and can be downloaded for free from the App Store or Google Play. One of the main features of Seeing AI is sign language recognition, which allows the user to translate gestures into text in several languages, including Ukrainian. To do this, you need to point the camera of your mobile device at the

person using the app, and the program will automatically recognize gestures and translate them into the text for chosen language.

Let's analyze the programs and technologies that allow you to translate text into SL (Text → SL). Some of them are used for online translation, while others are designed for use in educational institutions and rehabilitation centers. The most popular programs for translating text into SL include Signing Savvy, VL2 Signing Avatar, Spread The Sign, ProDeaf, and Hand Talk. Signing Savvy's online sign language dictionary allows users to look up and translate words and phrases into sign language, and provides videos demonstrating how to perform gestures and options for adjusting the video recording speed.

Let's take a look at some programs and technologies that use 3D animation to translate text into sign language. These programs typically use computer vision and machine learning technologies to create animated characters that can perform gestures and convey messages into sign language. The most popular programs of this type include SigningAvatar, JASigning, and Signily. SigningAvatar is a system that automatically translates text into SL using 3D animation. This technology allows users to enter text in natural language and receive videos with translation in the GUI. The main advantages of SigningAvatar are automatic translation, high accuracy, multilingualism, and ease of use. Disadvantages of SigningAvatar: American English-language version of the SL; limited set of gestures; limited user options (SigningAvatar does not allow users to create their own gestures or edit existing gestures).

Currently, there are several software-based sign language interpretation systems in Ukrainian that can be useful for people with hearing impairments (in particular, SignTalk, Ava, and SignAll). SignTalk is a software-based sign language interpreting system that uses SL to transmit messages between hearing impaired users and non-HSL users. SignTalk has a Ukrainian-language interface and can be used on computers and mobile devices. Ava is an application for mobile devices that uses machine learning to translate speech into text in real time and supports several languages, including Ukrainian. SignAll is a software system that uses modern gesture recognition technologies, including deep learning, to accurately translate speech into Ukrainian text in real time. The system uses three video cameras and one graphics card to collect and process information about hand movements and gestures.

It should be noted that mentioned systems, which are used to communicate between individuals with hearing loss, are constantly being improved using various modifications of artificial neural networks and 3D animation tools.

Let us consider some results of modern research concerning the practical implementation of communication methods for people with hearing impairments and confirming the relevance of the problem of creating software tools for automating sign language interpretation.

Article [5] presents an overview of some methods of segmentation techniques for recognizing hand gestures. This study is aimed at using hand gesture recognition for sign language interpretation in the process of human-computer interaction. Segmentation uses various hand detection schemes with the necessary morphological processing. Paper [6] describes a system capable of efficiently converting sign language gestures into spoken language. This system aims to provide voice output in various regional languages. In article [7], proposed model extracts temporal and spatial characteristics after capturing a video sequence. To extract the spatial characteristics, a program was used to determine the landmarks of the face, face, and pose using various types of RNNs (recurrent neural networks), including LSTM and GRU. Paper [8] proposes a variant of the perceptual computing user interface that allows computers to capture and interpret human gestures as commands. Paper [9] proposes a system for recognizing hand gesture images using modern image processing methods. During image segmentation, skin color detection and morphological operations are performed to accurately segment a part of the hand gestures. Then, a heuristic manta ray feature optimization (HMFO) technique is used to optimally select features by calculating the best fitness value. Paper [10] proposes a method for detecting 3D objects based on a point cloud and a graphical neural network (GNN) in combination with an attention mechanism. The paper [11] proposes a spatio-temporal GCN model for adaptive construction of spatio-temporal graphs that allow creating sign language recognition datasets based on a video skeleton. Article [12] provides an overview of research based on the use of hybrid recurrent neural networks (RNNs) to build a gesture recognition system that can reduce the number of classification errors and increase the stability of recognition. Article [13] provides an overview of research based on the use of hybrid recurrent neural networks (RNNs) to build a gesture recognition system that can reduce the number of classification errors and increase the stability of

recognition. In article [14], a convolutional neural network model for hand gesture recognition is proposed. In this paper, a one-time coding technique is used to convert categorical data values into binary form. Unimportant parameters are excluded from consideration, which improves classification accuracy. The paper [15] proposes the idea of static and dynamic capture and recognition of human gestures in real time based on a radial basis function neural network (RBFNN). Dynamic time warping (DTW) is used to select candidates for dynamic behavior, as well as to recognize gestures by comparing observed recordings with a series of pre-recorded reference data templates. In article [16], a segmentation method is proposed to identify hand gestures from an input image, which improves recognition accuracy. A comparison of hue and saturation segmentation methods for different background lighting conditions is presented. In article [17], proposed model can recognize hand gestures and signs using a convolutional neural network (CNN) and convert them to text.

Of course, the programs and technologies discussed are still evolving, and each has its own advantages and limitations. Nevertheless, they are an important step in providing access to sign language for people with hearing impairments and help make sign language more accessible to people without hearing impairments. Based on the results of the analysis of scientific publications and modern technologies devoted to the automation of sign language interpretation by software tools, we formulate the purpose and objectives of the proposed work.

The aim of the work is to develop and study methods of sign language interpretation automation that improve the level of communication for people with hearing impairments in accordance with the Sustainable Development Goals of Ukraine.

Research objectives:

- developing methods for converting sign language into text, converting text into sign language, and automating translation from one sign language to another sign language using modern intelligent technologies;
- testing of the proposed methods of sign language translation automation and determination of prospects for their application.

**Proposed technology
for automatic sign language interpretation**

The technology proposed in this paper involves the implementation of methods for converting sign language

into text (task A) and methods for converting text into sign language (task B) using neural networks and 3D animation.

Let's first consider the essence of the implementation of task A. Sign language to text (SL-to-Text) technologies are a set of various methods that can be used to convert gestures used in sign language into written text. The main goal of SL-to-Text methods is to help people with communication disabilities use the web as a means of interacting with the world around them by converting their gestures into text phrases. The proposed SL-to-Text method can be divided into three stages.

The first stage involves capturing the image of a gesture (or movement) of all parts of the body using video devices and software. At this stage, it is necessary to capture a video object using a webcam or other video device to obtain an image of a hand performing a gesture against the body background. Specialized software is used to capture and further process the image to be analyzed.

The second stage involves determining the contours of the hand and its location in space in the image using computer vision algorithms. In particular, to recognize the hand as an object, a color filter and markup algorithm is used. Skin color detection is achieved by detecting a skin area where skin pixels are spatially processed based on intensity and texture analysis. The resulting black-and-white image allows for the identification of the palm (in Fig. 1, its image is white). Then the palm is framed by a rectangular outline, which further helps to track the position of the hand (Fig. 1).



Fig. 1. An example of capturing the "Palm" object with a webcam

The third stage involves gesture recognition. Once the hand is detected in the image, the system identifies the intended gesture. For gesture recognition, a convolutional neural network (CNN) model is used, which is trained on imported and self-generated video gesture datasets. In general, a CNN is an artificial neural network that specializes in the ability to select or detect and interpret certain patterns. The use of existing patterns allows the CNN network to analyze frames of gesture

images. The CNN consists of: hidden convolutional layers with filters capable of detecting patterns and a nonlinear function (usually ReLU); pooling layers to reduce the number of parameters for too complex images; a fully-connected layer that transforms the combined image matrix into a vector and then applies the Softmax function to classify the object (Table 1).

Table 1. CNN architecture model

Stages / Operations	Formal models of operations implementation
Convolution	$z^l = h^{l-1} * W^l$ (1)
Pooling, Max-Pooling	$h_{xy}^l = \max_{i=0, \dots, s, j=0, \dots, s} h_{(x+i)(y+j)}^{l-1}$ (2)
Fully-connected layer	$z_l = W_l * h_{l-1}$ (3)
ReLU(Rectifier)	$\text{ReLU}(z_i) = \max(0, z_i)$ (4)
Softmax	$\text{softmax}(z_i) = e^{z_i} / \sum_j e^{z_j}$ (5)

The following notations are used in Table 1:

z^l – output of convolution layer l ;

h^l – activation of layer l ;

$*$ – discrete convolution operator;

W – training parameter.

The built model is tested on a pre-created test sample of gesture images obtained under different lighting conditions, on different backgrounds, etc. The result of using the model is text classified by gesture analysis. To ensure better gesture translation, it is advisable to add such operations as noise filtering, image alignment, adaptation to different lighting and camera movement to the basic SL-to-Text method procedure. In general, the above approach for implementing SL-to-Text using the CNN architecture is universal for solving gesture recognition tasks and processing corresponding images. However, it is important to pay attention to the specifics of a particular task and the context in which it is performed. An example of gesture classification using the Sign Language to Text method is shown in Fig. 2.

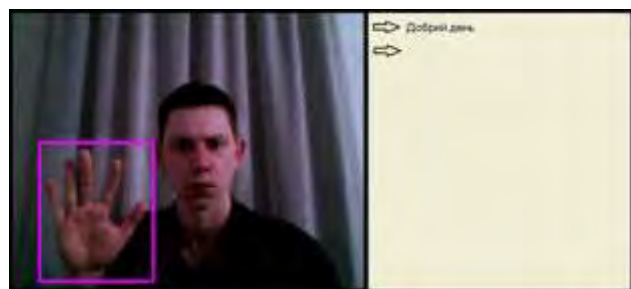


Fig. 2. An example of gesture classification using the Sign Language to Text method

Let us further consider the essence of text-to-sign language conversion according to the proposed approach (Text-to-SL and Speech-to-SL), which uses computer vision and machine learning methods to convert written or spoken text into SL. It should be noted that each SL has its own rules and grammar that differ significantly from the rules of written language, so it is impossible to make a direct word-for-word translation, but it is possible to convey the very idea of the statement, its keywords or phrases.

At the first stage of implementing the Text-to-SL method, the initial input text is obtained, which needs to be translated into a SL while preserving the idea behind the text. There are two ways to obtain the initial input text: entering text using the keyboard (Text to Text Summarization) or recording voice through a microphone and then recognizing speech, the sounds of which are converted into text (Speech-to-Text stage).

In the case of Speech-to-Text, the received audio file must be pre-processed, which involves noise reduction, filtering, and equalization. Next, various acoustic features are extracted from the audio signal, such as frequency, energy, sound pressure, etc. Based on the extracted acoustic features, speech is decoded using a hidden Markov model, which is widely used for pattern recognition in speech. The phoneme is divided into smaller sound elements, each of which corresponds to a state:

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad a_{ij} \leq 0, \quad \sum_{j=1}^N a_{ij} = 1,$$

$$B = \{b_j(k)\}, \quad b_j(k) = P[v_k \text{ at } t | q_t = S_j], \quad (6)$$

$$\pi_i = P[q_i = S_i], \quad 1 \leq i, j \leq N, \quad 1 \leq k \leq M,$$

where N is the number of states of the model;

S_j is the state of the model (at time t denoted as q_t);

M is the size of the discrete alphabet;

A is the probability transition matrix ($A = \{a_{ij}\}$);

B is the probability distribution of the appearance of observation symbols in state j ($B = \{b_j(k)\}$);

v_k is the observed symbol;

π is the initial probability distribution of states ($\pi = \{\pi_i\}$).

After choosing the values for N , M , A , B , and π , the neural network model can be used to generate a sequence of observations $O = O_1 O_2 \dots O_L$, where each observation O_i is a character from the alphabet V , and L is the number of characters in the observed sequence. Thus, by training the model on test recordings, it is

possible to obtain the primary text T_0 corresponding to certain sounds and use it in subsequent stages to generate the corresponding gestures. In the case of Text to Text Summarization, it is necessary to perform a semantic analysis of the primary text T_0 and generate a new reduced text T_1 , i.e., determine the meaning to be conveyed through gestures. To do this, we use Text Summarization methods that allow us to recognize language concepts based on context, phrase and phrase analysis. After the semantic analysis of the T_1 text, it is necessary to select gestures or sets of gestures that best match the semantics of the T_1 text concepts. Rule-based machine translation (RBMT) and the use of a dictionary of SL concepts allow for the selection of SL concepts. These concepts include the use of SL signs, gestures that reflect movement and space, and gestures that convey emotions and mood. That is, the result of this stage should be a word or phrase, the meaning of which can be further conveyed through gestures.

To solve this problem, it is proposed to use SL dictionaries, in particular, ASL Pro Dictionary, Lifepoint, Signing Saavy, etc.

The selection of gesture concepts is carried out according to an algorithm that involves the sequential implementation of points P1 – P7:

- P1. The input sentence is divided into fragments (words, phrases);
- P2. The selected fragments are searched in the available SL dictionaries;
- P3. If the fragment is present in the SL dictionaries, you go to P5;
- P4. If the fragment is present in the SL dictionaries, you go to P6;
- P5. The selected video fragment is saved in the current buffer of gesture concepts for further visualization (go to P7);
- P6. The fragment is excluded from the list of candidates for inclusion in the set of gesture concepts, and its SL dactylic translation is used to further form sets of gesture sequences that convey the semantics of the sentence, followed by a transition to P7;
- P7. A set of video gesture concepts is formed (according to P5) and a set of fingerprint displays of individual sentence fragments.

Dactylic translation of text into sign language implies the need to divide the analyzed sentence fragment into letters and symbols, followed by displaying the corresponding images of the position of the hand elements at intervals of several seconds.

It should be noted that the presence of a large number of fingerprint mappings (as a percentage) compared to the number of video gesture concepts is undesirable, as it can significantly increase the duration of real-time sign language interpretation. However, fingerprint mappings allow for a fairly accurate representation of some specific information (e.g., contact information and proper names of interlocutors).

After the SL concepts are selected (according to this algorithm), sets of gesture sequences that convey the semantics of the T1 text are generated (they should include determining the order of gestures, taking into account the tempo and intensity, and using gestures that emphasize important information). Once a sequence of gesture concepts has been formed, it is necessary to graphically visualize these gesture sets on device screens so that users can see them (the SL gesture concept visualization stage). You can display a demonstration of a pre-recorded video

gesture or an animated character. There are quite large databases of recorded video gestures for various SLs, as well as recommendations for the formation of a library of animated SL concepts (development of 3D characters by animation programmers based on a video sample).

In this paper, we propose to automate the process of filling the database of animated gesture concepts by generating an animated gesture (AG) by an AI model from a video file recorded by a person and saving it in a database (Fig. 3). It is necessary to develop a neural network model that will receive a video file as input and a csv (comma separated values) file as output, where each line contains the time and a set of hand points in space for each frame. Next, you need to bind each point to the corresponding point of the 3D character and change their position in space with each frame. The output is an AG with a description that is stored in the database of all animated gestures.



Fig. 3. Scheme of animated gesture generation

Using a dataset of video files with gesture demonstrations, the AI model was trained to generate a 3D gesture as a set of connected labeled points. The position of the hands (including the palm) in space during the gesture was tracked and recorded frame by frame. To do this, we used the method of Dung and Mizukawa, which implements a technique that can be used to extract the center of the hand and a group of feature pixels called distance feature pixels. For these pixels, the Hough transform is computed to detect all the extended fingers as lines. Once the lines are detected, the directions and positions of the fingers can be accurately determined. With a set of points previously saved in a csv file that correspond to the position of the hand in each frame, you can animate a 3D character and save the generated animated gesture to the database. The result of this step is a sequence of short animated hand gestures of a 3D character found in the database. At the same time, we get simplified phrases in SL according to the semantics of the T0 text that needed to be translated. An example of using the created software application "Text to Sign Language" (to translate a phrase into SL) is shown in Fig. 4. The trained AI model can be improved in such a way that 3D gestures are generated during communication, rather than from a database of animated gestures.

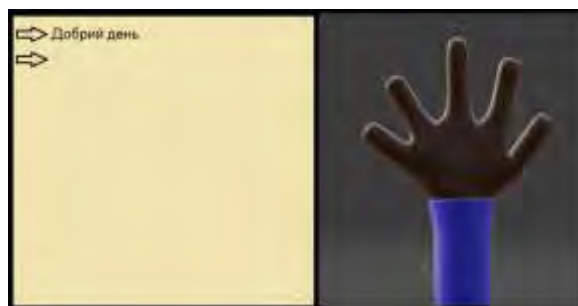


Fig. 4. An example of translating a phrase into SL using the "Text to Sign Language" method

The relevance of developing the Sign Language1 to Sign Language2 method is due to the problem of communication between people who speak different sign languages. Let us consider the features of the proposed variant of this method, the idea of which is to combine the technologies described above in order to obtain a general scheme of translation from one sign language to another:

- the interlocutor selects the SL1 sign language he or she speaks on the screen (Ukrainian, English, Spanish, German, etc.);
- the next step is to record a video of the interlocutor's gesture on the camera (the camera should be positioned so that the upper part of the body above the waist is visible);

- the video is then processed by the program, which results in the text Text1 appearing on the screen, which the program recognized from the gesture;
- the other person chooses the SL2 language in which he or she is communicating;
- Text1 is translated into SL2, resulting in the translated Text2;
- from Text2, the meaning of which must be understood by the other interlocutor in SL2, the corresponding gestures are generated and demonstrated on the screen using a 3D animated character in SL2.

An example of the implementation of this method for communication between hearing impaired users speaking different SLs is shown in Fig. 5.



Fig. 5. An example of sign language translation using the "Sign Language1 to Sign Language2" method

Such a software application can be especially relevant if a hearing impaired user travels to another country but does not know the specifics of its SL. With the proposed software application, the user will be able to use the camera of their device to recognize gestures and receive appropriate translations or interpretations of gestures in the language of the host country. This software application can also be useful for those who need translation or interpretation of gestures in various life situations (in particular, for communication between users with and without hearing impairments during a business trip).

Software implementation and testing results of the proposed automatic sign language interpretation technology

Let us justify the choice of programming language, libraries and development environment for the implementation of the Sign Language to Text and Text to Sign Language methods.

To solve the problem (SL-to-Text), we used the open source programming language Python. One of the advantages of this language is the availability of a wide range of libraries (including those for working with neural networks), tools, and frameworks that facilitate the process of quickly creating complex programs. PyCharm integrated development environment (IDE) was chosen to implement the Sign Language to Text method. This environment offers an extensive list of features and tools for Python developers, including code highlighting, code navigation, testing, and version control integration.

To implement the tasks of tracking the position of the hands and body in the proposed technology, the open source Mediapipe platform from Google was used. One of the main advantages of this platform is that it provides pre-built and optimized algorithms and models that can be easily integrated into applications, and offers a variety of tools that allow developers to create complex computer vision and machine learning applications.

Intel's OpenCV computer vision and machine learning library, which provides interfaces to Python, was used to render and display the contours of certain body parts found with Mediapipe. OpenCV contains a number of optimization algorithms that can be used to perform image and video processing tasks such as object detection, recognition, tracking, segmentation, and more.

TensorFlow and Keras libraries were used to build a neural network model for recognizing gestures from the screen. TensorFlow is an open source platform from Google that provides a wide range of built-in operations and functions for creating and training neural networks. One of the key advantages of the TensorFlow library is the ability to take advantage of GPUs to accelerate training. Keras is an open-source deep learning library used as an API for building and training deep neural networks. Keras is based on TensorFlow and provides a user-friendly interface for building and training models, which makes it easy to experiment with different architectures and hyperparameters. Keras also provides a variety of pre-built layers and models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which makes it easy to develop deep learning models. Keras also includes various tools for model visualization, as well as tools for data preparation and preprocessing.

To create a 3D model of the gesture demonstration, we used the 3D Unity engine, which allows us to support 3D graphics, as well as visualize the architecture and model various types of interactive media [10].

Let us consider the structure of projects for the software implementation of the Sign Language to Text and Text to Sign Language methods using the listed development environments, libraries, and programming languages.

The Sign Language to Text method and several stages of the Text to Sign Language method were implemented in Python in the PyCharm development environment using the libraries described above. The structure of the PyCharm project consists of the following directories and executive files:

- Dataset (the directory contains a set of points in space corresponding to each gesture on which the deep neural network model was trained);
- Logs (the directory contains log files during model training that allow visualizing the model training process and obtaining the desired statistics from certain stages of the process);
- main.py (this file contains the implementation of the Sign Language to Text method at all its stages: using a camera to record gestures, libraries to track body position, storing body position points in space, preparing, creating and training the neural network model itself, and displaying the results on the screen);
- sl_to_text.h5 (trained neural network model for further use when running the program);
- text-to-sl-csv.py (a file containing the implementation of the stage of collecting points in space to build a 3D model of the Text to Sign Language method);
- video (a directory containing the video files for the text-to-sl-csv.py file, from which the body position points for the 3D model are tracked).

The 3D model for displaying the gesture on the screen was created and processed in the Unity 3D development environment. The structure of this project consists of the event handler files BodyAnimation.cs, LeftHandAnimation.cs, RightHandAnimation.cs; the files for processing the display result for each frame LineCode.cs; the files pose.txt, lh.txt, rh.txt, which contain information about the position of the controlled body points; the files Blue.mat, Green.mat, Red.mat, which store the characteristics that were added to certain body display points. The project structure also includes the Scenes directory, which contains the system files for setting up the scene, as well as the pose_2test.txt, lh_2test.txt, rh_2test.txt files with the body position points of additional gestures that were added to test the developed program.

The Sign Language to Text method is implemented in 6 stages. Below is their description.

The first stage (importing libraries). To implement this stage, the PyCharm development environment uses a developed library import program.

Among the imported libraries are libraries for working with OpenCV, numpy data arrays, the os operating system, matplotlib graphical visualization, time, and the keras library for working with neural networks.

The second stage (using OpenCV to connect the camera and display it on the screen). First, you need to specify which camera will be used in the process (built-in or from an external device). Then, all the collected information from the camera is transferred to the loop, which changes the image on the screen with each frame. The camera can be closed by pressing the 'Q' key or simply by closing the camera window on the monitor screen. The program code for this stage was developed

The third stage (using the Mediapipe platform to track the position of body parts).

To implement this stage, we used the Mediapipe library module for recognizing the necessary body points and the module for displaying it on the screen. Mediapipe makes it possible to track hands, body posture, face, and all of the above separately.

In addition, to track the body with the desired parameters, a function was developed and programmatically implemented, the input parameters of which are the frame from which the necessary body parts are to be extracted and the Holistic model to which the desired changes are transferred.

Finally, to determine the position of the controlled body points themselves and their connection to each other for the purpose of clear visualization of the skeleton, an additional function was developed and programmatically implemented to adjust the display subjects (body pose, left hand, right hand) and their characteristics (thickness, size, color).

An example of the implementation of the functions of the third stage for tracking body and hand posture is shown in Fig. 6.

The fourth stage (creation of own dataset for further training of the neural network model). In the course of designing the software implementation, we demonstrated the Sign Language to Text method based on six randomly selected words: ambition, success, but, way, to, sun. For each of the words, 30 videos of 30 frames each were recorded, and for each frame, the position of the pose and hands in the three-dimensional xyz coordinate system was saved.



Fig. 6. An example of using the Mediapipe platform to track body and hand posture

The Mediapipe library tracks pose positions at 33 points and hand positions at 21 points. Thus, 258 points were recognized and saved with each frame of the video, and a total of 180 videos were recorded. To complete this stage, the following operations were sequentially implemented:

- creating directories for storing files with the position of body points in space;
- creating a function for extracting the pose, left and right hand position points from the Holistic model and entering them into an integer array;
- recording video of gestures and saving their position in space in files with the npy extension using the point extraction function from step 2 to the corresponding directories generated in step 1.

Thus, as a result of this step, a dataset with points of body position in space was obtained for further training of the neural network. Figure 7 shows an example of recording video of gestures and saving the result to the appropriate directory.

The fifth stage (creating and training a deep learning model). This is a stage of the Sign Language to Text method that requires testing the model for various parameters and determining their optimal values.

To implement this method, it was necessary to choose the types of neural networks that are capable of real-time classification of a sequence of SL image

```

model = Sequential()
model.add(LSTM(64, return_sequences=True, activation='relu', input_shape=(30, 258)))
model.add(LSTM(128, return_sequences=True, activation='relu'))
model.add(LSTM(64, return_sequences=False, activation='relu'))
model.add(Dense(64, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(actions.shape[0], activation='softmax'))
model.compile(optimizer='Adam', loss='categorical_crossentropy', metrics=['categorical_accuracy'])
model.fit(X_train, y_train, epochs=110, callbacks=[tb_callback])
print(model.summary())

```

Fig. 8. Program code for building a sequential model of a recurrent neural network

frames. Since each previous and subsequent frame is interconnected, it becomes necessary to use these connections to further train the model and obtain more successful results. Recurrent neural networks (RNNs) are the most suitable for this task, as they contain feedback and allow storing current information and passing it on to the next step of the procedure.



Fig. 7. An example of recording a gesture and saving the position of pose points and hands

The method was implemented and tested using the LSTM architecture (a modification of recurrent neural networks capable of learning long-term dependencies).

The primary dataset was divided into training and test samples in the ratio of 75% to 25%. Next, we built a sequential neural network model consisting of three LSTM layers and three Dense layers. Figure 8 shows the code for building this neural network.

The first LSTM layer has a dimension of 30 by 258 (each video consists of 30 frames for 258 points in the xyz coordinate system). The activation function `relu` was applied to each of the layers except the last one, and the `softmax` function was applied to the last layer Dense to normalize the results in the range from 0 to 1. During the network training (using the Adam optimizer and the loss function `categorical_crossentropy`), the number of epochs (110) was determined to minimize the overfitting. Fig. 9 shows a graph of the model's prediction accuracy (horizontal axis) as a function of the number of epochs (vertical axis).

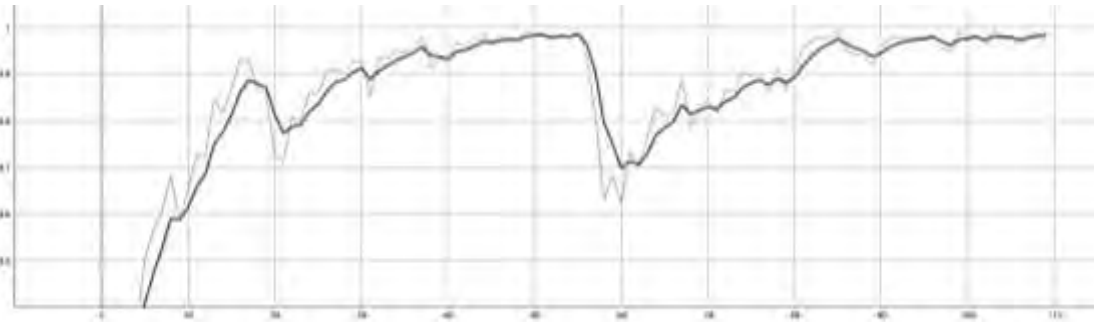


Fig. 9. Graph of forecasting accuracy depending on the number of epochs

The accuracy of the trained and optimized LSTM model is 98.52%. The parameters of this model are shown in Table 2. The results were obtained for the total number of trained parameters in the range (237, 530). Thus, after the fifth stage, a recurrent neural network model that implements the Sign Language to Text method was created, trained, saved, and ready for use.

Table 2. Parameters of the optimized LSTM model

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 64)	82688
lstm_1 (LSTM)	(None, 30, 128)	98816
lstm_2 (LSTM)	(None, 64)	49408
dense (Dense)	(None, 64)	4160
dense_1 (Dense)	(None, 32)	2080
dense_1 (Dense)	(None, 6)	198

The sixth stage (visualization of the results). This is the final stage of implementing the Sign Language to text method, which aims to display text classified from gestures in a user-friendly form (ticker) using a trained LSTM model and the OpenCV library.

To improve gesture recognition and correctly display the required text on the screen, a delay of 10 frames after gesture recognition was added. In this way, the user can change the position of their hands and prepare for the next gesture, thereby not interfering with recognition and not adding complexity to the process. An example of the program's operation is shown in Fig. 10.

The *Text to Sign Language* method is implemented in 4 stages. Below is a description of them.

The first step (tracking body position from video using the Mediapipe library). This stage is similar to the third stage of the Sign Language to Text method, but here the input parameter is a prepared video, not a camera that transmits images in real time. Next, the operations described above are implemented – creating an object to work with the Holistic module, using a function with the desired parameters to track the body, and displaying

connections between body points. An example of tracking the position of body points from a video gesture that means "Computer Science" is shown in Fig. 11.



Fig. 10. An example of the final result of the program for the Sign Language to Text method

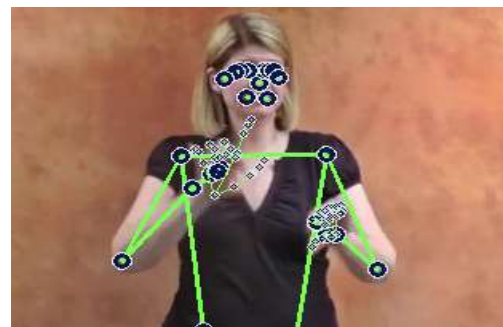


Fig. 11. An example of tracking the position of body points from a video gesture that means "Computer Science"

The second stage (saving pose and hand position points to txt files).

This step is necessary in order to transfer the saved points to the Unity 3D project to generate a 3D character, after which its position will change with each frame in accordance with the transferred points.

It should be noted that when the points are mapped to the Unity 3D project, they may be too close to each other, so their coordinates were multiplied by a thousand and rounded to the third decimal place. This stage allows

us to obtain 3 files of body position (right hand, left hand, and pose).

The program code for writing to the pose position file is implemented (writing to the right and left hand position files is done in the same way but using other variables).

The third stage (creating objects in the Unity 3D project).

To create a 3D character, you must first create sphere objects that will be responsible for displaying the corresponding points. To display the pose, only 21 points out of 33 were taken into account, because the gesture is shown above the waist, and therefore the points located at the level of the pelvis, knees, and feet do not carry important information in this case.

To represent the right and left hands, 42 sphere objects were created (21 objects for each hand). To display the connections between the created point-sphere objects, we created line objects connecting one point to another. In this case, 14-line objects connecting the pose points and 21 line objects for the right and left hands were taken into account. Additionally, 6 material colors were added for better visual perception of the graphic: red (body, nose, and mouth pose graphics); blue (ear graphics); green (graphics of connections between points); purple (hand graphics); black (eye graphics); and light brown (background graphics).

The fourth stage (programming of event handlers).

This stage is necessary to create the ability to read files with points in the corresponding classes of event handlers and change the position of sphere and line objects with each frame. Four event handlers were created: `BodyAnimationCode`, `LeftHandAnimationCode`, `RightHandAnimationCode`, `LineCode`. The first three handlers are designed to read different files (for pose, right and left hands, respectively).

The `BodyAnimationCode` event handler, using the `Start ()` method, which is called at the beginning of the scene initialization, implements reading a file with points into a list.

In order to change the position of the points with each frame, the `Update ()` method was used. Its idea is to divide the read file line into a comma-separated array, and then to parse each of the array elements, which are the coordinates of the points, into a floating-point type and assign them to the corresponding sphere objects. This sequence of actions is repeated until the file runs out of coordinates of the points being analyzed.

The `LineCode ()` event handler is implemented in a universal form, that is, it is used to display the graphic

of the connections of all the necessary points. The idea of this handler is to set the beginning and end of the line in the visual constructor by attaching sphere objects, and using the `Update ()` method to set the beginning and end of the line coordinates. Fig. 12 shows an example of a 3D model created (for displaying a gesture) with an indication of the relationships between points.

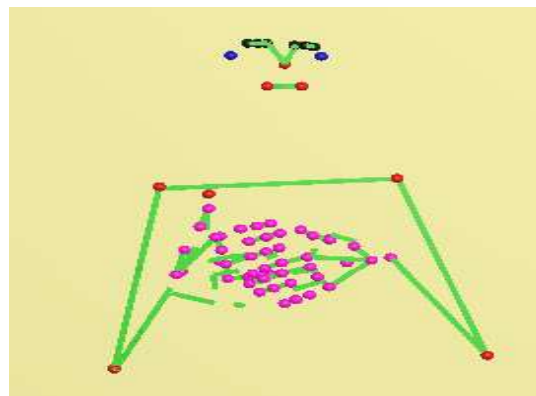


Fig. 12. An example of a 3D model of gesture display with an indication of the connections between points

Conclusions and prospects for further development

The following results were obtained: the main problems and tasks of social inclusion for people with hearing impairments were identified; a comparative analysis of modern methods and technologies of sign language translation was carried out; methods for converting sign language into text (SL-to-Text) and converting text into sign language (Text-to-SL) using neural networks and 3D animation were proposed; the possibility of using the developed methods to automate translation from one sign language to another sign language (Sign Language1 to Sign Language2) was substantiated.

For gesture recognition, a convolutional neural network model is used, which is trained using video gesture datasets imported and generated by the system. The trained model has a high recognition accuracy (98.52%).

The 3D model for displaying the gesture on the screen was created and processed in the Unity 3D environment. The structure of the project, executive and auxiliary files used to build 3D animation for generating sign language concepts includes: event handler files; display results according to which they contain information about the position of the tracked body points; files storing the characteristics of the materials that were added to certain body display points. The Scenes directory was added to

the general structure of the automatic sign language translation system, containing system files for setting up the scene, as well as files with body position points for additional gestures used for testing the developed program.

The practical significance of the results of the work is confirmed by the demo versions of the "Sign Language to Text" and "Text to Sign Language" software applications.

The combination of the developed models and technologies makes it possible to implement a comprehensive system of automatic sign language

translation that solves a number of the tasks discussed in this paper, namely: converting sign language into text, converting text into sign language and automating translation from one sign language to another sign language using modern intelligent technologies.

A promising area for further research on this topic is the improvement of SL1-to-SL2 methods using different neural network architectures, as well as the creation of open datasets with extended sets of video gestures and concept dictionaries for different sign languages.

References

1. Chumachenko, I.V. "Methods of human resources management in the formation World Health Organization. Deafness and hearing loss". 2023. URL: <https://www.who.int/health-topics/hearing-loss> (date of application: 10.04.2023).
2. "Web Accessibility Initiative". URL: <https://www.w3.org/WAI/> (date of application: 10.04.2023).
3. Decree of the President of Ukraine №722/2019 "On the Sustainable Development Goals of Ukraine for the period up to 2030" dated September 19, 2019. URL: <https://www.president.gov.ua/documents/7222019-29825> (date of application: 10.04.2023).
4. Kruglyk, O.P., Horlachev, O.S. (2023), "The importance of auditory perception in the process of sign language translation for people with hearing impairment" [Znachennya slukhovoho sprymannya v protsesi zdiysnennya surdoperekladu dlya osib z porushennyam slukhu], *Scientific journal. Series 19 – Correctional pedagogy and special psychology*, No. 43, P. 39–48. DOI: <https://doi.org/10.24919/2308-4863/56-2-26>
5. Gobhinath, S., Vignesh, T., Pavankumar, R., et al. (2020), "A Study of Hand Gesture Segmentation Techniques for Sign Languages", *Journal of Computational and Theoretical Nanoscience*, Vol. 17, No. 4, P. 1764–1769. DOI: <https://doi.org/10.1166/jctn.2020.8439>
6. Ghosh, P., Dutta, A., Topno, S. (2022), "Sign Language Hand Glove", *American Journal of Electronics & Communication*, Vol. 3, No. 1, P. 14–16. DOI: <https://doi.org/10.15864/ajec.3103>
7. Chakraborty, S., Prayosi, P., Sarkar, S., Chakraborty A. (2023), "Sign Language Recognition Using Landmark Detection, GRU and LSTM", *American Journal of Electronics & Communication*, Vol. 3, No. 3, P. 20–26. DOI: <https://doi.org/10.15864/ajec.3305>
8. Mahesh, R., Kumar, T., Kavin, R., Karthikeyan S. (2020), "Manipulation of Web Using Gestures", *Journal of Computational and Theoretical Nanoscience*, Vol. 17, No.8, P. 3782–3785. DOI: <https://doi.org/10.1166/jctn.2020.9320>
9. Khetavath, S., Sendhilkumar, N., Mukunthan P., et al. (2023), "An Intelligent Heuristic Manta-Ray Foraging Optimization and Adaptive Extreme Learning Machine for Hand Gesture Image Recognition", *Big Data Mining and Analytics*, No. 6(3), P. 321–335. DOI: <https://doi.org/10.26599/BDMA.2022.9020036>
10. Zhou, H., Wang, W., Liu G., et al. (2022), "PointGAT: Graph attention networks for 3D object detection", *Intelligent and Converged Networks*, No. 3(2), P. 204–216. DOI: <https://doi.org/10.23919/ICN.2022.0014>
11. Wuyan, L., Xiaolong, X., Fu, X. (2022), "Human gesture recognition of dynamic skeleton using graph convolutional networks", *Journal of Electronic Imaging*, Vol. 32, Issue 2, P. 1–21. DOI: <https://doi.org/10.1117/1.JEI.32.2.021402>
12. John, J., Deshpande, S. (2023), "Hand Gesture Identification Using Deep Learning and Artificial Neural Networks: A Review", *Computational Intelligence for Engineering and Management Applications. Lecture Notes in Electrical Engineering*, Vol. 984, Springer, Singapore. P. 389–403. DOI: <https://doi.org/10.1007/978-981-19-8493-830>
13. Tan, Y., Lim, K., Lee, C. (2021), "Hand gesture recognition via enhanced densely connected convolutional neural network", *Expert Syst. Appl.*, Vol. 175, No. 114797. P. 28569–28587 DOI: <https://doi.org/10.1016/j.eswa.2021.114797>
14. Gadekallu, T., Alazab, M., Kaluri, R., et al. (2021), "Hand gesture classification using a novel CNN-crow search algorithm", *Complex & Intell. Syst.*, Vol. 7, P. 1855–1868. DOI: <https://doi.org/10.1007/s40747-021-00324-x>
15. Zhang, Y., Huang, Y., Sun, X., et al. (2020), "Static and dynamic human arm/hand gesture capturing and recognition via multiinformation fusion of _exible strain sensors", *IEEE Sensors Journal*, Vol. 20, No. 12. P. 6450–6459. DOI: <https://doi.org/10.1109/JSEN.2020.2965580>
16. Rahim, M., Miah, A., Sayeed A. and Shin J. (2020), "Hand Gesture Recognition Based on Optimal Segmentation in Human-Computer Interaction", *Proceedings of the 3rd IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, Kaohsiung, Taiwan, P. 163–166. DOI: 10.1109/ICKII50300.2020.9318870
17. Bhavana, D., Kumar, K., Bipin Chandra, M., et al. (2021), "Hand Sign Recognition using CNN", *International Journal of Performance Analysis in Sport*, Vol. 17(3), P. 314–321. DOI: 10.23940/ijpe.21.03.p7.314321

Шовковий Євгеній Ігорович – Харківський національний університет радіоелектроніки, студент кафедри штучного інтелекту, Харків, Україна; e-mail: yevhenii.shovkovyi@nure.ua; ORCID ID: <https://orcid.org/0009-0007-5613-0946>

Гриньова Олена Євгенівна – Харківський національний університет радіоелектроніки, старший викладач кафедри штучного інтелекту, Харків, Україна; e-mail: olena.hrnyova@nure.ua; ORCID ID: <https://orcid.org/0000-0002-3367-8067>

Удовенко Сергій Григорович – доктор технічних наук, професор, Харківський національний економічний університет ім. С. Кузнеця, завідувач кафедри інформатики та обчислювальної техніки, e-mail: serhiy.udovenko@hneu.net; ORCID ID: <https://orcid.org/0000-0001-5945-8647>

Чала Лариса Ернестівна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри штучного інтелекту, Харків, Україна; e-mail: larysa.chala@nure.ua; ORCID ID: <https://orcid.org/0000-0002-9890-4790>

Shovkovyi Yevhenii – Kharkiv National University of Radio Electronics, student at the Department of Artificial Intelligence, Kharkiv, Ukraine.

Grinyova Olena – Kharkiv National University of Radio Electronics, Lecturer at the Department of Artificial Intelligence, Kharkiv, Ukraine.

Udovenko Serhii – Doctor of Sciences (Engineering), Professor, Simon Kuznets Kharkiv National University of Economics, Head at the Department of Informatics and Computer Engineering, Kharkiv, Ukraine.

Chala Larysa – Phd (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Artificial Intelligence, Kharkiv, Ukraine.

СИСТЕМА АВТОМАТИЧНОГО СУРДОПЕРЕКЛАДУ З ВИКОРИСТАННЯМ НЕЙРОМЕРЕЖНИХ ТЕХНОЛОГІЙ ТА 3D-АніМАЦІЇ

Упровадження програмних засобів автоматичного сурдоперекладу в процес соціальної інклюзії людей з вадами слуху є важливим завданням. Соціальна інклюзія для осіб із вадами слуху є нагальною проблемою, яку необхідно вирішувати з огляду на розвиток IT-технологій та законодавчі ініціативи, що забезпечують права людей з інвалідністю та їхні рівні можливості. Сказане обґрунтовує актуальність дослідження асистивних технологій у контексті програмних засобів, таких як процес соціального залучення людей з важкими порушеннями слуху в суспільство. **Предметом дослідження** є методи автоматизованого сурдоперекладу із застосуванням інтелектуальних технологій. **Мета роботи** – розроблення та дослідження методів автоматизації сурдоперекладу для поліпшення якості життя людей з вадами слуху відповідно до «Цілей сталого розвитку України» (в частині «Скорочення нерівності»). **Основними завданнями дослідження** є розроблення й тестування методів перетворення жестової мови в текст, перетворення тексту в жестову мову, а також автоматизація перекладу з однієї жестової мови іншою жестовою мовою із застосуванням сучасних інтелектуальних технологій. Для розв'язання цих завдань використовувались **методи** нейромережного моделювання та 3D-анімації. Унаслідок дослідження здобуто такі **результати**: виявлено основні проблеми й завдання соціальної інклюзії для людей з вадами слуху; здійснено порівняльний аналіз сучасних методів і програмних платформ автоматичного сурдоперекладу; запропоновано й досліджено систему, що об'єднує метод *SL-to-Text*; метод *Text-to-SL* з використанням 3D-анімації для генерації концептів жестової мови; метод генерації 3D-анімованого жесту з відеозаписів; метод реалізації технології *Sign Language1 to Sign Language2*. Для розпізнавання жестів застосовано модель згорткової нейронної мережі, що навчається за допомогою імпортованих і згенерованих системою датасетів відеожестів. Навчена модель має високу точність розпізнавання (98,52%). Створення 3D-моделі для відображення жесту на екран і його оброблення відбувалися у середовищі *Unity 3D*. Структура проєкту, виконавчих і допоміжних файлів, що застосовуються для побудови 3D-анімації з метою генерації концептів жестової мови, містить: файли обробників подій; результати відображення, що мають інформацію про положення відслідкованих точок тіла; файли, що зберігають характеристики матеріял, які були додані до тих чи інших точок відображення тіла. **Висновки**: запропоновані методи автоматизованого перекладу мають практичну значущість, що підтверджують демоверсії програмних застосунків *Sign Language to Text* і *Text to Sign Language*. Перспективним напрямом подальших досліджень з окресленої теми є вдосконалення методів *SL1-to-SL2*, створення відкритих датасетів відеожестів, залучення науковців і розробників для наповнення словників концептами різних жестових мов.

Ключові слова: автоматизація жестового мовлення; анімований персонаж; відслідковування положення тіла; люди з вадами слуху; мова жестів; нейронні мережі; розпізнавання жестів; українська жестова мова; сурдопереклад; скорочення нерівності.

Бібліографічні описи / Bibliographic descriptions

Шовковий Є. І., Гриньова О. Є., Удовенко С. Г., Чала Л. Є. Система автоматичного сурдоперекладу з використанням нейромережних технологій та 3D-анімації. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 108–121. DOI: <https://doi.org/10.30837/ITSSI.2023.26.108>

Shovkovyi, Y., Grinyova, O., Udovenko, S., Chala, L. (2023), "Data structures for deductive simulation of HDL conditional operators", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 108–121. DOI: <https://doi.org/10.30837/ITSSI.2023.26.108>

О. ШКІЛЬ, Д. РАХЛІС, І. ФІЛІПЕНКО, В. КОРНІЄНКО

ПРОЄКТУВАННЯ ТА САМОДІАГНОСТИКА КІБЕРФІЗИЧНИХ ПРИСТРОЇВ КЕРУВАННЯ НА ПЛАТФОРМІ SoC

Предметом дослідження в статті є моделі, методи та процедури проєктування та самодіагностики автоматних моделей пристроїв логічного керування, реалізованих в SoC. **Об'єкт роботи** – процедури автоматизованого проєктування та діагностування цифрових пристроїв на технологічній платформі SoC. **Метою дослідження** є розроблення моделей і процедур проєктування та самотестування в циклі автоматизованого проєктування автоматних систем логічного управління на технологічній платформі SoC, що суттєво підвищить надійність їх функціонування. У статті вирішуються такі **завдання**: розгляд процедур взаємодії процесорного ядра з програмованою логікою у складі SoC; удосконалення процедур проєктування та тестування програмно-апаратних систем на основі SoC; подальший розвиток процедур автоматизованого проєктування, верифікації та діагностування кіберфізичних систем логічного управління з використанням мов програмування та мов опису апаратури; реалізація процедури апаратного самотестування керуючих автоматів на технологічній платформі SoC. Упроваджуються такі **методи**: синтез керуючих автоматів на основі графових моделей, імплементація моделей керуючих автоматів мовою програмування C з використанням автоматного шаблону, діагностичний експеримент способом обходу графа переходів автомата. **Досягнуті результати**. На основі аналізу процедур взаємодії процесорного ядра та програмованої логіки на обраній платформі SoC спроектовано модель кіберфізичної системи логічного управління. Практичну реалізацію виконано на базі стеку інструментальних засобів САПР Vivado/Vitis/Vitis HLS. Реалізовано метод апаратного самотестування керуючих автоматів на технологічній платформі SoC ZYNQ-7000. **Висновки**. У статті проаналізовано принципи проєктування вбудованих кіберфізичних систем, що реалізуються в системах на кристалі. Розглянуто принципи побудови систем верифікації та вбудованої самодіагностики систем на кристалі, що містять програмну й апаратну частини. Розроблені методи апробовано на моделі пристрою логічного керування світлофором на технологічній платформі SoC FPGA сімейства ZYNQ-7000 фірми Xilinx. Керуючий автомат Мура реалізовано у блоці PL мовою програмування C, а операційний автомат – у блоці PS. Під час організації процесу самодіагностики здійснено неруйнівний діагностичний експеримент способом обходу всіх дуг графа переходів, починаючи з початкової вершини. Тестером у цьому разі був операційний автомат, еталонні логічні та часові значення якого зберігалися в пам'яті блока PS. Візуальне спостереження за виконанням діагностичного експерименту здійснювалося за допомогою панелі світлодіодів плати ZedBoard.

Ключові слова: кіберфізичні системи; вбудовані системи; логічне управління; проєктування систем на кристалі; FPGA; САПР; самотестування SoC; мова програмування C.

Вступ

Вбудовані цифрові пристрої та системи за останні 50 років зазнали значних змін не тільки в технологічному, а й в архітектурному аспекті. Якщо впродовж 60–70 років минулого століття вбудованою цифровою системою було прийнято називати спеціалізовані пристрої керування, спроектовані на логічних інтегральних схемах малого й середнього ступеня інтеграції, то сучасне розуміння вбудованої системи визначає її у вигляді замовної надвеликої інтегральної схеми (НВІС), що за своєю сутністю є спеціалізованою обчислювальною системою, технологічно виготовленою у вигляді системи на кристалі. Останні роки з'явилося розуміння, що більшість проблем вбудованих систем є наслідком складнощів в умовах взаємодії з фізичними процесами, а не через обмеженість

ресурсів. У 2006 р. було запропоновано термін "кіберфізичні системи" (*Cyber-Physical Systems, CPS*), що передбачає інтеграцію обчислень із фізичними процесами. Поняття позначає очікуваний у найближчому майбутньому якісний перехід у сприйнятті вбудованих систем і методів їх проєктування. Сутність CPS полягає в тому, що проєктування об'єкта керування та системи управління для цього об'єкта мають виконуватися в єдиному ключі, в єдиному комплексі інструментальних засобів, що тісно взаємодіють [1].

Система на кристалі (*System on Chip, SoC*) побудована на єдиному кристалі, у якій інтегруються такі елементи, як процесор (процесори, зокрема спеціалізовані), пам'ять, кілька периферійних пристроїв, спеціалізовані обчислювальні блоки та їх з'єднання. Усе згадане вище становить оптимальний набір для деякого задалегідь відомого

функціонала, наприклад, оброблення та передачі відеоінформації. Вислів "система на кристалі" не є насправді терміном. Це поняття відбиває загальну тенденцію до підвищення рівня інтеграції за допомогою інтеграції функцій [2].

SoC часто застосовуються як компоненти вбудованих систем, і для них використовуються одні й ті самі методи проектування [3]. Ідеалізований маршрут проектування *SoC* містить такі кроки: створення специфікації та моделювання, попередній аналіз перед поділом на апаратні та програмні частини, поділ на апаратні та програмні частини, аналіз, налагодження та усунення виявлених помилок після поділу, верифікація після поділу, реалізація апаратури, реалізація програмного забезпечення (ПЗ), верифікація реалізації.

Сучасні вимоги до проектування *SoC* привели до появи таких нових підходів, як сумісне проектування (співпроектування) та сумісна верифікація (співверифікація) [4]. Співпроектування – це процес паралельного проектування апаратних і програмних засобів, за умови якого оцінюється доцільність вибору апаратної або програмної реалізації певного фрагмента проекту. Цей процес також дає проектувальникам змогу побачити, як система могла б працювати з поділом обчислень між апаратурою та ПЗ. Співверифікація – це аналіз спільної роботи програмного забезпечення та апаратних засобів *SoC* з метою визначення, чи вони функціонуватимуть правильно разом. З іншого боку, під час цього процесу аналізується, чи однаково вирішуватимуться особливі завдання проекту в умовах апаратної чи програмної реалізації. Результатом роботи є фіксація обраного варіанта реалізації.

Однією з основних проблем проектування контролепридатних цифрових пристроїв є реалізація їх вбудованого самотестування. Самотестування програмно-апаратних систем на кристалі призначене для виявлення та діагностики потенційних помилок апаратного й програмного забезпечення всередині самого *SoC*. Запускаючи комплексне самотестування, розробники вбудованих систем можуть заздалегідь виявляти й усувати будь-які проблеми, здатні поставити під загрозу загальну надійність і функціональність системи. Ці тести охоплюють значну кількість різних сфер, зокрема пам'ять, порти введення-виведення, таймери, периферійні пристрої та навіть процесорну частину.

У багатьох галузях, особливо в тих, що пов'язані з критично важливими щодо безпеки застосунками,

діють суворі нормативні стандарти, яких необхідно дотримуватись. Дотримання стандартів є необхідним забезпеченням якості та надійності вбудованих систем. Тому самотестування відіграє вирішальну роль у виконанні цих вимог. Самотестування *SoC* допомагає в процесі налагодження, надаючи цінну інформацію про стан *SoC* і пов'язані з ним компоненти. Залучивши комплексні механізми самотестування, розробники можуть продемонструвати, що їх системи пройшли ретельне тестування та відповідають необхідним нормативним критеріям.

Аналіз останніх досліджень і публікацій

У роботі [5] визначено роль і місце систем логічного управління в класі реактивних систем, що активуються у відповідь на зовнішні події. Для опису частини керування систем логічного управління застосовують кінцеві автомати, зазвичай у формі моделей Мура. Для подання керуючих автоматів у системах логічного управління використовують графи переходів, що є не лише візуальним зображенням алгоритму роботи автомата, але і його повною математичною моделлю. У цій праці також запропоновано методику розроблення автоматних систем логічного управління з огляду на реальний час та оброблення зовнішніх подій.

Поняття часового автомата (*timed automata*) як засіб опису систем реального часу розглянуто в роботі [6]. Граф переходів автомата доповнюється скінченним набором таймерів, що приймають дійсні значення. Кожен таймер скидається на нуль у момент переходу та збільшує своє значення з кожним тактом автомата. З кожним переходом пов'язано часове обмеження (*clock constraint*), воно означає, що цей перехід може відбутися лише в разі, коли поточні значення таймера відповідають цьому обмеженню. З кожною позицією пов'язане обмеження на таймери, яке називається інваріантом; система може перебувати в цій позиції лише доти, доки виконується її інваріант.

У праці [7] досліджуються методи тестування, що беруть до уваги часові характеристики технічної системи. Для опису поведінки системи використовується модель часового кінцевого автомата *TFSM* (*Timed Finite State Machine*). У процесі розроблення тестів для часових автоматів розглядається модель *TFSM*, що зважає на тайм-аут (*timeout*) у станах і затримки вихідних сигналів щодо реалізації переходу в стан. Водночас береться до уваги, що якщо протягом тайм-ауту не надійшло

жодного вхідного сигналу, то автомат переходить у наступний стан.

Загалом модель часового автомата передбачає три типи часових параметрів: тайм-аут у станах, часове обмеження на прийом вхідних сигналів і час оброблення вхідного сигналу, тобто затримка вихідного сигналу щодо вхідного. У цьому разі можуть досліджуватися часові автомати з меншою кількістю параметрів. У роботі [8] розглядаються завдання мінімізації часових автоматів, перевірки їх еквівалентності та синтезу тестів для перевірки.

У праці [9] порушуються проблеми побудови апаратних подієвих систем логічного управління реальним часом на програмованих логічних інтегральних схемах (ПЛІС). Алгоритм управління реалізується на основі моделі часового автомата (*timed FSM*), поданого темпоральним графом переходів (*temporal state diagram*). Побудова моделі пристрою керування виконана мовою опису апаратури *VHDL* у формі трипроцесного шаблону з обробленням зовнішньої події. Функціональна верифікація моделі здійснювалася з використанням інструментальних засобів *Active-HDL*, синтез схемної реалізації виконаний на технологічній платформі ПЛІС *Spartan 3E* інструментальними засобами САПР *Xilinx ISE*. Проаналізовані апаратні витрати на схемну реалізацію пристрою керування.

У роботі [10] для реалізації систем логічного управління реального часу була запропонована модель часового автомата Мура з тайм-аутами та часовими обмеженнями з використанням додаткового лічильника для зберігання значення часової змінної. На основі цієї моделі був розроблений темпоральний граф переходів для системи управління дорожнім світлофором. На його основі розроблена двопроцесна *VHDL*-модель часового автомата Мура. За допомогою середовища *Xilinx ISE* виконані верифікація, синтез та імплементація розробленої *VHDL*-моделі. Синтез і моделювання до та після імплементації здійснювалися для мікросхем *CPLD XC9572XL-10-TQ100* та *FPGA XC3S500E-5fg320*. Результати синтезу та моделювання схеми після імплементації підтвердили працездатність і коректність розробленої *VHDL*-моделі.

У монографії [11] описано методи синтезу базових компонентів вбудованих цифрових пристроїв. Особливу увагу приділено синтезу вбудованих цифрових пристроїв з мікропрограмним управлінням. Подано аспекти впровадження елементів надійності на початкових стадіях проектування цифрових

пристроїв, зокрема й методи захисту цифрового контенту вбудованих систем. Розглянуті принципи побудови самотестованих цифрових пристроїв на основі вбудованих засобів для формування тестових послідовностей, аналізу реакцій із подальшим їх порівнянням з еталонними значеннями.

Деталі використання високорівневого синтезу (*High Level Synthesis, HLS*) для *FPGA* подано в роботі [12], докладно наведено приклади використання *HLS* інструментів та їх специфічні оптимізації для *FPGA*-платформ. Розглянуто використання *HLS* для розв'язання задач у сферах кодування інформації, оброблення зображень і аудіо в реальному часі.

У публікації [13] порушується проблема високопродуктивного потокового генерування випадкових чисел у діапазоні рівномірного й нормального розподілу в ПЛІС. Робота зосереджена на легкій реалізації, придатній для широкого спектра ПЛІС. Спочатку розглядаються наявні типи модулів генерації випадкових чисел і описано побудову розробленого генератора. Його розподілено на дві частини: реалізація потокового генератора рівномірних чисел і потоковий генератор гауссівських чисел на основі кумулятивного розподілу. У роботі модулі реалізовані за допомогою мови синтезу високого рівня (*C/C++*), протилежно до типових підходів на рівні мов опису апаратури (*HDL*).

У дослідженні [14] описуються інструменти високорівневого синтезу, що дають змогу розробляти спеціалізовані апаратні прискорювачі (*HWacc*). Проте етап верифікації все ще є найдовшим етапом у життєвому циклі розробки. На відміну від програмної індустрії, в інструментів *HLS* відсутні фреймворки тестування, які могли б охоплювати весь процес розроблення, особливо етап верифікації на бортовому рівні згенерованого опису рівня регістрових передач (*Register Transfer Level, RTL*). У статті запропоновано фреймворк для верифікації модулів, створених на основі *HLS*, з використанням реконфігурованої системи та контейнерів (*Docker*) з метою автоматизації процесу верифікації та збереження чистого середовища для тестування, зробивши тестову систему перехідною між різними етапами розроблення.

У дослідженні [15] розглядається використання потужних інтегральних схем на ринку для різноманітних застосувань із різними вимогами й завантаженнями обчислень. Особлива увага в роботі приділяється системам, що застосовуються в галузі високопродуктивних обчислень

(*High Performance Computing*). У цій галузі основними аспектами, що потрібно брати до уваги, є продуктивність (за визначенням) та споживання енергії (оскільки витрати на операційну діяльність переважають витрати на закупівлю).

У праці [16] оцінюється можливість використання *HLS* для ефективної імплементації загального коду на різних цільових платформах. Порівнювалися такі платформи: *ASIC* (система зі спеціалізованими інтегральними мікросхемами) та *FPGA* (програмовані логічні інтегральні схеми). Можливість використовувати одну кодову базу для обох платформ дозволить прототипування на *FPGA* та швидкий цикл розроблення. Для оцінювання придатності використання коду створюється фільтр із кінцевою імпульсною характеристикою.

Огляд публікацій доводить актуальність створення нових технологій проєктування та самодіагностики систем на кристалі.

Мета цього дослідження – розроблення моделей і процедур проєктування та самотестування в циклі автоматизованого проєктування автоматних систем логічного управління на технологічній платформі *SoC*, що суттєво підвищить надійність їх функціонування.

Для досягнення поставленої мети необхідно виконати такі завдання:

- розглянути процедури взаємодії процесорного ядра з програмованою логікою в складі *SoC*;
- удосконалити процедури проєктування та тестування програмно-апаратних систем на основі *SoC*;
- удосконалити процедури автоматизованого проєктування, верифікації та діагностування кіберфізичних систем логічного управління з використанням мов програмування та мов опису апаратури;
- реалізувати процедури апаратного самотестування керуючих автоматів на технологічній платформі *SoC*.

Апаратна платформа *SoC Zynq-7000*

Одним із прикладів програмованої системи на кристалі є *FPGA* сімейства *ZYNQ-7000* фірми *Xilinx Inc.* Для побудови *SoC* різного призначення платформа *Zynq-7000 EPP* містить такі функціональні блоки: процесорну підсистему, що містить процесорний модуль, інтерфейси пам'яті, периферійні інтерфейси, міжблокові інтерфейси та інтерфейси

до програмованої логіки, а також програмовану логіку. Процесор *ARM Cortex A9 MPCore* має вбудовану пам'ять, багатий набір периферійних пристроїв, інтерфейси до зовнішньої пам'яті. Взаємодія між двома процесорами може здійснюватися за допомогою міжпроцесорних переривань крізь ділянку загальної пам'яті способом передачі повідомлень [17].

Інтегральна схема *ZYNQ-7000 EPP (Extensible Processing Platform)* виконана за технологічним процесом *SRAM 28 нм* і є *FPGA* з упровадженими додатковими функціональними блоками. Крім 350 тис. логічних блоків (для *FPGA Z-7045*), кристал містить багатоядерний блок обчислювачів, побудований на базі двох ядер процесора *ARM Cortex-A9*, 2 Мбіт вбудованої багатопортової пам'яті, контролери зовнішніх динамічних оперативних запам'ятовувальних пристроїв (*DDR2* і *DDR3*), контролери пристроїв *flash*-пам'яті (*NAND* і *NOR*), два вбудовані блоки високошвидкісних 12-розрядних аналого-цифрових перетворювачів, вбудований контролер інтерфейсу *PCIe*, вбудовані контролери спеціалізованих інтерфейсів, таких як *I2C*, *USB 2 Ethernet*, *UART*, *CAN*, *SPI* тощо.

Платформа *Xilinx ZYNQ-7000* має архітектуру, де поєднано систему оброблення (*processing system, PS*) та програмовану логіку (*programmable logic, PL*). Взаємодія двох систем відбувається з допомогою інтерфейсу *AMBA* на основі протоколу *AXI*.

AXI – це двоточковий інтерфейс, розроблений для високопродуктивних і швидкодіючих мікроконтролерних систем. Протокол *AXI* оснований на каналі "точка-точка", що передбачає спільне використання шини та дає змогу збільшити пропускну здатність і зменшити затримку. *AXI* є найпопулярнішим серед усіх міжз'єднань інтерфейсу *AMBA*. Він забезпечує взаємодію різних блоків усередині кожного чипа. Механізм квітування гарантує, що інформація передається чітко й безперерійно. Це дає змогу різним компонентам взаємодіяти без будь-яких перешкод.

Маршрут проєктування *SoC*

Процес проєктування вбудованих систем, що реалізуються на базі кристалів програмованої логіки з архітектурою *FPGA* і розширюваних обчислювальних платформ сімейства *Zynq-7000 AP SoC*, загалом передбачає шість етапів [17].

1. Розроблення проєкту мікропроцесорної системи. На цьому етапі визначається архітектура

системи й розподіл функцій щодо виконання обчислень між апаратною та програмною частиною із використанням обраних критеріїв та метрик (енергоспоживання, швидкість реакції системи, оптимальність виконання блоків). Також додається необхідність вибору цільового процесора або *SoC* для подальших етапів проєктування.

2. Проєктування апаратної платформи системи, яка розробляється. Етап передбачає формування проєкту апаратної платформи та визначення IP-ядер, необхідних для конкретного завдання. Наступними завданнями на цьому етапі є реалізація апаратної платформи на базі обраного чипа та її верифікація.

3. Підготовка системних програмних засобів нижнього (апаратного) рівня. На цьому етапі відбуваються розроблення, налагодження первинного завантажувача системи (*First Stage BootLoader, FSBL*) та формування пакету підтримки плати (*Board Support Package, BSP*). Під час розроблення *FSBL* можуть бути розгорнуті базові засоби для оновлення (*Over-The-Air update, OTA*), що спрощує подальші етапи тестування системи. Створення *BSP* передбачає формування необхідного пакету драйверів системи та їх тестування / інтеграцію / адаптацію для обраного набору периферії.

4. Формування основного програмного забезпечення системи, що розробляється. На цьому етапі формується архітектура ПЗ системи та пишеться код застосунків з одночасним виконанням типового ітераційного створення ПЗ, орієнтованої на тестування (*Test Driven Development, TDD*).

5. Комплексне моделювання та налагодження. Під час цього етапу виконується інтеграція розроблених програмних і апаратних компонентів на цільовій платформі. Комплексне налагодження містить як перевірку інтеграції компонентів, так і налагодження граничних станів системи (режим сну, вихід із режиму сну, перехід у режим енергозбереження, навантажувальне тестування та перевірки безпеки системи загалом).

6. Генерація завантажувального образу та розгортання розробленої системи. Для обраного варіанта завантаження системи генерується образ, який може містити такі компоненти: образ для прошивання енергонезалежної вбудованої пам'яті (*Embedded Multimedia Memory Card, eMMC*), образ для розгортання системи із застосуванням завантажувача *NFS boot (Network File System)* для розробників, генерація та прошивання *fuse*-конфігурації для захисту системи.

Інструментальні засоби

Після вибору варіанта конфігурації процесорного блоку, що забезпечує необхідну продуктивність у процесі виконання функцій вбудованої системи, можна перейти безпосередньо до проєктування її апаратної платформи.

Розроблення вбудованого програмного забезпечення здійснюється з використанням *Vitis IDE*. Узагальнений процес проєктування ПЗ на базі *Vitis IDE* є типовим циклом для вбудованої платформи. Основна частина – це конфігурація платформи для цільової плати та генерація *Software Development Kit (SDK)*. Після цього етапу можна скористатися *Application Programming Interface (API)* для роботи в режимі *baremetal* або з додаванням *FreeRTOS*.

Вбудоване середовище розробки *Vitis IDE* містить готові базові приклади роботи з *Zynq-7000*, серед яких застосунок тесту *DDR* пам'яті, перевірка *LwIP* мережного стеку. Після створення апаратної частини у *Vivado* та генерації необхідних файлів доцільно перевірити цілісність пам'яті та виконати застосунок *memory_test*, результати якого будуть доступні в консолі послідовного порту, до якого підключено плату. Зазначимо, що для програмування цільової платформи може бути використаний як офіційний *Xilinx Platform Cable*, так і *Xilinx Virtual Cable Protocol*, що дає змогу прошити й налагодити віддалений пристрій.

Узагальнене проєктування IP-ядер виконується в середовищі *Vivado IDE* із застосуванням як готових блоків, так і користувацьких ядер. Після створення IP-ядра можна контролювати версії та, відповідно, оновлювати блоки, а також інтегрувати ядра через процес *System Block Designer*, де можна налаштувати параметри блоку та інтегрувати його разом до *PL*-частини. Одним із доступних варіантів генерації IP-ядер є використання *Vitis HLS* для роботи з високорівневим синтезом. Також може бути застосований класичний підхід з *Verilog/VHDL*-описом і відповідним *testbench*.

Після створення IP-ядра необхідно під'єднати основні комунікаційні інтерфейси, а саме *AXI*-шину та вхідні / вихідні порти. Для цього потрібно модифікувати згенерований шаблон мовою опису апаратури *Verilog/VHDL*.

Процес верифікації IP-блоку передбачає написання відповідного *testbench* та інтеграцію в *System Block Design*. Зазначимо, що після

проходження верифікації та збирання IP-блок стає доступним у репозиторії проекту, звідки може бути доданий до системи та налаштований. Основними параметрами налаштування можуть бути як константи, що змінюються для конкретного дизайну, так і параметри *AXI*-шини.

У типовому процесі дизайну для платформи *ZYNQ* доцільно використовувати інструментарій *Vitis HLS* для синтезу IP-блоків і *Vivado* – для проектування системи. Написання програмного забезпечення для *ARM* (*PS*-частини) здійснюється в середовищі *Vitis IDE*. На рис. 1 зображений типовий маршрут використання інструментів *Xilinx*. Цей процес є реалізацією окремих IP-блоків у середовищі *Vitis HLS* та написання відповідних *testbench* для них, після чого імпорт до середовища *Vivado*, де інтегруються блоки IP з *PS*-частиною *ZYNQ*. Після синтезу та отримання бітстріму для конфігурації *FPGA*-частини *ZYNQ* виконується операція *Export Hardware including bitstream* для отримання *xsa*-файлу. Далі, відповідно до створеної архітектури системи, можна або згенерувати образ *Linux* із використанням *petalinux* на базі *Yocto Framework* та *xsa*-файлу, або імпортувати отриманий *xsa*-файл до *Vitis IDE* й застосовувати *baremetal SDK* від *Xilinx*. Також можна проводити налагодження в середовищі *Vitis IDE* для крос-компільованого застосунку на базі *PetaLinux*.

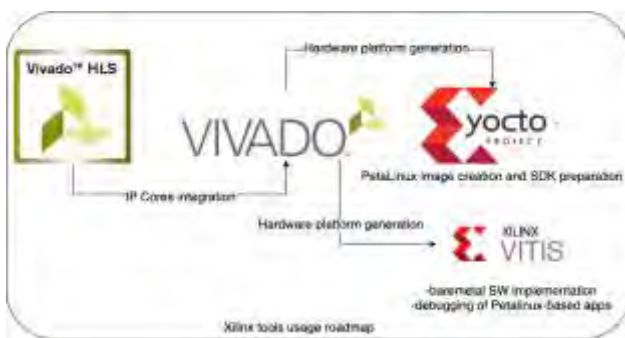


Рис. 1. Маршрут використання інструментів *Xilinx* для платформи *ZYNQ*

Пристрій логічного керування дорожнім світлофором

Розглянемо алгоритм функціонування дорожнього світлофора, що працює у двох режимах: денному й нічному. Множина вхідних сигналів $X = \{Onn, St, Btn\}$, де $Onn = \{0, 1\}$ – сигнал увімкнення дорожнього світлофора та запуску

нічного циклу; $St = \{0, 1\}$ – сигнал запуску денного циклу роботи світлофора; $Btn = \{0, 1\}$ – сигнал увімкнення зеленого світла на пішохідному переході. Отже, Onn і St є оповіщувальними сигналами, а Btn – подією.

Множина вихідних сигналів для світлофорів $Y = \{R1, YRG, YGR, G1, R2, G2\}$, де $R1$ – сигнал увімкнення червоного світла на основній дорозі; $G1$ – сигнал увімкнення зеленого світла світлофора на основній дорозі; YRG – сигнали увімкнення жовтого світла на основній дорозі (за умови зміни $R - G$); YGR – сигнали увімкнення жовтого світла на основній дорозі (у разі зміни $G - R$); $R2$ – сигнал увімкнення червоного світла на пішохідному переході; $G2$ – сигнал увімкнення зеленого світла на пішохідному переході.

Інтерфейс системи управління дорожнім світлофором зображений на рис. 2.

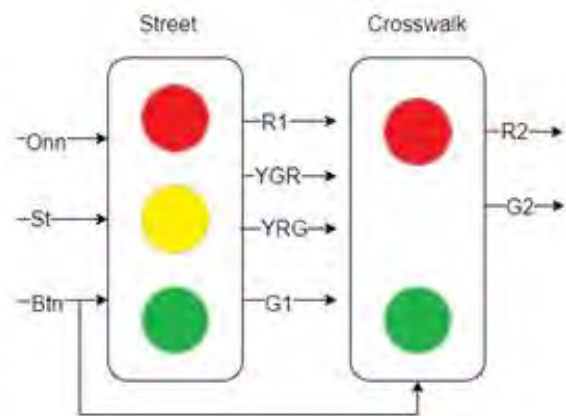


Рис. 2. Інтерфейс системи управління дорожнім світлофором

Визначимо стани керуючого автомата:

- стан $a1$ – увімкнення автомата, вихідних сигналів немає, затримка t_{01} ;
- стан $a2$ – жовтий за умови зміни $G - R$, виходи $\{YGR, R1, R2\}$, затримка t_{02} ;
- стан $a3$ – червоний на дорозі, зелений на переході, виходи $\{R1, G2\}$, затримка t_{03} ;
- стан $a4$ – жовтий за умови зміни $R - G$, виходи $\{YRG, R1, R2\}$, затримка t_{02} ;
- стан $a5$ – зелений на дорозі, червоний на переході, виходи $\{G1, R2\}$, затримка t_{03} ;
- стан $a6$ – увімкнення зеленого на переході та червоного на дорозі по кнопці Btn , затримка t_{06} ,

виходи $\{R1, G2\}$ із затримкою появи t_d , затримка у стані to_6 ;

- стан $a7$ – горить тільки жовтий, затримка to_1 .

Наведемо алгоритм роботи світлофора. Під час увімкнення пристрою керування ($Onn = 1$) запускається нічний цикл роботи світлофора, відбувається миготіння жовтого світла на основній дорозі ($a_1 - a_7$), світлофор на пішохідному переході не працює. Після запуску денного циклу ($St = 1$) реалізується така система переходів: $a_2 - a_3 - a_4 - a_5 - a_2$. У стані a_5 , коли на основній дорозі горить зелений, визначено вікно прийому (*time constraint*) t_c для зовнішньої події Btn (натиснення кнопки переходу). У процесі оброблення цієї події керуючий автомат переходить у стан a_6 . У цьому разі на основній дорозі загоряється червоний, а на пішохідному переході затримується червоний, а зелений загоряється із затримкою t_d (час на підготовку до переходу). За період t_c може бути прийнятий тільки один сигнал (зовнішня подія) Btn ; $t0$ – початок

вікна; t_c , t_d – затримка формування вихідних сигналів; t_c – вікно прийому зовнішньої події.

Отже, для системи управління дорожнім світлофором може бути побудований темпоральний граф переходів часового автомата Мура (рис. 3).

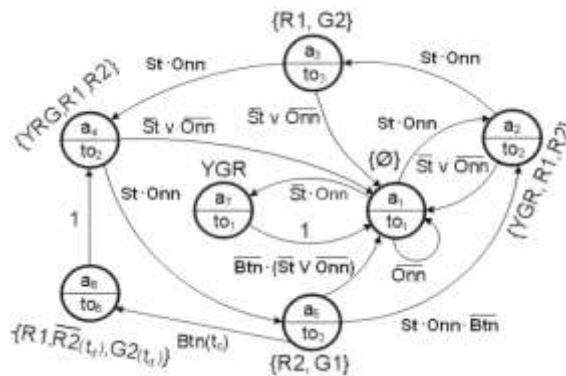


Рис. 3. Темпоральний граф переходів автомата Мура для системи управління дорожнім світлофором

На рис. 4 подана часова діаграма функціонування часового автомата Мура для системи управління дорожнім світлофором у денному циклі роботи, яка, власне, є специфікацією пристрою, що проектується.

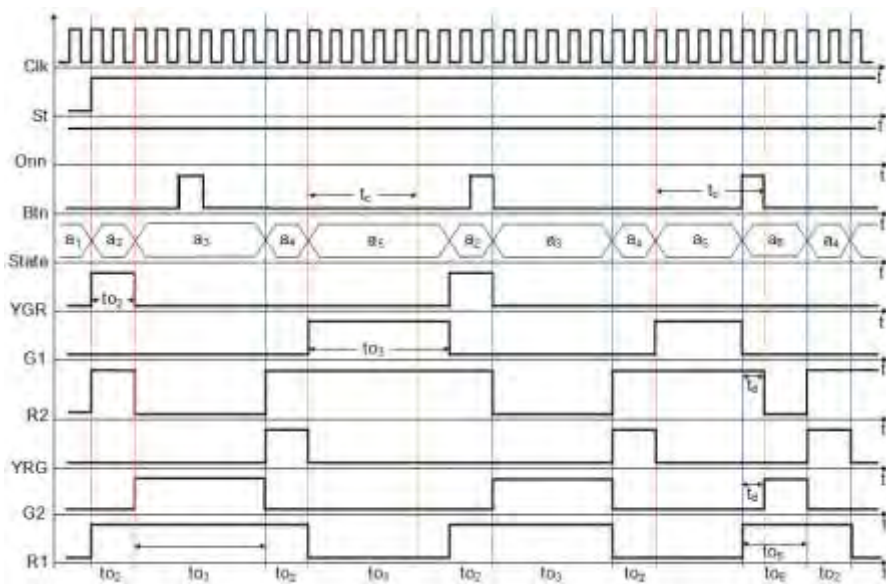


Рис. 4. Часова діаграма автомата Мура для системи управління дорожнім світлофором

Реалізація в SoC

У разі наявності графової моделі Мура для автоматного пристрою керування функціональне покриття специфікації може бути зведено до обходу всіх станів автомата Мура та реалізації всіх переходів у кожному стані. Під час проведення верифікації

автоматних моделей реалізується так званий "неруйнівний" діагностичний експеримент (ДЕ), у процесі проведення якого тест (обхід графа) починається з початкової вершини графа переходів і повертається до неї. В організації ДЕ із самодіагностики пристрою керування на SoC як тестер використовується операційний автомат (ОА),

що імітує сигнали Onn , St та Btn . Еталони зберігаються у пам'яті PS у вигляді значень $R1$, YGR , YRG , $G1$, $R1$, $G2$ у відповідні проміжки часу, а проміжки часу у вигляді констант T_i також зберігаються в пам'яті PS . Шляхи обходу графа переходів зберігаються в пам'яті PS і реалізуються способом задання параметра t у певні проміжки часу.

Керуючий автомат (КА) виробляє вихідні сигнали $R1$, YGR , YRG , $G1$, $R1$, $G2$, формує з них бінарний вектор та передає його до ОА. Там цей

вектор порівнюється з еталоном і результат порівняння крізь блок введення / виведення передається на контрольний світлодіод для візуального спостереження. Крім того, сигнали $R1$, YGR , YRG , $G1$, $R1$, $G2$ безпосередньо з КА крізь блок введення / виведення PL передаються на світлодіодну панель вихідних сигналів для візуального спостереження. Узагальнений вигляд системи наведено на рис. 5.

Операційний автомат у режимі проведення ДЕ імітує видачу вхідних сигналів для КА.

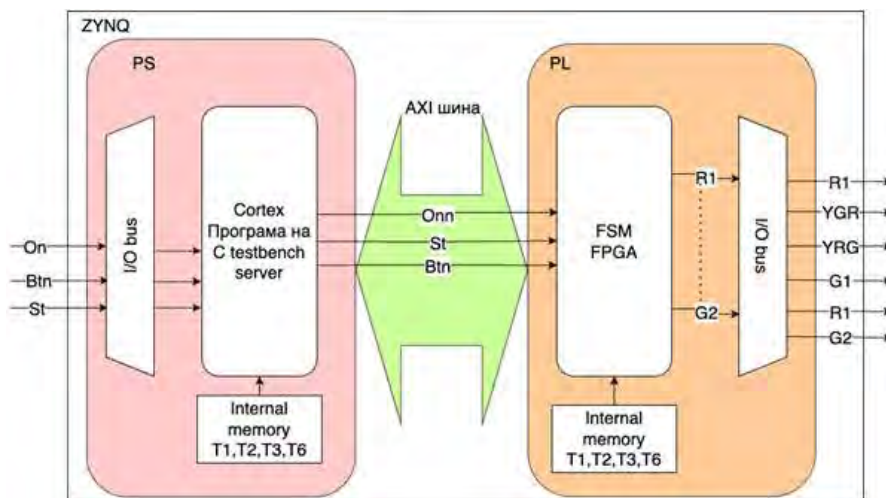


Рис. 5. Реалізація пристрою керування в SoC ZYNQ-7000

Побудова моделі пристрою у САПР Vivado

У середовищі *Vitis HLS* із підтримкою *High Level Synthesis* логіку КА реалізовано з використанням *pragma*-директив компілятора, що дають змогу призначити конкретний інтерфейс, що буде застосовано під час синтезу IP-блоку. Перед процесом синтезу необхідно вказати *top-level-function*, яка має бути використана для синтезу, й очікуваний *clock*-період блоку. Окремо потрібно визначити модель *SoC*, яка може бути застосована для *co-simulation* під час налагодження IP-блоку. Усі моделі логіки роботи *SoC* як блоку PS , так і блоків PL та AXI виконані мовою програмування C . Реалізація логіки FSM виконана у форматі автоматного шаблону. Як приклад наведемо фрагмент реалізації стану $a5$ у коді моделі FSM (рис. 6).

Після розроблення IP-блоку FSM виконується його синтез та експорт як IP до середовища *Vivado*. Інтеграція блоку до середовища *Vivado* виконана з допомогою типового процесу додавання користувальницького репозиторію IP-блоків та блоку

на *Block Design*. Усі розроблені блоки з використанням технології *Vitis HLS* окремо позначаються на діаграмі. Також система містить *ZYNQ Processing System*, IP-ядро генерації сигналу $RESET$ та IP-блок взаємодії з AXI -шиною *AXI Interconnect*. Блок *status leds* використано для візуального відображення інформаційних сигналів (рис. 7).

```

case fsm_states::A5:
  if (input_signals & Btn && m_counter >
      constraint_a5_l - 1 && m_counter <
      constraint_a5_h)
    (m_next_state = fsm_states::A6;)
  else if (m_counter < T3 - 1)
    {m_next_state = fsm_states::A5;
     ++m_counter;}
  else if (!(input_signals & OnSignal))
    {m_next_state = fsm_states::A1;}
  else if (!(input_signals & St))
    {m_next_state = fsm_states::A1;}
  else {m_next_state = fsm_states::A2; }
  break;

```

Рис. 6. Реалізація логіки перебування в стані $a5$ та переходів між відповідними станами

Реалізація логіки діагностування FSM виконана на ARM-частині з використанням мови програмування C. Vitis HLS дає змогу застосувати базову логіку з testbench IP-блоку. Налаштування

програмного забезпечення PS-частини виконується в середовищі Vitis IDE з використанням Xilinx Platform Cable, що робить можливим взаємодію з ZYNQ з допомогою JTAG-інтерфейсу.

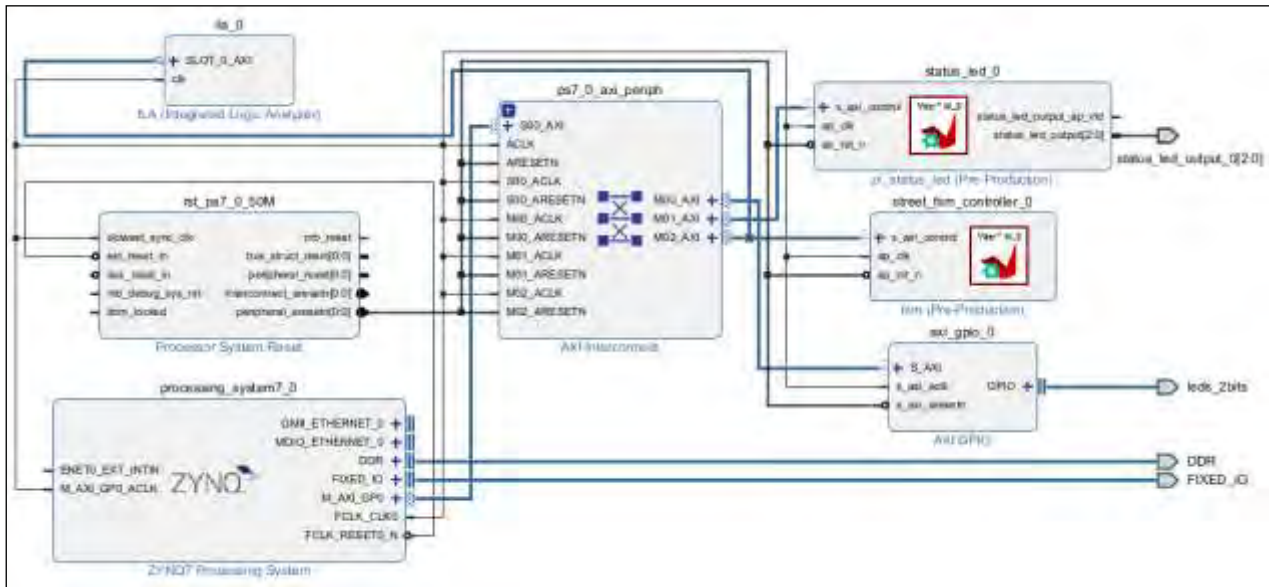


Рис. 7. Діаграма розробленого проекту SoC у середовищі Vivado ISE

Діагностичний експеримент

В організації процесу самодіагностики реалізується неруйнівний діагностичний експеримент (ДЕ) способом обходу всіх дуг графа переходів

керуючого автомата, починаючи з початкової вершини (рис. 8).

Під час виконання неруйнівного діагностичного експерименту було отримано часові діаграми з вбудованого логічного аналізатора IP-Core Xilinx Integrated Logic Analyzer.

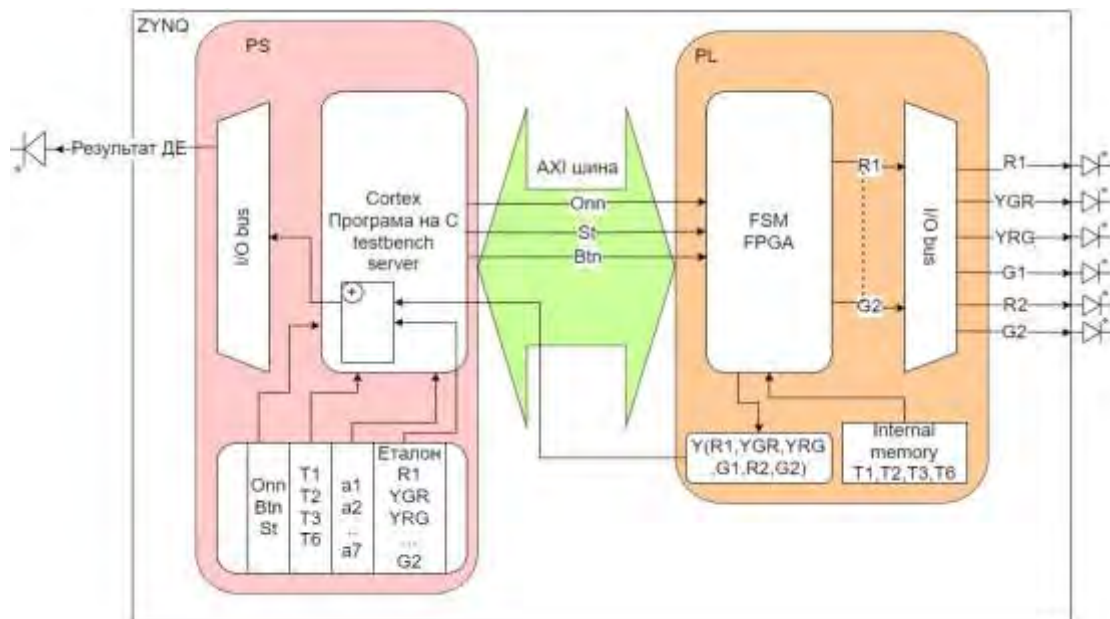


Рис. 8. Реалізація ДЕ пристрою керування для SoC Zynq-7000

Для перевірки коректності оброблення транзакцій AXI-шини було виконано під'єднання аналізатора до відповідної шини обміну в системі. У разі використання AXI-шини виходи закодовано через зсув одиниці у відповідному 32-бітному регістрі. Результати проведення ДЕ показані способом реалізації типового циклу роботи FSM $a_2 - a_3 - a_4 - a_5 - a_2$.

Відповідно, у стані a_3 активними сигналами керування є $R1$ та $G2$. Їх об'єднання дорівнює $R1 | G2$

або $1 \ll 0 | 1 \ll 5$, що у HEX дорівнює 0×21 . Відповідні дані отримані по AXI-шині на лінії *streetlight_fsm_bd_i/ps7_0_axi_periph_M02_AXI_RDATA[31:0]*.

На рис. 9 наведено стан виходів автомату для стану a_3 ($R1 = 1, G1 = 1$), отриманий за допомогою вбудованого логічного аналізатора.

Таким самим чином була отримана діаграма для стану a_5 після подачі відповідних тестових сигналів з боку PS-частини системи. На рис. 10 наведено діаграму з логічного аналізатора для стану a_5 .

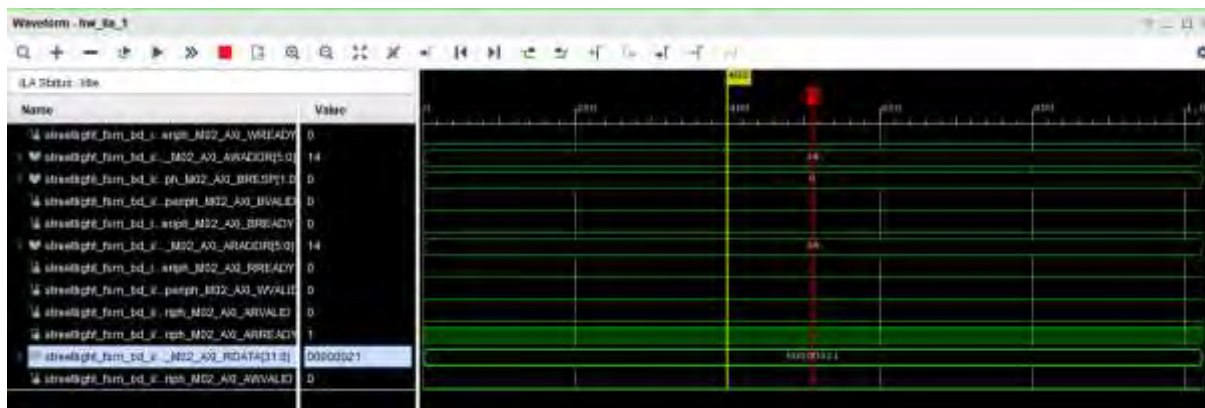


Рис. 9. Часова діаграма з логічного аналізатора сигналів керування для стану a_3

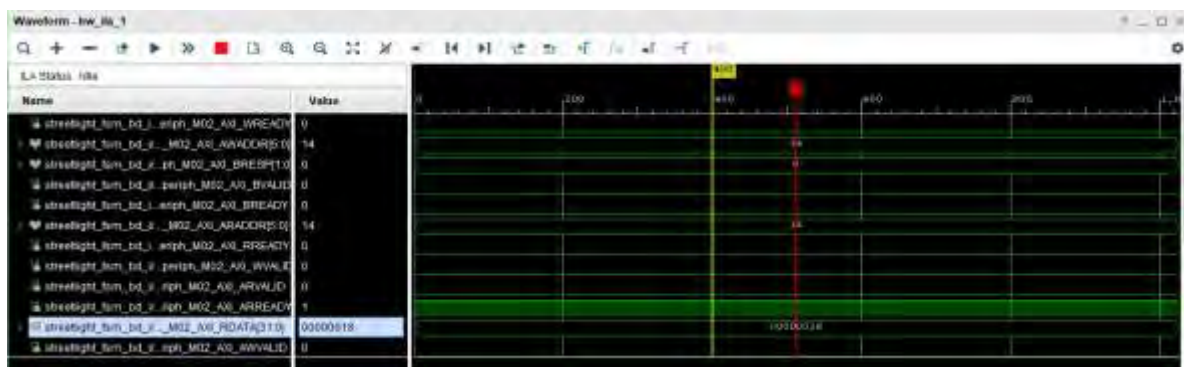


Рис. 10. Часова діаграма з логічного аналізатора сигналів керування для стану a_5

Порівняння отриманих часових діаграм з еталонною специфікацією (рис. 4) показало повний збіг, що підтверджує коректну працездатність спроектованого пристрою.

Висновки та перспективи подальшого розвитку

Унаслідок виконання роботи порушено питання побудови системи самодіагностики кіберфізичної системи логічного управління, реалізованої на технологічній платформі системи на кристалі. Проаналізовано принципи побудови вбудованих

кіберфізичних систем, які реалізуються в системах на кристалі. Розглянуто методи й маршрути проектування систем на кристалі, що містять програмну та апаратну частини. Описано принципи побудови систем верифікації та вбудованої самодіагностики SoC.

Технологічною платформою обрано програмовану систему на кристалі FPGA сімейства ZYNQ-7000 фірми Xilinx Inc. Платформа Xilinx ZYNQ-7000 має архітектуру, де поєднано систему оброблення PS і програмовану логіку PL. Взаємодія двох систем відбувається з допомогою інтерфейсу AMBA на основі

протоколу AXI. Як налагоджувальну плату використано недорого загальнодоступну плату *ZedBoard*, оснащену пристроєм *XC7Z020 Zynq*. Сформульовано загальну послідовність етапів проектування системи на кристалі на платформі *ZYNQ-7000*. Практичну реалізацію виконано на базі стеку інструментальних засобів САПР *Vivado/Vitis/Vitis HLS*. Це дає змогу виконати наскрізне проектування як частини, що синтезується з боку *PL*, так і користувацького програмного забезпечення з боку *PS*. Розроблені методи апробовано на моделі пристрою логічного керування світлофором. Як модель пристрою керування розглядається часовий автомат Мура, а його описом є темпоральний граф переходів. Під час реалізації на платформі *ZYNQ-7000* керуючий автомат реалізований у блоці *PL* мовою програмування *C*, а операційний автомат – у блоці *PS*.

В організації самодіагностики реалізується неруйнівний діагностичний експеримент способом обходу всіх дуг графа переходів, починаючи з початкової вершини. Тестером у цьому разі є операційний автомат, еталонні логічні та часові значення зберігаються в пам'яті блоку *PS*. Візуальне спостереження за виконанням ДЕ здійснюється на панелі світлодіодів плати *ZedBoard*.

Наукова новизна роботи полягає в подальшому розвитку моделей та методів проектування й самодіагностики автоматних пристроїв логічного керування, реалізованих на технологічній платформі систем на кристалі.

Перспективи подальшого розвитку досліджень у цьому напрямі полягають у розробленні моделей розподілу обчислень у *SoC* між блоками *PL* та *PS* залежно від критеріїв замовника.

Список літератури

1. Lee E. A., Seshia S. A. Introduction to embedded systems: A cyber-physical system approach. UC Berkeley, 2011. 499 p. URL: https://ptolemy.berkeley.edu/books/leeseshia/releases/LeeSeshia_DigitalV1_08.pdf
2. Greaves D. Modern system-on-chip design on ARM. Arm Education Media, 2021. 608 p. URL: <https://armkeil.blob.core.windows.net/developer/Files/pdf/ebook/arm-modern-soc.pdf>
3. Teich J. Hardware software codesign: the past, the present, and predicting the future. *Proc. IEEE*, Vol. 100. 2012. P. 1411–1430. DOI: [10.1109/JPROC.2011.2182009](https://doi.org/10.1109/JPROC.2011.2182009)
4. Bailey B., Martin G., Piziali A. ESL design and verification: A prescription for electronic system level methodology. *Morgan Kaufmann Publishers Inc.* 2007. P. 113–138. DOI: [10.1016/B978-012373551-5/50068-X](https://doi.org/10.1016/B978-012373551-5/50068-X)
5. Shalyto A. A. Software automation design: algorithmization and programming of problems of logical control. *Journal of Computer and System Sciences International*. Vol. 39, No. 6. 2000. P. 899–916. URL: https://www.researchgate.net/publication/297443303_Software_automaton_design_Algorithmization_and_programming_of_problems_of_logical_control
6. Alur R., Dill D.L. A theory of timed automata. *Theoretical Computer Science*. Vol.126, № 2. 1994. P. 183–235. DOI: [10.1016/0304-3975\(94\)90010-8](https://doi.org/10.1016/0304-3975(94)90010-8)
7. Zhigulin M., Yevtushenko N., Maag S., Cavalli A. R. FSM-based test derivation strategies for systems with time-outs. *Proceedings of the 11th International Conference on Quality Software (QSIC 2011)*, Madrid, Spain, 2011. P. 141–149. DOI: [10.1109/QSIC.2011.30](https://doi.org/10.1109/QSIC.2011.30)
8. El-Fakih K., Yevtushenko N., Simão A. A practical approach for testing timed deterministic finite state machines with single clock. *Science of Computer Programming*. Elsevier. Vol. 80. 2014. P. 343–355. DOI: [10.1016/j.scico.2013.09.008](https://doi.org/10.1016/j.scico.2013.09.008)
9. Miroshnyk M., Shkil A., Kulak E., Rakhlis D., Filippenko I., Hoha M., Malakhov M., Serhiienko V. Design of real-time logic control system on FPGA. *Proceedings of 2019 IEEE East-West Design & Test Symposium (EWDTS'19)*, September 13–16, Batumi, Georgia, 2019. P. 488–491. DOI: [10.1109/EWDTS.2019.8884387](https://doi.org/10.1109/EWDTS.2019.8884387)
10. Miroshnyk M. A., Shkil A. S., Kulak E. N., Rakhlis D. Y., Miroshnyk A. M., Malahov N. V. Design timed FSM with VHDL Moore pattern. *Radio Electronics, Computer Science, Control*. 2020. №2 (53). P. 137–148. DOI: [10.15588/1607-3274-2020-2-14](https://doi.org/10.15588/1607-3274-2020-2-14)
11. Мірошник М. А. Проектування діагностичної інфраструктури обчислювальних систем та пристроїв на ПЛІС: монографія. Х.: ХУПС, 2012. 188 с. URL: <http://lib.kart.edu.ua/bitstream/123456789/7162/1/%D0%9D%D0%B0%D0%B2%D1%87%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D0%B9%20%D0%BF%D0%BE%D1%81%D1%96%D0%B1%D0%BD%D0%B8%D0%BA.pdf>
12. Chi Y., Guo L., Lau J., Choi Y. K., Wang J., Cong J. Extending high-level synthesis for task-parallel programs. *Proceedings of 2021 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, 28 February – 2 March 2021. 225 p. DOI: [10.1145/3431920.3439470](https://doi.org/10.1145/3431920.3439470)
13. Gniazdowski T. et al. High-performance lightweight HLS generator module of normally distributed random numbers in FPGAs. *Electronics*. Vol. 12, No. 22. 2023. 4667 p. DOI: [10.3390/electronics12224667](https://doi.org/10.3390/electronics12224667)

14. Caba J., Rincón F., Barba J., De la Torre J. A., López J. C. FPGA-based solution for on-board verification of hardware modules using HLS. *Electronics*. Vol. 9, No. 12. 2020. 2024 p. DOI: [10.3390/electronics9122024](https://doi.org/10.3390/electronics9122024)
15. Roozmeh M. High performance computing via high level synthesis, *Xilinx FPGA*. PhD thesis. Turin, 2018. 125 p. DOI: [10.6092/POLITO/PORTO/2710706](https://doi.org/10.6092/POLITO/PORTO/2710706)
16. Heyden M. High Level Synthesis for ASIC and FPGA: master thesis. Sweden, Lund, 2023. 47 p. URL: <https://lup.lub.lu.se/student-papers/record/9126669/file/9128084.pdf> (дата звернення: 10.11.2023).
17. Crockett L. H., Elliot R. A., Enderwitz M. A., Stewart R. W. The Zynq Book: Embedded Processing with the ARM Cortex-A9 on the Xilinx Zynq-7000. All Programmable SoC. Strathclyde Academic Media, 2014. P. 484. URL: https://is.muni.cz/el/1433/jaro2015/PV191/um/The_Zynq_Book_ebook.pdf (дата звернення: 17.09.2023).

References

1. Lee E. A., Seshia S. A. "Introduction to embedded systems: A cyber-physical system approach, Berkeley". 2011. 499 p. available at: https://ptolemy.berkeley.edu/books/leeseshia/releases/LeeSeshia_DigitalV1_08.pdf
2. Greaves D. "Modern system-on-chip design on ARM, Arm Education Media". 2021. 608 p. available at: <https://armkeil.blob.core.windows.net/developer/Files/pdf/ebook/arm-modern-soc.pdf>
3. Teich, J. (2012), "Hardware/software codesign: the past, the present, and predicting the future", *Proc. IEEE*, Vol. 100. P. 1411–1430. DOI: [10.1109/JPROC.2011.2182009](https://doi.org/10.1109/JPROC.2011.2182009)
4. Bailey, B., Martin, G., Piziali, A. (2007), "ESL design and verification: A prescription for electronic system level methodology", *Morgan Kaufmann Publishers Inc.*, P. 113–138. DOI: [10.1016/B978-012373551-5/50068-X](https://doi.org/10.1016/B978-012373551-5/50068-X)
5. Shalyto, A. A. "Software automation design: algorithmization and programming of problems of logical control", *Journal of Computer and System Sciences International*, Vol. 39, No. 6, 2000. P. 899–916. available at: https://www.researchgate.net/publication/297443303_Software_automaton_design_Algorithmization_and_programming_of_problems_of_logical_control
6. Alur, R., Dill, D.L. (1994), "A theory of timed automata", *Theoretical Computer Science*, Vol.126, № 2, P. 183–235. DOI: [10.1016/0304-3975\(94\)90010-8](https://doi.org/10.1016/0304-3975(94)90010-8)
7. Zhigulin, M. Yevtushenko, N., Maag, S., Cavalli, A. R. (2011), "FSM-based test derivation strategies for systems with time-outs". *Proceedings of the 11th International Conference on Quality Software (QSIC 2011)*, Madrid, Spain, P. 141–149. DOI: [10.1109/QSIC.2011.30](https://doi.org/10.1109/QSIC.2011.30)
8. El-Fakih, K., Yevtushenko, N., Simão, A. (2014), "A practical approach for testing timed deterministic finite state machines with single clock", *Science of Computer Programming*, Elsevier, Vol. 80, P. 343–355. DOI: [10.1016/j.scico.2013.09.008](https://doi.org/10.1016/j.scico.2013.09.008)
9. Miroshnyk, M., Shkil, A., Kulak, E., Rakhlis, D., Filippenko, I., Hoha, M., Malakhov, M., Serhiienko, V. (2019), "Design of real-time logic control system on FPGA". *Proceedings of 2019 IEEE East-West Design & Test Symposium (EWDTS'19)*, September 13-16, Batumi, Georgia, P. 488–491. DOI: [10.1109/EWDTS.2019.8884387](https://doi.org/10.1109/EWDTS.2019.8884387)
10. Miroshnyk, M. A., Shkil, A. S., Kulak, E. N., Rakhlis, D. Y., Miroshnyk, A. M., Malahov, N. V. (2020), "Design timed FSM with VHDL Moore pattern", *Radio Electronics, Computer Science, Control*, №2 (53), P. 137–148. DOI: [10.15588/1607-3274-2020-2-14](https://doi.org/10.15588/1607-3274-2020-2-14)
11. Miroshnyk, M. A. (2012), "Design of diagnostic infrastructure of computing systems and devices on FPGA" ["Proektuvannya diagnostychnoyi infrastruktury obchyshlyval'nykh system ta prystryiv na PLIS"], monograph, Kharkiv, 331 p. available at: <http://lib.kart.edu.ua/bitstream/123456789/7162/1/%D0%9D%D0%B0%D0%B2%D1%87%D0%B0%D0%BB%D1%8C%D0%BD%D0%B8%D0%B9%20%D0%BF%D0%BE%D1%81%D1%96%D0%B1%D0%BD%D0%B8%D0%BA.pdf>
12. Chi, Y., Guo, L., Lau, J., Choi, Y. K., Wang, J., Cong, J. (2021), "Extending high-level synthesis for task-parallel programs". *Proceedings of 2021 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, 225 p. DOI: [10.1145/3431920.3439470](https://doi.org/10.1145/3431920.3439470)
13. Gniazdowski, T. et al. (2023), "High-performance lightweight HLS generator module of normally distributed random numbers in FPGAs", *Electronics*, Vol. 12, No. 22, 4667 p. DOI: [10.3390/electronics12224667](https://doi.org/10.3390/electronics12224667)
14. Caba, J., Rincón, F., Barba, J., De la Torre, J. A., López, J. C. (2020), "FPGA-based solution for on-board verification of hardware modules using HLS", *Electronics*, Vol. 9, No. 12, P. 2024. DOI: [10.3390/electronics9122024](https://doi.org/10.3390/electronics9122024)
15. Roozmeh, M. (2018), "High performance computing via high level synthesis". *Xilinx FPGA*. PhD thesis, Turin, 125 p. DOI: [10.6092/POLITO/PORTO/2710706](https://doi.org/10.6092/POLITO/PORTO/2710706)
16. Heyden, M. "High Level Synthesis for ASIC and FPGA": master thesis. Sweden, Lund, 2023. 47 p. available at: <https://lup.lub.lu.se/student-papers/record/9126669/file/9128084.pdf> (last accessed: 10.11.2023).
17. Crockett, L. H., Elliot, R. A., Enderwitz, M. A., Stewart, R. W. "The Zynq Book: Embedded Processing with the ARM Cortex-A9 on the Xilinx Zynq-7000. All Programmable SoC, Strathclyde Academic Media", 2014. P. 484. available at: https://is.muni.cz/el/1433/jaro2015/PV191/um/The_Zynq_Book_ebook.pdf (last accessed: 17.09.2023).

Відомості про авторів / About the Authors

Шкіль Олександр Сергійович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри автоматизації проектування обчислювальної техніки, Харків, Україна; e-mail: oleksandr.shkil@nure.ua; ORCID ID: <https://orcid.org/0000-0003-1071-3445>

Рахліс Дарія Юхимівна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри автоматизації проектування обчислювальної техніки, Харків, Україна; e-mail: dariia.rakhlis@nure.ua; ORCID ID: <https://orcid.org/0000-0002-6652-1840>

Філіпенко Інна Вікторівна – кандидат технічних наук, Харківський національний університет радіоелектроніки, доцент кафедри автоматизації проектування обчислювальної техніки, Харків, Україна; e-mail: inna.filipenko@nure.ua; ORCID ID: <https://orcid.org/0000-0002-3584-2107>

Корнієнко Валентин Русланович – Харківський національний університет радіоелектроніки, аспірант кафедри автоматизації проектування обчислювальної техніки, Харків, Україна; e-mail: valentyn.korniienko1@nure.ua; ORCID ID: <https://orcid.org/0000-0001-7070-5127>

Shkil Alexander – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associated Professor of Design Automation Department, Kharkiv, Ukraine.

Rakhlis Dariia – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associated Professor of Design Automation Department, Kharkiv, Ukraine.

Filipenko Inna – PhD (Engineering Sciences), Kharkiv National University of Radio Electronics, Associated Professor of Design Automation Department, Kharkiv, Ukraine.

Korniienko Valentyn – Kharkiv National University of Radio Electronics, PhD Student of Design Automation Department, Kharkiv, Ukraine.

DESIGN AND SELF-DIAGNOSTICS OF CYBERPHYSICAL CONTROL DEVICES ON *SoC* PLATFORM

The subject of research in this article is models, methods, and procedures for designing and self-diagnosing automated models of logic control devices implemented in *SoCs*. **The object of work** is the procedures for automated design and diagnosis of digital devices on the *SoC* technology platform. **The aim of the study** is to develop models and procedures for designing and self-testing in the cycle of automated design of automatic logic control systems on the *SoC* technology platform, which will significantly increase the reliability of their operation. The article solves the following **tasks**: consideration of the procedures for interacting the processor core with programmable logic as part of the *SoC*; improvement of the procedures for designing and testing software and hardware systems based on *SoC*; further development of procedures for automated design, verification, and diagnosis of cyber-physical logic control systems using programming languages and hardware description languages; implementation of the procedure for hardware self-testing of control automata on the *SoC* technology platform. The following **methods** are implemented: synthesis of control automata based on graph models, implementation of control automata models in the *C* programming language using an automata template, diagnostic experiment by traversing the automata transition graph. **Results** achieved. Based on the analysis of the procedures for the interaction of the processor core and programmable logic on the selected *SoC* platform, a model of a cyber-physical logic control system was designed. The practical implementation was carried out on the basis of the *Vivado/Vitis/Vitis HLS CAD* toolkit. The method of hardware self-testing of control automata on the technological platform of *SoC ZYNQ-7000* was implemented. **Conclusions.** The article analyzes the principles of designing embedded cyber-physical systems implemented in system-on-chip. The principles of building verification systems and embedded self-diagnostics of system-on-chip systems containing software and hardware are considered. The developed methods are tested on a model of a traffic light logic control device on the *SoC FPGA* platform of the *ZYNQ-7000* family by *Xilinx*. The Moore's control automaton is implemented in the *PL* block in the *C* programming language, and the operational automaton is implemented in the *PS* block. During the organization of the self-diagnosis process, a non-destructive diagnostic experiment was performed by traversing all arcs of the transition graph, starting from the initial vertex. In this case, the tester was an operational automaton, the reference logic and time values of which were stored in the memory of the *PS* unit. Visual observation of the diagnostic experiment was carried out using the LED panel of the *ZedBoard* board.

Keywords: cyber-physical systems; embedded systems; logic control; system-on-chip design; *FPGA*; *CAD*; *SoC* self-testing; *C* programming language.

Бібліографічні описи / Bibliographic descriptions

Шкіль О. С., Рахліс Д. Ю., Філіпенко І. В., Корнієнко В. Р. Проектування та самодіагностика кіберфізичних пристроїв керування на платформі *SoC*. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 122–134. DOI: <https://doi.org/10.30837/ITSSI.2023.26.122>

Shkil, A., Rakhlis, D., Filipenko, I., Korniienko, V. (2023), "Design and self-diagnostics of cyberphysical control devices on *SOC* platform", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 122–134. DOI: <https://doi.org/10.30837/ITSSI.2023.26.122>

I. NEVLIUDOV, M. VZHESNIEVSKYI, Y. ROMASHOV, O. CHALA

MATHEMATICAL MODELING OF MECHATRONIC SHUTTLES AS AUTOMATION OBJECTS FOR MULTILEVEL SYSTEMS OF INTRA-WAREHOUSE LOGISTICS

The subject of the study is the mathematical modeling of mechatronic shuttles of multilevel intra-warehouse logistics systems. **The purpose of the article** is to ensure energy efficiency and high productivity of multilevel warehouse logistics systems by developing mathematical models of mechatronic electromechanical drives as part of automated autonomous shuttle systems (*Pallet Shuttle*). To achieve this goal, it is necessary to solve the following tasks: to build a mathematical model of the processes occurring in the transportation shuttle as an electromechanical and mechatronic system; to build a mathematical model of the transportation shuttle as an automation object and to obtain a transfer function of the autonomous shuttle system for automatic control of its speed. **Conclusions.** The paper develops generalized approaches to the construction of mathematical models of electric drives as automation objects. The developed approach assumes that the linearized mathematical model of an electric drive as an automation object is represented as a linear differential equation with constant coefficients that relate the control and the controlled parameter. The construction of such a differential equation is proposed to be carried out by generalizing the properties inherent in electromechanical systems, and it is assumed that these properties are represented by Lagrange equations of the 2nd kind using electromechanical analogies. As a result of the research, it was shown that the use of this approach leads to a definite equation of the mathematical model of the mechatronic shuttle as an automation object. It is shown that the linearized mathematical model of the shuttle as an automation object should be determined by a differential equation of at least the second order. The research results prove the possibility of improving automation systems in design by taking into account the properties inherent in shuttles as electromechanical and mechatronic systems in more detail. It is shown that the use of this approach leads to a definite equation of the mathematical model of the mechatronic shuttle as an automation object, but it is generally almost impossible to obtain such an equation explicitly. To illustrate the proposed approach, we consider the modeling of a typical mechatronic electric drive of machines used in automated storages and present the results of modeling processes in the electric drive. It is shown that the linearized mathematical model of a mechatronic shuttle as an automation object should be determined by a differential equation of at least the second order. The research results show the possibility of improving automation systems during design due to a more detailed consideration of the properties inherent in shuttles as mechatronic electromechanical systems.

Keywords: automation; transfer function; cyber-physical systems; Industry 4.0; mechatronic electric drive; Pallet Shuttle; mathematical modeling.

Introduction

One of the key conditions for high labor productivity growth in the industry is the continuous improvement of the technological equipment fleet. This is achieved primarily by replacing outdated equipment with new, modern ones.

Logistical problems are relevant for every company that deals with inventory.

A modern trend in warehouse automation is the use of shuttle systems for warehouse maintenance and warehouse preparation areas. Traditionally, shuttles are usually realized with pulling cables. In terms of flexibility and the ability to move around the warehouse and production, this causes certain problems that can be solved by using autonomous shuttles.

The family of shuttles includes cable cars and automatic distribution vehicles. Cable cars and automatic

distribution vehicles are powered by electricity and run on horizontal rails. They are placed on the rails by a vehicle, the so-called carrier, and can only travel in a straight line. Changes of course are made by means of turnouts and shifting to parallel tracks using mobile bogies. Height differences are covered by lifts.

Electrically driven shuttles are usually connected to the carrier by means of a drag cable. Only in the latest developments are autonomous shuttles being designed, providing for new, expanded applications.

Thus, the task of modernizing mechatronic electric drives as part of multi-level intra-warehouse logistics systems with minimal financial and time costs is becoming urgent. A possible solution is to modernize equipment by synthesizing IIoT technologies and modern sensors.

The ability to significantly improve the efficiency of the existing equipment fleet, with low costs and quick payback, through modernization is the reason why such

work is paid to production in solving a number of intra-warehouse logistics issues.

Thus, by modernizing outdated equipment to perform specific functions of shuttle systems, replacing individual systems and equipment units, it is possible to bring production to modern Industry 4.0 standards, including elements of cyber-physical approaches, without huge capital expenditures [1]. Modernization works include structural modifications of equipment [2], which are aimed at expanding technological capabilities or changing the technological purpose of equipment [3–4]. These design changes also allow to extend the effective life of the equipment, and in some cases, even significantly increase its productivity and reduce energy consumption.

Analysis of recent research and publications

The concept of logistics automation, using Industry 4.0 cyber-physical approaches, is an important tool for increasing the productivity of logistics operators through the use of computer software, robotics, or robotics technologies. There are many different technologies related to automation in logistics. New hardware and software are constantly being developed. The main trends such as automation and digitalization of logistics processes, in particular, automation of warehouse complexes or their components using Industry 4.0 cyber-physical approaches, have been studied [5].

The main potential of logistics lies in the optimization of material flow management. According to research by UK scientists, approximately 70% of the final cost of goods that reach the consumer is the cost of logistics operations (transportation, warehousing, storage, packaging, etc.) associated with the movement of material flow [6].

Thus, the main object of logistics is material flows and ways to realize their movement [7].

The use of multi-level shuttle systems partially solves the problem of transportation and storage of large volumes of products in warehouses [8–9]. However, a significant drawback is the high cost of shuttle structures and their installation, maintenance, and low energy efficiency [10].

Automation and improvement of tent systems in logistics, through the use of cyber-physical approaches, is one of the shortest ways to increase the parameters of reducing the time of transportation, sorting, storage and increasing energy efficiency and flexibility of intra-warehouse logistics processes [11].

Despite the availability of publications on the modeling of mechatronic systems in logistics, they are either purely general materials or relate only to mechanical and system engineers [12]. At present, automation researchers lack objective recommendations on theoretical and practical ways to build mathematical models of mechatronic shuttles for multilevel systems, so their development is an urgent task. The paper proposes ways to improve automation systems in design by taking into account in more detail the properties inherent in shuttles as electromechanical and mechatronic systems based on mathematical modeling.

The aim of the work is to ensure energy efficiency and high performance of multilevel warehouse logistics systems by developing mathematical models of mechatronic electromechanical drives as part of automated autonomous shuttle systems (Pallet Shuttle).

Theoretical foundations of mathematical modeling of the mechatronic shuttle as an automation object

Mechatronic systems, as one of the types of dynamic systems [10], are designed to realize a given movement and are based on some kind of actuator and drive – electromechanical, hydraulic, or some other. Examples of modern mechatronic devices are machine tool modules and industrial robots, household appliances, shuttles in logistics warehouse systems, etc. Typically, mechatronic systems are modeled at the highest level of abstraction (macro level), where a distinction is made between structural (block) and physical multidomain approaches, each of which has its advantages and disadvantages [11]. At the same time, each of the two methods may have certain features of the mathematical core. The accuracy of modeling depends on the number of system properties taken into account, while the accuracy of simulation depends on the tools selected by the user and the consideration of the features of the developed model. Any mechatronic system requires control, and must be identified to build its model in order to create the necessary control device that provides a complex coordinated movement of the mechanical part [10]. The modeling of mechatronic systems includes the following stages of studying the characteristics of the system as a whole, the automatic control system and the automation object, and the construction and study of models of the mechanical part.

An automation object differs from any other object of mathematical modeling by the presence of a defined

controlled parameter x and the presence of a defined parameter u , which affects the value of the controlled parameter. The specificity of the mechatronic shuttle is the presence of an electric motor as a source of torque, which is formed by the applied electric supply voltage [13]. Therefore, taking into account these circumstances and the purpose of the mechatronic shuttle, the angular velocity of the rotor of the drive electric motor can be considered as a generalized controlled parameter x of the mechatronic shuttle, and the supply voltage of the drive electric motor can be taken as a generalized parameter u characterizing the control, which generally depends on the time t :

$$x = x(t), u = u(t). \quad (1)$$

Generalization is understood here in the sense that the controlled parameter and the control parameter (1) can be used regardless of the design, purpose, and operating modes of the mechatronic electric drive.

The mathematical model of the mechatronic shuttle as an automation object should generally determine the dependence of the controlled parameter x on the control parameter u and should be presented in a convenient form for further use in the design of the automatic control system of the electric drive. It is clear that, given the variety of designs and characteristics of mechatronic shuttles, it is quite difficult to propose a generalized form of the mathematical model of a mechatronic shuttle as an automation object, but from the point of view of convenient further use in the design of a mechatronic shuttle automatic control system, the most convenient form of such a model should correspond to a linear ordinary differential equation with constant coefficients and the required number of initial conditions [14]:

$$a_n \frac{d^n x}{dt^n} + a_{n-1} \frac{d^{n-1} x}{dt^{n-1}} + \dots + a_1 \frac{dx}{dt} + a_0 x = b_0 u, \quad (2)$$

$$x(t_0) = x_0^{(0)}, \frac{dx}{dt}(t_0) = x_0^{(1)}, \dots, \frac{d^{n-1} x}{dt^{n-1}}(t_0) = x_0^{(n-1)}, \quad (3)$$

where n and m – order of the differential equation and the control derivative;

$a_n, a_{n-1}, \dots, a_0, b_0$ – parameters of the mathematical model;

$$L \approx \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N m_{ij} \dot{q}_i \dot{q}_j - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N c_{ij} q_i q_j, R \approx \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N b_{ij} \dot{q}_i \dot{q}_j, Q_k \approx \sum_{j=1}^N p_{kj} \dot{q}_j + p_k u, \quad (5)$$

where m_{ij} , c_{ij} and b_{ij} – generalized characteristics of inertia, stiffness and viscosity;

t_0 – given moment in time;

$x_0^{(0)}, x_0^{(1)}, \dots, x_0^{(n-1)}$ – set values of the controlled

parameter and its derivatives at the initial time $t = t_0$.

The form (2), (3) is indeed the most convenient for formulating a mathematical model of a mechatronic shuttle as an automation object, since it allows solving all engineering issues related to the design of automated control systems, including the design of automatic controllers. At the same time, there is no general approach to constructing the differential equation (2) for modeling shuttles as automation objects, which in some way complicates the solution of engineering problems in the design of automated control systems.

To build the mathematical model (2), (3) of mechatronic shuttles as automation objects, it is proposed to consider the differential equation (2) as a form for generalizing the properties of processes inherent in shuttles as electromechanical systems. Taking into account the known electromechanical analogies, we have the most general form of a mathematical model of processes for a fairly wide class of mechatronic shuttles of various designs and purposes in the form of Lagrange equations of the 2nd kind:

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = - \frac{\partial R}{\partial \dot{q}_k} + Q_k, k = 1, 2, \dots, N, \quad (4)$$

where N – number of freedom degrees of the shuttle as an electromechanical system;

q_1, q_2, \dots, q_N and $\dot{q}_1, \dot{q}_2, \dots, \dot{q}_N$ – generalized coordinates and velocities of the shuttle as an electromechanical system;

L – Lagrange function (kinetic potential);

R – Rayleigh dissipative function;

Q_k – is a generalized driving force, which, of course, is neither potential nor dissipative.

It is clear that the peculiarities of the designs and schematic solutions of the mechanical and electrical parts of the shuttle are taken into account in equations (4) due to the appropriate proper form of generalized coordinates, Lagrange function, dissipative Rayleigh function, and generalized forces, which can be appropriately linearized and represented as:

p_{kj} and p_k ($k = 1, 2, \dots, N$) – generalized characteristics of the driving forces of the electric drive as an electromechanical system.

Taking into account the linearized expressions (5), the differential equations (4) will take the following form:

$$\sum_{j=1}^N m_{kj} \ddot{q}_j + \sum_{j=1}^N b_{kj} \dot{q}_j + \sum_{j=1}^N c_{kj} q_j = \sum_{j=1}^N p_{kj} \dot{q}_j + p_k u, \quad (6)$$

$$k = 1, 2, \dots, N.$$

Equation (6), of course, should be considered with initial conditions:

$$q_k(t_0) = q_k^{(0)}, \quad \dot{q}_k(t_0) = \dot{q}_k^{(0)}, \quad k = 1, 2, \dots, N, \quad (7)$$

where $q_k^{(0)}$ and $\dot{q}_k^{(0)}$ specified values of the generalized coordinate and velocity at the initial moment of time $t = t_0$.

In the form of (6), (7), we have a generalized linearized mathematical model of electric drive processes as in an electromechanical system, and such a model

$$\frac{dx}{dt} = \sum_{j=1}^N \frac{\partial \omega}{\partial q_j} \dot{q}_j + \sum_{j=1}^N \frac{\partial \omega}{\partial \dot{q}_j} \ddot{q}_j, \quad x(t_0) = \omega(q_1^{(0)}, \dots, q_N^{(0)}, \dot{q}_1^{(0)}, \dots, \dot{q}_N^{(0)}). \quad (9)$$

Thus, in the form of (6), (7), (9), we have an identical form of representation of the mathematical model (2), (3) of the mechatronic shuttle as an automation object. It is clear that by identical transformations of the differential equations (6), (9), we will have a differential equation of the form (2). It is also clear that the identical transformations of equations (6), (9) to the form (2) in the general case can be quite complicated, so that it will be practically impossible to perform them, although for some cases of equations (6), (9) such transformations are quite possible. Since in the form of (6), (9) we have a system N of differential equations of the second order and one differential equation of the first order, in general, we can state that the order of the differential equation (2) is as follows:

$$n = 2N + 1. \quad (10)$$

It should be noted that in some cases, the order of the differential equation (2) may be less than the value of (10) due to the form of differential equations (6) and relation (8). In general, relation (10) clearly shows that a more detailed modeling of processes in the shuttle as an electromechanical system by increasing the number of degrees of freedom and generalized coordinates [15] will lead to a complication of the mathematical model of the mechatronic shuttle as an automation object. Thus, we have opportunities to improve automation systems in design by taking into account in more detail the properties inherent in mechatronic shuttles as electromechanical systems [16].

is generalized in the sense that it covers a fairly wide class of mechatronic shuttles for various purposes. It is quite clear that the controlled parameter of the shuttle, introduced in relations (1), is determined by the current state of the shuttle, which can be mathematically written as follows:

$$x(t) = \omega(q_1(t), \dots, q_N(t), \dot{q}_1(t), \dots, \dot{q}_N(t)), \quad (8)$$

where $\omega(\dots)$ – some given function that determines the controlled parameter, which is generally the angular velocity of the rotor of an electric motor, through the generalized coordinates and speeds of the mechatronic shuttle as an electromechanical system. mechatronic shuttle as an automation object, we determine the derivative and the value at the initial time of the controlled parameter (8):

An example of mathematical modeling of the mechatronic shuttle of the electric drive of automated storage machines as an automation object

To further extend electromechanization and automation to a wider class of operations, various means are proposed [17], including automatic shuttles (Fig. 1, a) to ensure intra-warehouse transportation of goods in the spaces between rows of racks. It is clear that the automation of such transportation shuttles is fundamental in contrast to other warehouse equipment, since, in addition to various advantages, it also provides smaller dimensions due to the absence of a workplace for a human operator, which will reduce the space between rows of racks and, as a result, reduce the length of transportation routes. In such shuttles (Fig. 1, a), it is most convenient to use electric drives to provide traction, since they can be quite small in size and do not generate emissions into the environment, such as drives with internal combustion engines, which is actually quite important for closed warehouses. Thus, we have a transportation shuttle (Fig. 1, a) as an automation object in the form of a complex mechatronic electromechanical system, which can be considered as a special case of an industrial electric drive. It is clear that to ensure the automation of the shuttle, it is necessary to have its mathematical model as an automation object. Therefore, to illustrate the capabilities and techniques of using the proposed generalized approaches, we will further consider a simplified schematic diagram of the mechanical part (Fig. 1, c) and DC traction motors (Fig. 1, c) of the

mechatronic shuttle (Fig. 1, a) to simplify the identical mathematical transformations that are not fundamental to this study to obtain the differential equation (2) of the mathematical model of the mechatronic shuttle as an automation object.

It should be noted that the transportation shuttle (Fig. 1, a) is a rather complex automation object, since it should provide automation of both steady-state and transient modes in the presence of special requirements for positioning accuracy [18] and, possibly, smooth movements with the desired minimization of energy consumption. It is clear that solving each of these and other special automation tasks will require appropriate mathematical models of the shuttle as an automation

object, but here we will limit ourselves to considering the automation task of ensuring steady-state transportation of cargo by the shuttle at a given speed. In such a problem, the following should be chosen as parameters (1):

$$x(t) = \dot{s} - v, \quad u(t) = \tilde{U}_e, \quad (11)$$

where $s = s(t)$ – a rectilinear coordinate that determines the current position of the shuttle on the guides (Fig. 1, a and Fig. 1, b);

$v > 0$ – the desired speed of the shuttle during the transportation of goods;

$\tilde{U}_e = \tilde{U}_e(t)$ – supply voltage of traction electric motors (Fig. 1, c).

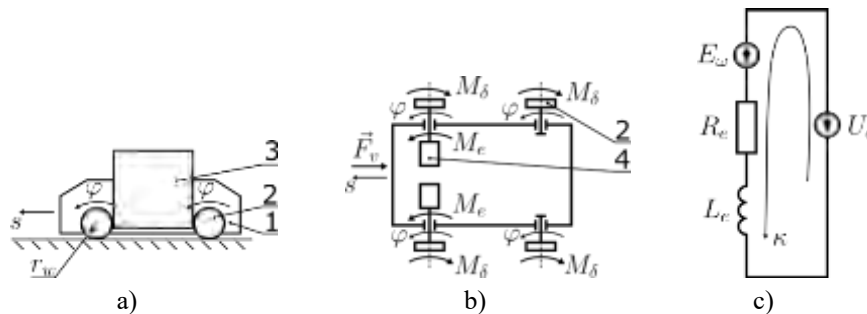


Fig. 1. General (a), kinematic (b) schemes and equivalent electrical circuit of electric motors (c) of the shuttle for intra-warehouse transportation (1 – shuttle; 2 – wheel; 3 – cargo; 4 – electric motor)

The mathematical model of the mechatronic shuttle as an automation object should make it possible to establish a connection between the parameters (11), and the most convenient form of such a model should correspond to (2), (3). Therefore, to obtain the mathematical model (2), (3) of the shuttle (Fig. 1) as an automation object with parameters (11), we consider the following processes inherent in the shuttle as an electromechanical system.

First, let's consider the mechanical part of the shuttle, which represents the automation object under study. So, we have a shuttle 1 that carries a load 3 on wheels 2 in such a way that the position of the shuttle on the guides is determined by the coordinate s , and the rotation of the wheels is determined by the angle of rotation φ , as shown in Fig. 1, a. We further assume that the movement of the shuttle is provided by two electric motors 4 of direct current, each of which is directly connected to the corresponding wheel, and the other two wheels of the shuttle are exclusively supporting, as shown in Fig. 1, б. We further assume that the coordinate and the angle of rotation are related by the following relation:

$$\dot{s} = r_w \dot{\varphi}, \quad (12)$$

where r_w – radius of the shuttle wheels (Fig. 1, a).

We take into account rolling friction and viscous resistance to shuttle movement:

$$|M_\delta| \approx \frac{1}{4} mg \delta, \quad |\vec{F}_v| = b |\dot{s}|, \quad (13)$$

where M_δ and \vec{F}_v – wheel rolling friction moment and total viscous drag force of the shuttle;

m – gross weight of the shuttle and $g = 9,81 \text{ m/c}^2$;

δ and b – the rolling friction coefficient of the wheels and the integral coefficient of the shuttle's viscous pore.

We further assume that the movement is provided by electromagnetic torques M_e (Fig. 1, b), which are formed as a result of supplying DC supply voltage to traction electric motors U_e and [20, 21] is determined by the electric charge of the rotor of an electric motor:

$$M_e = B_e \dot{\kappa}, \quad (14)$$

where M_e – electromagnetic torque generated by an electric motor;

B_e and $\kappa = \kappa(t)$ – electromechanical characteristic and electric charge in the rotor circuit of a DC traction electric motor.

It is clear that to determine the electromagnetic torque (14), we should consider the electrical part of the shuttle, which will be represented as an equivalent electrical circuit consisting of series-connected active resistances R_e , индуктивного опоры L_e , as well as sources E_ω and U_e electromotive force (Fig. 1, c). Source of electromotive force U_e is actually a source of energy that powers the shuttle, a source E_ω electromotive force is an electromotive force arising from the rotation of the rotor winding conductors with electric current in the magnetic field of an electric motor and is defined as follows:

$$E_\omega = B_e \dot{\varphi}. \quad (15)$$

It should be noted that the electromotive force E_ω counteracts the electromotive force U_e , which leads to the rotation of the rotor of the electric motor (Fig. 1, c).

Within the framework of the accepted schematization of the mechatronic shuttle as an electromechanical system (Fig. 1), taking into account relations (12)–(15), we can assume that the processes in such a system are determined by two generalized coordinates:

$$q_1 = \varphi, \quad q_2 = \kappa. \quad (16)$$

Thus, taking into account the adopted generalized coordinates (16), we have the following Lagrange function, Rayleigh function, and generalized forces:

$$L = \frac{1}{2} J \dot{q}_1^2 + 2 \frac{1}{2} L_e \dot{q}_2^2, \quad R = \frac{1}{2} B \dot{q}_1^2 + 2 \frac{1}{2} R_e \dot{q}_2^2, \quad (17)$$

$$Q_1 = -mg\delta + 2B_e \dot{q}_2, \quad Q_2 = -2B_e \dot{q}_1 + 2U_e, \quad (18)$$

where J and B – are the generalized moment of inertia and viscous drag coefficient reduced to the axis of rotation of the rotor of a traction electric motor.

Expressions (17), (18) and the values J and B contained in expression (17) can be obtained accordingly, which is not the main purpose of this study, so we present the final results here:

$$J = mr_w^2 + 2J_e + 4J_w, \quad B = br_w^2, \quad (19)$$

where J_e and J_w – moments of inertia of the rotor of the traction electric motor and the shuttle wheel relative to their axes of rotation.

It should be noted that the choice of the first generalized coordinate (16) corresponds to the generally accepted approaches of machine dynamics [19].

$$\frac{L_e J}{2B_e} \ddot{q}_1 + \frac{L_e B + R_e J}{2B_e r_w} \ddot{q}_1 + \left(B_e + \frac{R_e B}{2B_e} \right) \dot{q}_1 = -\frac{R_e mg \delta}{2B_e} + U_e(t). \quad (25)$$

Thus, taking into account expressions (17), (18), the Lagrange equations of the 2nd kind (4) will lead to a system of two second-order differential equations:

$$\begin{aligned} J \ddot{q}_1 &= -mg\delta - B \dot{q}_1 + 2B_e \dot{q}_2, \\ L_e \ddot{q}_2 &= -B_e \dot{q}_1 - R_e \dot{q}_2 + U_e(t). \end{aligned} \quad (20)$$

It is clear that equation (20) should be considered with the following initial conditions:

$$q_1(t_0) = q_1^{(0)}, \quad \dot{q}_1(t_0) = \dot{q}_1^{(0)}, \quad q_2(t_0) = q_2^{(0)}, \quad \dot{q}_2(t_0) = \dot{q}_2^{(0)}, \quad (21)$$

where $q_1^{(0)}$ and $\dot{q}_1^{(0)}$ set values of the angle of rotation and angular velocity of the shuttle wheels at the initial moment of time $t = t_0$;

$q_2^{(0)}$ and $\dot{q}_2^{(0)}$ set values of the electric charge and current strength of the rotor winding of the shuttle traction electric motors at the initial moment of time $t = t_0$.

Thus, the mathematical model of processes in a mechatronic shuttle as an electromechanical system, in this case, a transportation shuttle (Fig. 1), summarized in the form of (6), (7), has acquired a partial form [20, 21]. Generalized in the form of (8), the controlled parameter of the electric drive [20], in this case of the transportation shuttle (Fig. 1), taking into account relations (11), (12), (15), will take the form

$$x = r_w \dot{q}_1 - v. \quad (22)$$

Due to the relation (22), we have an additional differential equation for determining the controlled parameter, generally represented in the form of (9), as follows:

$$\dot{x} = -\frac{r_w mg}{J} \delta - \frac{r_w B}{J} \dot{q}_1 + \frac{r_w 2B_e}{J} \dot{q}_2, \quad x(t_0) = r_w \dot{q}_1^{(0)} - v. \quad (23)$$

Thus, the mathematical model of the transporting mechatronic shuttle (Fig. 1) as an automation object is indirectly represented in the form of (20), (21), (23), and this form of writing the mathematical model is consistent with the above generalized result (6), (7), (9). In this case, it is possible to obtain the equations of the mathematical model of the mechatronic shuttle as an automation object in an analytical form. To do this, first, using the first differential equation (20), we need to determine the rotor winding current of the electric motor:

$$\dot{q}_2 = \frac{mg\delta}{2B_e} + \frac{J}{2B_e} \ddot{q}_1 + \frac{B}{2B_e} \dot{q}_1. \quad (24)$$

Next, we take into account the relation (24) in the second differential equation (20) and obtain:

To reformulate the differential equation (25) with respect to the controlled parameter, we use the following relations (22):

$$\dot{q}_1 = (x + v)/r_w. \quad (26)$$

$$\frac{L_e J}{2B_e r_w} \ddot{x} + \frac{L_e B + R_e J}{2B_e r_w} \dot{x} + \frac{1}{r_w} \left(B_e + \frac{R_e B}{2B_e} \right) (x + v) = -\frac{R_e mg \delta}{2B_e} + U_e(t). \quad (27)$$

Consider further that the supply voltage of the electric motors to ensure a constant speed of the mechatronic shuttle is as follows:

$$U_e(t) = \bar{U}_e(v) + \tilde{U}_e(t), \quad (28)$$

where $\bar{U}_e(v)$ – Steady-state stress required to maintain a given desired speed v of cargo transportation;

$\tilde{U}_e(t)$ – is the alternating stress required to compensate for naturally occurring uncontrolled motion

$$x(t_0) = r_w \dot{q}_1^{(0)} - v, \quad \dot{x}(t_0) = -\frac{r_w mg}{J} \delta - \frac{r_w B}{J} \dot{q}_1^{(0)} + \frac{r_w 2B_e}{J} \dot{q}_2^{(0)}. \quad (31)$$

Thus, in the direct form of (30), (31), a mathematical model of the transporting mechatronic shuttle as an automation object was obtained, and this

$$n = 2, \quad a_2 = \frac{L_e J}{2B_e r_w}, \quad a_1 = \frac{L_e B + R_e J}{2B_e r_w}, \quad a_0 = \frac{1}{r_w} \left(B_e + \frac{R_e B}{2B_e} \right), \quad b_0 = 1, \quad (32)$$

$$x_0^{(0)} = r_w \dot{q}_1^{(0)} - v, \quad x_0^{(1)} = -\frac{r_w mg}{J} \delta - \frac{r_w B}{J} \dot{q}_1^{(0)} + \frac{r_w 2B_e}{J} \dot{q}_2^{(0)}. \quad (33)$$

It is noteworthy that the properties of the transport shuttle (29)-(31) as an automation object can be represented by a transfer function of the following form, taking into account the notation (32)

$$W(s) = \frac{1}{a_2 s^2 + a_1 s + a_0}, \quad (34)$$

where $W(s)$ – transfer function that defines the transporting mechatronic shuttle as an automation object;

s – is a complex variable generated by the integral Laplace transform.

It should be emphasized that the mathematical model (30), (31) of the transport shuttle as an automation object is obtained in the direct form (2), (3), (32), (33) only due to the form (20), (21) of the mathematical model of processes in the mechatronic shuttle as an electromechanical system. In the general case, it is impossible to obtain a mathematical model of the automation object in the direct form of (2), (3), and it will be necessary to use a mathematical model in an indirect

Taking into account expressions (23), (26), the differential equation (25) will take the following form:

disturbances to ensure a given desired speed of the transportation shuttle.

Relations (27)–(29), taking into account the control parameter adopted in the form of (11) and expressions (23), will lead us to the following final result:

$$\bar{U}_e(v) = \frac{v}{r_w} \left(B_e + \frac{R_e B}{2B_e} \right) + \frac{R_e}{2B_e} mg \delta, \quad (29)$$

$$\frac{L_e J}{2B_e r_w} \ddot{x} + \frac{L_e B + R_e J}{2B_e r_w} \dot{x} + \frac{1}{r_w} \left(B_e + \frac{R_e B}{2B_e} \right) x = u(t), \quad (30)$$

model is a special case of the model (2), (3) of the general form, in which

form, which contains a mathematical model of processes and the definition of the controlled parameter. It is clear that the mathematical model of the automation object in such an indirect form is more informative, but for the design of automation systems, a mathematical model that links the controlled parameter and control is more convenient.

Further research is planned to continue in the direction of modeling processes in the transportation shuttle. Such a study necessarily requires substantiation of the results obtained in the future to confirm the absence of errors in the software used and in the computer models developed, as well as the correct use of the software. It is clear that such substantiation involves both testing of individual software components and testing of the software as a whole. At the first stage, it is planned to develop and test a computer model based on the justification of the parameters of the processes that will occur in the mechatronic transportation shuttle during its acceleration from a stationary state by

a constant electric voltage instantly applied to the driving electric motors. This will make it possible to ensure energy efficiency and high productivity of multi-level warehouse logistics systems.

The team of authors would like to express their gratitude to Kapelou LLC for supporting the research of the Department of Computer-Integrated Technologies, Automation and Robotics of the Kharkiv National University of Radio Electronics (Kharkiv, Ukraine) in the field of automation, logistics of intra-warehouse systems and mechatronics and robotics.

Conclusions

The paper develops approaches to constructing mathematical models of mechatronic shuttles as automation objects. The developed approach assumes that a linearized mathematical model of a mechatronic shuttle as an automation object is presented in the form of a linear differential equation with constant coefficients that connects the control and the controlled parameter. The construction of such a differential equation is proposed to be carried out by generalizing the properties inherent in mechatronic electric drives as

electromechanical systems, and it is assumed that these properties are represented by Lagrange equations of the 2nd kind using electromechanical analogies. It is shown that the use of this approach leads to a definite equation of the mathematical model of the mechatronic shuttle as an automation object, but it is practically impossible to obtain such an equation explicitly in the general case. To illustrate the proposed approach, we consider the modeling of a typical mechatronic electric drive of machines used in automated storages and present the results of modeling processes in the electric drive. It is shown that a linearized mathematical model of a mechatronic shuttle as an automation object should be determined by a differential equation of at least second order. The results of the study show the possibility of improving automation systems in the design due to a more detailed consideration of the properties inherent in mechatronic shuttles as mechatronic electromechanical systems. Further research will be aimed at conducting computer modeling taking into account the specific parameters of the mechatronic shuttle, which in the future will make it possible to ensure energy efficiency and high productivity of multilevel warehouse logistics systems in general.

References

1. Liashuk, O. L., Hevko, R. B., Dziura, V. O. (2019), "The development and modernization of transport and technological mechanisms of machines and possession" ["Stvorennia ta modernizatsiia transportno-tehnolohichnykh mekhanizmiv mashyn i obladannia"] [Electronic resource]: monograph. NTU. 167 p. available at: <https://elartu.tntu.edu.ua/handle/lib/29468>
2. Nevliudov, I., Yevsieiev, V., Maksymova, S., Filippenko, I. (2020), "Development of an Architectural-Logical Model to Automate the Management of the Process of Creating Complex Cyber-Physical Industrial Systems". *Eastern-European Journal of Enterprise Technologies*, 4(3(106)), P. 44–52. DOI: 10.15587/1729-4061.2020.210761
3. Frolov, A.V. (2021), "The Experience of the Portal Machine Control System Modernizing for Automatic Arc Welding and Surfacing in a Shielded Gas". *Current Problems and Ways of Industry Development: Equipment and Technologies*, No. 7, P. 622–631. DOI: https://doi.org/10.1007/978-3-030-69421-0_66
4. Wąsik, M., Lis, K., Lech, K. and Mucha, Ł. (2018), "Model-Based Dynamic Structural Modification of Machine Tools". *Shock and Vibration*, Volume 2018. P.1-9. DOI: <https://doi.org/10.1155/2018/3469171>
5. Hryhorov, O. V. ta in. "Tekhnika materialnykh potokiv lohistychnykh system: navch. posib. "Kharkiv: NTU «KhPI», 2018. 496 p. available at: <https://repository.kpi.kharkov.ua/handle/KhPI-Press/39521>
6. Sanchez, Londono, D., Barbieri, G., Fumagalli, L. (2023), "Smart retrofitting in maintenance: a systematic literature review". *Journal of Intelligent Manufacturing*. Vol. 34(1). P. 1–9. DOI: 10.1007/s10845-022-02002-2
7. Alsudani, M.Q., et al. (2023), "Smart logistics with IoT-based enterprise management system using global manufacturing". *Journal of Combinatorial Optimization*. Vol. 45(2). 1179 p. DOI: 10.1007/s10878-022-00977-5
8. Akandere, G., Khajeh, E., Paksoy, T. (2023), "The internet of things and cyber physical systems". *Smart and Sustainable Operations and Supply Chain Management in Industry*. 4, P. 277–305. DOI:10.1201/9781003180302-14
9. Aleksiev, O., Aleksiev, V., Matsyi, O., & Matsyi, M. (2021), "End-to-end design of mechatronic industrial and automotive computer systems". *Bulletin of V.N. Karazin Kharkiv National University, Series «Mathematical Modeling. Information Technology. Automated Control Systems»*, 49, P. 6–15. DOI: <https://doi.org/10.26565/2304-6201-2021-49-01>

10. Nevliudov, I., Maksymova, S., Chala, O., Bronnikov, A., & Vzhesnievskyi, M. (2023), "Automated Logistics Processes Improvement in Logistics Facilities". *Multidisciplinary Journal of Science and Technology*, 3(3), P. 157–170. available at: <https://mjstjournal.com/index.php/mjst/article/view/202>

11. Yevsieiev, V. (2021), "Methods and models of cyber-physical process control in organizational and technical production facilities" ["Metody ta modeli kiber-fizychnoho keruvannya protsesamy v orhanizatsiino-tekhnichnykh vyrobnychkykh ob'ektakh"]: avtoref. dis. Kharkiv, NURE, 45 p. available at: <https://openarchive.nure.ua/handle/document/16097>

12. Mamalis, A.G., Nevliudov, I., Romashov, Yu. (2021), "An approach for numerical simulating and processing of measured electrical signals from board sensors installed on wheeled electro-mechanical platforms", *Journal of Instrumentation*. Vol. 16 (10). 10006 p. DOI: 10.1088/1748-0221/16/10/P10006

13. Alyokhina, S., Nevliudov, I., Romashov, Yu. (2021), "Safe Transportation of Nuclear Fuel Assemblies by Means of Wheeled Robotic Platforms", *Nuclear and Radiation Safety*, Vol. 3 (91), P. 43–50. DOI: [https://doi.org/10.32918/nrs.2021.3\(91\).05](https://doi.org/10.32918/nrs.2021.3(91).05)

14. Ge, Shuaishuai, Shuang, Hou, and Mingyao, Yao. (2023), "Electromechanical Coupling Dynamic Characteristics of the Dual-Motor Electric Drive System of Hybrid Electric Vehicles", *Energies* 16 (7), 3190 p. DOI: <https://doi.org/10.3390/en16073190>

15. Pletinckx, Cyril "Autonomous shuttles for a logistical warehouse : automatic design of the warehouse. Ecole polytechnique de Louvain, Université catholique de Louvain", 2021. Prom: Schaus, Pierre. available at: <http://hdl.handle.net/2078.1/thesis:30626>

16. Zhen, L., Wu, J., Li, H., Tan, Z., & Yuan, Y. (2023), "Scheduling multiple types of equipment in an automated warehouse". *Annals of Operations Research*, 322(2), P. 1119–1141. DOI: 10.1007/s10479-022-04935-6

17. Siciliano, G., Yu, Y., & Fottner, J. (2022), "Synergy between Shuttles and Stacker Cranes in Dynamic Hybrid Pallet Warehouses: Control Strategies and Performance Evaluation". *Anwendungen und Konzepte der Wirtschaftsinformatik*, Vol.15. P. 1–7. DOI: <https://doi.org/10.26034/lu.akwi.2022.3328>

18. Nevliudov, I., Romashov, Y. (2022), "Control of wheeled platforms straight motions taking into account jerk restrictions under speeding-up from the state of rest". *Visnyk of V. N. Karazin Kharkiv National University. Ser. Mathematics, Applied Mathematics and Mechanics*, Vol. 96, P. 4–22. <https://doi.org/10.26565/2221-5646-2022-96-01>

Received 30.11.2023

Відомості про авторів / About the Authors

Невлюдов Ігор Шакирович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, завідувач кафедри комп'ютерно-інтегрованих технологій, автоматизації та робототехніки, Харків, Україна; e-mail: igor.nevliudov@nure.ua; ORCID ID: <https://orcid.org/0000-0002-9837-2309>

Вжеснівський Максим Олегович – Харківський національний університет радіоелектроніки, аспірант кафедри комп'ютерно-інтегрованих технологій, автоматизації та робототехніки, Харків, Україна; e-mail: maksym.vzhesnievskyi@nure.ua; ORCID ID: <https://orcid.org/0000-0002-7585-628X>

Ромашов Юрій Володимирович – доктор технічних наук, доцент, Харківський національний університет радіоелектроніки, професор кафедри комп'ютерно-інтегрованих технологій, автоматизації та робототехніки; Харківський національний університет імені В.Н. Каразіна, професор кафедри прикладної математики, Харків, Україна; e-mail: yurii.romashov@nure.ua; ORCID ID: <https://orcid.org/0000-0001-8376-3510>

Чала Олена Олександрівна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри комп'ютерно-інтегрованих технологій, автоматизації та робототехніки, Харків, Україна; e-mail: olena.chala@nure.ua; ORCID ID: <https://orcid.org/0000-0003-2454-3774>

Nevliudov Igor – Doctor of Sciences (Engineering), Professor, Kharkiv National University of Radio Electronics, Head at the Department of Computer-Integrated Technologies, Automation and Robotics, Kharkiv, Ukraine.

Vzhesnievskyi Maksym – Kharkov National University of Radio Electronics, Post-Graduate Student at the Department of Computer-Integrated Technologies, Automation and Robotics, Kharkov, Ukraine.

Romashov Yurii – Doctor of Sciences (Engineering), Associate Professor, Kharkiv National University of Radio Electronics, Professor at the Department of Computer-Integrated Technologies, Automation and Robotics; V.N. Karazin Kharkiv National University, Professor at the Department of Applied Mathematics Department, Kharkiv, Ukraine.

Chala Olena – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Computer-Integrated Technologies, Automation and Robotics, Kharkiv, Ukraine.

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ МЕХАТРОННИХ ШАТЛІВ ЯК ОБ'ЄКТІВ АВТОМАТИЗАЦІЇ ДЛЯ БАГАТОРІВНЕВИХ СИСТЕМ ВНУТРІШНЬОСКЛАДСЬКОЇ ЛОГІСТИКИ

Предметом дослідження є математичне моделювання мехатронних шатлів багаторівневих систем внутрішньоскладської логістики. **Мета статті** – забезпечення енергоефективності та високої продуктивності багаторівневих систем складської логістики з допомогою розроблення математичних моделей мехатронних електромеханічних приводів у складі автоматизованих автономних шатлових систем (*Pallet Shuttle*). Для досягнення поставленої мети необхідно вирішити такі **завдання**: побудувати математичну модель процесів, що відбуваються в транспортувальному шатлі, як електромеханічної та мехатронної систем; розробити математичну модель транспортувального шатла як об'єкта автоматизації та отримання передатної функції автономної шатлової системи для автоматичного регулювання її швидкості. **Висновки**. У роботі розроблено узагальнені підходи щодо побудови математичних моделей електроприводів як об'єктів автоматизації. У запропонованому підході передбачено, що лінеаризована математична модель електроприводу як об'єкта автоматизації подана у вигляді лінійного диференційного рівняння зі сталими коефіцієнтами, що пов'язують керування та контрольований параметр. Побудову такого диференційного рівняння пропонується здійснювати способом узагальнення властивостей, притаманних електромеханічним системам, та передбачено, що ці властивості з використанням електромеханічних аналогій подані рівняннями Лагранжа 2-го роду. Унаслідок досліджень було показано, що використання такого підходу приводить до визначеного рівняння математичної моделі мехатронного шатла як об'єкта автоматизації. Показано, що лінеаризована математична модель шатла як об'єкта автоматизації має визначатися диференціальним рівнянням не нижче другого порядку. За результатами досліджень доведено можливість удосконалення систем автоматизації в проєктуванні способом більш деталізованого врахування властивостей шатлів як електромеханічних і мехатронних системам. Доведено, що використання такого підходу приводить до визначеного рівняння математичної моделі мехатронного шатла як об'єкта автоматизації, але отримати таке рівняння в явному вигляді практично не можливо. Для ілюстрації запропонованого підходу розглянуто моделювання типового мехатронного електроприводу машин, що використовуються в автоматизованих сховищах, та наведені результати моделювання процесів у електроприводі. Показано, що лінеаризована математична модель мехатронного шатла як об'єкта автоматизації має визначатися диференціальним рівнянням не нижче другого порядку. За результатами досліджень показано можливість удосконалення систем автоматизації під час проєктування завдяки більш деталізованому врахуванню властивостей шатлів як мехатронних електромеханічних систем.

Ключові слова: автоматизація; передатна функція; кіберфізичні ситеми; *Industry 4.0*; мехатронний електропривод; *Pallet Shuttle*; математичне моделювання.

Бібліографічні описи / Bibliographic descriptions

Невлюдов І. Ш., Вжесневський М. О., Ромашов Ю. В., Чала О. О. Математичне моделювання мехатронних шатлів як об'єктів автоматизації для багаторівневих систем внутрішньоскладської логістики. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 4 (26). С. 135–144. DOI: <https://doi.org/10.30837/ITSSI.2023.26.135>

Nevliudov, I., Vzhesnievskiy, M., Romashov, Y., Chala, O. (2023), "Mathematical modeling of mechatronic shuttles as automation objects for multilevel systems of intra-warehouse logistics", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 135–144. DOI: <https://doi.org/10.30837/ITSSI.2023.26.135>

К. БОНДАРЕНКО

АНАЛІЗ І ВИБІР РЕЛЕВАНТНОЇ МЕТРИКИ ВИЯВЛЕННЯ МЕРЕЖНИХ АНОМАЛІЙ

Метою дослідження є аналіз основних аспектів виявлення аномалій мережі та метрик їх оцінювання, що дає змогу вчасно виявляти кібератаки на мережу та значно підвищити рівень її безпеки. **Об'єктом дослідження** є виявлення мережних аномалій. **Завдання дослідження:** сформулювати принципи, що дозволяють здійснити узагальнення різних методів виявлення аномалій; проаналізувати метрики аномалій, зважаючи на заходи близькості формування оцінки поточного стану безпеки; обґрунтувати вибір релевантної міри близькості виявлення мережних аномалій. Сформульовано принципи, що дають змогу узагальнити різні методи виявлення аномалій. Для класифікації та полегшення виявлення мережних аномалій запропоновано метрики, що ґрунтуються на мірах близькості для типів даних, які характеризують аномалії. Визначено компоненти, що характеризують зазначену проблему, а саме: типи вхідних даних, прийнятність заходів близькості, маркування даних, класифікація методів, що ґрунтуються на використанні розмічених даних, виявлення відповідних особливостей та повідомлення про аномалії. Описано підхід, що дає змогу вчасно сформувати необхідний набір метрик, який забезпечить не тільки формування превентивних заходів протидії, а й дозволить оцінювати поточний стан системи безпеки загалом. Крім цього, забезпечується можливість формування багатоконтурних систем безпеки з огляду на вплив (комплексування) цільових (змішаних) атак на елементи інфраструктури, а також здатність їх синтезу з методами соціальної інженерії. **Висновки:** сформульовано принципи, що дають змогу виконати узагальнення різних методів виявлення аномалій; наведено види, показники та приклади мережних аномалій; запропоновано міри близькості для числових, категоріальних і змішаних типів даних з метою полегшення виявлення мережних аномалій; обґрунтовано вибір міри близькості Махаланобіса як основи метрики аномалій.

Ключові слова: мережна аномалія; система виявлення вторгнення; міра близькості; класифікація атак.

Вступ

Виявлення аномалій є важливою здатністю будь-якої схеми класифікації сигналів. Зважаючи на те, що ми ніколи не зможемо навчити систему машинного навчання на всіх класах об'єктів, з даними яких система може зіткнутися, стає важливим, щоб вона могла розрізняти інформацію про відомі та невідомі об'єкти під час тестування. Виявлення мережних аномалій є надзвичайно складним завданням. Саме з цієї причини існує кілька моделей виявлення аномалій, що добре зарекомендували себе на різних даних. Цілком очевидно, що не існує єдиної найкращої моделі для виявлення мережних аномалій, і успіх залежить не тільки від типу використовуваного методу, а й від статистичних властивостей даних, що обробляються. Попри значний прогрес та великий обсяг роботи все ще існує чимало можливостей для розвитку сучасних технологій виявлення та запобігання мережних атак. Спроба вторгнення чи загроза – це навмисна та несанкційна спроба (I) отримати доступ до інформації, (II) маніпулювати нею або (III) зробити систему ненадійною

чи непридатною для використання. Наприклад, (а) атака типу "відмова в обслуговуванні" (DoS) намагається позбавити хост ресурсів, необхідних для правильного функціонування під час оброблення; (б) хробаки та віруси застосовують інші хости з допомогою мережі; (в) зломи дають змогу отримати привілейований доступ до хосту, скориставшись перевагами відомих вразливостей.

Термін "виявлення вторгнень у мережі на основі аномалій" належить до проблеми визначення виняткових закономірностей у мережному трафіку, які не відповідають очікуваній нормальній поведінці. Ці невідповідні закономірності часто називають аномаліями, викидами, винятками, абераціями, сюрпризами, особливостями чи суперечливими спостереженнями в різних галузях застосування. "Аномалії" та "викиди" – два терміни, що найчастіше використовуються в контексті виявлення вторгнень у мережі на основі аномалій.

Можна стверджувати, що швидке виявлення вторгнень (аномалій, відхилень від нормальної роботи) може забезпечити вчасне формування превентивних заходів та/або необхідний рівень безпеки [1].

2. Аналіз літературних джерел і постановка проблеми

Одним із типів аналізу даних, який шукає незвичайні стани в системі, є виявлення аномалій, також відомі терміни "виявлення викидів" або "виявлення подій". Алгоритми виявлення аномалій є контрольними точками вхідного трафіку на різних етапах – від рівня мережі до центру оброблення інформації. В останньому випадку існує висока потреба в надійному виявленні для очищення даних [2] та цілей класифікації [3].

Виявлення аномалій у роботі мережі має широкий спектр застосунків, деякі з яких більш розвинені (наприклад, мережна безпека), а інші мають потенціал для зростання.

Аномалія – це точка даних, яка не пов'язана з прогнозованою поведінкою в системі, що моделюється. Аномалії – це рідкісні події або спостереження, які значно відхиляються від звичайної поведінки чи закономірностей, що спостерігаються в одній точці даних, у певному контексті або інтервалі часу (наприклад, сезон чи квартал) або весь набір даних. Здебільшого аномалії виникають унаслідок зовнішніх факторів, таких як відмова датчика або зовнішній напад, і мета алгоритму виявлення полягає в тому, щоб визначити, де відбулася аномалія, та класифікувати / обчислити причину. За умови бінарної класифікації аномалії вирішальне значення має апроксимаційна модель, яка найкраще відповідає очікуваній поведінці даних. Складність багатьох ситуацій потребує окремої стратегії виявлення кожного застосунку [4, 5].

У роботах [6, 7] подано чотири категорії методів виявлення мережних аномалій. Запропоновано класифікацію методів залежно від підходу до проблеми, способу застосування, типу методу та орієнтації алгоритму.

Описані в дослідженні [8] статистичні методи, зокрема метод мінімального обсягу, намагаються моделювати нормальні дані з використанням математичних моделей і розподілів. Підхід мінімального обсягу спрямований на створення n -вимірного симплекса навколо певної хмари даних (наземні справжні дані), де цільова функція полягає в тому, щоб мінімізувати займаний обсяг за умови максимізації точок наземних істинних даних. Аномалія визначається як будь-які дані, що не відповідають симплексу. У праці [9] наведено метод прогнозування, що називається експоненційним

згладжуванням. Зазначений метод прогнозує майбутню точку даних, використовуючи попередні точки та параметр згладжування. Аномальні дані, отримані статистичними методами, – це ті, що відхиляються від установленної моделі. Традиційні геометричні та статистичні методи підкріплені значним обсягом досліджень і покладаються на глибоке розуміння процесів, що відбуваються. Автори наголошують на необхідності рішень машинного навчання на основі даних та глибокого навчання, які дають змогу здійснювати більш гнучкі модифікації класичного методу прогнозування.

Для підкатегорії методів, пов'язаних із моделями машинного навчання та глибокого навчання, у роботі [10] стверджується, що характер даних, які надаються, визначає вибір моделі. Наприклад, моделі з довгою короткостроковою пам'яттю (*LSTM*) і трансформатори віддають перевагу послідовному введенню даних, зокрема аудіо, відео та часові ряди [11]. До того ж у статті [12] зазначено, що згортова нейронна мережа (*CNN*) і автоенкодер (*AE*) віддають перевагу непослідовним типам даних, зокрема введенню зображень. У студіях [13–15] увагу приділено алгоритмам, що намагаються розрізнити нормальну й аномальну поведінку, установлюючи межу прийняття рішення, наприклад, за допомогою класифікатора машини опорних векторів (*SVM*) [13] або майбутніх значень прогнозування поточкових даних [14] з мережами *LSTM* [15]. У роботі [16] показано, що залежності від наявності міток навчання ці підходи бувають контрольованими, напівконтрольованими, самоконтрольованими або повністю не контрольованими.

У публікаціях виокремлено три способи категоризації аномалій: конструктивний, руйнівний та очищення даних. Конструктивні застосунки мають продуктивний характер. У дослідженні [17] виконано порівняння продуктивності багатопарового перцептрона (*MLP*) і k -найближчих сусідів (*KNN*) та класифікатори *SVM* для широкого кола застосунків. Інші статті описують використання навчання з підкріпленням для різних застосунків безпілотних літальних апаратів (БПЛА) [18], а в роботі [19] використовується підхід машинного навчання для застосунків "розумного будинку". На відміну від конструктивних застосунків, деструктивні призначені для порушення повсякденної роботи з метою отримання сумнівної фінансової вигоди, наміру завдати шкоди мережі та програмам,

що проходять крізь мережу Інтернету речей, або порушити критично важливі бізнес-практики. У дослідженні [20] розглядаються різні кібератаки Інтернету речей та останні розробки в сфері безпеки Інтернету речей.

У статті [21] описано один із найбільш відомих типів аномалій, що залежить від обставин, – точковий, контекстуальний і колективний. Прикладом може бути виявлення шахрайства з кредитними картками [22].

Затримка й масштабованість алгоритму виявлення визначають, чи виконуватиметься метод "на льоту" на етапі збору даних або на більш пізньому етапі зберігання. Онлайн-алгоритм може послідовно обробляти інформацію за допомогою однієї точки даних або вікна, не маючи доступу до всіх вхідних даних. Традиційні геометричні та статистичні онлайн-підходи передбачають застосування раніше згаданих методів, основаних на відстані, щільності та відхиленні, а також методи, основані на кутах. У роботі [23] наведено приклад онлайн-методу, в якому використовуються нечіткі *C*-середні, а в статті [24] розглянуто ансамблевий підхід для виявлення аномалій.

Отже, аналіз показав, що попри широту охоплення різноманітних методів, предметних сфер і завдань виявлення мережних аномалій менше уваги приділяється ключовому питанню – аналізу метрик мережних аномалій та обґрунтуванню вибору релевантної метрики в різних ситуаціях.

3. Мета й завдання дослідження

Метою цієї роботи є дослідження основних аспектів виявлення аномалій мережі, що дасть змогу сформулювати метрики виявлення та забезпечити оцінювання рівня безпеки.

Для досягнення окресленої мети необхідно вирішити такі завдання:

- сформулювати принципи, що дозволять узагальнити різні методи виявлення аномалій;
- проаналізувати метрики аномалій, зважаючи на заходи близькості формування оцінки поточного стану безпеки;
- обґрунтувати вибір релевантної міри близькості виявлення мережних аномалій.

4. Аналіз та вибір метрик аномалій на основі заходів подібності

4.1. Формування принципів

узагальнення методів виявлення аномалій

Виявлення аномалій широко застосовується в таких галузях, як виявлення шахрайства з кредитними картками, встановлення вторгнень із метою кібербезпеки та військового спостереження за діями супротивника. Наприклад, аномальний характер трафіку в комп'ютерній мережі може означати, що зламаний комп'ютер надсилає конфіденційну інформацію неавторизованому хосту.

Вторгнення – це комплекс дій, спрямований на порушення безпеки комп'ютерних і мережних компонентів з погляду конфіденційності, цілісності та доступності. Це може бути зроблено внутрішнім або зовнішнім агентом для отримання несанкційного доступу та контролю за механізмом безпеки. Для захисту інфраструктури мережних систем системи виявлення вторгнень (*IDS*) надають механізми, що добре зарекомендували себе, які збирають і аналізують інформацію з різних сфер усередині хосту або мережі для виявлення можливих порушень безпеки.

Функції виявлення вторгнень передбачають:

- 1) моніторинг та аналіз дій користувача, системи й мережі;
- 2) налаштування систем для створення звітів про можливі вразливості;
- 3) оцінювання цілісності системи та файлів;
- 4) розпізнавання шаблонів типових атак;
- 5) аналіз аномальної активності та
- 6) відстеження порушень політики користувача.

IDS використовує оцінку вразливостей для оцінювання безпеки хосту чи мережі. Виявлення вторгнень ґрунтується на припущенні, що дії вторгнень помітно відрізняються від звичайних дій системи, отже, їх можна виявити.

Прийнято визначати два типи зловмисників: зовнішні та внутрішні. Зовнішні є неавторизованими користувачами машин, що зловмисники атакують, тоді як внутрішні мають дозвіл на доступ до системи, але не мають привілеїв для режиму *root* або суперкористувача. Маскарадний внутрішній зловмисник входить до системи як інший користувач, що має законний доступ до конфіденційної інформації, тоді як таємний внутрішній зловмисник найнебезпечніший, має право відключити контроль аудиту для себе.

У комп'ютерних системах існують різні класи вторгнень чи атак. Подамо їх у табл. 1.

Таблиця 1. Класи комп'ютерних атак: характеристики та приклад

Назва атаки	Характеристики	Приклад
Вірус	– Самовідтворювана програма, що заражає систему без відома користувача. – Збільшує ймовірність зараження мережної файлової системи, якщо до системи звертається інший комп'ютер.	<i>Trivial. 88.D, Polyboot.B, Tuareg</i>
Хробак	– Самовідтворювана програма, що поширюється через мережні служби комп'ютерних систем без утручання користувача. – Може завдати серйозної шкоди мережі, споживаючи пропускну спроможність мережі.	<i>SQL Slammer, Mydoom, Code Red, Nimda</i>
Троян	– Шкідлива програма, що не здатна копіювати себе, але може спричинити серйозні проблеми безпеки в комп'ютерній системі. – Виглядає як корисна програма, але насправді вона має секретний код, що здатний створити бекдор у системі, дозволяючи їй легко робити що-небудь у системі, і може бути викликаний, коли хакер отримує контроль над системою без дозволу користувача.	<i>Example-Mail Bomb, phishing attack</i>
Відмова в обслуговуванні (DoS)	– Спроби заблокувати доступ до системних або мережних ресурсів. – Втрата обслуговування – це нездатність конкретної мережі або хост-сервісу, наприклад електронної пошти, для роботи. – Реалізується способом примусового перезавантаження цільового комп'ютера (комп'ютерів), або способом споживання ресурсів. – Передбачувані користувачі не можуть адекватно спілкуватися через недоступність послуги або перешкоди засобів зв'язку.	<i>Buffer overflow, ping of death (PoD), TCP SYN, smurf, teardrop</i>
Мережна атака	– Будь-який процес, що використовується для зловмисної спроби поставити під загрозу безпеку мережі, починаючи з рівня каналу передачі інформації та завершуючи прикладним рівнем за допомогою різних засобів, зокрема маніпулювання мережними протоколами. – Незаконне використання облікових записів і привілеїв користувачів, виконання дій з вилучення мережних ресурсів та пропускну спроможності, виконання дій, що перешкоджають законним авторизованим користувачам отримати доступ до мережних служб та ресурсів.	<i>Packet injection, SYN flood</i>
Фізична атака	Спроба пошкодження фізичних компонентів мереж або комп'ютерів.	<i>Cold boot, evil maid</i>
Парольна атака	Метою є отримання пароля протягом нетривалого часу, що проявляється як серія невдалих спроб входу в систему.	Атака за словником, атака за допомогою SQL-ін'єкції
Атака для збору інформації	Збирає інформацію або знаходить відомі вразливості з допомогою сканування або перевірки комп'ютерів чи мереж.	Сканування SYS, сканування FIN, сканування XMAS
Атака для отримання root-прав користувача (U2R)	– Може використовувати вразливості для отримання привілеїв суперкористувача системи під час запуску в системі як звичайний користувач. – До вразливостей належать перехоплення паролів, атака за словником або соціальна інженерія.	Руткіт, завантажувальний модуль, Perl
Віддалена атака на локальний комп'ютер (R2L)	– Можливість відправляти пакети у віддалену систему з допомогою мережі, не маючи облікового запису в цій системі, або отримати доступ як користувач або root у системі та виконувати шкідливі операції. – Здійснення атаки на загальнодоступні служби (наприклад, HTTP і FTP) або в процесі підключення захищених сервісів (таких як POP та IMAP).	<i>Warezclient, warezmaster, imap, ftp write, multihop, phf, spy</i>
Зонд	– Сканує мережі для визначення дійсних IP-адрес та збору інформації про хост (наприклад, які послуги вони пропонують, використовується операційна система). – Надає зловмисникові інформацію зі списком потенційних уразливостей, які згодом можуть бути використані для атаки на обрані системи та служби.	перевірка IP та портів

Виявлення вторгнень, зважаючи на неправомірне використання, зазвичай шукає відомі шаблони вторгнень, але виявлення вторгнень з огляду на аномалії намагається виявити незвичайні шаблони. Методи виявлення вторгнень можна поділити на три типи залежно від механізму виявлення. До них належать: (I) основані на неправильному використанні; (II) основані на аномаліях та

(III) гібридні (див. табл. 2). Сучасні дослідники здебільшого зосереджуються на виявленні мережних вторгнень на основі аномалій, оскільки вони можуть встановлювати як відомі, так і невідомі атаки.

Щоб забезпечити відповідне рішення щодо виявлення мережних аномалій, необхідно сформувані концепцію нормальності. Ідея нормальності зазвичай вводиться за допомогою формальної моделі, яка

виражає відношення між фундаментальними якщо ступінь його відхилення щодо профілю або змінними, що беруть участь у динаміці системи. поведінки системи, заданої моделлю нормальності, Отже, подія чи об'єкт розпізнається як аномальне, досить високий.

Таблиця 2. Характеристики та типи методів виявлення вторгнень

Метод	Характеристики
Оснований на неправильному використанні	<ul style="list-style-type: none"> – Виявлення ґрунтується на наборі правил або сигнатур відомих атак. – Може виявити всі відомі шаблони атак на основі довідкових даних. – Написання сигнатури, що охоплює всі можливі варіанти відповідної атаки, є складним завданням.
Оснований на аномаліях	<ul style="list-style-type: none"> – Основне припущення: всі дії щодо вторгнення обов'язково є аномальними. – Такий метод створює нормальний профіль активності та перевіряє, чи стан системи відрізняється від встановленого профілю на статистично значущу величину, щоб повідомити про спроби вторгнення. – Аномальні дії, які не є вторгненням, можуть бути позначені як вторгнення. Це хибні спрацьовування. – Необхідно обирати порогові рівні так, щоб жодна з двох вищезгаданих проблем не була необґрунтовано посилена, а вибір функцій для моніторингу не було оптимізовано. – Обчислювально витратно через накладні витрати та можливе оновлення кількох матриць профілів системи.
Гібридний	<ul style="list-style-type: none"> – Використовує переваги як методів неправильного використання, так і методів виявлення аномалій. – Спроби виявити як відомі, так і невідомі атаки.

Наприклад, візьмемо систему виявлення аномалій S , яка використовує контрольований підхід. Її можна розглядати як пару $S = (M, D)$, де M – модель нормальної поведінки системи, а D – міра близькості, що дозволяє за даними запису активності визначити ступінь відхилення, який ця діяльність має щодо моделі M . Отже, кожна система має здебільшого два модулі: модуль моделювання та модуль виявлення. Системи навчаються отримати модель нормальності M . Отримана модель згодом використовується модулем виявлення для оцінювання

нових подій, об'єктів чи трафіку як аномальних або викидів. Саме вимір відхилення дає змогу класифікувати події чи об'єкти як аномальні або викиди. Зокрема модуль моделювання має бути адаптивним, щоб справлятися з динамічними сценаріями.

Існують дві широкі категорії мережних аномалій: (а) аномалії, пов'язані з функціонуванням, і (б) аномалії, пов'язані з безпекою. Аномалії, пов'язані з безпекою, поділяються на три типи: точкові, контекстні та колективні. Ця схема класифікації описана в табл. 3.

Таблиця 3. Аномалії: види, характеристики та приклади

Типи	Характеристики	Приклад
Точкова аномалія	Зразок окремих даних, який виявився аномальним щодо інших даних.	Ізольований екземпляр мережного трафіку від звичайних екземплярів у певний час.
Контекстна аномалія	<ul style="list-style-type: none"> – Зразок даних, визнаний аномальним у певному контексті. – Контекст визначається структурою набору даних. – Для визначення контексту використовуються два набори атрибутів: (а) контекстуальні та (б) поведінкові. 	Інтервал часу між транзакціями під час шахрайства з кредитними картками.
Колективна аномалія	<ul style="list-style-type: none"> – Множина пов'язаних екземплярів даних, які виявилися аномальними щодо всього набору даних. – Сукупність подій є аномалією, але окремі події не є аномаліями, якщо вони відбуваються окремо в послідовності. 	Послідовність, подібна до такої: ... <i>http – web</i> , переповнення буфера, <i>http – web</i> , <i>http – web</i> , <i>ftp</i> , <i>httpweb</i> , <i>ssh</i> , <i>http – web</i> , <i>ssh</i> , переповнення буфера...

Розгляд основних аспектів виникнення, прояву мережних аномалій, а також характеристик і наявних методів виявлення вторгнень дає змогу сформулювати принципи узагальнення методів виявлення аномалій (див. табл. 4).

Сформульовані принципи можна розглядати як сукупність критеріїв та обмежень, що дозволяють оцінювати наявні та проектувати нові методи й технології виявлення мережних аномалій.

Таблиця 4. Принципи узагальнення методів виявлення аномалій

№	Принцип	Опис
1	Принцип надійності та компромісу	Метод виявлення новизни має забезпечувати надійну роботу з тестовою інформацією, що дозволяє максимально вилучити нові зразки та звести до мінімуму вилучення відомих зразків. Цей компроміс має бути обмеженою мірою передбачуваним і під експериментальним контролем.
2	Принцип одноманітного масштабування даних	Щоб полегшити виявлення новизни, має бути можливість, щоб усі тестові дані та дані навчання після нормалізації розміщувалися в одному діапазоні.
3	Принцип мінімізації параметрів	Метод виявлення новизни має бути спрямований на мінімізацію кількості параметрів, що встановлюються користувачем.
4	Принцип узагальнення	Система має бути здатною до узагальнення, не плутаючи узагальнену інформацію з новою.
5	Принцип незалежності	Метод виявлення новизни не має залежати від кількості доступних ознак і класів та має демонструвати прийнятну продуктивність у контексті незбалансованого набору даних, невеликої кількості вибірок і шуму.
6	Принцип адаптивності	Важливо, щоб система, яка розпізнає нові зразки під час тестування, мала змогу використовувати цю інформацію для повторного навчання.
7	Принцип обчислювальної складності	Низка застосунків для виявлення аномалій є в мережі, тому обчислювальна складність механізму виявлення аномалій має бути якнайменшою.

4.2. Аналіз метрик аномалій на основі мір близькості

Проблема виявлення мережних аномалій є завданням класифікації або кластеризації, що визначається такими компонентами [25]:

- типи вхідних даних;
- прийнятність мір близькості;
- маркування даних;
- класифікація методів, основаних на використанні розмічених даних;
- виявлення відповідних особливостей;
- повідомлення про аномалії.

Типи вхідних даних. Ключовим аспектом будь-якого методу виявлення вторгнень мереж на основі аномалій є характер вхідних даних, що використовуються для аналізу. Вхідні дані зазвичай є набором екземплярів даних (так званими об'єктами, записами, точками, векторами, шаблонами, подіями, випадками, вибірками, спостереженнями, об'єктами). Кожен екземпляр даних можна описати за допомогою набору атрибутів двійкового, категоріального або числового типу. Кожен екземпляр даних може складатися лише з одного атрибута (одномірний) або кількох атрибутів (багатомірний). У разі екземплярів багатовимірних даних усі атрибути можуть бути одного типу або бути поєднанням типів даних. Природа атрибутів визначає застосовність методів виявлення аномалій.

Доцільність заходів близькості. Заходи близькості (подібності чи відмінності) необхідні для розв'язання багатьох проблем розпізнавання образів для класифікації та кластеризації. Відстань – це кількісний ступінь того, наскільки далеко один

від одного є два об'єкти. Заходи відстані, що задовольняють метричні властивості, називаються просто метрикою, тоді як інші неметричні заходи відстані іноді називають дивергенцією. Вибір міри близькості залежить від типу вимірювання чи подання об'єктів.

Як правило, міри близькості – це функції, що приймають аргументи у вигляді пар об'єктів і повертають числові значення, які стають вищими залежно від того, як об'єкти стають більш схожими. Міра близькості зазвичай визначається так.

Визначення: міра близькості D – це функція, $X \times X \rightarrow \mathbb{R}$, що має такі властивості:

- позитивність $\forall x, y \in X, S(x, y) \geq 0$;
- симетричність $\forall x, y \in X, S(x, y) = S(y, x)$;
- максимальність $\forall x, y \in X, S(x, x) \geq S(x, y)$,

де X – простір даних (інша назва «всесвіт»), а x, y – пара k -мірних об'єктів.

Найбільш поширені заходи близькості для числових, категоріальних і змішаних типів даних перелічені в табл. 5. Для числових даних передбачається, що вони подані у вигляді векторів, які містять реальні числа. Значення атрибутів належать до безперервної ділянки. Передбачається, що є два об'єкти: $x = x_1, x_2, x_3, \dots, x_d$, $y = y_1, y_2, y_3, \dots, y_d$ і Σ^{-1} , що є коваріацією даних з атрибутами, тобто розмірністю.

Для категоріальних даних обчислення мір подібності чи близькості не є простим через те, що немає чіткого поняття впорядкування категоріальних значень. Найпростіший спосіб знайти подібність

між двома категоріальними атрибутами – присвоїти подібність 1, якщо значення ідентичні, і 0, якщо значення не ідентичні. У табл. 5 функція $S_k(x_k, y_k)$ є подібністю

за атрибутами. Вага атрибута w_k для атрибута k обчислюється, як показано у таблиці, IOF – зворотна частота виникнення, а OF – частота виникнення.

Таблиця 5. Міра близькості для даних числового, категоріального та змішаного типу

Numeric			
Name	Measure, $S_i(x_i, y_i)$	Name	Measure, $S_i(x_i, y_i)$
Euclidean	$\sqrt{\sum_{i=1}^d x_i - y_i ^2}$	Weighted Euclidean	$\sqrt{\sum_{i=1}^d \alpha_i x_i - y_i ^2}$
Squared Euclidean	$\sum_{i=1}^d x_i - y_i ^2$	Squared-chord	$\sum_{i=1}^d (\sqrt{x_i} - \sqrt{y_i})^2$
Squared X^2	$\sum_{i=1}^d \frac{(x_i - y_i)^2}{x_i + y_i}$	City block	$\sum_{i=1}^d x_i - y_i $
Minkowski	$\sqrt[p]{\sum_{i=1}^d x_i - y_i ^p}$	Chebyshev	$\max_i x_i - y_i $
Canberra	$\sum_{i=1}^d \frac{ x_i - y_i }{x_i + y_i}$	Cosine	$\frac{\sum_{i=1}^d x_i y_i}{\sqrt{\sum_{i=1}^d x_i^2} \sqrt{\sum_{i=1}^d y_i^2}}$
Jaccard	$\frac{\sum_{i=1}^d x_i y_i}{\sum_{i=1}^d x_i^2 + \sum_{i=1}^d y_i^2 - \sum_{i=1}^d x_i y_i}$	Bhattacharyya	$-\ln \sum_{i=1}^d \sqrt{(x_i y_i)}$
Pearson	$\sum_{i=1}^d (x_i - y_i)^2$	Divergence	$2 \sum_{i=1}^d \frac{(x_i - y_i)^2}{(x_i + y_i)^2}$
Mahalanobis	$\sqrt{(x - y)^t \Sigma^{-1} (x - y)}$	–	–
Categorical			
$w_k, k = 1 \dots d$	Measure, $S_k(x_k, y_k)$	$w_k, k = 1 \dots d$	Measure, $S_k(x_k, y_k)$
1/2	$Overlap = \begin{cases} 1 & \text{if } x_k = y_k \\ 0 & \text{otherwise} \end{cases}$	1/d	$Eskin = \begin{cases} 1 & \text{if } x_k = y_k \\ \frac{n_k^2}{n_k^2 + 2} & \text{otherwise} \end{cases}$
1/d	$IOF = \begin{cases} 1 & \text{if } x_k = y_k \\ \frac{1}{1 + \log f_k(x_k) \times \log f_k(y_k)} & \text{otherwise} \end{cases}$	1/d	$OF = \begin{cases} 1 & \text{if } x_k = y_k \\ \frac{1}{1 + \log \frac{N}{f_k(x_k)} \times \log \frac{N}{f_k(y_k)}} & \text{otherwise} \end{cases}$
Mixed			
Name	Measure	Name	Measure
General Similarity Coefficient	$s_{gsc}(x, y) = \frac{\sum_{k=1}^d w(x_k, y_k) s(x_k, y_k)}{\sum_{k=1}^d w(x_k, y_k)}$ Для числових атрибутів, $s(x_k, y_k) = 1 - \frac{ x_k - y_k }{R_k}$, де R_k – діапазон k -го атрибуту; $w(x_k, y_k) = 0$, якщо x чи y мають пропущені значення для k -го атрибуту; інакше $w(x_k, y_k) = 1$. Для категоріальних атрибутів, $s(x_k, y_k) = 1$, якщо $x_k = y_k$; інакше $s(x_k, y_k) = 0$; $w(x_k, y_k) = 0$, якщо точка даних x чи y має пропущене значення k -го атрибуту; інакше $w(x_k, y_k) = 1$.	General Distance Coefficient	$d_{gdc}(x, y) = \left(\frac{\sum_{k=1}^d w(x_k, y_k) d^2(x_k, y_k)}{\sum_{k=1}^d w(x_k, y_k)} \right)^{1/2}$, де $d^2(x_k, y_k)$ – квадратна відстань для k -го атрибуту; $w(x_k, y_k)$ – аналогічно як у <i>General Similarity Coefficient</i> . Для числових атрибутів, $d(x_k, y_k) = \frac{ x_k - y_k }{R_k}$, де R_k – діапазон значень k -го атрибуту. Для категоріальних атрибутів, $d(x_k, y_k) = 0$, якщо $x_k = y_k$; інакше $d(x_k, y_k) = 1$.

Нарешті, дані змішаного типу містять категоріальні й числові значення. Звичайною практикою кластеризації змішаного набору даних є перетворення категоріальних значень на числові значення з подальшим використанням алгоритму числової кластеризації. Інший підхід полягає в прямому порівнянні категоріальних значень, за умови якого два різні значення дають відстань, що дорівнює 1, а ідентичні значення – відстань, що дорівнює 0. Звичайно, можна використовувати й інші заходи для категоріальних даних. Дві добре відомі міри близькості, загальний коефіцієнт подібності й загальний коефіцієнт відстані даних змішаного типу показані в табл. 5. Такі методи можуть брати до уваги інформацію про подібність, укладену в категоріальних значеннях. Отже, кластеризація не спроможна точно виявити структуру подібності множини даних.

Маркування даних. Позначка, пов'язана з екземпляром даних, вказує, чи є цей екземпляр нормальним або аномальним. Необхідно зазначити, що отримання точних даних як нормального, так і аномального типу здебільшого обходиться непомірно дорого. Маркування часто виконується вручну фахівцями-людьми, і, отже, для отримання маркованого набору навчальних даних потрібні значні зусилля. Крім того, аномальна поведінка часто має динамічний характер, наприклад, можуть виникати нові типи аномалій, для яких немає позначених навчальних даних.

Класифікація методів, основана на використанні розмічених даних. Залежно від ступеня доступності міток методи виявлення аномалій можуть працювати в трьох режимах – контрольованому, напівконтрольованому й неконтрольованому.

У контрольованому режимі передбачається наявність набору навчальних даних, у якому зазначені екземпляри як нормального, так і аномального класу. Типовий підхід у таких випадках – побудова прогнозу моделі для класів нормальних та аномальних. Будь-який невидимий екземпляр даних порівнюється з моделлю, щоб визначити, до якого класу він належить. У разі контрольованого виявлення аномалій виникають дві основні проблеми. По-перше, аномальних випадків у навчальних даних набагато менше, ніж звичайних. По-друге, отримання точних і репрезентативних міток, особливо для класу аномалій, зазвичай є складним завданням. Низка методів вводять штучні аномалії у звичайний набір даних для отримання позначеного набору навчальних даних.

Напівконтрольовані методи припускають, що дані навчання позначають екземпляри тільки для класу. Оскільки їм не потрібні мітки класу аномалій, їх легше використовувати, на відміну від контрольованих методів.

Нарешті, методи без учителя не вимагають навчальних даних, отже, потенційно найбільш застосовувані. Методи цієї категорії неявно припускають, що нормальні ситуації трапляються набагато частіше, ніж аномалії тестових даних. Коли це припущення не правильне, такі методи страждають від високого рівня хибних тривог. Багато напівконтрольованих методів можна адаптувати для роботи в неконтрольованому режимі, використовуючи вибірку немаркованого набору даних як навчальних даних. Така адаптація передбачає, що тестові дані містять дуже мало аномалій і навчена модель стійка до цих небагатьох аномалій.

Ідентифікація релевантної функції. Вибір функцій відіграє важливу роль у виявленні мережних аномалій. Методи вибору функцій використовуються у сфері виявлення вторгнень для унеможливлення неважливих або нерелевантних функцій. Вибір ознак знижує обчислювальну складність, усуває надмірність інформації, підвищує точність алгоритму виявлення, полегшує розуміння даних та покращує узагальнення. Процес вибору функцій передбачає три основні етапи: (а) створення підмножини, (б) оцінювання підмножини та (в) перевірка. Три різні підходи до генерації підмножини: повний, евристичний і випадковий. Функції оцінки поділяються на п'ять окремих категорій: на основі оцінок, на основі ентропії чи взаємної інформації, на основі кореляції, на основі узгодженості та на основі точності виявлення. Моделювання та реалізація в реальному світі – два способи перевірки оціненої підмножини.

Алгоритми вибору ознак поділяються на три типи: методи-оболонки, фільтри та гібридні методи. Тоді як методи-оболонки намагаються оптимізувати деякі зумовлені критерії щодо набору функцій у межах процесу вибору, методи фільтрації покладаються на загальні характеристики навчальних даних для вибору функцій, які не залежать один від одного й сильно залежать від вихідних даних. Гібридний метод вибору ознак намагається використати суттєві особливості методів обгортки та фільтра.

Повідомлення про аномалії. Важливим аспектом будь-якого методу виявлення аномалій є спосіб повідомлення про аномалії. Зазвичай вихідні дані, отримані за допомогою методів виявлення аномалій,

бувають двох типів: (а) оцінка, яка є значенням, що поєднує (I) відстань або відхилення з посиленням на набір профілів або сигнатур; (II) вплив більшості в його околиці та (III) явне домінування відповідного підпростору; (б) позначка, що є значенням (нормальним або аномальним), присвоєним кожному екземпляру тесту. Зазвичай маркування екземпляра залежить від (I) розміру груп, створених за допомогою неконтрольованого методу; (II) компактності групи (груп); (III) голосування більшості на основі результатів, отриманих за допомогою декількох індексів або (IV) явне домінування підмножини ознак.

Аналіз запропонованої таблиці дає змогу зробити висновок, що серед безлічі поданих заходів близькості суттєво виокремлюється міра Махаланобіса. Її особливість пояснюється тим, що вона єдина, яка бере до уваги значення коефіцієнтів кореляції між різними вимірами, отриманими під час моніторингу активності мережі. Це змушує звернути увагу та більш детально проаналізувати саме цей захід близькості для виявлення мережних аномалій.

4.3. Обґрунтування вибору релевантної міри близькості виявлення мережних аномалій

Вибір міри близькості між спостереженнями дає підстави для обґрунтування метрики аномалій мережі, що виявляються в режимі онлайн. Нехай $A = \{a_1, \dots, a_n\}$ – набір атрибутів мережної активності, що відстежуються, в довільний момент часу. Контрольовані атрибути можуть збиратися як внутрішніми, так і зовнішніми датчиками системи виявлення вторгнень (IDS). Дані періодично ростуть кожні t мілісекунд. Вхідний вектор $\vec{i}_t = \{i_{t,1}, \dots, i_{t,n}\}$ надається онлайн, де $i_{t,j} \in \mathbb{R}$ позначає значення атрибута a_j за певний час t . З кожним заданим \vec{i}_t рішення має бути прийняте негайно, незалежно від того, чи є \vec{i}_t аномальним, чи ні.

Також доступні попередні дані H (передбачені нормальним функціонуванням мережі). H – це $m \times n$ матриця, де стовпці позначають n атрибутів, що відстежуються, а рядки зберігають значення цих атрибутів за m періодів часу. H може бути записаний на основі повної роботи, яка свідомо є номінальною (наприклад, трафік без відомих збоїв), або він може бути створений на основі останніх вхідних даних m , які були отримані в режимі онлайн, тобто $H = \{\vec{i}_{t-m-1}, \dots, \vec{i}_{t-1}\}$.

Розглянемо для виявлення аномалій дві найбільш популярні міри близькості спостережень за умови виявлення мережних аномалій – евклідову метрику й метрику Махаланобіса. Їх принципова різниця полягає в тому, що евклідова відстань може бути визначена для кожної пари спостережень, і, на відміну від відстані Махаланобіса, жодним чином не зважає на кореляцію, що існує між елементами всього вектора спостережень.

Відстань Махаланобіса – це n -мірний Z -показник. Він обчислює відстань між n -вимірною точкою та групою інших точок в одиницях стандартних відхилень. На відміну від звичайної n -вимірної евклідової відстані, відстань Махаланобіса також зважає на розподіл точок. Отже, якщо група точок є спостереженнями, то відстань Махаланобіса вказує, чи є нова точка викидом порівняно зі спостереженням. Точка зі значеннями, аналогічними точкам спостереження, розташована в багатовимірному просторі, в щільній ділянці й матиме меншу відстань Махаланобіса. Однак якщо викид буде розташований за межами щільної ділянки, тоді для нього буде більша відстань Махаланобіса.

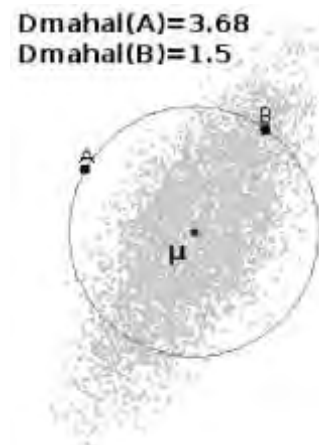


Рис. 1. Евклідова відстань порівняно з відстанню Махаланобіса

Формально відстань Махаланобіса обчислюється так: нехай $\vec{i}_t = \{i_{t,1}, i_{t,2}, \dots, i_{t,n}\}$ – вектор поточних вхідних даних n атрибутів, що відстежуються, а H матриця розмірністю $m \times n$ – група номінальних значень цих атрибутів. Визначаємо середнє значення H за значенням вектора $\mu = (\mu_1, \mu_2, \dots, \mu_n)$, а S – матриця H . Відстань Махаланобіса, D_{mahal} , від \vec{i}_t до H визначається так:

$$D_{mahal}(\vec{i}_t, H) = \sqrt{(\vec{i}_t - \vec{\mu})^T S^{-1} (\vec{i}_t - \vec{\mu})}$$

Приклад зображено на рис. 1, на якому можна побачити, що, хоча A і B мають однакову евклідову відстань від центроїда μ , відстань Махаланобіса для $A(3,68)$ більша, ніж для $B(1,5)$, оскільки екземпляр B імовірніший, ніж екземпляр A щодо інших точок.

Завдяки природі відстані Махаланобіса можемо використовувати його для виявлення аномалій у мережному оточенні. Кожен із n атрибутів домену (наприклад, трафіку) відповідає виміру. Вхідний вектор \vec{i}_i – це n -вимірний вектор, який вимірюється відстанню Махаланобіса щодо H . Відстань Махаланобіса потім використовується для вказівки того, чи кожна нова вхідна точка \vec{i}_i є викидом щодо H .

5. Аналіз досягнутих результатів

Вибір міри близькості Махаланобіса як підстави метрики виявлення мережних аномалій пояснюється тим, що тільки міра близькості Махаланобіса бере до уваги корельованість спостережень і, відповідно до цього зважає на геометрію розкиду спостерігачів нормального режиму роботи, що дає більш обґрунтовані оцінки для віднесення спостереження.

Використовуючи відстань Махаланобіса, можна легко виявити три загальні категорії аномалій.

1) *Точкові аномалії*: неприпустимі екземпляри даних, що відповідають неприпустимим значенням \vec{i}_i .

2) *Контекстуальні аномалії*: екземпляри даних, які є аномальними лише щодо певного контексту, але не інакше. У нашому підході контекст забезпечується змінними даними ковзного вікна.

3) *Групові аномалії*, що є пов'язаними екземплярами даних, які допустимі окремо, але аномальні, коли

вони зустрічаються разом. Це досягається завдяки багатовимірності точок, що визначаються відстанню Махаланобіса.

Аномалія будь-якого типу може призвести до того, що репрезентативна точка виявиться окремо від номінальних точок у відповідному вимірі, що помістить її за межі щільної ділянки. Це призведе до значної відстані Махаланобіса і, зрештою, спрацьовування тривоги.

Висновок

1. У статті розглянуто основні аспекти виявлення мережних аномалій. Сформульовано принципи, що дають змогу узагальнити різні методи виявлення аномалій. Описано атаки, з якими зазвичай стикаються системи виявлення мережних вторгнень, а також характеристики й типи методів їх виявлення. Мережні аномалії розглянуто як прояви мережних атак, що дозволяє виконати класифікацію аномалій.

2. Подано види, показники та приклади мережних аномалій. Для класифікації та полегшення виявлення мережних аномалій пропонуються міри близькості для числових (що характеризують аномалії), категоріальних і змішаних типів даних.

3. Аргументовано вибір міри близькості Махаланобіса як основи метрики аномалій. Обґрунтовано, що тільки міра близькості Махаланобіса бере до уваги корельованість спостережень і, отже, враховує геометрію розкиду спостережень нормального режиму роботи й відповідно дає більш повні оцінки для визначення спостереження як аномального.

Список літератури

1. Yevseiev S., Zviertseva N., Pribyliev Y., Lezik O., Komisarenko O., Nalyvaiko A., Pogorelov V., Katsalap V., Husarova I. Development of the concept for determining the level of critical business processes security. *Eastern-European Journal of Enterprise Technologies*. 2023. Vol. 1/9. No. 121. P. 21–40. DOI: 10.15587/1729-4061.2023.274301
2. Blazquez-Gartfa, A., Conde A., Mori U., Lozano J. A review on outlier/anomaly detection in time series data, *ACM Comput. Surv.* 2021. Vol. 54. No. 3. DOI: <http://dx.doi.org/10.1145/3444690>
3. Arif I., Ackovska N. IoT aided smart home architecture for anomaly detection, in: *Data Science and Internet of Things: Research and Applications at the Intersection of DS and IoT*. Springer International Publishing, Cham. 2021. P. 1–19. DOI: http://dx.doi.org/10.1007/978-3-030-67197-6_1
4. Lin X., Yeh E., Lin P. Anomaly detection for IoT systems, in: *Encyclopedia of Wireless Networks*. Springer International Publishing, Cham. 2020. P. 18–20. DOI: http://dx.doi.org/10.1007/978-3-319-78262-1_183
5. Pei J., Zhong K., Jan M., Li J. RETRACTED: Personalized federated learning framework for network traffic anomaly detection. *Computer Networks*. 2022. Vol. 209, P. 1389–1286. DOI: <https://doi.org/10.1016/j.comnet.2022.108906>

6. Fahim M., Sillitti A. Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review. *IEEE Access* 7. 2019. P. 81664–81681. DOI: <http://dx.doi.org/10.1109/ACCESS.2019.2921912>
7. Cook A., Misirli G., Fan Z. Anomaly detection for IoT time-series data: A survey. *IEEE Internet Things J.* 2020. Vol. 7. No. 7. P. 6481–6494. DOI: <http://dx.doi.org/10.1109/JIOT.2019.2958185>
8. O'Reilly C., Gluhak A., Imran M. Distributed anomaly detection using minimum volume elliptical principal component analysis. *IEEE Trans. Knowl.* 2016. Vol. 28. No. 9. P. 2320–2333. DOI: <http://dx.doi.org/10.1109/TKDE.2016.2555804>
9. Mahajan S., Chen L., Tsai T. Short-term PM2.5 forecasting using exponential smoothing method: a comparative analysis. *Sensors*. 2018. Vol. 18. No. 10. 3223 p. DOI: <http://dx.doi.org/10.3390/s18103223>
10. Charles, A. Interpreting deep learning: the machine learning rorschach test? 2018. URL: [arXiv:1806.00148](https://arxiv.org/abs/1806.00148)
11. Chen Z., Chen D., Zhang X., Yuan Z., Cheng X. Learning graph structures with transformer for multivariate time series anomaly detection in IoT. *IEEE Internet Things J.* 2021. No. 1. P. 1–12. DOI: <http://dx.doi.org/10.1109/JIOT.2021.3100509>
12. Ukil A., Bandyopadhyay S., Puri C., Pal A. IoT healthcare analytics: The importance of anomaly detection. *30th International Conference on Advanced Information Networking and Applications. AINA*. 2016. P. 994–997. DOI: <http://dx.doi.org/10.1109/AINA.2016.158>
13. Yang K., Kpotufe S., Feamster N. An efficient one-class SVM for anomaly detection in the internet of things. 2021. URL: [arXiv:2104.11146](https://arxiv.org/abs/2104.11146)
14. Dunne M., Gracioli G., Fischmeister S. A comparison of data streaming frameworks for anomaly detection in embedded systems. *Proceedings of the 1st International Workshop on Security and Privacy for the Internet-of-Things IoTSec. 2018. Orlando, FL, USA*. URL: <https://uwaterloo.ca/embedded-software-group/publications/comparison-data-streaming-frameworks-anomaly-detection>
15. Wu D., Jiang Z., Xie X., Wei X., Yu W., Li R. LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT. *IEEE Trans. Ind. Inf.* 2020. Vol 16. No 8. P. 5244–5253. DOI: <http://dx.doi.org/10.1109/TII.2019.2952917>
16. Fahim M., Sillitti A. Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review. *IEEE Access*. 2019. Vol. 7. P. 81664–81681. DOI: <http://dx.doi.org/10.1109/ACCESS.2019.2921912>
17. Galvao Y., Albuquerque V., Fernandes B., Valenka M. Anomaly detection in smart houses: Monitoring elderly daily behavior for fall detecting. *Latin American Conference on Computational Intelligence. la-CCI*. 2017. P. 1–6. DOI: <http://dx.doi.org/10.1109/LA-CCI.2017.8285701>
18. Lu H., Li Y., Mu S., Wang D., Kim H., Serikawa S. Motor anomaly detection for unmanned aerial vehicles using reinforcement learning. *IEEE Internet Things J.* 2018. Vol. 5. No. 4. P. 2315–2322. DOI: <http://dx.doi.org/10.1109/JIOT.2017.2737479>
19. Nguyen T., Marchal S., Miettinen M., Fereidooni H., Asokan N., Sadeghi A. Diot: A federated self-learning anomaly detection system for IoT. *39th International Conference on Distributed Computing Systems. ICDCS*. 2019. P. 756–767. DOI: <http://dx.doi.org/10.1109/ICDCS.2019.00080>
20. Alsheikh M., Konieczny L., Prater M., Smith G., Uludag S. State and trends of IoT security: Unequivocal appeal to cybercriminals, onerous to defenders. *IEEE Consum. Electr. Mag.* 2021. Vol. 1. P. 1–17. DOI: <http://dx.doi.org/10.1109/MCE.2021.3079635>
21. Munir M., Siddiqui S., Dengel A., Ahmed S. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series. *IEEE Access*. 2019. Vol. 7. P. 1991–2005. DOI: <http://dx.doi.org/10.1109/ACCESS.2018.2886457>
22. Srikanth P. An efficient approach for clustering and classification for fraud detection using bankruptcy data in IoT environment. *Int. J. Inf. Technol.* 2021. P. 1–7. URL: <https://www.x-mol.net/paper/article/1442394146737025024>
23. Hafeez I., Antikainen M., Ding A., Tarkoma S. IoT-KEEPER: Detecting malicious IoT network activity using online traffic analysis at the edge. *IEEE Trans. Netw. Serv. Manag.* 2020. Vol. 17. No. 1. P. 45–59. DOI: <http://dx.doi.org/10.1109/TNSM.2020.2966951>
24. Bosman H., Lacca G., Tejada A., Wortche H., Liotta A. Ensembles of incremental learners to detect anomalies in ad hoc sensor networks. *Ad Hoc Netw.* 2015. No. 35. P. 14–36. DOI: <http://dx.doi.org/10.1016/j.adhoc.2015.07.013>
25. Milov O., Yevseiev S., Opirskyy I., Dunaievskya O., Huk O., Pogorelov V., Bondarenko K., Zviertseva N., Yevgen Melenti Y., Tomashevsky B. Development of concepts for the cyber security metrics classification. *Eastern-European Journal of Enterprise Technologies*. 2022. Vol. 4/4. No. 118. P. 6–18, DOI: <https://doi.org/10.15587/1729-4061.2022.263416>

References

1. Yevseiev, S., Zviertseva, N., Pribyliev, Y., Lezik, O., Komisarenko, O., Nalyvaiko, A., Pogorelov, V., Katsalap, V., Husarova, I. (2023), "Development of the concept for determining the level of critical business processes security", *Eastern-European Journal of Enterprise Technologies*, No. 1/9 (121). P. 21–40. DOI: 10.15587/1729-4061.2023.274301

2. Blazquez-Gartfa, A., Conde, A., Mori, U., Lozano, J. (2021), "A review on outlier/anomaly detection in time series data", *ACM Comput.* No. 54 (3). DOI: <http://dx.doi.org/10.1145/3444690>
3. Arif, I., Ackovska, N. (2021), "IoT aided smart home architecture for anomaly detection, in: Data Science and Internet of Things: Research and Applications at the Intersection of DS and IoT", *Springer International Publishing, Cham*, P. 1–19. DOI: http://dx.doi.org/10.1007/978-3-030-67197-6_1
4. Lin, X., Yeh, E., Lin, P. (2020), "Anomaly detection for IoT systems, in: Encyclopedia of Wireless Networks", *Springer International Publishing, Cham*, P. 18–20. DOI: http://dx.doi.org/10.1007/978-3-319-78262-1_183
5. Pei, J., Zhong, K., Jan, M., Li, J. (2022), "RETRACTED: Personalized federated learning framework for network traffic anomaly detection", *Computer Networks*, Vol. 209, 108906, ISSN 1389–1286. DOI: <https://doi.org/10.1016/j.comnet.2022.108906>
6. Fahim, M., Sillitti, A. (2019), "Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review", *IEEE Access* No.7, P. 81664–81681. DOI: <http://dx.doi.org/10.1109/ACCESS.2019.2921912>
7. Cook, A., Misirli, G., Fan, Z. (2020), "Anomaly detection for IoT time-series data: A survey", *IEEE Internet Things J.*, Vol. 7, No. 7, P. 6481–6494. DOI: <http://dx.doi.org/10.1109/JIOT.2019.2958185>
8. O'Reilly, C., Gluhak, A., A. Imran, M. (2016), "Distributed anomaly detection using minimum volume elliptical principal component analysis", *IEEE Trans. Knowl.* Vol. 28, No. 9, P. 2320–2333. DOI: <http://dx.doi.org/10.1109/TKDE.2016.2555804>
9. Mahajan, S., Chen, L., Tsai, T. (2018), "Short-term PM2.5 forecasting using exponential smoothing method: a comparative analysis", *Sensors*, Vol. 18, No. 10. 3223 p. DOI: <http://dx.doi.org/10.3390/s18103223>
10. Charles, A. (2018), "Interpreting deep learning: the machine learning roschach test?" available at: [arXiv:1806.00148](https://arxiv.org/abs/1806.00148)
11. Chen, Z., Chen, D., Zhang, X., Yuan, Z., Cheng, X. (2021), "Learning graph structures with transformer for multivariate time series anomaly detection in IoT", *IEEE Internet Things J.* No. 1. P. 1–12. DOI: <http://dx.doi.org/10.1109/JIOT.2021.3100509>
12. Ukil, A., Bandyopadhyay, S., Puri, C., Pal, A. (2016), "IoT healthcare analytics: The importance of anomaly detection, in: 2016 IEEE", *30th International Conference on Advanced Information Networking and Applications, AINA*, P. 994–997. DOI: <http://dx.doi.org/10.1109/AINA.2016.158>
13. Yang, K., Kpotufe, S., Feamster, N. (2021), "An efficient one-class SVM for anomaly detection in the internet of things", available at: [arXiv:2104.11146](https://arxiv.org/abs/2104.11146)
14. Dunne, M., Gracioli, G., Fischmeister, S. (2018), "A comparison of data streaming frameworks for anomaly detection in embedded systems", *Proceedings of the*, available at: <https://uwaterloo.ca/embedded-software-group/publications/comparison-data-streaming-frameworks-anomaly-detection>
15. Wu, D., Jiang, Z., Xie, X., Wei, X., Yu, W., Li, R. (2020), "LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT", *IEEE Trans.* Vol. 16, No. 8, P. 5244–5253. DOI: <http://dx.doi.org/10.1109/TII.2019.2952917>
16. Fahim, M., Sillitti, A. (2019), "Anomaly detection, analysis and prediction techniques in IoT environment: A systematic literature review", *IEEE Access* No. 7, P. 81664–81681. DOI: <http://dx.doi.org/10.1109/ACCESS.2019.2921912>
17. Galvao, Y., Albuquerque, V., Fernandes, B., Valenka, M. (2017), "Anomaly detection in smart houses: Monitoring elderly daily behavior for fall detecting", *IEEE Latin American Conference on Computational Intelligence, la-CCI*, P. 1–6. DOI: <http://dx.doi.org/10.1109/LA-CCI.2017.8285701>
18. Lu, H., Li, Y., Mu, S., Wang, D., Kim, H., Serikawa, S. (2018), "Motor anomaly detection for unmanned aerial vehicles using reinforcement learning", *IEEE Internet Things J.*, Vol. 5, No. 4, P. 2315–2322. DOI: <http://dx.doi.org/10.1109/JIOT.2017.2737479>
19. Nguyen, T., Marchal S., Miettinen, M., Fereidooni, H., Asokan, N., Sadeghi, A. "Diot: A federated self-learning anomaly detection system for IoT", *39th International Conference on Distributed Computing Systems, ICDCS*, P. 756–767. DOI: <http://dx.doi.org/10.1109/ICDCS.2019.00080>
20. Alsheikh, M., Konieczny, L., Prater, M., Smith, G., Uludag, S. "State and trends of IoT security: Unequivocal appeal to cybercriminals, onerous to defenders", *IEEE Consum. Electr. Mag.* Vol. 1, P. 1–17. DOI: <http://dx.doi.org/10.1109/MCE.2021.3079635>
21. Munir, M., Siddiqui, S., Dengel, A., Ahmed, S. (2019), "DeepAnT: A deep learning approach for unsupervised anomaly detection in time series", *IEEE Access*, Vol. 7, P. 1991–2005. DOI: <http://dx.doi.org/10.1109/ACCESS.2018.2886457>
22. Srikanth, P. (2021), "An efficient approach for clustering and classification for fraud detection using bankruptcy data in IoT environment", *Int. J. Inf. Techno.*, available at: <https://www.x-mol.net/paper/article/1442394146737025024>
23. Hafeez, I., Antikainen, M., Ding, A., Tarkoma, S. (2020), "IoT-KEEPER: Detecting malicious IoT network activity using online traffic analysis at the edge", *IEEE Trans. Netw. Serv. Manag.* Vol. 17, No. 1, P. 45–59. DOI: <http://dx.doi.org/10.1109/TNSM.2020.2966951>

24. Bosman, H., Lacca, G., Tejada, A., Wortche, H., Liotta, A. (2015), "Ensembles of incremental learners to detect anomalies in ad hoc sensor networks", *Ad Hoc Netw.*, No. 35, P. 14–36. DOI: <http://dx.doi.org/10.1016/j.adhoc.2015.07.013>

25. Milov, O., Yevseiev, S., Opriskyu, I., Dunaievska, O., Huk, O., Pogorelov, V., Bondarenko, K., Zviertseva, N., Melenti, Y., Tomashevsky, B. (2022), "Development of concepts for the cyber security metrics classification", *Eastern-European Journal of Enterprise Technologies*. Vol. 4/4, No. 118, P. 6–18, DOI: <https://doi.org/10.15587/1729-4061.2022.263416>

Received 07.12.2023

Відомості про авторів / About the Authors

Бондаренко Кирило Олександрович – Національний технічний університет "Харківський політехнічний інститут", аспірант, Харків, Україна; e-mail: bond.kirill.alexandrovich@gmail.com, ORCID ID: <https://orcid.org/0000-0002-2168-155X>

Bondarenko Kyrylo – National Technical University "Kharkiv polytechnic institute", PhD student, Kharkiv, Ukraine.

ANALYSIS AND SELECTION OF RELEVANT NETWORK ANOMALY DETECTION METRICS

The object of the study is the detection of network anomalies - an important and dynamically developing area of research. The article discusses the main aspects of network anomaly detection. Principles are formulated that allow generalization of various anomaly detection methods. The attacks that network intrusion detection systems typically face are presented, along with the characteristics and types of intrusion detection methods. Network anomalies are considered as manifestations of network attacks, which makes it possible to classify anomalies. The analysis of iterative sources showed that, despite the breadth of coverage of various methods, subject areas and tasks for detecting network anomalies, less attention is paid to the key issue – the analysis of network anomaly metrics and the rationale for choosing the relevant metric in a particular case. The paper presents types, characteristics and examples of network anomalies. To classify and facilitate the detection of network anomalies, metrics are proposed that are based on proximity measures for numeric, categorical, and mixed data types that characterize anomalies. The network anomaly detection problem is presented as a classification or clustering problem. The components that characterize this problem are identified, namely types of input data, acceptability of proximity measures, data labeling, classification of methods based on the use of labeled data, identifying relevant features and reporting anomalies. An approach is described that allows you to timely generate the required set of metrics, which will ensure not only the formation of preventive countermeasures, but also allow you to assess the current state of the security system as a whole. In addition, it provides the possibility of forming multi-circuit security systems, taking into account the influence (integration) of targeted (mixed) attacks on infrastructure elements, as well as the possibility of their synthesis with social engineering methods.

Keywords: network anomaly; intrusion detection system; proximity measure; attack classification.

Бібліографічні описи / Bibliographic descriptions

Бондаренко К. О. Аналіз і вибір релевантної метрики виявлення мережних аномалій. *Сучасний стан наукових досліджень та технологій в промисловості. 2023. № 4 (26). С. 145–157. DOI: <https://doi.org/10.30837/ITSSI.2023.26.145>*

Bondarenko, K. (2023), "Analysis and selection of relevant network anomaly detection metrics", *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (26), P. 145–157. DOI: <https://doi.org/10.30837/ITSSI.2023.26.145>

АЛФАВІТНИЙ ПОКАЖЧИК

Бакурова А. В.	5
Барковська О. Ю.	16
Бондаренко К. О.	145
Вжеснєвський М. О.	135
Гаврашенко А. О.	16
Гавриленко О. В.	24
Гринченко М. А.	33
Гриньова О. Є.	108
Коваленко А. А.	58
Коломійцев О. В.	78
Корнієнко В. Р.	122
Ляшик В. А.	45
Михайленко І. В.	78
Мягкий М. Ю.	24
Невлюдов І. Ш.	135
Порошенко А. І.	58
Рахліс Д. Ю.	122
Роговий М. А.	33
Романків Н. Д.	68
Ромашов Ю. В.	135
Савранська А. В.	5
Ситніков Д. Е.	68, 87
Стрижак О. Е.	95
Терещенко Е. В.	5
Третяк В. Ф.	78
Удовенко С. Г.	108
Федорченко В. М.	78
Філіпенко І. В.	122
Чала Л. Є.	108
Чала О. О.	135
Чернишов Д. В.	87
Шаповалов В. Б.	95
Шевчук М. В.	5
Широкоград Д. В.	5
Шкіль О. С.	122
Шматко О. В.	78
Шовковий Є. І.	108
Шубін І. Ю.	45

ALPHABETICAL INDEX

Bakurova Anna	5
Barkovska Olesia	16
Bondarenko Kyrylo	145
Vzhesnievskyi Maksym	135
Havrashenko Anton	16
Gavrilenko Olena	24
Grinchenko Marina	33
Grinyova Olena	108
Kovalenko Andriy	58
Kolomiitsev Oleksii	78
Korniienko Valentyn	122
Liashyk Volodymyr	45
Mykhailenko Iryna	78
Miahkyi Mykhailo	24
Nevliudov Igor	135
Poroshenko Anton	58
Rakhlis Dariia	122
Rohovyi Mykyta	33
Romankiv Nazarii	68
Romashov Yurii	135
Savranska Alla	5
Sytnikov Dmytro	68, 87
Stryzhak Oleksandr	95
Tereschenko Elina	5
Tretiak Viacheslav	78
Udoenko Serhii	108
Fedorchenko Volodymyr	78
Filippenko Inna	122
Chala Larysa	108
Chala Olena	135
Chernyshov Dmytro	87
Shapovalov Viktor	95
Shevchuk Mark	5
Shyrokograd Dmytro	5
Shkil Alexander	122
Shmatko Oleksandr	78
Shovkovyi Yevhenii	108
Shubin Ihor	45

НАУКОВЕ ВИДАННЯ

SCIENTIFIC PUBLICATION

**СУЧАСНИЙ СТАН
НАУКОВИХ ДОСЛІДЖЕНЬ
ТА ТЕХНОЛОГІЙ
В ПРОМИСЛОВОСТІ**

**INNOVATIVE
TECHNOLOGIES
AND SCIENTIFIC SOLUTIONS
FOR INDUSTRIES**

Щоквартальний науковий журнал

Quarterly scientific journal

№ 4 (26), 2023

№ 4 (26), 2023

Відповідальний секретар журналу *І.Г. Перова*
Відповідальний за випуск *А.А. Коваленко*
Відповідальний за ліцензування *В.В. Косенко*
Редактор *Л.В. Кузьміна*
Комп'ютерна верстка *Л.Ю. Светайло*

Responsible secretary of journal *I. Perova*
Responsible for release *A. Kovalenko*
Responsible for licensing *V. Kosenko*
Editor *L. Kuzmina*
Computer layout *L. Svietailo*

Формат 60×84/8. Умов. друк. арк. 12,5.
Тираж 150 прим.

Format 60×84/8. Conventional printed sheets 12,5.
Edition of 150 copies.

Віддруковано з готових оригінал-макетів
в типографії ФОП Андреев К.В.
Єдиний державний реєстр юридичних осіб
та фізичних осіб-підприємців.
Запис №24800170000045020 від 30.05.2003.

Printed from ready-made original layouts
in the printing house
of Individual Entrepreneur Andreev K.V.
Unified State Register of Legal Entities
and Individual Entrepreneurs.
Entry No. 24800170000045020 of 30.05.2003.

61157, Харків, вул. Акад. Богомольця, 9, кв. 50,
тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com

fl. 50, 9, Acad. Bogomolets Str., Kharkiv, 61157,
тел. +38 (063) 993-62-73
e-mail: ep.zakaz@gmail.com