

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**

№ 3 (25), 2023

No. 3 (25), 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:

Артюх Роман Володимирович, канд. техн. наук,
професор;
Бабенко Віталіна Олексіївна, д-р екон. наук, канд. техн. наук,
професор;
Безкорвайний Володимир Валентинович, д-р техн. наук,
професор;
Гасімов Юсіф, д-р мат. наук, професор (Азербайджан);
Гопсєнко Віктор, д-р техн. наук, професор (Латвія);
Го Цян, д-р техн. наук, професор (КНР);
Джавад Хамісабаді, канд. техн. наук, доцент (Іран);
Зайцева Єлєна, 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 Галіна 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;
Семенєв Валерій Васильович, д-р техн. наук, професор;
Семенов Сергій, д-р техн. наук, професор (Польща);
Сєтлак Галіна, д-р техн. наук, професор (Польща);
Тєрзіян Ваган Якович, д-р техн. наук, професор (Фінляндія);
Тєлєтов Олександр Сергійович, д-р екон. наук, професор;
Тімєфєєв 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);
Ястрємська 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).**

Свідчення про державну реєстрацію журналу Серія KB № 22696-12596P від 04.05.2017 р.

ЗМІСТ

Інформаційні технології

- 5 **Барковська О. Ю., Воропаєва К. А., Руських О. В.**
Обґрунтування вибору нейромережного лінгвістичного класифікатора (eng)
- 15 **Каратаєв О. А., Ситніков Д. Е.**
Метод повторного використання знань у формі логічних рівнянь (ua)
- 27 **Козирєв А. Д., Шубін І. Ю.**
Метод планування завдань оброблення даних у розподілених системах з обмеженою інформацією про доступні ресурси. (ua)
- 40 **Ляшик В. А., Шубін І. Ю.**
Методи автоматизації та інтерпретації якості освітніх тестів у системах розподіленого віртуального навчання (ua)
- 52 **Можасєв О. О., Гнусов Ю. В., Манжсай О. В., Струков В. М., Носов В. В., Радченко В. В., Єнгаличев С. О.**
Стеганографічний метод захисту акустичної інформації у системах критичного застосування (eng)
- 64 **Переяславська С. О., Смагіна О. О.**
Проектування рівня маршрутизації в мікросервісних архітектурах на платформі Spring (ua)
- 79 **Пономаренко О. Є., Горбачов В. О.**
Програмна платформа для оцінювання ефективності агрегації структурної моделі складних систем (ua)
- 88 **Резанов Б. М., Кучук Г. А.**
Модель розподілу елементарних потоків даних у туманній платформі підтримки інтернету речей (ua)
- 98 **Шкіль О. С., Мірошник М. А., Рахліс Д. Ю., Трифанов О. В.**
Структури даних для дедуктивного моделювання умовних операторів HDL (ua)
- 114 **Шпак О. І., Федорка П. П., Пригара М. П.**
Розумні міста та Інтернет речей: вплив розробок у сфері ІТ на розвиток міст і покращення якості життя (ua)

Сучасні технології управління підприємством

- 129 **Довгопол Н. В., Пересада О. В., Прібильнова І. Б.**
Модель прогнозування обсягу ринку розважальних послуг за недетермінованих умов (ua)
- 138 **Іваненко В. А.**
Покращення аспектно-орієнтованого аналізу фінансових настроїв за допомогою контрастного навчання (eng)
- 148 **Ястремська О. М., Малярець Л. М., Ястремська О. О., Бараннік І. О.**
Сучасні визначальні характеристики інноваційного розвитку економіки України: теоретичний та практичний аспекти (eng)

Інженерія та промислові технології

- 163 **Мосьян Д. В.**
Автоматизація технологічного процесу формоутворення деталей із прямолінійними рифтами (ua)
- 174 **Цзяо Ханкунь, Аврунін О. Г.**
Аналіз можливості переміщення імпланту по дуговій траєкторії під безконтактним контролем у магнітній стереотаксичній системі (eng)
- 183 **Алфавітний покажчик**

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

CONTENTS

Information Technology

- 5 **Barkovska O., Voropaieva K., Ruskikh O.**
Justifying the selection of a neural network linguistic classifier (eng)
- 15 **Karataiev O., Sytnikov D.**
The method of reuse of knowledge in the form of logical equations (ua)
- 27 **Kozyriev A., Shubin I.**
Method of planning data processing tasks in distributed systems
with limited information about available resources (ua)
- 40 **Liashyk V., Shubin I.**
Methods of automation and interpretation of the quality educational tests
in distributed virtual learning systems (ua)
- 52 **Mozhaiev O., Gnusov Y., Manzhai O., Strukov V., Nosov V., Radchenko V., Yenhalychev S.**
Steganographic method of acoustic information protection in critical applications systems (eng)
- 64 **Pereiaslavskaya S., Smahina O.**
Designing the routing level in microservice architectures on the Spring platform" (ua)
- 79 **Ponomarenko O., Gorbachov V.**
The software platform for evaluation of the effectiveness
of structural model aggregation of complex systems (ua)
- 88 **Rezanov B., Kuchuk H.**
Model of elemental data flow distribution in the internet of things supporting fog platform (ua)
- 98 **Shkil A., Miroshnyk M., Rakhlis D., Trifanov O.**
Data structures for deductive simulation of HDL conditional operators (ua)
- 114 **Shpak O., Fedorka P., Prygara M.**
Smart cities and the internet of things: the impact of it developments
on the development of cities and improving the quality of life (ua)

Modern Enterprise Management Technologies

- 129 **Dovgopol N., Peresada O., Pribylnova I.**
A model for forecasting the volume of the entertainment services market
under non-determinative conditions (ua)
- 138 **Ivanenko V.**
Enhancing aspect-based financial sentiment analysis through contrastive learning (eng)
- 148 **Iastremska O., Malyarets L., Iastremska O., Barannik I.**
Modern defining characteristics of innovative development of the ukrainian economy:
theoretical and practical aspects (eng)

Engineering & Industry Technologies

- 163 **Mospan D.**
Automation of the technological process of forming details with straight rifts (ua)
- 174 **Hunkun J., Avrunin O.**
Feasibility analysis of implant movement along arc trajectory
under non-contact control in magnetic stereotaxic system (eng)
- 183 **Alphabetical index**

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

O. BARKOVSKA, K. VOROPAIEVA, O. RUSKIKH

JUSTIFYING THE SELECTION OF A NEURAL NETWORK LINGUISTIC CLASSIFIER

The **subject matter** of this article revolves around the exploration of neural network architectures to enhance the accuracy of text classification, particularly within the realm of natural language processing. The significance of text classification has grown notably in recent years due to its pivotal role in various applications like sentiment analysis, content filtering, and information categorization. Given the escalating demand for precision and efficiency in text classification methods, the evaluation and comparison of diverse neural network models become imperative to determine optimal strategies. The **goal** of this study is to address the challenges and opportunities inherent in text classification while shedding light on the comparative performance of two well-established neural network architectures: Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). To achieve the goal, the following **tasks** were solved: a comprehensive analysis of these neural network models was performed, considering several key aspects. These aspects include classification accuracy, training and prediction time, model size, data distribution, and overall ease of use. By systematically assessing these attributes, this study aims to provide valuable information about the strengths and weaknesses of each model and enable researchers and practitioners to make informed decisions when selecting a neural network classifier for text classification tasks. The following **methods** used are a comprehensive analysis of neural network models, assessment of classification accuracy, training and prediction time, model size, and data distribution. The following **results** were obtained: The LSTM model demonstrated superior classification accuracy across all three training sample sizes when compared to CNN. This highlights LSTM's ability to effectively adapt to diverse data types and consistently maintain high accuracy, even with substantial data volumes. Furthermore, the study revealed that computing power significantly influences model performance, emphasizing the need to consider available resources when selecting a model. **Conclusions.** Based on the study's findings, the Long Short-Term Memory (LSTM) model emerged as the preferred choice for text data classification. Its adeptness in handling sequential data, recognizing long-term dependencies, and consistently delivering high accuracy positions it as a robust solution for text analysis across various domains. The decision is supported by the model's swift training and prediction speed and its compact size, making it a suitable candidate for practical implementation.

Keywords: text classification; neural networks; LSTM; CNN; classification accuracy; model comparison; sequential data.

Introduction

Text classification is an important task in today's information society, as it allows you to automatically process and classify large amounts of textual information. This is especially true in the digital revolution, when the amount of text data is constantly growing, and with it the need for efficient methods of analysing and disseminating information. Text classification is widely used in many areas of life, such as medicine, finance, marketing, social media, Internet search, and many others, as shown in Figure 1. For example, text classification in medicine can help to automatically determine diagnoses based on patient symptoms or filter out malicious content on social media [1].

Text classification plays an important role in the field of natural language processing (NLP) and is central to other NLP tasks. Text classification helps to determine whether texts belong to certain categories or topics or solve other problems related to the distribution of textual information [2].

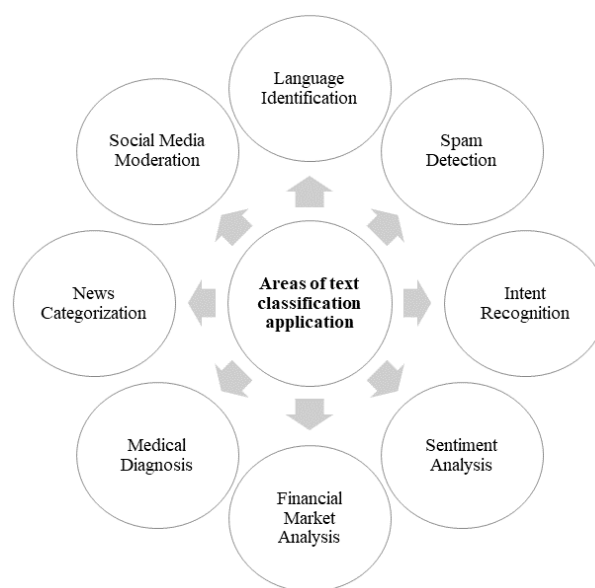


Fig. 1. Areas of text classification task application

For text classification, various tools are used to achieve a high level of accuracy in this task,

often combining existing methods that have proven themselves [3]. First of all, rules and heuristic methods based on predefined rules and expert knowledge are used. These methods are especially effective when the data structure is simple and the relationships between categories are already known. Machine learning methods are also used for classification. With machine learning, you can create models that automatically recognize patterns in text data and perform classification based on a set of training data. In this field, methods such as Naive Bayes, Support Vector Machine (SVM) [4], Decision Trees, and others are widely used.

Neural networks are a special kind of machine learning methods inspired by the structure of the brain's neural network. Neural network models show impressive results and advantages over traditional methods shown in Figure 2, such as Naive Bayes, Support Vector Machine (SVM), Decision Trees, and others, as they allow to automatically detect internal patterns in text data and classify with high accuracy, as shown in [5].

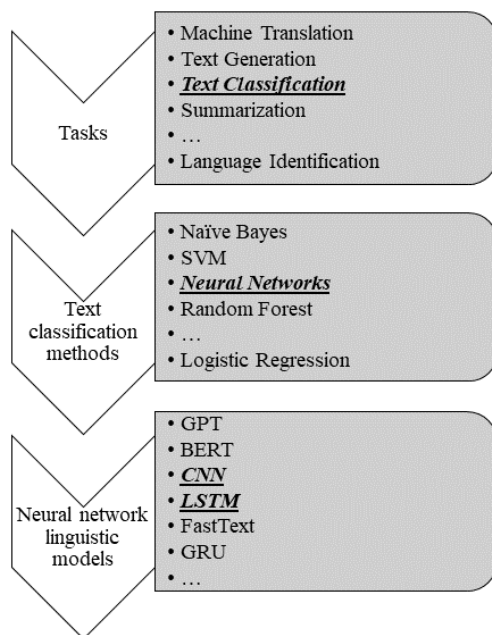


Fig. 2. Generalized analysis of the problem area

When solving text classification tasks, it is important to choose a neural network model that best suits the specifics of the task and the needs of the research. Choosing the best model requires a careful comparison of different architectures and their characteristics to ensure optimal performance and accuracy. To select the best model, a comparative analysis of various model characteristics is necessary.

For this analysis, we have selected the following key characteristics: classification accuracy, learning rate, prediction rate, model size, data dissemination, and overall ease of use. In the following, we report on the comparison of these models based on these characteristics to find out their effectiveness and suitability for use in different text classification scenarios.

Analysis of last achievements and publications

Work [6] shows that an important role in improving classification efficiency is played by the use of Word Embedding, which allows words to be converted into vectors of numbers with small sizes while preserving the semantic connections between them. This enables neural networks to work with both text and numeric data, allowing for higher classification accuracy and reduced computing costs. In addition, in recent years, new and more powerful methods have emerged, such as Contextual Embedding, which allow for more accurate capture of the semantic context of words and sentences. As an extension of Word Embedding, Contextual Embedding takes into account the context of each word in the text and thus allows for a deeper understanding of textual information using neural networks. These trends in the use of Word Embedding and Contextual Embedding in neural network models have become an important means of improving text classification results (Table 1).

In addition, some studies emphasize the importance of achieving state-of-the-art results in text classification tasks. A high level of accuracy in text classification can be achieved by using pre-training and fine-tuning models. Such approaches are becoming more and more relevant, which increases the variety of applications and the development of text classification for various tasks.

In our study, we will focus on analyzing and comparing different approaches using neural networks for text classification, in particular, we will compare the performance of models using traditional Word Embedding [7] and modern Contextual Embedding [8–9]. Such an analysis will allow us to better understand current trends in this field and identify the most promising areas of research in text classification using neural networks.

Comparison of neural network models for text classification is a critical task that allows us to determine the most efficient and accurate approaches to specific tasks. It is especially interesting to compare models based on Contextual Embedding and static Word Embedding.

Such a comparison will show the pros and cons of different approaches to text vectorization and emphasize the importance of taking context into account when analyzing text data. A comparative study of neural network models for text classification has several advantages. It allows us to determine which models are more effective for different types of text data, which can improve the quality of classification and the accuracy of the results. Knowing the advantages and limitations of different models also helps to choose the best classifier for a particular text classification task.

However, performing a neural network comparison can cause certain problems. For example, it may require significant computing resources and time, as such models usually have a large number of parameters. It is also important to choose the right metrics to evaluate the results and avoid training the models repeatedly on the same dataset. Taking all these factors into account, studying the comparison of neural network models for text classification remains an interesting and important task that contributes to the development and improvement of this field.

Table 1. Ways to improve text classification accuracy

Strategy	Description
Contextual vectors	Using contextual word vectors such as BERT, GPT, or ELMO to capture context
Model selection	Selecting a suitable classification model, such as Naive Bayes, SVM, LSTM, CNN, etc.
Tuning of hyperparameters	Adjusting hyperparameters to optimize the performance of the selected model
Data augmentation	Generating synthetic data to expand the training set and improve model generalization
Ensemble approach	Combining forecasts from several models to improve overall accuracy
Transfer of learning	Use of pre-trained models and their fine-tuning for a specific classification task
Cross-validation	Evaluating model performance using methods that take into account multiple trials for accuracy
Text preprocessing	Text cleaning and normalization, redundant word removal, stemming, and special character processing

One of the first known successful applications of neural networks for text classification was based on the Convolutional Neural Network (CNN) architecture. The CNN model [10-11] has been successfully applied to image analysis, but has also proven to be effective in processing text data. Using CNNs and Word Embedding, which transform words into numerical vectors [12-14], impressive results have been achieved in text classification tasks, in particular in determining the tone of the text. However, recent trends indicate that there are more powerful and flexible approaches to text classification, including Contextual Embedding. Models such as Transformer, BERT, or GPT are very popular in the world of applied machine learning because they can analyze texts in context, taking into account the semantic relationships between words and sentences.

As a first step, we empirically compare the performance of contextual embeddings with classical embeddings such as word2vec [15] and GloVe [16]. Contextual Embedding is a type of Word Embedding in which the vector values take into account the context in which the words appear in the text. This gives you a more accurate representation of the word depending on the context. Context-aware neural network models, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained

Transformer), can achieve high results in various classification tasks. On the other hand, static Word Embeddings [2], such as Word2Vec and GloVe, assign a fixed vector to each word that is independent of context. These models work well for many tasks, but do not take into account semantic dependencies between words in a sentence, which can lead to less accurate text classification results.

For this reason, we analyze the impact of using Contextual Embedding on text classification accuracy. Using this analysis, we can evaluate how effective Contextual Embeddings are compared to traditional static Word Embeddings in various classification scenarios. They check whether contextual vector representations provide better results and more accurate classification, especially in complex and diverse text data analysis scenarios. This approach allows us to understand which word vectorization methods are more effective for different types of textual information and helps to draw conclusions about the advantages and limitations of each approach in text classification tasks.

The purpose of this study is to conduct a comparative analysis of two important neural network architectures – Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) – to provide recommendations for the selection of a neural network

linguistic classifier. The study includes an analysis of various aspects of both models, including classification accuracy, training and prediction time, and model size, taking into account three different training sample sizes. For applications in areas such as news and content classification, reviews and testimonials classification, social media analysis, and others.

To achieve this goal, the following key tasks must be solved, which are discussed in detail in this study:

- analysis of the problem area and justification of the relevance of the topic;
- comparative analysis of existing text corpora;
- formulation of criteria and requirements for neural network classifiers;
- performing experimental studies on samples of different sizes to compare the training and prediction time, prediction accuracy, and model size when using LSTM and CNN neural network models as a classifier;

- analysis of the obtained results;
- justification for choosing the most effective model.

Materials and methods

There are many different corpora for training text classification models that have become an important resource for research and applications in the field of natural language processing [17, 18]. These corpora represent different types of text data from different sources and cover a wide range of topics. To ensure a successful comparison of neural network models for text classification, it is important to carefully select an appropriate corpus that meets the research goals and task characteristics. The most popular cases include the ones listed in Table 2: IMDB corpus, Reuters news corpus, PubMed scientific article corpus, Twitter sentiment analysis corpus.

Table 2. Text bodies for text classification in NLP tasks

№	Corpus	Description	Classification task
1	IMDB corpus	Contains movie reviews from IMDB, divided into positive and negative categories	Binary text classification (positive/negative reviews)
2	Reuters news corpus	Contains news articles organized by topic	Categorical classification of texts
3	PubMed scientific articles corpus	Contains scientific articles from medical sources categorized by topic or importance	Categorical text classification
4	Twitter sentiment analysis corpus	Contains Twitter messages categorized by sentiment (positive, negative, or neutral)	Sentiment classification in short texts

Given the complexity of the text classification task and the desire to perform a qualitative comparison of neural network models, it was important to choose a suitable corpus for training the models. The most convenient and suitable corpus for our research purposes was the IMDB corpus.

The reasons for choosing the IMDB corpus are its diversity, accessibility, and representativeness. Due to the large amount of data, IMDB can provide a sufficient number of examples for training and testing models, which is important for reliable comparison of their performance. In addition, IMDB contains textual reviews with emotional coloring, which is an important feature for solving the binary classification task. Another important advantage of IMDB is its accessibility, which allows researchers from all over the world to use this corpus for their studies. This contributes to the wide applicability of the results and the possibility of comparison with other scientific studies. In addition, the variety of text lengths in the IMDB corpus allows testing the ability of models to work with sequences of varying complexity and length,

which is crucial for realistic analysis of neural network performance on a variety of input data.

The results of the comparison of neural network models for text classification find application in many promising areas and tasks, such as:

1. Sentiment analysis of product and service reviews – this will help companies understand customer satisfaction, identify problem areas, and improve their products.
2. Content classification in web services – this will help organizations and platforms automatically filter content to ensure safety and a positive user experience.
3. Emotion analysis in social media – this will help to understand the reaction to news, events and publications, which is important for advertisers and marketers.
4. Monitoring brands and companies to track and analyze public opinion, helping managers respond to changes in the perception of goods and services.

Analyzing the accuracy and effectiveness of different approaches is an important step to better

understand the potential advantages and limitations of contextual embedding compared to traditional methods. This data can serve as an important guide in choosing the optimal model for specific text classification tasks. The goal of our research is to improve the quality and accuracy of classification results and to contribute to the development and implementation of new methods in this field.

Choosing the right neural network architecture is an important step in solving the text classification problem. The appropriate architecture can affect the efficiency and accuracy of the model in the text classification task from different perspectives.

The architecture of a neural network determines how it solves the problem of analyzing text data. Different architectures may have different approaches to pattern recognition, word dependency detection, and interpretation of textual information. The right architecture can help solve specific problems in the text classification task and provide more accurate and reliable results. In this study, we chose two popular models for comparative analysis: Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), because these networks represent two different approaches to analyzing text data and have their own features that can be useful for different types of texts and classification tasks (Table 3).

Table 3. Comparative analysis of LSTM and CNN models for text classification

Characteristics	LSTM Model	CNN Model
Basic architecture	Recurrent neural network with LSTM layer	Convolutional neural network with Conv1D layers
Typical tasks	Analysis of sequential data, text data	Analysis of sequential data, images
Features	Takes into account the context of word dependencies	Detects local patterns in data
Learning algorithms	Backward error propagation, Adam optimizer	Back propagation of error, Adam optimizer
Activation functions	Tanh, Sigmoid	ReLU
Memory usage	Uses short-term and long-term memory	Does not use memory
Application	Sequential text analysis, language translation	Image, video analysis
Implementation libraries	TensorFlow, Keras	TensorFlow, Keras

The CNN model is unique and specialized for recognizing patterns in images, but it can also be successfully used to process text data, where it recognizes local dependencies and important features of the textual context. On the other hand, the LSTM model is a part of recurrent neural networks and has the ability to store and use information from previous steps, which allows it to work efficiently with sequential data, especially text sequences, and take into account the context in texts.

In order to make the right choice of a neural network linguistic model for classifying input texts in future studies, this section presents the results of a comparative analysis of the selected models according to the following criteria: classification accuracy, training speed, prediction speed, model size, data distribution, and overall usability.

Accuracy is an important indicator for determining the performance of classification models. This characteristic measures the proportion of correct predictions made by the model out of the total number of predictions. In our experiment, accuracy allows us to understand how accurately the selected models identify positive and negative reviews.

Learning and prediction times are important aspects when implementing neural models. The learning rate

indicates the time it takes for the model to adapt to the data during training. It can affect the overall training time of the model. Prediction speed indicates the time it takes for the model to predict new input data. In our study, we measure these parameters for each architecture and sample size to understand which model can be more efficient in terms of computational complexity.

Model size reflects the number of parameters used to hold information in the model. In our experiment, we can use this metric to understand which architecture has more parameters to store information. A large model size can affect memory and computation requirements.

By comparing two models based on data distribution characteristics, we evaluate how well each model adapts to different types of data. Possible changes in the distribution can affect the training results, as certain models may be more sensitive to changes in the data distribution. In our experiment, this metric is used to understand how well each model is able to generalize the knowledge gained during training to new and unknown data. It is important to keep in mind that real-world data that a model encounters in solving practical problems may contain variations and diversity. Thus, the importance of adapting to different data

distributions emphasizes the need to choose a model that demonstrates stable and reliable performance even under variable input conditions.

The importance of evaluation parameters and metrics lies in their ability to provide an objective and complete assessment of model performance. The accuracy rating helps to understand which of the architectures is better at solving the classification task. Measuring training and prediction time gives us an idea of how fast the models work in real time. Model size is important in practical applications where computing resources may be limited.

In addition, the experiment investigated the effect of the training set size on the time to obtain the result, namely

- **Small Training Set.** In this case, lower-level computing resources were used, which are characterized by limited capacity and processing of a limited amount of data;
- **Medium Training Set.** For this category, medium computing resources were used, which allowed working with a larger amount of data and provided higher computing power;
- **Large Training Set.** To ensure the efficiency of calculations, powerful computing resources were used in this case, which allowed us to process large amounts of data quickly and efficiently.

By analyzing the experimental results and considering the impact of the computing base on the performance of models for different training set sizes, we can understand the importance of the role of computing resources in determining the efficiency of neural networks under different operating conditions.

The results allow us to better understand how computing resources affect key model characteristics, such as training time, prediction time, and classification accuracy. This analysis helps to select the most efficient models for specific work scenarios. For example, for tasks with limited computational resources, it may be important to favor models that perform best under constrained conditions. On the other hand, if there is a lot of computing power available, it may make sense to use more sophisticated models with higher accuracy and the ability to adapt to different types of data.

In the Results section, we present detailed results obtained during the experiments with LSTM and CNN text classification models. The experiments were conducted with three different sizes of training samples: small (1000 samples), medium (10000 samples), and large (50.000 samples). Table 4 shows the results of the comparative analysis of two models – Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) – for different training set sizes.

Research results

All models presented in Table 4 were trained on the same hardware computing platform with the same characteristics. This ensured an adequate comparison of model performance, as the possibility of hardware differences affecting the results was excluded. In addition, the same input data was used for all experiments. The dataset was divided into training and test sets, ensuring an even distribution of classes in each set. This helped to avoid bias in the results due to class imbalance.

Table 4. Results of comparative analysis of LSTM and CNN models on different sizes of training samples

Model/Sample size	Training time (sec)	Prediction time per sample (sec)	Training accuracy	Accuracy on test data	Model size (parameters)
LSTM Small Training Set	144.22	0.00173	0.8860	0.7288	689.473
CNN Small Training Set	143.05	0.00140	0.7500	0.5002	804.225
LSTM Medium Training Set	272.19	0.00171	0.9346	0.8377	689.473
CNN Medium Training Set	188.91	0.00139	0.9877	0.8422	804.225
LSTM Big Training Set	483.21	0.00173	0.9489	0.8705	689.473
CNN Big Training Set	443.11	0.00099	0.9848	0.8652	804.225

The data preprocessing methods shown in Figure 3 were also identical for both models. This allowed us to create a common initial context for all models and compare their performance under the same conditions. This approach to conducting experiments helped to avoid biases when comparing different models and

ensured objectivity in choosing the most effective neural network classifier.

The table shows the metrics such as training time, inference time on a single sample, training accuracy, testing accuracy, and model size for each of the training sample sizes. These results help us

to better understand how different models respond to different conditions and data sizes, which may indicate their effectiveness and suitability for specific text classification tasks.

To better understand the results of the study and comparative analysis of LSTM and CNN models in the context of text classification, we present Figure 4, which illustrates the key indicators of each model.

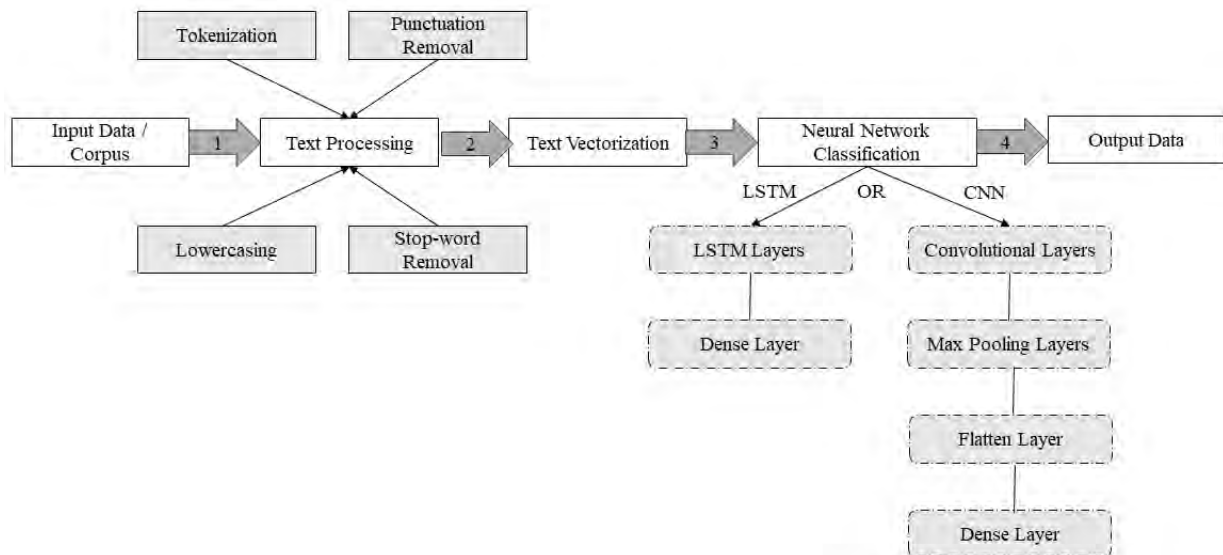


Fig. 3. Generalized requirements for the experiment

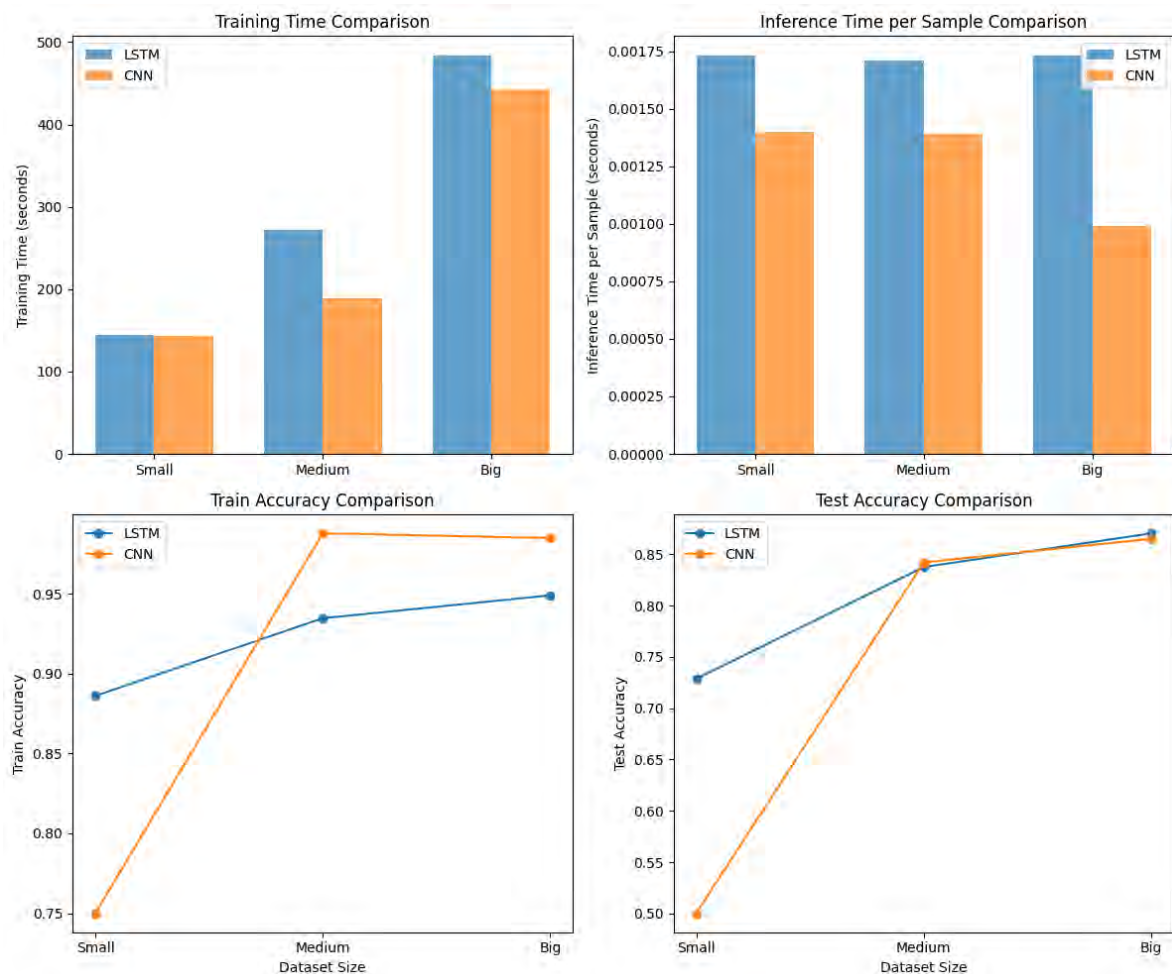


Fig. 4. Comparison of LSTM and CNN models on different sizes of training samples

The graphs allow you to quickly evaluate the performance of each model based on various aspects such as training time, prediction time, and classification accuracy. The graphs show results for three training sample sizes: small, medium, and large. Each graph shows the training time and classification accuracy for LSTM and CNN models on their respective training sets. This approach allows you to compare the performance of both models depending on the size of the training set and the computing base.

The analysis of experimental results provides an understanding of the performance and properties of LSTM and CNN models in the context of text classification. It is important to note that the LSTM model demonstrates competitive accuracy at different training set sizes, which indicates its ability to analyze long-term dependencies in text data. On the other hand, the CNN model shows improved performance as the training set size increases, indicating its effectiveness in recognizing local features and patterns. In addition, different training and prediction times provide insight into the computational efficiency of each model. In general, the CNN model shows faster training times, especially with large training samples. On the other hand, the LSTM model shows stable performance with different training sample sizes. In terms of model size, both LSTM and CNN have the same number of parameters regardless of the training sample size. This aspect emphasizes their scalability for different amounts of data.

Conclusions

Given the results of our study, we chose the Long Short-Term Memory (LSTM) model for further experiments in text classification. This decision is based on several important factors that confirm the advantages of LSTM in this context:

1. LSTM is a recurrent neural network specially designed to work with sequential data such as text. This allows it to recognize complex relationships and dependencies between words in a text, which is crucial for accurate classification.

2. One of the key advantages of LSTM is its ability to identify long-term dependencies in sequential

data. This can be an important factor for text analysis, where the relationships between words can be very scattered.

3. Our experimental results show that LSTM has high classification accuracy for different training sample sizes. This demonstrates its ability to effectively learn and generalize patterns in the data.

4. The LSTM showed a stable size regardless of the size of the training samples. This can be an important aspect for further research, as it avoids unnecessary computational effort when expanding the dataset.

All these justifications confirm that LSTM is an excellent choice for solving the problem of text classification. Its ability to work with sequential data, detect long-term dependencies, and demonstrate high accuracy makes it an important tool for analyzing text data in various fields.

Taking into account the results of the experiments and the analysis of the data obtained, we can conclude that the best model for text data classification can be chosen. Our study included a comparative analysis of two popular neural network architectures – LSTM (Long Short-Term Memory) and CNN (Convolutional Neural Network) – to determine their effectiveness in solving the text classification task. Comparing the models by various characteristics, we obtained convincing results that demonstrate their advantages and disadvantages. Analyzing the classification accuracy, we found that the LSTM model performs better than CNN for all three training sample sizes. It is able to effectively adapt to different types of data and demonstrates consistently high accuracy even with large amounts of data. In addition, analyzing the impact of the computing base on the results, we found that computing power can significantly affect the performance of models. The choice of model should be justified not only by the results but also by the available resources.

So, taking into account all the data and analysis results, we can conclude that the LSTM model is more suitable for our text data classification task. Its high accuracy, training and prediction speed, and compact size make it the best choice for further research and implementation in practical applications.

References

1. Serdechnyi, V., Barkovska, O., Rosinskiy, D., Axak, N., & Korablyov, M. (2020), "Model of the internet traffic filtering system to ensure safe web surfing". *In Lecture Notes in Computational Intelligence and Decision Making: Proceedings*

of the XV International Scientific Conference "Intellectual Systems of Decision Making and Problems of Computational Intelligence" (ISDMCI'2019), Ukraine, May 21–25, 2019, 15 P. 133–147. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-030-26474-1_10

2. Barkovska, O., Pyvovarova, D., Kholiev, V., Ivashchenko, H., & Rosinskiy, D. "Information Object Storage Model with Accelerated Text Processing Methods". In COLINS, 2021, P. 286–299, available at: <https://csitjournal.khmn.edu.ua/index.php/csit/article/download/182/112/559>

3. Zhang, J. et al. (2018), "LSTM-CNN hybrid model for text classification", *IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. – IEEE. P. 1675–1680. DOI: <https://doi.org/10.1109/IAEAC.2018.8577620>

4. WANG, Haitao, HE, Jie, ZHANG, Xiaohong, et al., (2020), "A Short Text Classification Method Based on N-Gram and CNN", *Chinese Journal of Electronics*, Vol. 29, No. 2, P. 248–254. DOI: <https://doi.org/10.1049/cje.2020.01.001>

5. Mohammad, A.H., Alwada'n, T., Al-Momani, O., (2019), "Arabic Text Categorization Using Support Vector Machine.", *Naïve Bayes and Neural Network*, P. 930–933. DOI: https://doi.org/10.5176/2251-3043_4.4.360

6. Wang, Congcong & Nulty, Paul & Lillis, David. (2020), "A Comparative Study on Word Embeddings in Deep Learning for Text Classification". *Conference on Natural Language Processing and Information Retrieval*. P. 37–46. DOI: <https://doi.org/10.1145/3443279.3443304>

7. Selva, Birunda S., Kanniga, Devi R. (2021), "A review on word embedding techniques for text classification", *Innovative Data Communication Technologies and Application: Proceedings of ICIDCA*. P. 267–281. DOI: https://doi.org/10.1007/978-981-15-9651-3_23

8. Mars, M. (2022), "From word embeddings to pre-trained language models: A state-of-the-art walkthrough", *Applied Sciences*. – Vol. 12. № 17. P. 8805. DOI: <https://doi.org/10.3390/app12178805>

9. Patil, Rajvardhan & Boit, Sorio & Gudivada, Venkat & Nandigam, Jagadeesh. (2023), "A Survey of Text Representation and Embedding Techniques in NLP". *IEEE Access*. P 1–10. DOI: <https://doi.org/10.1109/ACCESS.2023.3266377>

10. Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012), "ImageNet Classification with Deep Convolutional Neural Networks". *Neural Information Processing Systems*. 25 p. DOI: <https://doi.org/10.1145/3065386>

11. Wang, S. et al. (2018), "Densely connected CNN with multi-scale feature attention for text classification" *IJCAI*. Vol. 18. P. 4468–4474. DOI: <https://doi.org/10.24963/ijcai.2018/621>

12. Jang B. et al. (2020), "Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism". *Applied Sciences*. Vol. 10. No 17. 5841 p. DOI: <https://doi.org/10.3390/app10175841>

13. Gao, M., Li T., Huang P. (2019), "Text classification research based on improved Word2vec and CNN". *Service-Oriented Computing–ICSOC 2018 Workshops: ADMS, ASOCA, ISYyCC, CloTS, DDBS, and NLS4IoT*, Hangzhou, China, November 12–15, Revised Selected Papers 16. Springer International Publishing. P. 126–135. DOI: https://doi.org/10.1007/978-3-030-17642-6_11

14. Zhou, H. (2022), "Research of text classification based on TF-IDF and CNN-LSTM". *Journal of Physics: Conference Series*. – IOP Publishing, Vol. 2171. No 1. 2021 p. DOI: <https://doi.org/10.1088/1742-6596/2171/1/012021>

15. Mikolov, Tomas & Chen, Kai & Corrado, G.s & Dean, Jeffrey "Efficient Estimation of Word Representations in Vector Space", *Proceedings of Workshop at ICLR*. 2013. available at: https://www.researchgate.net/publication/234131319_Efficient_Estimation_of_Word_Representations_in_Vector_Space

16. Jeffrey, Pennington, Richard, Socher, and Christopher, Manning (2014), "GloVe: Global Vectors for Word Representation". In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar. Association for Computational Linguistics. P. 1532–1543. DOI: <https://doi.org/10.3115/v1/D14-1162>

17. Panchenko, D. et al. (2021), "Ukrainian news corpus as text classification benchmark". *International Conference on Information and Communication Technologies in Education, Research, and Industrial Applications*. Cham: Springer International Publishing, P. 550–559. DOI: <https://doi.org/10.1007/978-3-319-76168-8>

18. Schwenk, H., & Li, X. "A Corpus for Multilingual Document Classification in Eight Languages". 2018. available at: <https://arxiv.org/abs/1805.09821>

Received 14.09.2023

Відомості про авторів / About the Authors

Барковська Олесь Юрійвна – 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>

Воропаєва Ксенія Андріївна – Харківський національний університет радіоелектроніки, магістрантка кафедри електронних обчислювальних машин, Харків, Україна; e-mail: ksenii.voropaieva@nure.ua; ORCID ID: <https://orcid.org/0009-0003-1270-4857>

Руських Олександр Валерійович – Харківський національний університет радіоелектроніки, магістрант кафедри електронних обчислювальних машин, Харків, Україна; e-mail: oleksandr.ruskikh@nure.ua; ORCID ID: <https://orcid.org/0000-0001-9931-4226>

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

Voropaieva Kseniia – Kharkiv National University of Radio Electronics, student of Department of Electronic Computers, Kharkiv, Ukraine.

Ruskikh Oleksandr – Kharkiv National University of Radio Electronics, student of Department of Electronic Computers, Kharkiv, Ukraine.

ОБҐРУНТУВАННЯ ВИБОРУ НЕЙРОМЕРЕЖНОГО ЛІНГВІСТИЧНОГО КЛАСИФІКАТОРА

Предметом статті є дослідження архітектури нейронних мереж для підвищення точності класифікації тексту, зокрема у сфері оброблення природної мови. Значення класифікації тексту помітно зросло в останні роки, що пов'язано з її ключовою роллю в різних програмах, зокрема аналіз налаштувань, фільтрація вмісту та категоризація інформації. З огляду на зростання попиту на точність та ефективність методів класифікації тексту, оцінювання та порівняння різноманітних моделей нейронних мереж стають обов'язковими для визначення оптимальних стратегій. **Метою дослідження** є порівняльний аналіз двох важливих архітектур нейронних мереж – довгострокової короткочасної пам'яті (*LSTM*) та згорткової нейронної мережі (*CNN*) – для формування рекомендацій щодо вибору нейромережного лінгвістичного класифікатора. Для досягнення мети були розв'язані такі **завдання**: проаналізовано проблемні сфери, зокрема обґрунтування актуальності теми, порівняння наявних текстових корпусів; сформовано критерії та вимоги до роботи нейромережних класифікаторів; проведено дослідження на вибірках різних розмірів з метою порівняння часу навчання та передбачення, точності передбачення в процесі використання нейромережних моделей *LSTM* і *CNN* як класифікатора; проаналізовано здобуті результати; обґрунтовано вибір найефективнішої моделі. Оцінювання таких параметрів, як точність класифікації, час навчання та прогнозування, розмір моделі, розподіл інформації та простота використання, надає обґрунтовані показники про переваги й недоліки кожної моделі та дає змогу дослідникам і практикам приймати рішення щодо вибору нейромережного лінгвістичного класифікатора. Застосовано такі **методи**: попереднє оброблення текстової інформації (методи токенизації, вилучення стоп-слів), векторизація тексту, нейромережні методи аналізу вхідних даних, статистичні методи. Здобуто конкретні **результати**. Модель *LSTM* продемонструвала вищу точність класифікації для всіх трьох розмірів навчальної вибірки порівняно з *CNN*. Це доводить здатність *LSTM* ефективно адаптуватися до різних типів інформації та стабільно підтримувати високу точність, навіть із великими обсягами даних. Крім того, дослідження показало, що обчислювальна потужність значно впливає на продуктивність моделі та необхідно враховувати доступні ресурси у виборі моделі. **Висновки**. З огляду на результати дослідження *LSTM*-модель визнана кращим вибором для класифікації текстової інформації. Вправність зазначеної моделі в обробленні послідовних даних, розпізнаванні довгострокових залежностей і стабільному забезпеченні високої точності позиціонує її як надійне рішення для аналізу тексту в різних сферах. Швидке навчання, ефективність прогнозування та компактний розмір моделі роблять її придатною для практичного застосування.

Ключові слова: класифікація тексту; нейронні мережі; *LSTM*; *CNN*; точність класифікації; порівняння моделей; послідовні дані.

Бібліографічні описи / Bibliographic descriptions

Барковська О. Ю., Воропаєва К. А., Руських О. В. Обґрунтування вибору нейромережного лінгвістичного класифікатора. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 5–14. DOI: <https://doi.org/10.30837/ITSSI.2023.25.005>

Barkovska, O., Voropaieva, K., Ruskikh, O. (2023), "Justifying the selection of a neural network linguistic classifier", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 5–14. DOI: <https://doi.org/10.30837/ITSSI.2023.25.005>

О. КАРАТАЄВ, Д. СИТНИКОВ

МЕТОД ПОВТОРНОГО ВИКОРИСТАННЯ ЗНАТЬ У ФОРМІ ЛОГІЧНИХ РІВНЯНЬ

Предмет дослідження – процеси побудови та поповнення баз знань, повторного використання знань, і створення програмних систем на основі баз знань, інтерпретація знань як один із підходів до повторного їх застосування, що полягає у виведенні нових знань на основі наявних фактів у базі знань. **Мета дослідження** – розроблення методу повторного використання знань за допомогою вирішення логічних рівнянь скінченних предикатів для певної предметної галузі. Для досягнення поставленої мети визначено такі **завдання**: дослідити описативні підходи до логічного моделювання предметної галузі, що дають змогу повторного використання знань, заданих системою логічних рівнянь у межах алгебри скінченних предикатів; розробити метод поповнення бази знань у формі предикатних рівнянь за допомогою вилучення змінних із логічних систем, де рівняння є складними логічними зв'язками між дискретними ознаками об'єктів або процесів. Використано такі **методи**: алгебра скінченних предикатів, кванторні операції з предикатами для інтерпретації знань. Здобуто такі **результати**: проаналізовано описативні методи повторного використання знань; запропоновано метод поповнення бази знань у формі логічних рівнянь з метою спрощення подальшого використання неявних зв'язків між дискретними ознаками; розглянуто приклад поповнення бази знань медичної галузі, що дає змогу повторного використання знань, заданих неявно. **Висновки**. Запропонований метод дає змогу поповнювати базу знань у формі логічних рівнянь способом додавання предикатних рівнянь, що пов'язують окремі набори дискретних змінних, які цікавлять ученого або практика. Метод поповнення бази знань оснований на додаванні нових спрощених рівнянь; спрощені предикатні рівняння моделюють логічні закономірності, що неявно містяться в основній базі знань; проведено експериментальне дослідження.

Ключові слова: програмна інженерія; бази знань; повторне використання знань; алгебра скінченних предикатів; логічні рівняння.

Вступ

Повторне використання знань є актуальним питанням, що стосується процесу побудови та поповнення баз знань на основі аналізу та перетворення наявних знань. Однак повторного використання знань не відбувається або принаймні бажані результати не досягаються. Зазвичай припускають, що як тільки знання будуть належним чином збережені, їх повторне застосування відбудеться автоматично. Отже, питання, на яке потрібно відповісти, полягає в тому, як можна створювати бази знань, щоб їх надалі ефективно використовувати. Для цього необхідно проаналізувати системи повторного застосування знань, запропонувати підхід щодо створення бази знань та оцінити запропоноване рішення.

Проблема повторного використання знань щодо предметної галузі, де багато дискретних ознак пов'язані складними логічними залежностями, які важко розглянути в явній формі, полягає в тому, щоб використати неявні зв'язки між дискретними предикатними змінними для подання цільових логічних зв'язків у явній формі з метою спрощення наявних логічних залежностей. Це має дати змогу

багато разів використовувати такі спрощені залежності без необхідності аналізувати складну систему з несуттєвими дискретними змінними.

Мета дослідження – розроблення методу повторного використання знань за допомогою вирішення логічних рівнянь для певної предметної галузі. Метод поповнення бази знань має бути оснований на додаванні нових спрощених рівнянь, що неявно містяться в основній базі знань; дослідити процес спрощення повторного використання знань завдяки додаванню більш простих рівнянь до бази знань. Для досягнення поставленої мети мають бути виконані такі **завдання**: дослідити описативні підходи до логічного моделювання предметної галузі, що дають змогу повторного використання знань, заданих системою логічних рівнянь у межах алгебри скінченних предикатів, що є узагальненням алгебри логіки; змінні, які містить система логічних рівнянь, мають приймати значення зі скінченних алфавітів; розробити метод поповнення бази знань у формі предикатних рівнянь за допомогою вилучення змінних із логічних систем, де рівняння є складними логічними зв'язками між дискретними ознаками об'єктів або процесів; метод поповнення бази знань має бути оснований на додаванні нових спрощених

рівнянь, отриманих дедуктивним методом із застосуванням кванторних операцій; спрощені предикатні рівняння мають моделювати логічні закономірності, що неявно містяться в основній базі знань; дослідити на прикладі медичної діагностики, де класифікація групи ризиків спрощується завдяки додаванню більш простих рівнянь до бази знань, полегшуючи процес повторного використання знань, заданих неявно системою предикатних рівнянь.

Аналіз проблеми

Як гілка символічного штучного інтелекту, подання знань і логічне виведення спрямовані на проектування програмного забезпечення систем, що інтерпретують подання світу, схоже на інтерпретацію людини. Системи, основані на знаннях, мають обчислювальну модель предметної галузі, де символи відіграють роль артефактів реального світу, таких як фізичні об'єкти, події та відносини. Предметна галузь може охоплювати будь-яку частину реального світу або гіпотетичну систему, щодо якої є бажання подати та інтерпретувати знання для цілей проектування відповідних програмних систем.

У багатьох сучасних дослідницьких роботах аналізуються теоретичні та практичні аспекти окресленої теми [1, 2]. Широкий спектр можливостей повторного використання знань для підвищення ефективності розроблення програмного забезпечення запропоновано в роботах [3, 4]. Оскільки повторне використання знань є одним із найважливіших етапів процесу управління знаннями, у літературі можна знайти кілька досліджень про цей етап [5]. Компанії в усьому світі витрачають час, зусилля та гроші, намагаючись залучити найефективніші системи управління знаннями. Розглядаються не лише програмні засоби, а й культурні та стратегічні аспекти.

Повторне використання знань в інформаційних системах часто описується дескриптивними логічними мовами [6, 7]. Дескриптивний підхід насамперед належить до формалізмів логіки предикатів першого порядку. Його важливість випливає з того факту, що здебільшого всі формалізми символічного подання знань можуть бути зведені до логіки предикатів першого порядку. Логіка першого порядку відтворює сутність мислення людини введенням поняття логічного підсумку. Вона також дає змогу вводити поняття універсальної правильності в тому сенсі, що логічне твердження може бути правильним незалежно від будь-яких передумов. Логічний

підсумок та універсальна правильність можуть бути описані в термінах моделювання семантики тверджень. Модель для логічної теорії описує умови, за яких теорія правильна. Логічний підсумок є твердженням, правильним для всіх моделей теорії. Далі дедуктивне виведення пояснює, як підсумки виводяться з теорії. Дедукція дає доступ до знань, що неявно подані в теорії. У термінах предикатних рівнянь [8] це означає, що розв'язання рівняння щодо невідомих значень змінних чи відношень між ними дає нам додаткові знання, не подані в явному вигляді в системі рівнянь бази знань.

Додавання рівнянь, отриманих способом дедуктивного виведення з наявної бази знань, є поповненням бази знань та уможливує повторне використання знань. Це має багато практичних переваг, зокрема підвищення ефективності, якості та інноваційної діяльності, зменшення витрат часу й ресурсів, а також сприяння навчанню та розвитку. Проблеми повторного використання знань мають велике значення для різних галузей науки та практики.

Система логічних предикатних рівнянь здатна застосовувати знання предметної галузі, отримані з меншої проблеми, до більш складних проблем тієї самої або спорідненої галузі, але наразі переважній більшості логічних та еволюційних обчислювальних методів не вистачає цієї здатності. Відсутність здатності застосувати вже набуті знання про предметну галузь призводить до споживання більшої кількості ресурсів і часу для вирішення складніших проблем предметної галузі [8]. Залежно від того, як проблема збільшується в розмірах, її стає важко, а іноді навіть непрактично (якщо не неможливо) розв'язати через відсутність ресурсів і часу. Тому необхідна система, що має здатність повторно використовувати отримані знання про проблемну галузь для масштабування в цій галузі.

Система, основана на знаннях, підтримує базу знань, що зберігає символи обчислювальної моделі у формі тверджень щодо галузі, яка розглядається. У нашій ситуації ці твердження записуються як рівняння алгебри скінченних предикатів. Така система здійснює логічні виведення з наявних рівнянь, результати яких можуть поповнювати наявну базу знань, забезпечуючи цим повторне використання знань. Програмні застосунки можуть приймати рішення на основі запитів до бази знань.

Основна мета наукового напрямку полягає в тому, щоб розробити систему, здатну автономно масштабувати навчання, починаючи від незначних

проблем і завершуючи складнішими проблемами тієї самої чи спорідненої сфери, у поведінці, подібній до людської. Для того, щоб автономно масштабувати проблемну галузь, потрібно вміти формувати багаторазово використовувані логічні блоки знань. Для вилучення та повторного використання інформації з логічних рівнянь у проблемній галузі потрібне розширене кодування. Простір пошуку може розширитися, наприклад, у деяких формах генетичного програмування машинного навчання [9–11].

Як правило, агент містить еволюційні обчислення та машинне навчання для вирішення певного завдання в невідомому середовищі. Правила мають форму "якщо умова, то дія". Зазвичай умова подана бітовим рядком фіксованої довжини, визначеним у потрібному алфавіті $\{0,1,\#\}$, де $\#$ – символ, що означає "невизначено", тобто або 0 або 1, а дія подана числовою константою. Техніка класифікаційної системи навчання може масштабуватися в проблемних сферах, але щоразу її потрібно вивчати із самого початку. Крім того, збільшення розмірності проблеми, що призводить до розширення простору пошуку, вимагає великого простору пам'яті, значно збільшує час навчання та зрештою обмежує систему розміром проблеми [12]. За допомогою явної передачі знань домену в системі класифікації можна досягти масштабованості, але це додає упередженості та обмежує використання в кількох доменах.

У роботі [13] досліджується процес формування правил за допомогою алгебри скінчених предикатів і предикатних операцій. Аналізуються двійкові предикати, що використовуються для структурування наявних фактів. Розроблено метод підвищення ефективності повторного використання знань.

Мета цього дослідження – розробити метод повторного використання знань за допомогою вирішення логічних рівнянь для певної предметної галузі, наприклад, медицини. Об'єктом є процес побудови та поповнення баз знань, а предметом – логічне моделювання знань і логічне виведення. Основні завдання роботи: розробити метод поповнення бази знань за допомогою логічних рівнянь із суттєвими змінними, провести дослідження та оцінити результати.

Вирішення завдання

Теорія інтелекту з огляду на своє визначення має обмежитися вивченням власне машиноподібних функцій людського інтелекту, що характеризуються

детермінованістю, дискретністю та скінченністю. До того ж виникають питання про те, якою мірою введені обмеження звужують клас функцій людського інтелекту, що допускають моделювання в межах теорії інтелекту.

Будь-яка обчислювальна система має такі обмеження: 1) алфавіт літер, з яких будуються слова для будь-якої конкретної обчислювальної системи, завжди скінченний; 2) довжина слів, які здатна сприймати та формувати ця система, обмежена деяким кінцевим, наперед заданим числом букв, що визначається конструкцією системи, її швидкодією та терміном служби; 3) реакції обчислювальної системи чітко детерміновані: повторне подання вхідного слова завжди призводить до формування системою того самого вихідного слова.

Проблема повторного використання має три основні причини. Перша стосується розуміння контексту та застосованого рішення, інша – змісту знань, що документуються, і остання стосується всієї системи бази знань або програмного забезпечення, яке використовується для її підтримки. Коли говорять про обмін і повторне використання знань, існує припущення, що знання є такими, що їх можна відтворювати та переміщувати з місця на місце; сутність, що можна отримати від людини-експерта й перенести з однієї комп'ютерної системи чи програми в іншу. Але як можна запропонувати поділитися або повторно використовувати те, що не має таких властивостей, як локальність і стійкість? Знання розглядається як абстракція, що неможливо записати й ніколи не можна мати в руках. Знання – це те, що спостерігач пояснив би розумному агенту, зі свого боку це дає змогу агенту раціонально моделювати свою поведінку для досягнення певних цілей, що сприймаються, відповідно до того, що він дізнався від спостерігача.

Якщо припустити, що все вищезазначене виконано, то проблема ІТ-підтримки процесу управління знаннями все ще потребує вирішення. Метою поточних баз знань є сприяння обміну знаннями та повторному використанню. Для подання знань необхідно враховувати значну кількість факторів. Термінології, онтології та методи вирішення проблем – це декілька з них [14]. Проблема полягає в тому, щоб мати змогу поділитися знаннями, які містяться в різних базах знань, оскільки всі ці фактори відрізняються від однієї бази до іншої. Несумісність систем і форматів також унеможливило об'єднання двох баз знань.

Можна виокремити чотири основні причини.

1. Неоднорідність подання. Існує кілька підходів до подання знань, але один формалізм подання не може бути безпосередньо доданий до іншого. Не існує універсального формалізму подання знань, що ідеально відповідав би всім вимогам, і тому обмін знаннями передбачає переклад змісту однієї бази в іншу.

2. Мовні діалекти. Обмін знаннями між системами може бути дуже складним, якщо знання закодовані різними діалектами. Це може повністю змінити зміст повідомлення або його інтерпретацію.

3. Відсутність визначених правил спілкування. Теоретично окремі системи можуть спілкуватися одна з одною і таким чином мати користь від обміну знаннями, навіть без спільної бази. Але зазвичай це неможливо, оскільки бракує узгодженого протоколу, що б давав змогу системам взаємодіяти та запитувати одна одну.

4. Невідповідності моделі на рівні знань. Навіть у разі усунення всіх перешкод проблеми з термінологією також стануть перешкодою для ефективного спілкування між різними базами. Відсутність спільного словникового запасу не дає змоги співвідносити знання однієї бази з іншою.

Отже, знання розглядається як здатність реагувати певним чином, а не як матеріальна субстанція. Навіть інформацію, що застосовується для подання знань, не можна вважати такою; правила, символи та фрейми не здатні генерувати розумну поведінку. Для того щоб мати змогу математично описувати функції інтелекту, необхідно створити формальну мову, якою можна було б робити такий опис. Формальна мова має обиратися в такий спосіб, щоб нею можна було в зручній формі записати будь-який скінченний алфавітний оператор. Такою мовою є алгебра скінченних предикатів [8].

Поняття скінченного предиката було введено таким чином. Нехай A – скінченний алфавіт, що складається з k літер a_1, a_2, \dots, a_k , а Σ – множина, яка містить два елементи, що позначаються символами 0, 1 і названі відповідно – похибкою та істиною. Змінну, задану на множині A , називають літерною, а змінну, задану на множині Σ , – логічною.

Скінченним n -місцевим предикатом над алфавітом A називається будь-яка функція $f(x_1, x_2, \dots, x_n) = t$ від n буквених аргументів x_1, x_2, \dots, x_n , заданих на множині A , і яка приймає логічні значення t . Іноді скінченний предикат f

називається k -ічним. Цим наголошується, що його алфавіт A складається з k літер.

Кожному скінченному алфавітному оператору може відповідати певний власний скінченний предикат. Зробити це можна таким чином. Нехай $F(x_1, x_2, \dots, x_m) = y_1 y_2 \dots y_m$ – довільно обраний скінченний алфавітний оператор, що перетворює вхідні слова x_1, x_2, \dots, x_m довжини m у вихідні слова y_1, y_2, \dots, y_m тієї самої довжини, які складаються з літер алфавіту A .

Пропонується такий метод поповнення бази знань новими логічними рівняннями, що спрощують повторне використання логічних зв'язків між дискретними ознаками. У загальному вигляді розглядається система рівнянь алгебри скінченних предикатів, що інтерпретує логічну структуру наявних знань щодо предметної галузі. Така система завжди може бути подана у вигляді одного рівняння:

$$f(x_1, x_2, \dots, x_n) = 1,$$

де x_1, x_2, \dots, x_n – всі наявні дискретні ознаки, а формула f відображає логічну структуру бази знань.

Обираються набори ознак, для яких передбачається повторне використання в логічних висновках: $x_{i1}, x_{i2}, \dots, x_{im}$. Для кожного такого набору всі змінні, що не входять до нього, вилучаються квантором існування. Якщо можливо, застосовуються властивості квантора, що допомагають уникнути складних обчислень. Кожне отримане рівняння додається до бази знань.

Крім того, у процесі вирішення практичних задач виникають запитання:

- як конкретні значення цієї ознаки, підставлені в логічне рівняння, вплинуть на зв'язки між іншими ознаками?
- наскільки сильним є логічний зв'язок між двома (або більше) заданими ознаками?

Для відповіді на перше запитання природно виокремити ті предикати (і, відповідно, рівняння), які під час підставлення певного значення ознаки перетворюються на предикати, що приводять до сильнішого зв'язку між змінними, а також такі предикати, підставлення в які цього значення спричиняє послаблення логічного зв'язку між ознаками [13].

Щоб отримати відповідь на друге запитання, необхідно вилучити з вихідного рівняння за допомогою квантора існування всі змінні, крім тих, що розглядаються, і дослідити отримане рівняння

з меншою кількістю змінних, яке описує всі допустимі множини значень досліджуваних ознак.

Коли виникає необхідність повторно використати логічні зв'язки між ознаками конкретного набору, вирішується не початкове складне рівняння, а те рівняння, що пов'язує обрані дискретні змінні.

Для спрощення обчислювальних процедур важливо вилучити з рівняння всі змінні, крім розглянутих, і дослідити отримане рівняння з меншою кількістю змінних. Ці додаткові рівняння поповнюють базу знань і дають змогу уникнути використання значної кількості обчислювальних ресурсів для логічних висновків у базі знань. У роботі [14] розглядається досить широкий клас предикатів, для яких можна задати ефективний алгоритм елімінації змінних без збільшення розміру вихідної формули. Розглянемо такі властивості квантора існування:

1. $\exists x x^a = 1$.
2. $\exists x \neg x^a = 1$.
3. $\exists x (\neg (P(x) \& Q(x))) = \exists x \neg P(x) \vee \exists x \neg Q(x)$.
4. $\exists x ((P(x) \vee Q(x))) = \exists x P(x) \vee \exists x Q(x)$.
5. $\exists x (P(x) \& Q(y)) = \exists x P(x) \& Q(y)$.
6. $\exists y (P(x) \rightarrow Q(y)) = P(x) \rightarrow \exists y Q(y)$.
7. $\exists y (P(x) \rightarrow Q(y)) = P(x) \rightarrow \exists y Q(y)$.
8. Припустимо, $P_i(x) \& P_j(x) = 0, i \neq j, i, j = 1, 2, \dots, k$, тоді:

$$\exists y ((P_1(x) \rightarrow Q_1(y)) \& (P_2(x) \rightarrow Q_2(y)) \& \dots \& (P_k(x) \rightarrow Q_k(y))) = (P_1(x) \rightarrow \exists y Q_1(y)) \& (P_2(x) \rightarrow \exists y Q_2(y)) \& \dots \& (P_k(x) \rightarrow \exists y Q_k(y)).$$
9. Якщо тотожність $P_i(x) \equiv 0$ не є істинною для будь-якої $i = 1, 2, \dots, k$ і $P_i(x) \& P_j(x) = 0$ for $i \neq j, i, j = 1, 2, \dots, k$, тоді

$$\exists x ((P_1(x) \rightarrow Q_1(y)) \& (P_2(x) \rightarrow Q_2(y)) \& \dots \& (P_k(x) \rightarrow Q_k(y))) = Q_1(y) \vee Q_2(y) \vee \dots \vee Q_k(y).$$

Перелічені вище властивості дають змогу описати широкий клас скінченних предикатів (відповідно до рівнянь), визначених на множині змінних $\{x, y, \dots, z\}$, для яких легко знайти зв'язки між вибраними змінними без збільшення розміру вихідних формул. Визначимо такий клас рекурсивно.

1. Усі "розпізнавання" x^a, x^b, \dots, x^c (a, b, \dots, c – символи, що належать домену для змінної x) належать до Δx .

2. Усі заперечення $\neg x^a, \neg x^b, \dots, \neg x^c$ належать до Δx .

3. Якщо предикати $\neg P(x), \neg Q(x)$ належать до Δx , то предикат $\neg (P(x) \& Q(x))$ належить до Δx .

4. Будь-який предикат, що не залежить від змінної x , належить до множини Δx .

5. Якщо предикати P_1 та P_2 належать до Δx , тоді предикат $P = P_1 \vee P_2$ належить до Δx .

6. Якщо предикат P_1 належить до Δx і предикат P_2 не залежить від x , тоді предикат $P = P_1 \& P_2$ належить до Δx .

7. Якщо предикат P_1 не залежить від x і предикат P_2 належить до Δx тоді предикат $P = P_1 \rightarrow P_2$ належить до Δx .

8. Нехай предикати P_1, P_2, \dots, P_k не залежать від x ; $P_i \& P_j = 0$ для $i \neq j, i, j = 1, 2, \dots, k$, предикати Q_1, Q_2, \dots, Q_k належать до Δx , тоді

$$P = (P_1 \rightarrow Q_1) \& (P_2 \rightarrow Q_2) \& \dots \& (P_k \rightarrow Q_k)$$

належить до Δx .

9. Якщо предикати P_1, P_2, \dots, P_k залежать тільки від x , $P_i \& P_j = 0$ для $i \neq j, i, j = 1, 2, \dots, k$; для будь-якого $i = 1, 2, \dots, k$ тоді тотожність $P_i \equiv 0$ не є істинною; предикати Q_1, Q_2, \dots, Q_k не залежать від x ; тоді предикат

$$P = (P_1 \rightarrow Q_1) \& (P_2 \rightarrow Q_2) \& \dots \& (P_k \rightarrow Q_k)$$

належить до Δx .

Також може виникнути потреба вилучити зайві змінні з допомогою універсального квантора. У цьому разі можемо скористатися такими властивостями цього квантора:

1. $\forall x x^a = 0$
2. $\forall x \neg x^a = 0$
3. $\forall x \neg (P(x) \vee Q(x)) = \forall x \neg P(x) \& \forall x \neg Q(x)$
4. $\forall x (P(x) \& Q(x)) = \forall x P(x) \& \forall x Q(x)$
5. $\forall x (P(x) \vee Q(x)) = \forall x P(x) \vee \forall x Q(x)$
6. $\forall y (P(x) \& Q(y)) = P(x) \& \forall y Q(y)$

7. Припустимо,

$$P_i(x) \& P_j(x) = 0, i \neq j, i, j = 1, 2, \dots, k,$$

тоді:

$$\begin{aligned} & \forall y ((P_1(x) \& Q_1(y)) \vee (P_2(x) \& Q_2(y)) \vee \dots \\ & \vee (P_k(x) \& Q_k(y))) = (P_1(x) \& \forall y Q_1(y)) \vee \\ & \vee (P_2(x) \& \forall y Q_2(y)) \vee \dots \vee (P_k(x) \& \forall y Q_k(y)). \end{aligned}$$

8. Якщо ідентичність $P_i(x) \equiv 0$ не є істинною для будь-якої $i = 1, 2, \dots, k$ і $P_i(x) \& P_j(x) = 0$ для $i \neq j, i, j = 1, 2, \dots, k$, тоді

$$\begin{aligned} & \forall x ((P_1(x) \& Q_1(y)) \vee (P_2(x) \& Q_2(y)) \vee \dots \\ & \vee (P_k(x) \& Q_k(y))) = Q_1(y) \& Q_2(y) \& \dots \& Q_k(y). \end{aligned}$$

Визначимо рекурсивно клас предикатів Σ_x , з якого можна вилучити змінну x без збільшення розміру формули.

1. Усі "розпізнавання" x^a, x^b, \dots, x^c належать Σ_x .

2. Усі заперечення $\neg x^a, \neg x^b, \dots, \neg x^c$, які не залежать від x , належать Σ_x .

3. Якщо $\neg P_1$ і $\neg P_2$ належать до Σ_x , тоді $\neg(P_1 \vee P_2)$ належить до Σ_x .

4. Якщо предикати P_1 та P_2 належать до Σ_x , тоді предикат $P = (P_1 \& P_2)$ належить до Σ_x .

5. Якщо предикат P_1 належить до Σ_x і предикат P_2 не залежить від x , тоді предикат $P = P_1 \vee P_2$ належить до Σ_x .

6. Якщо предикат P_1 не залежить від x і предикат P_2 належить до Σ_x , тоді предикат $P = P_1 \& P_2$ належить до Σ_x .

7. Припустимо, що предикати P_1, P_2, \dots, P_k не залежать від x , $P_i \& P_j = 0$ для $i \neq j, i, j = 1, 2, \dots, k$ предикати Q_1, Q_2, \dots, Q_k належать до Σ_x , тоді

$$P = (P_1 \& Q_1) \vee (P_2 \& Q_2) \vee \dots \vee (P_k \& Q_k)$$

належить до Σ_x .

8. Якщо предикати P_1, P_2, \dots, P_k залежать тільки від x , $P_i \& P_j = 0$ для $i \neq j, i, j = 1, 2, \dots, k$ для будь-якого $i = 1, 2, \dots, k$ тотожність $P_i \equiv 0$ не є істинною, предикати Q_1, Q_2, \dots, Q_k не залежать від x , тоді предикат $P = (P_1 \& Q_1) \vee (P_2 \& Q_2) \vee \dots \vee (P_k \& Q_k)$ належить до множини Σ_x .

Універсальним способом розв'язання систем рівнянь алгебри скінченних предикатів є приведення предиката, заданого системою рівнянь і початковими умовами, до досконалої диз'юнктивної нормальної форми. Однак така процедура передбачає перебір багатьох проміжних розв'язків, а її практична реалізація потребує значних затрат комп'ютерного часу. Для деяких типів предикатних рівнянь, зважаючи на особливості їх структури, можна розробити простіші алгоритми розв'язання.

Отже, показано, що для репрезентації та інтерпретації знань за умови подання у формальному вигляді інформації про об'єкти й процеси в базах знань використовуються різні методи алгебри скінченних предикатів і предикатних операцій. Так, логічні методи розпізнавання образів передбачають складання та розв'язання логічних рівнянь зі змінними, що набувають значень 1 та 0 залежно від того, чи має об'єкт певну властивість, чи ні. Розв'язання таких рівнянь дає змогу або ідентифікувати об'єкт за наявними наборами значень атрибутивних змінних, або встановити невідомі властивості цього об'єкта. У багатьох практичних задачах, пов'язаних із семантичним обробленням медичних даних, природно-мовної інформації, відомостей про клієнтів, немає необхідності отримувати всі набори значень семантичних ознак, але потрібно отримати один або кілька наборів значень ознак (цільових змінних), що цікавлять користувача. Часто необхідно знайти значення цільових змінних за заданих початкових умов, які є фіксованим набором значень інших ознак. У процесі розв'язання таких задач інші змінні, які відсутні в початкових умовах і не є цільовими, вилучаються з рівняння способом зв'язування їх з екзистенціальними кванторами.

Приклад

Розглянемо медичний приклад поповнення бази знань новими рівняннями. План експерименту такий. Ми використовуємо реальні медичні дані та кодуємо їх за допомогою предикатних рівнянь. Зауважимо, що хоча деякі змінні можуть набувати значення "невідомо", а втім це випадок закритого світу, оскільки "невідомо" означає лише значення з алфавіту, на якому визначена змінна. Отже, кожна галузь для будь-якої змінної є закритою. Після того, як ми написали систему рівнянь за допомогою

експертів, починаємо вилучати змінні, які вважаємо несуттєвими на цей момент. Це не означає, що в іншому разі інші змінні вважатимуться несуттєвими. Істотні змінні – це ті, для яких ми хочемо визначити логічні зв'язки. На виході отримуємо рівняння, де несуттєві змінні вилучено. Отримане рівняння простіше, ніж початкова система, і можна проаналізувати зв'язки між основними змінними більш простим способом. Так, ми поповнюємо базу знань додатковими рівняннями, отриманими як результати дедуктивного виведення з основної бази знань.

Якщо розглядати інформаційний скринінг медичних показників для оцінювання розвитку та профілактики захворювань серця й судин [15–17], то можна виокремити набір ознак для формалізації скринінгових процедур. Розглянемо такі ознаки та їх значення.

$$\text{Стать: } X_1 = \{x_1^1, x_1^2\},$$

де x_1^1 означає жінка; x_1^2 – чоловік.

$$\text{Вік: } X_2 = \{x_2^1, x_2^2, x_2^3\},$$

де x_2^1 – менше ніж 40 років; x_2^2 – від 40 до 50 років; x_2^3 – понад 50 років.

$$\text{Цукровий діабет: } X_3 = \{x_3^1, x_3^2, x_3^3, x_3^4\},$$

де x_3^1 – так; x_3^2 – ні (фактичний діагноз); x_3^3 – ні (діагноз не встановлений); x_3^4 – невідомо.

$$\text{Артеріальна гіпертензія: } X_4 = \{x_4^1, x_4^2, x_4^3, x_4^4\},$$

де x_4^1 – так; x_4^2 – ні (фактичний діагноз); x_4^3 – ні (діагноз не встановлений); x_4^4 – невідомо.

$$\text{Проблеми з нирками: } X_5 = \{x_5^1, x_5^2, x_5^3\},$$

де x_5^1 – так; x_5^2 – ні; x_5^3 – невідомо.

$$\text{Тахікардія: } X_6 = \{x_6^1, x_6^2, x_6^3, x_6^4, x_6^5\},$$

де x_6^1 – так (фактичний діагноз); x_6^2 – так (діагноз не встановлений); x_6^3 – ні (справжній діагноз); x_6^4 – ні (діагноз не встановлений); x_6^5 – невідомо.

Спадковість щодо захворювань серця та судин:

$$X_7 = \{x_7^1, x_7^2, x_7^3\},$$

де x_7^1 – так; x_7^2 – ні; x_7^3 – невідомо.

$$\text{Паління: } X_8 = \{x_8^1, x_8^2, x_8^3\},$$

де x_8^1 – так; x_8^2 – ні; x_8^3 – невідомо.

$$\text{Проблеми з алкоголем: } X_9 = \{x_9^1, x_9^2, x_9^3\},$$

де x_9^1 – так; x_9^2 – ні; x_9^3 – невідомо.

$$\text{Гіподинамія } X_{10} = \{x_{10}^1, x_{10}^2, x_{10}^3\},$$

де x_{10}^1 – так; x_{10}^2 – ні; x_{10}^3 – невідомо.

Ці особливості дають змогу розробити модель ідентифікації діагностичних параметрів, за допомогою якої можна визначити групу здоров'я пацієнта $R = \{r_1, r_2, r_3, r_4\}$, де r_1 – низький ризик захворювань серця та судин; r_2 – помірний ризик; r_3 – високий ризик; r_4 – дуже високий ризик.

Для визначення групи здоров'я використовується набір агрегованих ознак $Q_1 - Q_3$, де Q_1 виражається через X_1 та X_2 ; Q_2 виражається через X_7 до X_{10} ; Q_3 виражається через $X_3 - X_6$.

Значення кожної групи здоров'я та кожної агрегованої ознаки розподілено на чотири класи згідно з відповідною медико-технологічною документацією (уніфікованим клінічним протоколом і локальними протоколами щодо профілактики хвороб серця й судин).

Наприклад, для формування ознаки Q_2 можна скласти таку систему предикатних рівнянь:

$$\left\{ \begin{aligned} q_2^1 &= x_7^2 x_8^2 (x_9^2 \vee x_9^3 (x_{10}^2 \vee x_{10}^3)) \vee x_7^2 x_8^3 x_9^2 x_{10}^2 \vee x_7^3 x_8^2 x_{10}^2 (x_9^2 \vee x_9^3) \\ q_2^2 &= x_7^2 ((x_8^1 (x_9^1 x_{10}^2 \vee x_9^2) \vee x_9^3 (x_8^1 x_{10}^2 \vee x_8^2 x_{10}^1)) \vee (x_7^2 (x_8^2 x_9^1 \vee x_8^3 x_9^2) \vee (x_7^2 x_9^3 \vee x_7^3 x_9^1) x_8^3 x_{10}^2 \vee \\ &\quad \vee (x_7^2 x_8^3 \vee x_7^3 x_8^2) x_9^1 (x_{10}^2 \vee x_{10}^3)) \vee x_7^3 x_8^2 (x_9^2 \vee x_9^3)) (x_{10}^1 \vee x_{10}^3) \vee x_7^3 x_8^3 (x_9^1 x_{10}^2 \vee x_{10}^3) \\ q_2^3 &= x_7^1 x_{10}^2 (x_8^1 x_9^2 \vee x_8^2 (x_9^1 \vee x_9^2)) \vee (x_7^1 x_9^3 (x_8^1 \vee x_8^2) \vee (x_7^1 x_8^3 \vee x_7^3 x_8^1) x_9^1) (x_{10}^2 \vee x_{10}^3) \vee x_7^1 x_8^3 (x_9^2 \vee x_9^3) \vee \\ &\quad \vee (x_7^2 (x_8^1 (x_9^1 \vee x_9^2) \vee x_8^3 x_9^3) \vee (x_7^2 x_8^3 \vee x_7^3 x_8^2) x_9^1 x_{10}^1 \vee x_7^3 (x_8^1 x_9^2 \vee x_8^3 x_9^1)) (x_{10}^1 \vee x_{10}^3) \vee x_7^3 x_8^3 (x_9^2 x_{10}^1 \vee x_9^3) \\ q_2^4 &= x_7^1 x_9^3 x_{10}^1 (x_8^1 \vee x_8^2) \vee (x_7^1 x_8^3 \vee x_7^3 x_8^1) x_9^1 x_{10}^1 \vee (x_7^1 x_8^2 x_9^1 \vee x_7^3 x_9^2 (x_8^1 \vee x_8^2)) (x_{10}^1 \vee x_{10}^3) \end{aligned} \right.$$

Остаточна класифікація може бути виражена

такою системою:

$$\begin{cases} r_1 = q_1^1 q_2^1 (q_3^1 \vee q_3^2) \vee (q_1^1 q_2^2 \vee (q_1^2 \vee q_1^3) q_2^1) q_3^1 \\ r_2 = q_1^1 (q_2^1 q_3^3 \vee q_2^2 q_3^2) \vee (q_1^1 (q_2^3 \vee q_2^4) \vee q_1^2 (q_2^2 \vee q_2^3) \vee q_1^3 q_2^2 \vee q_1^4 (q_2^1 \vee q_2^2)) (q_3^1 \vee q_3^2) \vee (q_1^2 \vee q_1^3) q_2^1 (q_3^2 \vee q_3^3) \vee \\ \vee (q_1^2 q_2^4 \vee (q_1^3 \vee q_1^4) q_2^3) q_3^1 \\ r_3 = q_2^1 q_3^4 \vee (q_1^1 \vee q_1^2 \vee q_1^3) (q_2^3 \vee q_2^2) (q_3^3 \vee q_3^4) \vee q_1^3 q_3^2 (q_2^3 \vee q_2^4) \vee (q_1^3 \vee q_1^4) q_2^4 q_3^1 \vee (q_1^1 q_2^4 \vee q_1^4 (q_2^1 \vee q_2^2)) q_3^3 \vee \\ \vee (q_1^2 q_2^4 \vee q_1^4 q_2^3) (q_3^2 \vee q_3^3) \\ r_4 = (q_1^1 \vee q_1^2) q_2^4 q_3^4 \vee q_1^3 q_2^4 (q_3^3 \vee q_3^4) \vee q_1^4 q_3^4 (q_2^2 \vee q_2^3) \vee q_1^4 q_2^4 (q_3^2 \vee q_3^3 \vee q_3^4) \end{cases}$$

Дослідимо логічні зв'язки між дискретними ознаками $x_1 - x_{10}$. Насамперед перепишемо систему

предикатних рівнянь у такому вигляді:

$$\begin{aligned} P(q_2, x_1, \dots, x_{10}) = & \\ = & q_2^1 (x_7^2 x_8^2 (x_9^2 \vee x_9^3 (x_{10}^2 \vee x_{10}^3)) \vee x_7^2 x_8^3 x_9^2 x_{10}^2 \vee x_7^3 x_8^2 x_{10}^2 (x_9^2 \vee x_9^3)) \vee q_2^2 (x_7^2 (x_8^1 (x_9^1 x_{10}^2 \vee x_9^2) \vee x_9^3 (x_8^1 x_{10}^2 \vee x_8^2 x_{10}^1)) \vee \\ & \vee (x_7^2 (x_8^2 x_9^1 \vee x_8^3 x_9^2) \vee (x_7^2 x_9^3 \vee x_7^3 x_9^1) x_8^3 x_{10}^2 \vee (x_7^2 x_8^3 \vee x_7^3 x_8^2) x_9^1 (x_{10}^2 \vee x_{10}^3) \vee x_7^3 x_8^2 (x_9^2 \vee x_9^3)) (x_{10}^1 \vee x_{10}^3) \vee \\ & \vee x_7^3 x_8^1 (x_9^2 x_{10}^2 \vee x_9^3) \vee x_7^3 x_8^3 x_9^2 (x_{10}^1 \vee x_{10}^2)) \vee \\ & \vee q_2^3 (x_7^1 x_{10}^2 (x_8^1 x_9^2 \vee x_8^2 (x_9^1 \vee x_9^2)) \vee (x_7^1 x_9^3 (x_8^1 \vee x_8^2) \vee (x_7^1 x_8^3 \vee x_7^3 x_8^1) x_9^1) (x_{10}^2 \vee x_{10}^3) \vee x_7^1 x_8^3 (x_9^2 \vee x_9^3) \vee \\ & \vee (x_7^2 (x_8^1 (x_9^1 \vee x_9^2) \vee x_8^3 x_9^3) \vee (x_7^2 x_8^3 \vee x_7^3 x_8^2) x_9^1 x_{10}^1 \vee x_7^3 (x_8^1 x_9^2 \vee x_8^3 x_9^1)) (x_{10}^1 \vee x_{10}^3) \vee x_7^3 x_8^3 (x_9^2 x_{10}^3 \vee x_9^3)) \vee \\ & \vee q_2^4 (x_7^1 x_9^3 x_{10}^1 (x_8^1 \vee x_8^2) \vee (x_7^1 x_8^3 \vee x_7^3 x_8^1) x_9^1 x_{10}^1 \vee (x_7^1 x_8^2 x_9^1 \vee x_7^1 x_9^2 (x_8^1 \vee x_8^2)) (x_{10}^1 \vee x_{10}^3) \vee x_7^1 x_8^1 x_9^1) = 1 \end{aligned}$$

Видно, що цей предикат належить до класу Δ_{x_7} .

Це вилучення дасть нам зв'язок між змінними

Дослідимо зв'язок між усіма змінними, крім x_7 .

$q_2, x_1, \dots, x_6, x_8, x_9, x_{10}$:

$$\begin{aligned} F = \exists x_7 P(q_2, x_1, \dots, x_{10}) = & \\ = & q_2^1 (x_8^2 (x_9^2 \vee x_9^3 (x_{10}^2 \vee x_{10}^3)) \vee x_8^3 x_9^2 x_{10}^2 \vee x_8^2 x_{10}^2 (x_9^2 \vee x_9^3)) \vee q_2^2 ((x_8^1 (x_9^1 x_{10}^2 \vee x_9^2) \vee x_9^3 (x_8^1 x_{10}^2 \vee x_8^2 x_{10}^1)) \vee \\ & \vee ((x_8^2 x_9^1 \vee x_8^3 x_9^2) \vee (x_9^3 \vee x_9^1) x_8^3 x_{10}^2 \vee (x_8^3 \vee x_8^2) x_9^1 (x_{10}^2 \vee x_{10}^3) \vee x_8^2 (x_9^2 \vee x_9^3)) (x_{10}^1 \vee x_{10}^3) \vee x_8^1 (x_9^2 x_{10}^2 \vee x_9^3) \vee \\ & \vee x_8^3 x_9^2 (x_{10}^1 \vee x_{10}^2)) \vee q_2^3 (x_{10}^2 (x_8^1 x_9^2 \vee x_8^2 (x_9^1 \vee x_9^2)) \vee (x_9^3 (x_8^1 \vee x_8^2) \vee (x_8^3 \vee x_7^3 x_8^1) x_9^1) (x_{10}^2 \vee x_{10}^3) \vee \\ & \vee x_8^3 (x_9^2 \vee x_9^3) \vee ((x_8^1 (x_9^1 \vee x_9^2) \vee x_8^3 x_9^3) \vee (x_8^3 \vee x_8^2) x_9^1 x_{10}^1 \vee (x_8^1 x_9^2 \vee x_8^3 x_9^1)) (x_{10}^1 \vee x_{10}^3) \vee x_8^3 (x_9^2 x_{10}^3 \vee x_9^3)) \vee \\ & \vee q_2^4 (x_9^3 x_{10}^1 (x_8^1 \vee x_8^2) \vee (x_8^3 \vee x_8^1) x_9^1 x_{10}^1 \vee (x_8^2 x_9^1 \vee x_9^2 (x_8^1 \vee x_8^2)) (x_{10}^1 \vee x_{10}^3) \vee x_8^1 x_9^1) = 1 \end{aligned}$$

Необхідно зазначити, що розмір формули не збільшився, тому що предикат $P(q_2, x_1, \dots, x_{10})$ належить до Δ_{x_7} .

Припустимо, нас цікавить зв'язок між q_2, x_9, x_{10} .

Вилучимо інші ознаки з предиката $F(x_1, \dots, x_{10})$:

$$\begin{aligned}
G(q_2, x_9, x_{10}) &= \exists x_1 \exists x_2 \exists x_3 \exists x_4 \exists x_5 \exists x_6 \exists x_7 \exists x_8 P(q_2, x_1, x_{10}) = \\
&= q_2^1 \left((x_9^2 \vee x_9^3 (x_{10}^2 \vee x_{10}^3)) \vee x_9^2 x_{10}^2 \vee x_{10}^2 (x_9^2 \vee x_9^3) \right) \vee q_2^2 \left(((x_9^1 x_{10}^2 \vee x_9^2) \vee x_9^3 (x_{10}^2 \vee x_{10}^1)) \vee \right. \\
&\vee \left((x_9^1 \vee x_9^2) \vee (x_9^3 \vee x_9^1) x_{10}^2 \vee x_9^1 (x_{10}^2 \vee x_{10}^3) \vee (x_9^2 \vee x_9^3) \right) (x_{10}^1 \vee x_{10}^3) \vee (x_9^2 x_{10}^2 \vee x_9^3) \vee x_9^2 (x_{10}^1 \vee x_{10}^2) \vee \\
&\vee q_2^3 \left(x_{10}^2 (x_9^1 \vee x_9^2) \vee x_9^3 (x_{10}^2 \vee x_{10}^3) \vee (x_9^2 \vee x_9^3) \vee ((x_9^1 \vee x_9^3) \vee x_9^1 x_{10}^1 \vee (x_9^2 \vee x_9^1)) (x_{10}^1 \vee x_{10}^3) \vee (x_9^2 x_{10}^3 \vee x_9^3) \right) \vee \\
&\vee q_2^4 \left(x_9^3 x_{10}^1 \vee x_9^1 x_{10}^1 \vee (x_9^1 \vee x_9^2) (x_{10}^1 \vee x_{10}^3) \vee x_9^1 \right) = 1.
\end{aligned}$$

Отже, скоротили початкову формулу й маємо більш просту залежність між обраними медичними ознаками. Отримавши необхідну залежність, можна розв'язати рівняння з однією або декількома цільовими змінними. Останнє рівняння додаємо до бази знань, щоб нові знання, отримані внаслідок наведених вище обчислень, поповнили наявну базу знань. Розглянемо можливі рішення цього рівняння.

Наприклад, $x_9 = 1$, тобто є проблеми з алкоголем. Тоді отримаємо таку залежність між ознаками:

$$q_2^2 x_{10}^2 \vee q_2^3 x_{10}^2 \vee q_2^4 = 1.$$

З останнього рівняння видно, що за умови $x_{10}^2 = 1$, q_2^4 має дорівнювати 4, тобто якщо є проблеми з алкоголем та невідомо, чи є гіподинамія, у класифікації пацієнтів необхідно використовувати значення агрегованої ознаки, що дорівнює 4.

Отже, вилучаючи різні набори змінних з рівнянь, що задовольняють наведені умови, поповнюємо базу знань і повторно використовуємо знання, які були неявно задані в початковій базі даних. Це спрощує процес логічного виведення та масштабує логічну систему подання знань. У цьому разі неявно задані знання стають явними.

Висновки

Дескриптивні підходи до моделювання процедур повторного використання знань ефективні у ситуаціях, коли розглядаються дискретні ознаки об'єктів, які пов'язані між собою складними логічними умовами. Алгебра скінченних предикатів дає змогу формувати бази знань у формі логічних рівнянь із дискретними змінними. Застосування кванторних операцій допомагає поповнювати наявні бази знань новими рівняннями, що спрощує повторне використання знань, заданих неявно.

Зручність диз'юнктивної алгебри скінченних предикатів у тому, що її мовою стисло й зручно записуються закони істинності та хибності, які відіграють особливу роль, оскільки вони задають

вимоги, необхідні й достатні для коректного введення змінних ознак на скінченних множинах. Закон істинності задає область зміни змінної, а закон хибності забезпечує попарну різницю всіх елементів множини, на якій задана змінна. Алгебра скінченних предикатів дає змогу інтерпретувати знання в чіткій математичній формі, де різні ознаки та їх значення пов'язані між собою за допомогою булевих і предикатних операцій.

Для ефективного розв'язання рівнянь із невідомими скінченними предикатами доцільно спочатку дослідити рівняння, записані за допомогою операцій булевої алгебри, оскільки саме ці операції найчастіше мають місце під час запису довільних предикатних рівнянь.

Показано, що в процесі побудови системи ознак для бази знань доцільно використовувати алгебро-логічну модель предметної галузі. Запропоновано підхід, оснований на алгебрі скінченних предикатів. Рекомендовано виокремити ті предикати (i , відповідно, рівняння), які за умови підставлення певного значення ознаки перетворюються на предикати, що дають сильніший зв'язок між змінними, а також такі предикати, підставлення в які цього значення призводить до послаблення логічного зв'язку між ознаками.

Запропоновано алгоритм вилучення несуттєвих змінних у розв'язанні алгебро-логічних рівнянь, який призводить до зменшення обсягу необхідних обчислень.

Розглянуто скінченні предикатні рівняння різних типів. Наведено задачі класифікації на основі предикатних рівнянь. Задача класифікації об'єктів на основі ознак, що набувають дискретних значень, описана математично як розв'язок предикатних рівнянь.

Запропоновано метод поповнення бази знань у формі логічних рівнянь для спрощення процедури повторного використання знань. Спочатку вилучаються змінні, що вважаються несуттєвими. База знань поповнюється простішими рівняннями, які можуть бути багаторазово використані за потреби. У практичних задачах, пов'язаних з інтерпретацією

знань, що здебільшого подані в текстовій формі, немає необхідності отримувати всі набори значень семантичних ознак, але потрібно отримати один або кілька суттєвих наборів значень ознак (цільових змінних), які є фіксованим набором значень інших ознак. У розв'язанні таких задач інші змінні, що не містяться в початкових умовах і не є цільовими, вилучаються з рівняння способом зв'язування їх з екзистенціальними кванторами.

Використання математичного апарату теорії інтелекту, методу компараторної ідентифікації та інструментарію алгебри скінченних предикатів дає змогу побудувати модель ідентифікації медично-діагностичних параметрів у вигляді системи

предикатних рівнянь, у процесі розв'язання яких маємо інтерпретацію медичних знань у певній галузі.

Описано широкий клас предикатів, з яких можна вилучити зайві змінні й зосередитися на зв'язках між суттєвими змінними. Запропоновано метод вилучення несуттєвих змінних способом застосування екзистенціального квантора й продемонстровано його на реальному медичному прикладі.

Розглянуто практичний приклад рішення цієї задачі. Подальші дослідження планується присвятити аналізу предикатних операцій, що дають змогу якомога швидше отримувати нові предикатні рівняння для автоматичної генерації логічних правил і визначати важливість інформаційних ознак.

Список літератури

1. L. He and P. Jiang Manufacturing Knowledge Graph: A Connectivism to Answer Production Problems Query With Knowledge Reuse, in *IEEE Access*, Vol. 7, 2019. P. 101231–101244. DOI: 10.1109/ACCESS.2019.2931361
2. B. Zhou, J. Bao, Y. Liu and D. Song BA-IKG: BiLSTM Embedded ALBERT for Industrial Knowledge Graph Generation and Reuse, *2020 IEEE 18th International Conference on Industrial Informatics (INDIN)*, Warwick, United Kingdom, 2020. P. 63–69. DOI: 10.1109/INDIN45582.2020.9442198
3. L. He and P. Jiang P. SaaS: knowledge service-oriented manufacturing workflow model for knowledge collaboration and reuse, *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, Hong Kong, China, 2020, P. 570–575. DOI: 10.1109/CASE48305.2020.9216974
4. D. Sapra and A. D. Pimentel Deep Learning Model Reuse and Composition in Knowledge Centric Networking, *29th International Conference on Computer Communications and Networks (ICCCN)*, Honolulu, HI, USA, 2020. P. 1–11. DOI: 10.1109/ICCCN49398.2020.9209668
5. N. Sharonova et al. Issues of Fact-based Information Analysis. *International Conference on Computational Linguistics and Intelligent Systems*, 2018. URL: <https://www.semanticscholar.org/paper/Issues-of-Fact-based-Information-Analysis-Sharonova-Doroshenko/f923b77b8561736202388db853e51df9bb7b9301>
6. W. Adjandra, Y. B. Putrapratama, A. Wiraguna, D. I. Sensuse and N. Safitri Systematic Literature Review Knowledge Reuse in Software Development, *International Conference on Computer Science and Engineering (IC2SE)*, Padang, Indonesia, 2021, P. 1–7. DOI: 10.1109/IC2SE52832.2021.9792093
7. Y. B. Putrapratama, W. Adjandra, A. Wiraguna, D. I. Sensuse and N. Safitri Knowledge Reuse Evaluation in Software Development: A Case Study on a Startup Company, *Sixth International Conference on Informatics and Computing (ICIC)*, Jakarta, Indonesia, 2021. P. 1–7. DOI: 10.1109/ICIC54025.2021.9632904
8. I. Shubin Development of conjunctive decomposition tools. *CEUR Workshop Proceedings*, 2021, Vol. 2870, P. 890–900 URL: <https://ceur-ws.org/Vol-2870/>
9. K. U. Sarker, A. B. Deraman and R. Hasan, Descriptive Logic for Software Engineering Ontology: Aspect Software Quality Control, *4th International Conference on Computer and Information Sciences (ICCOINS)*, Kuala Lumpur, Malaysia, 2018. P. 1–5. DOI: 10.1109/ICCOINS.2018.8510585
10. N. Kamide Sequential Fuzzy Description Logic: Reasoning for Fuzzy Knowledge Bases with Sequential Information, *IEEE 50th International Symposium on Multiple-Valued Logic (ISMVL)*, Miyazaki, Japan, 2020. P. 218–223. DOI: 10.1109/ISMVL49045.2020.000-2
11. R. Rawal, K. Goel and C. Gupta COVID-19: Disease Pattern Study based on Semantic-Web Approach using Description Logic, *IEEE International Conference for Innovation in Technology (INOCON)*, Bangluru, India, 2020. P. 1–5. DOI: 10.1109/INOCON50539.2020.9298278
12. I. Shubin, S. Snisar and S. Litvin Formalization and Application of Algebraic Methods in Automated Intelligent Systems, *IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*, Kharkiv, Ukraine. 2021. P. 67–70. DOI: 10.1109/PICST54195.2021.9772174
13. Каратаєв О.А., Шубін І.Ю. Проблеми повторного використання знань при проєктуванні програмних систем. *Сучасний стан наукових досліджень та технологій в промисловості*, 2023. No. 2 (24). С. 25–34. DOI: 10.30837/ITSSI.2023.24.025

14. 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>
15. Melnik K., Cherednichenko O., Glushko V. Towards medical screening information technology: the healthgrid-based approach *Information Systems: Methods, Models, and Applications*. Heidelberg: Springer, 2013. P. 202–204. DOI: 10.1007/978-3-642-38370-0_21
16. H. Kandil et al. Analysis of The Importance of Systolic Blood Pressure Versus Diastolic Blood Pressure in Diagnosing Hypertension: MRA Study. *IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, 2020. P. 443–447. DOI: 10.1109/ICIP40778.2020.9190990
17. M. Zhang, T. Zhang and Y. Cheng Intelligent Method of Non-Invasive Detection and Grading of Hypertension Using Fingertip Photoplethysmography. *3rd International Conference on Applied Machine Learning (ICAML)*, Changsha, China, 2021. P. 418–422. DOI: 10.1109/ICAML54311.2021.00094

References

1. He, L., Jiang, P. (2019), "Manufacturing Knowledge Graph: A Connectivism to Answer Production Problems Query With Knowledge Reuse", in *IEEE Access*, Vol. 7. P. 101231–101244. DOI: 10.1109/ACCESS.2019.2931361
2. Zhou, B., Bao, J., Liu, Y., Song, D. (2020), "BA-IKG: BiLSTM Embedded ALBERT for Industrial Knowledge Graph Generation and Reuse," *2020 IEEE 18th International Conference on Industrial Informatics (INDIN)*, Warwick, United Kingdom. P. 63–69, DOI: 10.1109/INDIN45582.2020.9442198
3. He, L., Jiang, P. (2020), "P-SaaS: knowledge service-oriented manufacturing workflow model for knowledge collaboration and reuse," *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, Hong Kong, China. P. 570–575. DOI: 10.1109/CASE48305.2020.9216974
4. Sapra, D., Pimentel, A. (2020), "Deep Learning Model Reuse and Composition in Knowledge Centric Networking," *29th International Conference on Computer Communications and Networks (ICCCN)*, Honolulu, HI, USA. P. 1–11. DOI: 10.1109/ICCCN49398.2020.9209668
5. Sharonova, N. et al. "Issues of Fact-based Information Analysis". *International Conference on Computational Linguistics and Intelligent Systems*, 2018. available at: <https://www.semanticscholar.org/paper/Issues-of-Fact-based-Information-Analysis-Sharonova-Doroshenko/f923b77b8561736202388db853e51df9bb7b9301>
6. Adjandra, W., Putrapratama, Y., Wiraguna, A., Sensuse, D. and Safitri, N. (2021), "Systematic Literature Review Knowledge Reuse in Software Development," *International Conference on Computer Science and Engineering (IC2SE)*, Padang, Indonesia. P. 1–7. DOI: 10.1109/IC2SE52832.2021.9792093
7. Putrapratama, Y., Adjandra, W., Wiraguna, A., Sensuse, D. and Safitri, N. (2021), "Knowledge Reuse Evaluation in Software Development: A Case Study on a Startup Company", *Sixth International Conference on Informatics and Computing (ICIC)*, Jakarta, Indonesia. P. 1–7. DOI: 10.1109/ICIC54025.2021.9632904
8. Shubin, I. "Development of conjunctive decomposition tools". *CEUR Workshop Proceedings*, 2021, Vol. 2870, P. 890–900 available at: <https://ceur-ws.org/Vol-2870/>
9. Sarker, K., Deraman, A. and Hasan, R. (2018), "Descriptive Logic for Software Engineering Ontology: Aspect Software Quality Control", *4th International Conference on Computer and Information Sciences (ICCOINS)*, Kuala Lumpur, Malaysia. P. 1–5. DOI: 10.1109/ICCOINS.2018.8510585
10. Kamide, N. (2020) "Sequential Fuzzy Description Logic: Reasoning for Fuzzy Knowledge Bases with Sequential Information", *IEEE 50th International Symposium on Multiple-Valued Logic (ISMVL)*, Miyazaki, Japan. P. 218–223. DOI: 10.1109/ISMVL49045.2020.000-2
11. Rawal, R., Goel, K., Gupta, C. (2020), "COVID-19: Disease Pattern Study based on Semantic-Web Approach using Description Logic", *IEEE International Conference for Innovation in Technology (INOCON)*, Bangluru, India. P. 1–5. DOI: 10.1109/INOCON50539.2020.9298278
12. Shubin, I., Snisar S., Litvin, S. (2021), "Formalization and Application of Algebraic Methods in Automated Intelligent Systems", *IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*, Kharkiv, Ukraine. P. 67–70. DOI: 10.1109/PICST54195.2021.9772174
13. Karataiev, O., Shubin, I. (2023), "Reuse of Information Based on the Interpretation of Knowledge". ["Problemy povtornoho vykorystannia znan pry proiektuvanni prohrannykh system"]. *Technologies and Scientific Solutions for Industries*. No. 2(24). P. 25–34. DOI: 10.30837/ITSSI.2023.24.025
14. 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>

15. Melnik, K., Cherednichenko, O., Glushko, V. (2013), "Towards medical screening information technology: the healthgrid-based approach". *Information Systems: Methods, Models, and Applications*. Heidelberg: Springer. P. 202-204. DOI: 10.1007/978-3-642-38370-0_21
16. Kandil, H. et al. (2020), "Analysis of The Importance of Systolic Blood Pressure Versus Diastolic Blood Pressure in Diagnosing Hypertension: MRA Study", *IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates. P. 443–447. DOI: 10.1109/ICIP40778.2020.9190990
17. Zhang, M., Zhang, T. and Cheng, Y. (2021), "Intelligent Method of Non-Invasive Detection and Grading of Hypertension Using Fingertip Photoplethysmography". *3rd International Conference on Applied Machine Learning (ICAML)*, Changsha, China. P. 418–422. DOI: 10.1109/ICAML54311.2021.00094

Надійшла 15.09.2023

Відомості про авторів / About the Authors

Каратаєв Олександр Анатолійович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна; e-mail: tosanik@gmail.com; ORCID ID: <https://orcid.org/0009-0007-6654-1327>

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>

Karataiev Oleksandr – Kharkiv National University of Radio Electronics, Postgraduate, Kharkiv, Ukraine.

Sytnikov Dmytro – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Professor at the Department of System Engineering, Kharkiv, Ukraine.

THE METHOD OF REUSE OF KNOWLEDGE IN THE FORM OF LOGICAL EQUATIONS

The subject of research is the processes of building and replenishing knowledge bases, the processes of knowledge reuse and the creation of software systems based on knowledge bases, the interpretation of knowledge is one of the approaches to their repeated application, which consists in deriving new knowledge based on the available facts in the knowledge base. The **purpose** of the research is to develop a method of knowledge reuse by solving logical equations of finite predicates for a certain subject area. The **tasks** were performed: to investigate descriptive approaches to logical modeling of the subject area, which make it possible to reuse the knowledge given by the system of logical equations within the algebra of finite predicates, to develop a method of replenishing the knowledge base in the form of predicate equations by extracting variables from logical systems, where equations represent complex logical connections between discrete features of objects or processes. The **methods** were used to solve the listed problems: the algebra of finite predicates, quantifier operations with predicates for the interpretation of knowledge. The **results** were obtained: descriptive methods of knowledge reuse were considered; a method of replenishing the knowledge base in the form of logical equations is proposed simplify the further use of implicit connections between discrete features; an example of replenishment of the knowledge base of the medical field is considered, which makes it possible to reuse knowledge given implicitly. **Conclusions:** the proposed method makes it possible to replenish the knowledge base in the form of logic equations by adding predicate equations connecting separate sets of discrete variables that are of interest to a scientist or a practitioner. The method of replenishing the knowledge base is based on adding new simplified equations; simplified predicate equations model logical patterns implicitly contained in the main knowledge base; an experimental study was conducted.

Keywords: software engineering; knowledge bases; reuse of knowledge; algebra of finite predicates; logical equations.

Бібліографічні описи / Bibliographic descriptions

Каратаєв О. А., Ситніков Д. Е. Метод повторного використання знань у формі логічних рівнянь. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 15–26. DOI: <https://doi.org/10.30837/ITSSI.2023.25.015>

Karataiev, O., Sytnikov, D. (2023), "The method of reuse of knowledge in the form of logical equations", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 15–26. DOI: <https://doi.org/10.30837/ITSSI.2023.25.015>

А. КОЗИРСЬ, І. ШУБІН

МЕТОД ПЛАНУВАННЯ ЗАВДАНЬ ОБРОБЛЕННЯ ДАНИХ У РОЗПОДІЛЕНИХ СИСТЕМАХ З ОБМЕЖЕНОЮ ІНФОРМАЦІЄЮ ПРО ДОСТУПНІ РЕСУРСИ

У сучасному цифровому ландшафті розподілені системи оброблення даних (PCOD) стають усе більш критично важливими для забезпечення ефективного оброблення, аналізу та управління великими обсягами інформації. Ці системи часто використовуються в комерційних, наукових і соціальних доменах для оброблення комплексних даних в реальному часі або пакетному режимі. Одним із ключових складників таких систем є планування завдань, що є надзвичайно складним процесом, зокрема коли інформація про ресурсні потреби не є повною або точною. **Предметом дослідження** є алгоритми, методи та підходи, які використовуються для планування завдань між вузлами в розподілених системах. **Мета роботи** – створення оптимізованого методу планування завдань у PCOD з обмеженою інформацією про доступні ресурси. **Завдання дослідження:** проаналізувати недоліки сучасних методів для планування завдань у розподілених системах оброблення даних; оптимізувати метод планування завдань на основі метаданих між вузлами PCOD, який ґрунтується на методології пошуку найближчих сусідів із використанням методу локалізованого хешування та алгебри скінченних предикатів; розробити архітектуру програмного рішення та його реалізацію на основі оптимізованого методу; протестувати алгоритм на прикладі завдання декодування відео. Застосовано такі **методи:** статистичні алгоритми та техніки, зокрема класифікація та кластерний аналіз, використані для прогнозування потреб у ресурсах; візуалізаційні методи допомогли в аналізі та інтерпретації результатів. **Результати роботи:** проаналізовано недоліки сучасних методів для розподілу завдань у розподілених системах оброблення даних; створено оптимізований метод планування завдань на основі метаданих у PCOD, який ґрунтується на методології пошуку найближчих сусідів із використанням методу локалізованого хешування та алгебри скінченних предикатів; деталізовано процеси в модифікованому алгоритмі пошуку найближчих сусідів; розроблено архітектуру програмного рішення, що інтегрує оптимізований метод планування завдань на основі метаданих та алокації ресурсів; за допомогою практичного сценарію здійснено валідацію програмного рішення – використання створеного алгоритму в задачі планування для декодування відеоінформації. **Висновки.** Запропонований метод, що ґрунтується на методології локалізованого хешування та на застосуванні алгебри скінченних предикатів, є ефективним навіть у разі недостатньої або обмеженої інформації про ресурсні потреби. Це підтверджує можливість використання динамічних стратегій планування для адаптації до мінливих умов навантаження та доступності ресурсів.

Ключові слова: програмна інженерія; розподілені системи; оброблення даних; база даних; планування завдань; алгебра скінченних предикатів.

Вступ

У сучасній цифровій екосистемі величезні масиви даних є результатом активності низки джерел, зокрема IoT-пристроїв, систем машинного навчання, соціальних медіа, платформ електронної комерції та інших технологічних засобів. Відповідно до цього стрімкого приросту інформаційних обсягів, сучасні методи зберігання та оброблення даних мають утілювати найновітніші технологічні розробки. У цьому контексті методи планування завдань для оброблення даних у розподілених системах оброблення даних (далі PCOD) посідають особливо важливе місце.

Розподілені системи, що можуть функціонувати на кластерах серверів або в хмарних обчислювальних середовищах, є особливим викликом з погляду

планування завдань. Оскільки дані та обчислювальні ресурси розподілені між різними фізичними або віртуальними вузлами, необхідно розробляти алгоритми, що можуть динамічно виділяти завдання, мінімізувати час очікування й оптимізувати використання ресурсів.

Планування завдань у розподілених системах передбачає розподіл обчислювальних задач між мережею взаємопов'язаних вузлів, часто з метою оптимізації продуктивності, використання ресурсів і пропускної здатності системи. У цьому контексті "завдання" означає обчислювальну задачу, яка може варіюватися від запитів до бази даних до алгоритмів машинного навчання, тоді як "розподілена система оброблення даних" охоплює мережну архітектуру взаємопов'язаних програмних засобів та баз даних, що можуть обробляти,

зберігати й отримувати дані в децентралізований спосіб.

Незважаючи на те, що за умови взаємодії з окремими обчислювальними вузлами в межах розподілених систем можуть бути прийняті локально оптимальні рішення, це не гарантує глобальної оптимальності відображення цілої множини завдань на множину доступних вузлів. Тобто, якщо кожен вузол незалежно вирішує, яке завдання він виконає наступним, можливо, відсутність координації призведе до неефективного розподілу ресурсів на більш високому, системному рівні.

Такий підхід, імовірно, спричинить низку проблем, зокрема затримки в обробленні, незбалансоване використання ресурсів та взаємні блокування завдань. Відтак, щоб досягти оптимальної ефективності, процедури планування завдань та алокації вузлів мають бути інтегровані в одному процесі планування. Це означає, що замість того, щоб розглядати кожне завдання або вузол ізольовано, система має зважати на глобальний стан, а саме: на доступність ресурсів, поточну завантаженість вузлів та залежності між завданнями – для прийняття більш інформованих та оптимізованих рішень.

Метою дослідження є розроблення методу планування завдань у розподілених системах оброблення даних в умовах неповноти інформації про ресурсні вимоги.

Аналіз останніх досліджень і публікацій

Упродовж останнього десятиліття спостерігається відчутне зростання потреби в розподілених обчислювальних ресурсах у різноманітних сферах застосування [1]. Серед таких завдань можна виокремити електронну комерцію, фінансову аналітику, аналіз даних у соціальних мережах, а також оброблення та дистрибуцію мультимедійного контенту [2].

У широкому використанні зараз перебувають системи паралельного та розподіленого оброблення даних, зокрема обчислювальні кластери [3], ґрід-системи [4] та хмарні платформи [5, 6]. Використання перелічених систем у наукових дослідженнях та для вирішення конкретних завдань призводить до подальшого збільшення навантаження на апаратні ресурси. Це зумовлено постійним зростанням складності завдань, які вимагають все більше обчислювальної та просторової потужності. Необхідно зазначити, що ці завдання

часто визначаються нерівномірними вимогами до обчислювальних ресурсів.

Завдання є сутністю, яка надходить на вхід системи планування й містить набір завдань [7]. Потік завдань надходить у вхідну чергу, завдання є атомарною одиницею планування, й у межах одного завдання вони можуть бути незалежними або організованими в дерево залежностей [8]. Відповідно до цих параметрів підзавдання обробляються або паралельно в разі відсутності залежностей, або в заданому порядку, якщо такі залежності існують. Залежності слугують керівними принципами для послідовності виконання та перенесення даних між підзавданнями.

Низка наукових досліджень [5, 9–11] вказує на те, що сучасні завдання у сфері оброблення даних демонструють значну динамічність. З одного боку, можна спостерігати численні "легкі" завдання, що визначаються мінімальними вимогами до ресурсів, наприклад, замовлення на оброблення невеликих текстових файлів або зображень. З іншого боку, існують "важкі" завдання, такі як оброблення великих обсягів даних або виконання складних математичних розрахунків, що потребують значно більших обчислювальних можливостей.

У такому різноманітному обчислювальному середовищі інтелектуальне планування розподілу завдань між доступними обчислювальними вузлами набуває критичної важливості. Це питання актуальне не тільки для забезпечення ефективності роботи конкретного вузла, але й для оптимального використання всієї обчислювальної інфраструктури загалом.

Методи планування можуть бути категоризовані за різними критеріями. Наприклад, існують методи, що ґрунтуються на апіорному знанні про ресурсні потреби кожного завдання. Тут можуть використовуватися алгоритми оптимізації, які попередньо аналізують завдання та призначають їх вузлам на основі детального ресурсного профілю [12]. З іншого боку, є підходи, де завдання розподіляються динамічно, без попереднього аналізу [13]. У такому разі може застосовуватися загальний принцип оброблення, де завдання просто надходять у чергу й обслуговуються за принципом "першим прийшов – першим обслуговано" [14], або за деякими іншими евристичними правилами.

Основна дилема в плануванні завдань полягає в розподілі обчислювальних ресурсів та встановленні порядку, за яким завдання з вхідної черги будуть

реалізовані на цих ресурсах [7]. Системи, де ресурси керуються в централізованій спосіб, відомі як системи управління ресурсами, і ці системи зазвичай тісно взаємодіють з механізмами планування завдань.

Комбінація системи управління ресурсами та планувальника завдань формує проміжний програмний модуль. З одного боку, цей модуль відповідає за управління ресурсами на базовому рівні, а з іншого – адаптується до високорівневих вимог прикладних завдань. Для досягнення цієї мети система постійно аналізує актуальний стан використання ресурсів та далі ділить їх між заявленими завданнями [15].

У системі управління ресурсами часто інтегрований інформаційний сервіс, який відповідальний за моніторинг стану завдань і ресурсів. В окремих ситуаціях цю роль виконує зовнішня система моніторингу. Планувальник оперує на основі даних, які надає цей інформаційний сервіс, і використовує конкретні методики для управління завданнями та ресурсами.

Політика планування завдань (*task scheduling policy*) – це комплекс правил, за якими визначається момент та спосіб вибору нового завдання для оброблення [16]. Основний механізм вибору завдань ґрунтується на процедурі відбору завдань щодо їх пріоритетів.

Зазвичай для планування завдань використовується одна з дисциплін обслуговування. Ці дисципліни можуть або базуватися на детальних ресурсних вимогах завдань, або оперувати без попереднього знання, тобто обробляти всі завдання універсальним методом.

Визначення не вирішених раніше частин загальної проблеми

Недоліки методів розподілення запланованих завдань без попереднього знання полягають у тому, що вони не адаптуються до особливих ресурсних потреб, зосереджуються на виборі конкретних етапів для оброблення на основі інших критеріїв, припускають однорідність і використовують однотипні обчислювальні вузли. Це може призвести до неефективного використання ресурсів, особливо коли вузли та завдання мають різні характеристики або коли один вузол може паралельно обробляти декілька завдань.

Використання цих методів планування ресурсів часто не є оптимальним у таких сценаріях:

- коли обчислювальні вузли в розподілених системах оброблення даних мають різну потужність, що вказує на гетерогенність РСОД;
- коли завдання відрізняються за обчислювальною та просторовою (за пам'яттю) складністю, що вимагає різних ресурсних потреб;
- коли вузли РСОД можуть обробляти декілька завдань паралельно.

Додаткова характеристика згаданих методів планування завдань полягає в можливості зберігання стану виконуваних завдань. Це дає змогу призупинити поточне завдання й перейти до іншого, що може бути більш пріоритетним. Методи планування, зокрема *Time Sharing*, *Least Attained Service* та *Shortest Remaining Processing Time*, належать до цього класу алгоритмів, завдяки чому на одному обчислювальному вузлі може одночасно виконуватися декілька завдань на різних етапах оброблення.

Збереження стану завдань є критичним компонентом у системах планування процесів, що активно використовуються в сучасних операційних системах. Ця методика дає змогу декільком процесам поділити доступ до обмеженого ресурсу процесорного часу, симулюючи цим паралельне виконання завдань. Поділ процесорного часу здійснюється способом присвоєння кожному процесу короткого кванта часу, після якого процес може бути призупинений, його стан збережений, ресурс переданий іншому процесу.

Однак цей підхід має декілька недоліків. По-перше, він передбачає додаткові оперативні витрати, наприклад, на збереження та відновлення стану процесу, а також на переключення контексту між різними процесами. Ці дії можуть стати затратними з погляду обчислювальної ефективності, особливо в системах із значною кількістю активних процесів. По-друге, для завдань, що вимагають високої обчислювальної інтенсивності та синхронізації, переривання для збереження стану може призвести до суттєвого падіння продуктивності. У таких ситуаціях необхідність у перериванні та збереженні стану може знизити загальну пропускну здатність системи та збільшити латентність виконання завдань.

Далі в процесі планування для кожного обраного завдання застосовується політика розміщення на вузлах. Ця політика обирає оптимальний комплекс ресурсів для конкретного завдання оброблення

даних. У системі ресурсного управління беруться до уваги характеристики кожного вузла або ресурсу в розподіленій інфраструктурі, а також задовольняються попередньо встановлені критерії від застосунків.

Методики призначення ресурсів на завдання можуть відрізнятися, зокрема, залежно від наявності попередньої інформації. Це може бути вибір на менш завантаженому або недавно використаному вузлі, випадковий вибір вузла, вибір на основі метрик черги, вибір вузла з найкоротшим прогнозованим часом виконання завдання або вибір вузла, що найкраще відповідає прогнозуванню споживання ресурсів.

Однак основна проблема наявних методів планування полягає в тому, що вибір завдання та вибір вузла для його виконання розглядаються як два незалежних етапи. Спершу визначається завдання з найвищим пріоритетом, а потім для нього підбирається найбільш підходящий вузол.

На основі проведеного огляду можна визначити необхідність створення інноваційного методу планування завдань у розподілених системах оброблення даних, особливо в разі обмеженої інформації про потреби в ресурсах. Це має потенціал зменшити часові затрати на реалізацію конкретних завдань оброблення інформації.

Для досягнення цієї мети необхідно вивчити взаємозв'язок між атрибутами (метаданими) завдань та метричними показниками використання обчислювальних ресурсів. У дослідженні увага зосереджена на методах планування непов'язаних завдань, що не передбачають підтримки збереження стану виконання, та розробляються в контексті обмеженої доступності інформації про ресурсні потреби завдань.

Матеріали й методи

Стратегія планування завдань, що ґрунтується на метаданих і метричних показниках ресурсів, передбачає пошук найближчих сусідів для даних, що підлягають обробленню. Оскільки ці дані мають як чисельні, так і категоріальні атрибути, традиційні метрики відстані, як-от евклідова або хемінгова відстань, не можна застосувати безпосередньо. Замість цього використовуються гетерогенні функції відстані, що дають змогу порівняти такі різномірні об'єкти даних. Відстані для чисельних і категоріальних атрибутів обчислюються окремо. Після цього здійснюється нормалізація та зважування

отриманих значень для мінімізації впливу шуму та розбіжностей у шкалах.

Основний недолік такого типу пошуку полягає в його великій розмірності. Для кожного заданого об'єкта (завдання) потрібно визначити відстань до всіх n об'єктів з історії виконань, що призводить до лінійної асимптотичної складності одного запиту, або $O(n)$.

Під час планування набору завдань потрібно виміряти ресурсну складність для всіх можливих пар (вузол-процесор, завдання). Це призводить до загальної асимптотичної складності, що формулюється як

$$O(h-v-n) = h-v-O(n) = O(n), \quad (1)$$

де h – кількість доступних для планувальника вузлів-оброблювачів, v – кількість завдань, що очікують розроблення.

Хоча теоретична складність є лінійною, у практичному застосуванні константи h і v також суттєво впливають на обчислювальну вартість процесу планування. Наприклад, у разі наявності 33 обчислювальних вузлів і мінімум 33 завдань у черзі очікування загальна кількість запитів становитиме 1089. Кожний із цих запитів має лінійну складність, що корелює з розміром історичних даних.

Зменшення розмірності в завданні пошуку найближчих сусідів (*Nearest Neighbor Search, NNS*) є критичною проблемою, особливо у великих датасетах із високою розмірністю. Алгебра скінченних предикатів, що є математичною теорією, може бути використана для вирішення цієї проблеми. Сутність полягає у формалізації атрибутів або властивостей елементів набору даних за допомогою логічних виразів або предикатів.

Одним із способів реалізації може бути застосування алгебри скінченних предикатів. Основна ідея полягає в тому, щоб використовувати логічні оператори для розбиття високовимірного простору на підпростори, що спрощує пошук.

Візьмемо декартовий простір \mathbb{R}^d , де d – розмірність простору. Цільова функція відстані між двома точками $x, y \in \mathbb{R}^d$ може бути евклідовою відстанню:

$$D(x, y) = \sqrt{\sum_{i=1}^d (x_i - y_i)^2}. \quad (2)$$

За допомогою алгебри скінченних предикатів можна створити систему логічних умов, які розділять простір на менші підпростори [17]. Наприклад,

можна ввести предикати P_1, P_2, \dots, P_n , що розділять простір на n підпросторів:

$$P_i(x) : \mathbb{R}^d \rightarrow \{0, 1\}. \quad (3)$$

Тепер, замість того, щоб шукати найближчого сусіда в усьому просторі \mathbb{R}^d , можна обмежитися пошуком у підпросторах, у яких є точка запиту, тобто тільки в тих підпросторах, для яких $P_i(x) = 1$.

Застосування алгебри скінченних предикатів уможливить створення більш гнучких та оптимізованих алгоритмів планування. Ці математичні інструменти можуть бути використані для моделювання обмежень і вимог до системи, які традиційно важко виразити алгебраїчно.

Розглянемо систему з H обчислювальними вузлами та V завданнями, що потрібно запланувати. Для кожного вузла h та завдання v можна ввести предикат $P_{h,v}$, який оцінює, чи може завдання v бути ефективно виконаним на вузлі h :

$$P_{h,v}(x) : \{h, v\} \rightarrow \{0, 1\}. \quad (4)$$

Щоб зменшити розмірність задачі знаходження k найближчих сусідів для об'єктів із різноманітними характеристиками, рекомендовано використовувати алгоритм, оснований на методі локалізованого хешування (*Locality-Sensitive Hashing, LSH*). Такий підхід виявляється масштабованим і дає змогу виконати деяку частину пошуку в константний час $O(1)$.

Задача пошуку найближчих сусідів у високовимірних просторах часто виявляється вимогливою щодо обчислювальних ресурсів. Один із способів оптимізації цієї задачі полягає у використанні алгоритмів локалізованого хешування, які намагаються згрупувати близькі точки в одному або декількох хеш-бакетах, забезпечуючи швидкий пошук.

Алгебра скінченних предикатів може бути інтегрована в алгоритм *LSH* для додаткової оптимізації. Предикати можуть бути використані для вибору оптимальних функцій хешування або для динамічної адаптації алгоритму до структури даних.

$$P_1(x_0, x_1) \wedge \dots \wedge P_n(x_{n-1}, x_n) \wedge a_1(x_{i_1}) \wedge \dots \wedge a_m(x_{i_m}) \rightarrow p(x_0, x_n). \quad (7)$$

Атрибути часто корелюють із значною кількістю об'єктів, формуючи цим особливі "концентратори" взаємодії. Ідентифікацію таких концентраторів рекомендується проводити на основі метрики

У межах алгебри скінченних предикатів можна додати ряд предикатів P_1, P_2, \dots, P_n , що слугують критеріями для хешування:

$$P_i(x, y) : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \{0, 1\}. \quad (5)$$

Ці предикати можуть вимірювати різні аспекти подібності між точками x та y . Застосування предикатів до пар точок може бути використано для динамічного вибору функцій хешування в алгоритмі *LSH*. Наприклад, для пар точок, для яких $P_i(x, y) = 1$, можна застосовувати функцію хешування h_i , що є оптимальною для цього типу подібності.

$$h(x) = \begin{cases} h_1(x), & \text{if } P_1(x, y) = 1 \\ h_2(x), & \text{if } P_2(x, y) = 1 \\ \vdots \\ h_k(x), & \text{if } P_k(x, y) = 1. \end{cases} \quad (6)$$

Цей підхід забезпечує більшу гнучкість у виборі функцій хешування та, відповідно, може підвищити ефективність пошуку найближчих сусідів у розподілених системах.

Ключовим аспектом методу локалізованого хешування є обережний вибір атрибутів даних, які будуть використовуватися для розрахунку хеш-значення. Обрані атрибути визначають, який сегмент пошукового процесу може бути оптимізований до константного часу, а який залишиться з лінійною складністю. Ці атрибути мають відображати подібність між цими об'єктами, та їх вибір має ґрунтуватися на особливостях конкретного прикладного домену.

Варто зазначити, що атрибути зазвичай об'єднують обмежену кількість сутностей для подання характеристик, які корелюють із значною кількістю інших сутностей. У такий спосіб атрибути можуть слугувати як фокусні точки й природно лягати в основу для класифікації за принципом еквівалентності. Сучасні методи дослідження часто не роблять чіткого розмежування між атрибутами сутності та їх взаємозв'язками [18].

Сутність e пов'язана з іншою сутністю e' через скінченний двійковий предикат p :

щільності їх з'єднань [18]. Щільність концентратора можна визначити як

$$den_{in}(e, p) = \frac{\#e' : p(e', e)}{\#(e', e'') : p(e', e'')}. \quad (8)$$

На основі поданих методів у цій роботі запропонована модифікована версія алгоритму знаходження найближчих сусідів, що ґрунтується на методі локалізованого хешування з використанням алгебри скінченних предикатів. Ця модифікація дає змогу здійснити деяку частину пошуку в константний час $O(1)$, а іншу частину – в сублінійний час $O(t)$.

Алгоритм знаходження найближчих сусідів, оснований на методі локалізованого хешування, передбачає такі етапи:

Крок 1. Селекція атрибутів для хеш-функції, а саме:

- згрупувати атрибути на числові та категоріальні типи;
- оцінити значущість кожного атрибуту за допомогою навчального датасету;
- обрати n найбільш ресурс-інтенсивних категоріальних атрибутів та n числових атрибутів із найвищою значущістю.

Крок 2. Конструювання хеш-таблиці, основаної на тренувальному датасеті:

- розрахувати хеш-значення для вибраних категоріальних та числових атрибутів;

- використовувати згенерований агрегований хеш як індекс для розміщення об'єктів у відповідному "кошику" хеш-таблиці;

- зберегти унікальний ідентифікатор об'єкта в хеш-таблиці за відповідною адресою.

Крок 3. Пошук найближчих сусідів для нової данини:

- обчислити хеш-значення для нового об'єкта на основі атрибутів, обраних на першому кроці;
- використати згенерований агрегований хеш для доступу до відповідного "кошика" в хеш-таблиці;
- виміряти відстань до всіх об'єктів, чії ідентифікатори зберігаються в цьому "кошику";
- упорядкувати згенерований набір відстаней;
- обрати k об'єктів із найменшою вимірною відстанню.

Така модифікація алгоритму спрямована на оптимізацію пошуку найближчих сусідів і водночас забезпечує більш ефективне використання ресурсів і швидкість пошуку.

На рис. 1 зображено фінальний етап роботи алгоритму, а саме методику знаходження найближчих сусідів, основану на вибраних атрибутах.



Рис. 1. Процес пошуку найближчих сусідів

Запропонована програмна архітектура застосовує зазначений метод, що передбачає детальний аналіз використовуваних атрибутів даних, показників застосування обчислювальних ресурсів, а також методики оцінювання ресурсних витрат на основі попередніх досліджень.

Планувальник завдань і ресурсів відповідає за організацію оброблення даних і розподіл обчислювальних ресурсів, інтегруючи модулі для оцінювання ресурсних потреб і розподілу ресурсів. Модуль оцінювання ресурсних потреб аналізує історичні експериментальні дані та реалізує

функціонал `fust.2` з пропонованого методу планування. Це передбачає обчислення коефіцієнтів ресурсних витрат і створення набору кортежів вигляду $\{t, k, c\} \in Ghist$ на основі отриманих метрик `Mhist`. Цей модуль також слугує для оцінювання ресурсних потреб нових завдань із використанням попередньо побудованої моделі для знаходження схожих завдань `Thist.nearest`, чії ресурсні потреби усереднюються для нового завдання. У підсумку модуль генерує матрицю ресурсних потреб S .

Модуль розподілу ресурсів приймає цю матрицю ресурсних потреб C від модуля оцінювання та вирішує проблеми розподілу, шукаючи найефективніший спосіб розподілу наявних завдань на доступні обчислювальні вузли. Результатом є матриця розподілу ресурсів W .

Система управління ресурсами та модуль для збирання ресурсних метрик функціонують як додаткові компоненти. Система управління ресурсами отримує матрицю розподілу W від планувальника завдань, а потім видає інструкції для запуску конкретних завдань на відповідних обчислювальних вузлах. Додатково ця система має функціонал для сповіщення про завершення раніше ініціалізованих завдань і локацію, де зберігаються результати оброблення даних.

Модуль збирання метрик надає дані про використання системних ресурсів, таких як пам'ять, процесор, графічний процесор, системи введення-виведення тощо, на обчислювальних вузлах. Цей модуль також видає інформацію про часові характеристики виконання завдань і дає змогу отримувати деталі конфігурації кожного обчислювального вузла в системі. Він відіграє роль інформаційного сервісу в контексті сучасних систем управління ресурсами.

Модуль оброблення метаданих відповідає за операції з витягу, перетворення та збереження метаданих.

У межах засобу планування завдань існує об'єкт класу *metadata mapper*, який здійснює конвертацію метаданих до формату, що підходить для виконання пошуку найближчих сусідів згідно з рекомендованим алгоритмом. Також тут розміщується об'єкт класу *ModelBuilder*, який відповідає за побудову моделі прогнозування.

Для експериментального вивчення методів планування завдань додано клас *Task Generator*, що може створювати набори завдань із різними режимами інтенсивності: низький, середній та високий. Він також здатний моделювати різні рівні трудомісткості завдань за допомогою чотирьох типів розподілу: константний, рівномірний, випадковий та зважений дискретний.

Як приклад застосування модифікованого методу планування завдань розглянемо завдання декодування відео для мобільних пристроїв. Традиційні відеохостинги використовують заздалегідь оброблені профілі для декодування відео, але *Amazon Elastic*

Transcoder дозволяє динамічну оренду ресурсів для цієї мети. Це забезпечує баланс між користувацькими потребами та оптимізацією ресурсів. Застосовується метод планування на основі метаданих і метрик для ефективного розподілу завдань.

У розподіленому виконавчому середовищі функціонує 33 обчислювальних вузли, конфігурація яких описана в табл. 1.

Таблиця 1. Конфігурація PCOD

Тип вузла	Одиниць	Cores	Sockets	RAM
transcoding.node.a	7	1	1	1024
transcoding.node.b	8	2	1	1024
transcoding.node.c	8	2	2	1024
transcoding.node.d	5	3	2	1024
transcoding.node.e	5	4	2	1024

Для різних груп метаданих існують окремі життєві цикли, які керують їх збиранням, перетворенням, збереженням і використанням. Властивості відеопотоку визначаються наперед, коли завантажуються нові вихідні відеофайли, тоді як властивості користувацького обладнання та мережі стають відомими тільки в разі надходження завдань. Метадані відео збираються за допомогою утиліти *MediaInfo* та зберігаються в централізованій базі даних. Щодо обладнання та мереж, то система планується з урахуванням їх типових характеристик, для яких задаються декодувальні профілі на основі утиліти *Handbrake*.

Для цілей експерименту завдання щодо оброблення даних генеруються програмним методом за допомогою генератора завдань і далі надсилаються до координатного вузла (обидва компоненти розміщені на одній фізичній машині). Тут відбувається розподіл завдань на оброблювальні вузли. Генератор завдань випускає набір кортежів, що складаються з типу пристрою, типу мережного з'єднання та ідентифікатора відеофайлу, з фіксованим часовим інтервалом T .

Після надходження таких кортежів у систему відбувається збір метаданих, їх перетворення та безпосереднє оброблення інформації. Запит для оброблення ініціюється користувачем через вибір відеофайлу за допомогою команди *show_video(id)*. Після цього програмний клієнт доповнює запит метаданими, що містять інформацію про тип обладнання та характеристики мережного з'єднання через функцію *add_metadata()*. Зрештою відеодекодувальний сервіс отримує завдання

у вигляді кортежа, що містить тип пристрою, тип мережного з'єднання та ідентифікатор відеофайлу.

Після отримання завдання сервіс ініціює процес відображення вхідних даних у відповідний набір метаданих, які керують операцією декодування, через функцію *map_metadata()*. Цей процес відбувається в декілька стадій. Насамперед за ідентифікатором відеофайлу завантажуються відповідні метадані вихідного відеопотоку з бази даних. Далі за ідентифікатором типу обладнання з бази даних завантажуються профіль декодування, що визначає характеристики вихідного відеопотоку.

Також на основі ідентифікатора типу мережного з'єднання з бази даних завантажуються окремі профіль декодування, який ставить обмеження на якість вихідного відеопотоку з огляду на характеристики мережного з'єднання. Якість декодування, зокрема відеобітрейт, визначається відповідно до трьох параметрів: значення властивості у вихідному відеофайлі (*psrc*), характеристика, специфічна для певного типу обладнання (*Pdevice*), і верхнє обмеження, основане на швидкості мережного з'єднання (*Pconnection*).

Отже, складність операції декодування формується на підставі характеристик вихідного відеофайлу, апаратних можливостей типу обладнання для відтворення та пропускну здатності мережного з'єднання. Наприклад, незалежно від того, наскільки потужне обладнання, недоречно надавати відеопотік високої якості в разі обмеженої пропускну здатності мережі, оскільки це може викликати нестабільність у передачі даних, зокрема джитер (*jitter*).

Після конвертації метаданих сервіс інтегрує нове завдання в чергу планувальника за допомогою функції *add_task(data)*. Планувальник, зазвичай відомий як *Scheduler*, забезпечує динамічне розміщення завдань на вузлах-оброблювачах завдяки методу *assign_task(task_data)*. Взамін вузол-оброблювач декодує відеодані, а URL для доступу до вихідного відеопотоку надсилається на клієнтське обладнання.

Найбільш звичайний підхід до блокування завдань між вузлами-оброблювачами ґрунтується

на використанні пріоритетної черги *FIFO (First In, First Out)* та алгоритму виділення завдань на найменш завантажений вузол, зазвичай позначеного як *LLF (Least Loaded First)*.

Цей метод, відомий як *FIFO_LLF*, був обраний як базовий для аналітичного порівняння переважно тому, що він, так само як і модифікований метод планування на основі метаданих (*Meta_Sched*), не вимагає попереднього знання про ресурсні потреби. Додатково цей підхід використовується для акумуляції емпіричних даних, що пізніше можна використати для оптимізації методу *Meta_Sched*.

Метод *Meta_Sched* вимагає розроблення моделі, що здатна визначати обчислювальні витрати для нових завдань на основі раніше зібраної статистичної інформації про виконані завдання. У процесі виконання різноманітних завдань, крім отримання результатів декодування, система також збирає метрики, пов'язані з використанням обчислювальних ресурсів, і зберігає їх в базі даних. До кожного запису також додаються метадані про саме завдання й характеристики обчислювального вузла.

Специфічні метрики й параметри обчислювальних вузлів, що збираються, містять середнє завантаження процесорів вузла під час декодування (*avg_cpu_load_percent*), кількість оперативної пам'яті, використаної процесом декодування (*ram_load_kb*), та низку часових штампів, зокрема момент уведення завдання в систему (*submit_timestamp*) та початок декодування (*start_processing_timestamp*). Додатково збирається інформація про час, потрібний для оброблення даних (*elapsed_seconds*), доступний обсяг оперативної пам'яті на вузлі (*node_free_ram*) і кількість вільних процесорів на вузлі (*node_free_processors*).

Ці показники дають змогу оцінювати не тільки обчислювальні й часові затрати, пов'язані з обробленням інформації, але й затрати часу на організаційні аспекти завдань, а саме: період очікування та тривалість виконання.

Для кожного запису обчислюється нормалізована величина затрат за формулою:

$$\begin{aligned} cost = & W1 \times cpu_cost(avg_ram_load_kb, node_free_ram) + \\ & + W2 \times ram_cost(avg_cpu_load, node_free_processors), \end{aligned} \quad (9)$$

де *cpu_cost* – функція оцінювання процесорних затрат; *ram_cost* – функція оцінки затрат пам'яті;

W1, *W2* – вагові коефіцієнти значущості обчислювального ресурсу.

У формулі (9) вагові фактори налаштовуються так, щоб надати перевагу ключовим ресурсам залежно від типу завдання:

- для обчислювально інтенсивних завдань увага переноситься на процесор;
- для завдань, що потребують великих обсягів пам'яті, увага переноситься на пам'ять;
- для підсистеми введення-виведення;
- для графічної компоненти;
- для мережних операцій.

Задача декодування відео належить до категорії обчислювально інтенсивних, тому коефіцієнт важливості для процесора $W1$ дорівнює 1. За паралель, споживання пам'яті є мінімальним, із ваговим коефіцієнтом $W2$ рівним 0,3.

Функції cpu_cost і ram_cost аналізують вартість використання конкретних ресурсів, водночас беручи до уваги ступінь завантаженості або перевантаження певного ресурсу.

Для вимірювання відстаней до всіх потенційних об'єктів з отриманого набору поатрибутні відстані коригуються залежно від важливості кожного числового параметра. На основі оцінки використання ресурсів для всіх задач і оброблювальних вузлів формується матриця ресурсних затрат. Створена матриця ресурсних затрат застосовується для вирішення проблеми розподілу, що приводить до отримання матриці призначень, згідно з якою завдання оброблення даних будуть делеговані на відповідні вузли.

Для оцінювання ефективності методів планування завдань розподілена обчислювальна система тестується в різних обставинах. Сценарії тестування варіюються за розподілом складності вхідного потоку завдань та за рівнем завантаженості системи.

Результати досліджень та їх обговорення

Для тестування модифікованого методу планування завдань здійснено планування та оброблення завдань з декодування відеоінформації впродовж 10 хв. Після цього на основі агрегованих статистичних показників якості виконання методів планування $FIFO_LLF$ та $Meta_Sched$ оцінюється за допомогою таких метрик:

- рівень використання процесорних ресурсів обчислювального вузла;
- середній час очікування на оброблення завдання, у секундах;
- загальний час перебування завдання в системі, у секундах (час очікування та час оброблення);
- середній коефіцієнт затримки оброблення завдання, у секундах (час очікування поділено на час оброблення);
- флуктуація загального часу виконання завдань для різних сценаріїв;
- коливання середнього часу очікування на оброблення завдання в різних сценаріях.

Компоненти програмної системи зображені на діаграмі компонентів і діаграмі класів (рис. 2 та 3).

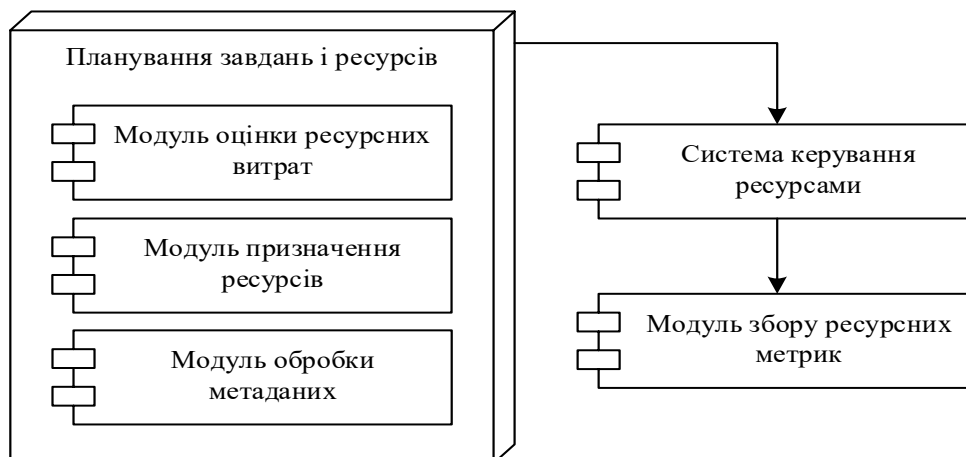


Рис. 2. Діаграма компонентів системи планування завдань і ресурсів РСОД

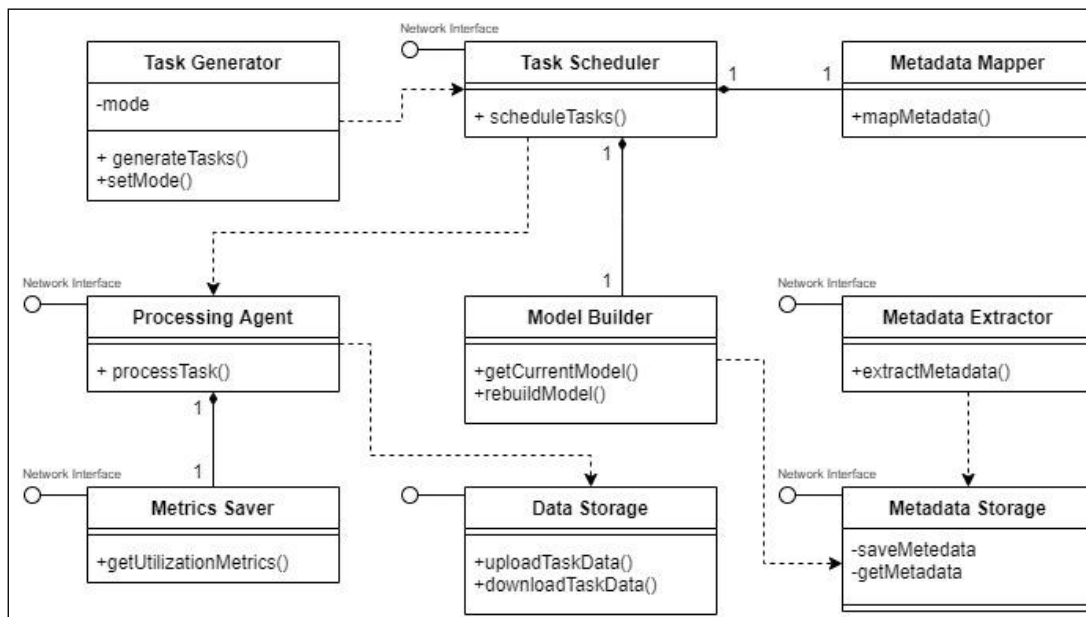


Рис. 3. Діаграма класів

У цьому контексті рівень використання процесорних ресурсів обчислювального вузла визначається з огляду на відсоток процесорного часу, який отримує паралельне завдання, а також зважаючи на кількість доступних процесорів.

За умови середнього навантаження метод *FIFO_LL*F незначно випереджає метод *Meta_Sched*. Це сталося тому, що *Meta_Sched* некоректно алокував деякі завдання, унаслідок чого ресурси стали перевантаженими. Крім того, в оптимальному діапазоні [0,8, 1] метод *Meta_Sched* незначно відстає від методу *FIFO_LL*F, якщо звернути увагу на кількість точок (70 проти 90).

Для адекватної інтерпретації отриманих показників важливо визначити ключові критерії, що оцінюють придатність розглянутого методу планування для реального впровадження. Основні оцінювальні параметри передбачають: аналіз часових показників оброблення завдань, зокрема середній час перебування заявки в системі, середня затримка й середній час оброблення. Другий критерій – це ефективність використання ресурсів. З погляду планування ідеальним вважається зайнятість ресурсу в діапазоні [0,8, 1]. Значення, нижчі від 0,8, свідчать про недостатнє використання ресурсів, тоді як значення понад 1 вказують на їх перевантаження.

За умови високого рівня завантаження метод планування, оснований на метаданих і ресурсних метриках, демонструє відчутну перевагу. Зокрема набір точок, який спочатку потрапив у діапазон [2, 6],

був зміщений до діапазону [0, 2], що свідчить про зниження перевантаження системи.

Тимчасові показники, зокрема середній час перебування заявки в системі, середня затримка й середній час очікування на оброблення, є величинами, що мають схожий масштаб для обох розглянутих методів планування.

Щодо ефективності використання обчислювальних можливостей, то це питання буде досліджено для кожного тестового сценарію. В умовах низької навантаженості метод *FIFO_LL*F демонструє вищу кількість точок у діапазоні [0, 0,8), що свідчить про недостатнє використання ресурсів. Також цей метод має завдання, для яких виділено недостатньо обчислювальної потужності (з використанням процесора понад 2). У діапазоні [0,8, 1] обидва методи планування показують подібні результати. Запропонована архітектура програмної системи для планування завдань у розподіленій обчислювальній системі втілює метод, оснований на метаданих і ресурсних метриках.

Аналітичний огляд результатів свідчить про таке: застосування методу планування, оснований на метаданих, призводить до зменшення загального часу перебування заявок у системі на 20 % за умови високої інтенсивності вступу завдань, однак цей показник збільшується на 4,5 % у разі середньої інтенсивності та на 10,7 %, якщо інтенсивність низька. Також середній час очікування на оброблення заявок скорочується на 7,58 %, коли висока

інтенсивність вступу завдань, але зростає на 10,55 % за умови середньої інтенсивності та на дуже значні 70,6 %, якщо інтенсивність низька.

Висновки

Розглянуто системи управління ресурсами та сучасні методи планування завдань у розподілених системах оброблення даних, що вказало на необхідність розроблення нових програмних рішень, які ґрунтуються на плануванні в умовах недостатньої інформації про ресурсні потреби.

Розглянуто недоліки методів розподілення запланованих завдань без попереднього знання про ресурсні потреби.

Створено метод планування завдань на основі метаданих, оснований на модифікованому алгоритмі пошуку найближчих сусідів через локалізоване

хешування з використанням алгебри скінченних предикатів, що бере до уваги типи атрибутів та їх важливість для ресурсопотреби, щоб спростити процес пошуку найближчих завдань.

З метою тестування зазначеного методу створено програмну архітектуру для системи планування завдань у РСОД. Визначено критерії для характеристик даних (метаданих) та метрик ресурсопотреби, взаємозв'язок між якими дає змогу створювати прогнози моделі для оцінювання потреб нових завдань, зважаючи на історію виконання попередніх завдань.

Реалізовано програмне рішення з використаного методу на прикладі практичного сценарію – у завданні планування декодування відеоінформації. Проведено серію експериментів, результати яких показали, що середній час виконання завдань зменшився приблизно на 20 % за умови високого рівня надходження запитів.

Список літератури

1. Endo, P., de Almeida Palhares, A., Pereira, N., Goncalves, G., Sadok, D., Kelner, J., Melander, B. & Mangs, J.E. Resource allocation for distributed cloud: concepts and research challenges. *IEEE Network*. 2011. Vol. 25, No. 4. P. 42–46. DOI: 10.1109/mnet.2011.5958007
2. Sunyaev A. Principles of Distributed Systems and Emerging Internet-Based Technologies. *Internet Computing*. Cham: Springer International Publishing, 2020. Vol. XVIII. 413 p. DOI: 10.1007/978-3-030-34957-8
3. Siva Prasad B. V., Sucharitha G., Venkatesan K. G., Patnala T. R., Murari T., Karanam S. R., Optimisation of the Execution Time Using Hadoop-Based Parallel Machine Learning on Computing Clusters. *Computer Networks, Big Data and IoT*. Singapore, 2022. P. 233–244. DOI: 10.1007/978-981-19-0898-9_18
4. Kamran M. Fundamentals of Smart Grid Systems. Elsevier Science & Technology Books, 2023. 500 p. URL: <https://shop.elsevier.com/books/fundamentals-of-smart-grid-systems/kamran/978-0-323-99560-3>
5. Hasimi L., Penzel D. A Case Study on Cloud Computing: Challenges, Opportunities, and Potentials. *Studies in Systems, Decision and Control*. Cham, 2023. P. 1–25. DOI: 10.1007/978-3-031-27506-7_1
6. Zolotariov D. Microservice Architecture for Building High-Availability Distributed Automated Computing System in A Cloud Infrastructure. *Сучасний стан наукових досліджень і технологій в промисловості*. 2021. No. 3 (17). P. 13–22. DOI: 10.30837/itssi.2021.17.013
7. Tom L., Bindu V. R. Task Scheduling Algorithms in Cloud Computing: A Survey. *Inventive Computation Technologies*. Cham, 2019. P. 342–350. DOI: 10.1007/978-3-030-33846-6_39
8. Chen C., Shi H., Wang Z., Yu Z. A Task Scheduling Algorithm Based on Big.LITTLE Architecture in Cloud Computing". In: *2020 6th International Conference on Big Data and Information Analytics (BigDIA)*, 4–6 December 2020, Shenzhen, China IEEE. 2020. P. 94–99. DOI: 10.1109/bigdia51454.2020.00023
9. Chan C., Cooper B. Debugging incidents in Google's distributed systems. *Communications of the ACM*. 2020. Vol. 63, No. 10. P. 40–46. DOI: 10.1145/3397880
10. Langhnoja H. K., Hetal A Joshiyara P. Multi-Objective Based Integrated Task Scheduling In Cloud Computing. *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 12–14 June 2019. 2019. DOI: 10.1109/iceca.2019.8821912
11. Zolotariov D. Automated deployment of a software environment for microservices in a rapidly changing technology stack. *Сучасний стан наукових досліджень і технологій в промисловості*. 2021. No. 4 (18). P. 23–30. DOI: 10.30837/itssi.2021.18.023.
12. Danielsson J., Seceleanu T., Jagemar M., Behnam M., Sjodin M., Resource Dependency Analysis in Multi-Core Systems. *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, Madrid, Spain, 13–17 July 2020. 2020. DOI: 10.1109/compsac48688.2020.00021
13. Liu F., Guo W. Optimized Min-Min Dynamic Task Scheduling Algorithm in Grid Computing. *Advances in Intelligent Systems and Computing*. Cham, 2019. P. 745–752. DOI: 10.1007/978-3-030-25128-4_92

14. Stavrinides G. L., Karatza H. D. Scheduling Single-Task Jobs along with Bag-of-Task-Chains in Distributed Systems. *ICFNDS '19: 3rd International Conference on Future Networks and Distributed Systems*, Paris France. New York, NY, USA, 2019. P. 1–6. DOI: 10.1145/3341325.3342023
15. Yadav S., Mohan R., Yadav P. K. Fuzzy based task allocation technique in distributed computing system. *International Journal of Information Technology*. 2018. Vol. 11, No. 1. P. 13–20. DOI: 10.1007/s41870-018-0172-6
16. Li C., Liu F., Wang B., Philip Chen C. L., Tang X., Jiang J., Liu J. Dependency-Aware Vehicular Task Scheduling Policy for Tracking Service VEC Networks. *IEEE Transactions on Intelligent Vehicles*. 2022. P. 1–15. DOI: 10.1109/tiv.2022.3224057
17. Kozyriev A., Litvin S. Methods of Creating Service-oriented Software Systems. *CEUR Workshop Proceedings*. Vol. 3171, 2022. P. 763–774. URL: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85134768124&partnerID=40&md5=6ed7a3d1bcc4e527548c2ade3019562d>
18. Shubin I., Karataiev O. Reuse of Information Based on The Interpretation of Knowledge. *Сучасний стан наукових досліджень і технологій в промисловості*. 2023. No. 2 (24). P. 62–71. URL: <https://doi.org/10.30837/itssi.2023.24.062> (дата звернення: 01.10.2023).

References

1. Endo, P., de Almeida Palhares, A., Pereira, N., Goncalves, G., Sadok, D., Kelner, J., Melander, B. & Mangs, J.E. (2011), "Resource allocation for distributed cloud: concepts and research challenges". *IEEE Network*. Vol. 25(4), P. 42–46. DOI: 10.1109/mnet.2011.5958007
2. Sunyaev A. (2020), "Principles of Distributed Systems and Emerging Internet-Based Technologies". *Internet Computing. Cham: Springer International Publishing*, 2020. Vol. XVIII. 413 p. DOI: 10.1007/978-3-030-34957-8
3. Siva Prasad, B. V., Sucharitha, G., Venkatesan, K. G., Patnala, T. R., Murari, T., Karanam, S. R. (2022), "Optimisation of the Execution Time Using Hadoop-Based Parallel Machine Learning on Computing Clusters". *Computer Networks, Big Data and IoT*. Singapore, 2022. P. 233–244. DOI: 10.1007/978-981-19-0898-9_18
4. Kamran, M. "Fundamentals of Smart Grid Systems". Elsevier Science & Technology Books, 2023. 500 p. available at: <https://shop.elsevier.com/books/fundamentals-of-smart-grid-systems/kamran/978-0-323-99560-3>
5. Hasimi L., Penzel D. (2023), "A Case Study on Cloud Computing: Challenges, Opportunities, and Potentials". *Studies in Systems, Decision and Control*. Cham. P. 1–25. DOI 10.1007/978-3-031-27506-7_1
6. Zolotariov, D. (2021), "Microservice Architecture for Building High-Availability Distributed Automated Computing System in A Cloud Infrastructure". *Сучасний стан наукових досліджень і технологій в промисловості*. No. 3 (17). P. 13–22. DOI: 10.30837/itssi.2021.17.013
7. Tom, L., Bindu, V. R. (2019), "Task Scheduling Algorithms in Cloud Computing: A Survey". *Inventive Computation Technologies*. Cham. P. 342–350. DOI: 10.1007/978-3-030-33846-6_39
8. Chen, C., Shi, H., Wang, Z., Yu, Z. (2020), "A Task Scheduling Algorithm Based on Big.LITTLE Architecture in Cloud Computing". In: *2020 6th International Conference on Big Data and Information Analytics (BigDIA)*, 4–6 December 2020, Shenzhen, China IEEE. P. 94–99. DOI: 10.1109/bigdia51454.2020.00023
9. Chan, C., Cooper, B. (2020), "Debugging incidents in Google's distributed systems". *Communications of the ACM*. 2020. Vol. 63, No. 10. P. 40–46. DOI: 10.1145/3397880
10. Langhnoja, H.K., Hetal, A., Joshiyara, P. (2019), "Multi-Objective Based Integrated Task Scheduling In Cloud Computing". *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*, Coimbatore, India, 12–14 June 2019. DOI: 10.1109/iceca.2019.8821912
11. Zolotariov, D., (2021). "Automated Deployment of a Software Environment for Microservices in a Rapidly Changing Technology Stack". *Innovative Technologies and Scientific Solutions for Industries*. Vol. 4 (18), P. 23–30. DOI: 10.30837/itssi.2021.18.023
12. Danielsson, J., Seceleanu, T., Jagemar, M., Behnam, M., Sjodin, M. (2020), "Resource Dependency Analysis in Multi-Core Systems". In: *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 13–17 July 2020, Madrid, Spain. DOI: 10.1109/compsac48688.2020.00021
13. Liu, F., Guo, W. (2019), "Optimized Min-Min Dynamic Task Scheduling Algorithm in Grid Computing". *Advances in Intelligent Systems and Computing*. Cham: Springer International Publishing. P. 745–752. DOI: 10.1007/978-3-030-25128-4_92
14. Stavrinides, G. L., Karatza, H. D. (2019), "Scheduling Single-Task Jobs along with Bag-of-Task-Chains in Distributed Systems". *ICFNDS '19: 3rd International Conference on Future Networks and Distributed Systems*, Paris France. New York, NY, USA. P. 1–6. DOI: 10.1145/3341325.3342023
15. Yadav, S., Mohan, R., Yadav, P. K. (2018), "Fuzzy based task allocation technique in distributed computing system". *International Journal of Information Technology*. Vol. 11, No. 1. P. 13–20. DOI: 10.1007/s41870-018-0172-6
16. Li, C., Liu, F., Wang, B., Philip Chen, C. L., Tang, X., Jiang, J. & Liu, J. (2022), "Dependency-Aware Vehicular Task Scheduling Policy for Tracking Service VEC Networks". *IEEE Transactions on Intelligent Vehicles*. P. 1–15. DOI: 10.1109/tiv.2022.3224057

17. Kozyriev, A., Litvin, S. "Methods of Creating Service-oriented Software Systems". CEUR Workshop Proceedings. Vol. 3171, 2022. P. 763–774. available at: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85134768124&partnerID=40&md5=6ed7a3d1bcc4e527548c2ade3019562d>

18. Shubin, I., Karataiev, O. (2023), "Reuse of Information Based on The Interpretation of Knowledge". *Сучасний стан наукових досліджень і технологій в промисловості*. No. 2 (24). P. 62–71. available at: <https://doi.org/10.30837/itssi.2023.24.062> (last accessed: 01.10.2023)

Надійшла 28.08.2023

Відомості про авторів / About the Authors

Козирев Андрій Дмитрович – Харківський національний університет радіоелектроніки, аспірант кафедри програмної інженерії, Харків, Україна; e-mail: andrii.kozyriev@nure.ua; ORCID ID: <https://orcid.org/0000-0001-6383-5222>

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>

Kozyriev Andrii – Kharkiv National University of Radio Electronics, Postgraduate, Kharkiv, Ukraine.

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

METHOD OF PLANNING DATA PROCESSING TASKS IN DISTRIBUTED SYSTEMS WITH LIMITED INFORMATION ABOUT AVAILABLE RESOURCES

In today's digital landscape, distributed data processing systems (DDPs) are becoming increasingly critical to efficiently process, analyze, and manage large volumes of data. These systems are often used in commercial, scientific and social domains to process complex data in real-time or batch mode. One of the key components of such systems is task scheduling, which is an extremely complex process, particularly when information about resource requirements is not complete or accurate. The **subject** of research are algorithms, methods and approaches used for scheduling tasks between nodes in distributed systems. The **purpose** of the study is to create an optimized method of task planning in the RSOD with limited availability of information about available resources. The **task of the research**: to analyze the limitations of modern methods for scheduling tasks in distributed data processing systems (DDS); optimize the method of scheduling tasks based on metadata between RSOD nodes, based on the methodology of searching for nearest neighbors using the method of localized hashing and the algebra of finite predicates; develop the architecture of the software solution and its implementation based on the optimized method; test the algorithm on the example of a video decoding task. The following **methods** were used: statistical algorithms and techniques such as classification and cluster analysis were used to predict resource requirements, and visualization techniques assisted in the analysis and interpretation of results. As a **result** of the work: the limitations of modern methods for the distribution of tasks in distributed data processing systems (DDPs) were analyzed; an optimized method of task planning based on metadata in RSOD was created, based on the methodology of searching for nearest neighbors using the method of localized hashing and the algebra of finite predicates; the processes in the modified nearest neighbor search algorithm are detailed; the architecture of the software solution was developed, which integrates an optimized method of task planning based on metadata and resource allocation; validation of the software solution was carried out with the help of a practical scenario – the use of the created algorithm in the planning task for decoding video information. The **conclusions** of this study confirmed that the proposed method, based on the methodology of localized hashing and the use of finite predicate algebra, is effective even with insufficient or limited information about resource needs. This highlights the possibility of using dynamic scheduling strategies to adapt to changing load conditions and resource availability.

Keywords: software engineering; distributed systems; data processing; database; task scheduling; algebra of finite predicates.

Бібліографічні описи / Bibliographic descriptions

Козирев А. Д., Шубін І. Ю. Метод планування завдань оброблення даних у розподілених системах з обмеженою інформацією про доступні ресурси. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 27–39. DOI: <https://doi.org/10.30837/ITSSI.2023.25.027>

Kozyriev, A., Shubin, I. (2023), "Method of planning data processing tasks in distributed systems with limited information about available resources", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 27–39. DOI: <https://doi.org/10.30837/ITSSI.2023.25.027>

В. Ляшик, І. Шубін

МЕТОДИ АВТОМАТИЗАЦІЇ ТА ІНТЕРПРЕТАЦІЇ ЯКОСТІ ОСВІТНІХ ТЕСТІВ У СИСТЕМАХ РОЗПОДІЛЕНОГО ВІРТУАЛЬНОГО НАВЧАННЯ

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 студентів (суб'єктів навчання) за умови впровадження концепції віртуального розподіленого навчання. **Мета роботи** – створення методів автоматизації оцінювання якості освітніх тестів та інтерпретації результатів, подання математичного апарату, що допомагає якнайкраще описувати складні інтелектуальні процеси й методи оцінювання комплектів тестових завдань під час перевірки рівня знань суб'єктів навчання. 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**: математична статистика й інтелектуальний аналіз інформації, алгебра скінченних предикатів і операцій. Здобуто такі **результати**: сформульовано принципи інтелектуального аналізу значення коефіцієнтів надійності, коефіцієнтів валідності, коефіцієнта дискримінативності; визначено індекс важкості завдань для оцінювання знань суб'єктів навчання. **Висновки**. Застосування методів формалізації оцінювання тестів підтверджують необхідність введення в практику освіти кількісних методів оцінювання знань студентів. Упровадження зазначених методів передбачає правильне визначення цілей контролю, предмета вимірювання та вибір засобів вимірювання. Застосування розробленого методу аналізу наборів тестових завдань сприяє ефективній реалізації всіх функцій контролю, відповідає основним його принципам для вирішення завдань оцінювання якості тестів і лежить в основі моделей розподіленого віртуального навчання та методів аналізу успішності суб'єктів навчання.

Ключові слова: програмна інженерія; бази знань; алгебра скінченних предикатів; надійність тестів; використання знань; логічні правила.

Вступ

У зв'язку з переходом на рейтингову систему контролю знань студентів, що є системою взаємопов'язаних правил і яка останнім часом все ширше використовується на різних рівнях освіти, зростають вимоги до якості педагогічного контролю. Актуальним стає розроблення педагогічного програмного інструментарію, що дає змогу об'єктивно та всебічно проводити безперервний контроль знань студентів. Використання рейтингової системи без тестування взагалі неможливе, оскільки саме на тестових оцінках ґрунтується ця система. Під адаптивним тестовим контролем розуміють комп'ютеризовану систему науково обґрунтованої перевірки й оцінювання результатів навчання, яка має високу ефективність завдяки оптимізації процедур генерації, подання й оцінювання результатів виконання адаптивних тестів і яка оснований на методах побудови та оптимізації логічних мереж. Алгоритми підбору й подання завдань будуються за принципом зворотного зв'язку, коли в разі правильної відповіді чергове завдання обирається важчим, а за умови неправильної

відповіді пропонується наступне завдання, легше, ніж попереднє, на яке учасник тестування не відповів. Також існує функція додаткових запитань із тем, які суб'єкт навчання знає не дуже добре. Це потрібно для більш точного з'ясування рівня знань у певній науковій галузі. Вибір алгоритмів тестування наразі фактично обмежений формами подання тестових завдань і оцінювання результатів. Досягнення більш високих результатів і підвищення мотивації навчання в остаточному підсумку є основною метою тестування.

Аналіз створення локальних інформаційно-освітніх середовищ дав змогу сформулювати основні принципи професійної дистанційної освіти: універсальність; інваріантність; масштабованість; відкритість; технологічність процесів створення, зберігання й використання ресурсів; спадкоємність тощо. Огляд гіпермедійних засобів адаптивного навчання показав, що це новий дослідницький напрям у сфері інженерії програмного забезпечення підтримки дистанційного навчання, завдання якого – додати до дистанційних навчальних систем можливості індивідуалізації. Аналітичний огляд проблем електронного навчання, зокрема

інтенсифікація, індивідуалізація, послідовність викладення матеріалів, адаптивність і гнучкість, дав змогу виокремити основні підходи до їх вирішення: формування індивідуальних траєкторій навчання, розроблення моделей і суб'єктів навчання, а також методів адаптивного навчання.

Контроль знань або тестування – процес, що проводиться для визначення рівня знань суб'єкта навчання [2]. Це найбільш стандартизований і об'єктивний метод контролю й оцінювання знань, умінь і навичок, який позбавлений традиційних недоліків інших методів контролю знань, наприклад, неоднорідності вимог, суб'єктивності екзаменаторів, невизначеності системи оцінювання тощо. Рівні знань зазвичай дискретизуються. За такого підходу тестування може розглядатися як деякий діагностичний процес, а стани, що визначають оцінки знань суб'єкта навчання, – як діагностичні стани. Тести є ефективним засобом перевірки якості знань, що набули студенти, і оперативного контролю етапів навчання [3].

У роботі запропоновано завдання підвищення ефективності інформаційно-аналітичного забезпечення підсистеми контролю знань суб'єктів освітнього процесу за умови побудови індивідуальної траєкторії навчання, що передбачає такі етапи: 1) розроблення концептуальної моделі електронного адаптивного навчання в розподіленому віртуальному середовищі; 2) створення моделі побудови індивідуальної освітньої траєкторії; 3) побудова математичної моделі навчального об'єкта; 4) розроблення моделі визначення тестових об'єктів перевірки засвоєння навчальних матеріалів відповідно до індивідуальних вимог користувача; 5) розроблення методу побудови адаптивної освітньої траєкторії; 6) створення прототипів програмних рішень для реалізації інформаційної технології побудови індивідуальних траєкторій навчання.

Для визначення базових алгоритмів необхідно навести алгебро-логічні моделі системи, в основі яких лежить модель приймання іспиту викладачем у студента (суб'єкта навчання) як модель адаптивного тестування. Такий вибір сценарію роботи системи зумовлений тим, що, по-перше, ця процедура історично добре формалізована, по-друге, у процесі проектування тестів розробнику необхідно спиратися на загальноприйняті, відомі та використовуваним ним методи з мінімальною модифікацією.

Аналіз проблеми й наявних методів

Виокремлюють такі функції педагогічного контролю: контролююча, навчальна (освітня), діагностична, прогностична, розвивальна, орієнтувальна, виховна. Вони, як правило, тісно взаємозв'язані [1]. У навчальному процесі ці функції виявляються в різному ступені та в неоднакових поєднаннях. Реалізація перелічених функцій робить контроль більш результативним, а також ефективнішим стає і сам освітній процес.

Навчальна функція контролю полягає в удосконаленні та систематизації знань і вмінь. У процесі перевірки студенти повторюють і закріплюють вивчений матеріал. Вони не тільки відтворюють раніше вивчене, але й застосовують знання та вміння в нових ситуаціях. Перевірка допомагає виокремити основні положення досліджуваного матеріалу, зробити знання та вміння більш чіткими й точними. Контроль сприяє також узагальненню та систематизації знань [2].

Сутність діагностичної функції полягає в отриманні інформації про помилки, недоліки й прогалини в знаннях і вміннях студентів, про причини, що породжують труднощі в опануванні навчального матеріалу, і про кількість, характер помилок. Результати діагностичних перевірок допомагають обрати найбільш інтенсивну методику, а також уточнити напрям подальшого вдосконалення змісту, методів і засобів навчання.

Прогностична функція слугує отриманню випереджальної інформації про навчально-виховний процес. За результатами перевірки здобувається прогноз про певний етап навчального процесу: чи достатньо сформовані конкретні знання, уміння й навички для засвоєння наступного навчального матеріалу (розділу, теми); розробляється модель подальшої поведінки студента.

Розвивальна функція контролю полягає в стимулюванні пізнавальної активності студентів, формуванні їхніх творчих здібностей. Контроль має виняткові можливості для розвитку студентів. У процесі контролю розвивається мова, пам'ять, увага, уява, воля та мислення молодої людини.

Сутність орієнтувальної функції контролю – отримати інформацію про ступінь досягнення мети навчання окремим студентом і групою загалом, тобто рівень вивчення навчального матеріалу. Розкриваючи прогалини, помилки й недоліки

в підготовленні студентів, визначаються напрями щодо вдосконалення знань і вмінь. Контроль допомагає суб'єкту навчання краще пізнати самого себе, оцінити свої знання й можливості [4].

Виховна функція контролю формує в студентів відповідальне ставлення до навчання, дисципліну, акуратність, чесність. Перевірка спонукає до більш серйозного й регулярного виконання завдань.

Система перевірки результатів навчання й виховання містить три основних види контролю: поточний, проміжний і підсумковий [5].

Для опису процесу вибору індивідуальної траєкторії навчання необхідні засоби інтелектуального оброблення інформації. Теорія інтелекту, засоби моделювання інтелектуальних функцій людини на основі методу компараторної ідентифікації та інструментарію алгебри скінченних предикатів, розроблені науковою школою професора Ю. Шабанова-Кушнарєнка, дають таку можливість та добре зарекомендували себе у вирішенні завдань інтелектуального оброблення інформації, зокрема текстового та візуального [6].

Ця мета в створенні методів і алгоритмів побудови віртуального розподіленого навчального середовища насамперед пов'язана з визначенням якості засвоєння студентами навчального матеріалу – рівня набуття знань, умінь і навичок, передбачених програмою. По-друге, конкретизація основної мети контролю пов'язана з навчанням прийомів взаємоконтролю та самоконтролю та формуванням потреби в них. По-третє, ця мета передбачає виховання в студентів таких особистісних якостей, як відповідальність за виконану роботу, вияв ініціативи.

Мета статті – створення методів автоматизації оцінювання якості освітніх тестів і впровадження цих методів у системи підтримки дистанційного навчання.

Вирішення завдання

Серед форм контролю знань у процесі впровадження віртуального навчального середовища все більшого розвитку набуває тестовий контроль, рекомендований Міністерством освіти і науки України та документами Болонської конвенції. Тестовий контроль у цих документах розглядається як один із засобів удосконалення процесу навчання. Актуальними є питання узагальнення, систематизації наявного досвіду й вироблення рекомендацій у сфері створення тестових завдань і проведення тестового контролю знань, умінь і навичок [6].

Для подання знань у інформаційних системах використовуються різні формальні мови, зокрема обчислення предикатів. Вона має однозначну формальну семантику й операційну підтримку у вигляді досконалого механізму виведення. Мовою алгебри предикатів будь-яка множина $U = \{a_1, a_2, \dots, a_n\}$ може бути записана у вигляді рівняння $x^{a_1} \vee x^{a_2} \vee \dots \vee x^{a_n} = 1$, де x – предметна змінна [7]. Сукупність усіх коренів цього рівняння збігається з множиною U . Рівняння є формальним записом твердження $x \in U$. Методи ідентифікації та подання знань із використанням алгебри предикатів можуть одночасно подати знання у вигляді рівнянь алгебри предикатів [9].

Алгебра предикатів описує лише знання про факти. Алгебра операцій над предикатами або алгебра предикатних операцій має формалізувати операції над знаннями, поданими як відношення на деякому предметному просторі. Алгебра предикатів визначає декларативний складник знань, а алгебра предикатних операцій – процедурний складник набуття знань, умінь і компетенцій суб'єктом навчання в умовах індивідуалізації освітнього процесу в розподіленому віртуальному середовищі [9].

Комп'ютерне тестування має свої переваги й недоліки. Саме об'єктивність отриманих оцінок сприяє розширенню сфери впровадження комп'ютерного тестування – воно успішно застосовується в школах, вишах, під час складання іспитів, співбесіди тощо.

До таких рішень належать: програма *NetTest* для комп'ютерного тестування знань у мережі [9]; *MyTestXPro* – система програм для створення та проведення комп'ютерного тестування, збору та аналізу їх результатів [10]; комп'ютерна система тестування знань *OpenTest* [11]; конструктор тестів *Keepsoft* – універсальної системи перевірки знань [12]; програма *x-TLS* [13]. Основні відмінності таких систем: централізований збір і аналіз результатів тестування за наявності мережі; виставлення оцінки із використанням довільної системи та шкали оцінювання; організація локального та мережного тестування; стеження за процесом тестування. Застосовується підтримка п'яти типів запитань; використання необмеженої кількості тем, запитань і відповідей; наявність запитань, що містять текст необмеженої довжини; застосування в запитаннях гіпермедійних і мультимедійних об'єктів; можливість декільком

користувачам проходити тестування на одному комп'ютері; збір і аналіз результатів за допомогою локальної мережі та синхронізації бази даних; копіювання/резервне копіювання та збереження й аналіз результатів. Після розгляду аналогів, присутніх на ринку, було прийнято рішення розробити новий продукт для аналізу якості тестів, у якому буде реалізовано можливість складання тестів, проведення тестування за допомогою мережі Інтернет, з подальшим упровадженням таких моделей у розподілені віртуальні навчальні системи.

Тестові технології контролю знань і вмінь широко використовують у ЗВО України на проміжних етапах навчального процесу, але на кінцевому етапі використовуються нечасто. Упровадження тестування на іспитах стримується через відсутність інформації про якість тестів [14].

Тільки якісний тест забезпечує відповідний контроль знань і дає змогу скласти індивідуальні програми навчання. Належна структура та помірна важкість тестових завдань суттєво розширюють інформативність здобутих результатів як в плані необхідності корегування методик викладання дисциплін (розділів), так і в гуманістичному напрямі, допомагаючи студентам визначитися щодо ефективності своєї реалізації у відповідній дисципліні (зокрема порівняти рівень своєї підготовки з іншими студентами групи).

Для мінімізації похибок тест має розроблятися за певними правилами та процедурами, що, наприклад, передбачають перевірку якості як самого тесту, так і його окремих завдань [13].

Тому процес оцінювання якості освітніх тестів має достатню кількість повторюваних кроків і потребує значних матеріальних та інтелектуальних витрат, що важливо звести до мінімуму, а це, безумовно, визначає необхідність автоматизації процесу [14].

Методи автоматизації та інтерпретації якості освітніх тестів мають виконувати такі завдання: формувати тест (добирати тестові завдання з дисципліни); коректувати матриці результатів, а саме: вилучати невдалі завдання, відповіді студентів, знання яких не можуть бути оцінені з допомогою набору тестових завдань, що підлягає аналізу; оцінювати якість тестів, зокрема розраховувати коефіцієнти надійності тесту, коефіцієнт валідності тесту та якість окремого завдання з допомогою статистичних показників моди й медіани, дисперсії,

середньоквадратичного відхилення, коефіцієнтів асиметрії та ексцесу.

Також як результати застосування методів можна отримати коефіцієнти дискримінативності, індекси та коефіцієнти важкості завдань.

Методи інтерпретації наборів тестових завдань мають виконуватися в проведенні тестування знань у режимі реального часу.

Вхідною інформацією є тест і відповіді студентів на тест, оцінки експертів. Вихідна інформація – це значення коефіцієнта надійності, коефіцієнта валідності, коефіцієнта дискримінативності, індекс важкості завдання та оцінки студентів – учасників тестування.

Сучасна теорія тестування (*IRT*) [15] має такі основні припущення: існують приховані параметри особистості, не доступні для безпосереднього спостереження (у тестуванні це рівень підготовленості студента й рівень важкості завдання); залежно від моделі існують індикаторні змінні, пов'язані з прихованими параметрами й доступні для безпосереднього спостереження. За значеннями таких індикаторів можна судити про значення прихованих параметрів.

Якщо обрано тестування з пріоритетом "швидке навчання", тоді рішення додати навчальний об'єкт до набору тестових завдань проводиться за формулою:

$$l_{oj} = \begin{cases} 1, & l_{oj_{time}} \leq 1, l_{oj_{importance}} > 2, l_{oj_{level}} \geq K_{cu_{level}}, \\ 0, & l_{oj_{time}} > 1, l_{oj_{importance}} \leq 2, l_{oj_{level}} < K_{cu_{level}}. \end{cases} \quad (1)$$

Кількість часу на вивчення навчальних об'єктів має бути мінімальна: $\sum_{j \in J} l_{oj_{time}} \rightarrow \min$.

Якщо обрано тестування з пріоритетом на "глибоке навчання", тоді рішення додати навчальний об'єкт до тренінгу проводиться за формулою:

$$l_{oj} = \begin{cases} 1, & l_{oj_{level}} \geq K_{cu_{level}}, \\ 0, & l_{oj_{level}} < K_{cu_{level}}. \end{cases} \quad (2)$$

Кількість навчальних об'єктів має бути максимальна: $\sum_{j \in J} l_{oj_{time}} \rightarrow \max$.

Якщо обрано пріоритет "лише найважливіше", тоді рішення додати навчальний об'єкт до тренінгу проводиться за формулою:

$$l_{oj} = \begin{cases} 1, & l_{oj_{importance}} > 2, l_{oj_{level}} \geq K_{cu_{level}}, \\ 0, & l_{oj_{importance}} \leq 2, l_{oj_{level}} < K_{cu_{level}}. \end{cases} \quad (3)$$

Кількість часу на вивчення навчальних об'єктів має бути мінімальна, а обсяг інформації має бути максимальним:

$$\begin{cases} \sum_{j \in J} lo_{j_{time}} \rightarrow \min \\ \sum_{j \in J} lo_{j_{time}} \rightarrow \max. \end{cases} \quad (4)$$

Розроблений метод містить і методи розрахунку надійності тестів – жодний з наявних тестів не є ідеальним щодо надійності.

Показники надійності тесту можуть істотно змінюватись у разі незначних, на перший погляд, варіацій умов проведення тестування, зміни характеру тестових завдань, а також залежно від ступеня важкості завдань для суб'єкта навчання.

На ознаки надійності суттєво впливає характер вибірки, що використовується для перевірки тесту за показниками надійності. Особливе значення тут має діапазон розбіжності в оцінках. Так, якщо оцінки студента розміщуватимуться у відносно вузькому діапазоні значень і будуть близькі одна до одної, можна сподіватися, що й під час повторного дослідження ці оцінки також розміщуватимуться щільною гомогенною групою. Імовірні зміни

рангових місць окремих піддослідних будуть зовні незначними. Коефіцієнт надійності у такому разі буде завищений. Таке саме завищення коефіцієнтів може виникнути під час аналізу надійності, що проводиться на матеріалі вибірки, яка складається з контрастних груп, наприклад, з осіб, що мають найвищі та найнижчі оцінки з тесту. Тоді ці віддалені один від одного результати не перекриватимуться випадковими чинниками.

Оцінювання надійності тестів проводиться різними методами, які можна поділити на дві групи: оцінка, що основана на двократному тестуванні, й оцінка, яка передбачає однократне тестування. Дворазове тестування проводиться за допомогою одного й того самого тесту або двох паралельних тестів (рис. 1).

Оцінка надійності найчастіше будується на обчисленні кореляції між двома наборами результатів виконання одного й того самого тесту або двох паралельних форм. У цьому разі логіка міркувань ґрунтується на принципі: що вища кореляція, то більш надійним є тест [19].

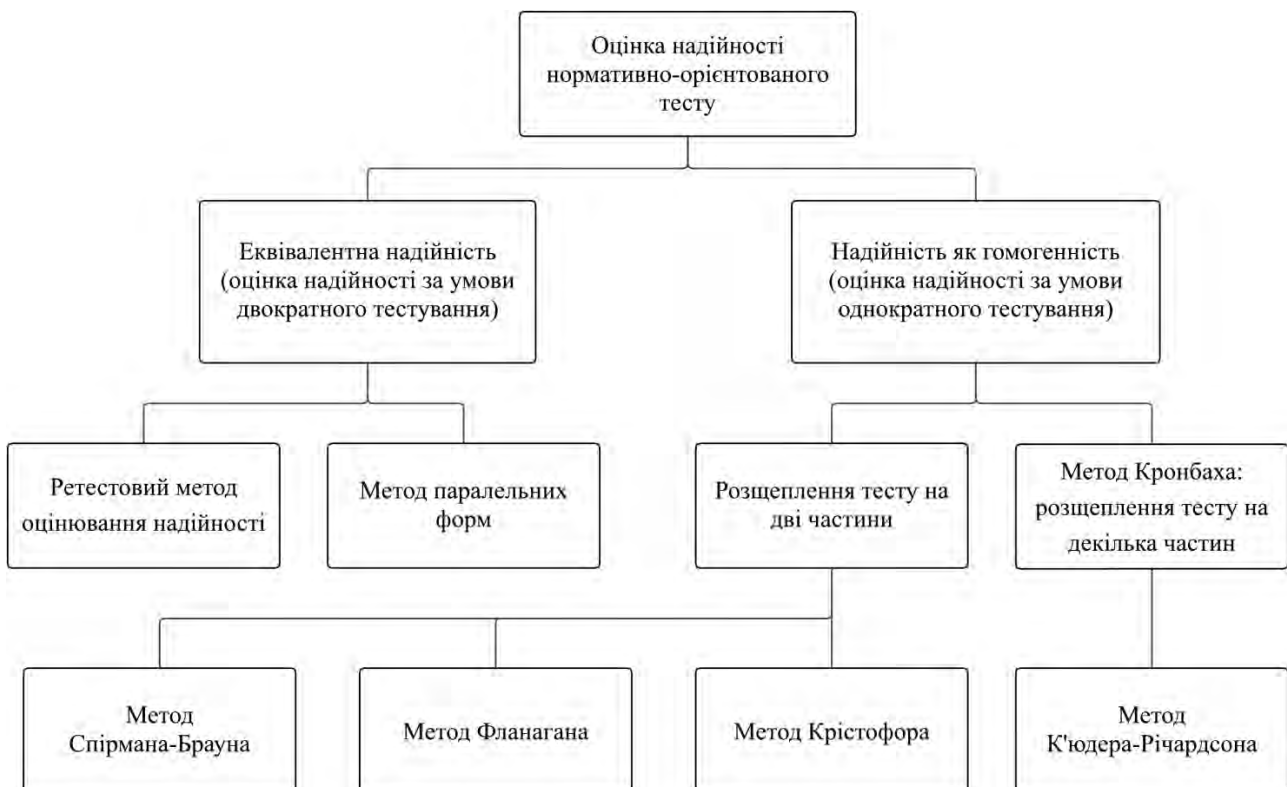


Рис. 1. Методи оцінювання надійності тесту

Розмаїття характеристик і показників надійності тесту так само велике, як і розмаїття умов, що можуть впливати на результат тесту, але найширшого практичного застосування набули:

- ретестова надійність:

$$r_n^{rem} = \frac{n \sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{\sqrt{n \left(\sum_{i=1}^n X_i \right)^2 - \left(\sum_{i=1}^n X_i \right)^2} \times \sqrt{n \left(\sum_{i=1}^n Y_i \right)^2 - \left(\sum_{i=1}^n Y_i \right)^2}}; \quad (6)$$

- надійність паралельних форм тесту:

$$r_n = \frac{2(r_n^{расц})}{1 + r_n^{расц}}; \quad (7)$$

– надійність частин тесту – розщеплення завдань тесту на рівноцінні половини є лише частковим випадком аналізу надійності частин тесту. Дійсно, можливе розщеплення на три, чотири й більше частин. У граничному випадку кількість таких частин дорівнює кількості пунктів тесту. Тоді аналіз надійності, власне, стає аналізом внутрішньої узгодженості, де $r_n^{расц}$ – кореляція між частинами тесту.

Для мінімізації похибок тест має розроблятися за певними правилами й процедурами, що, зокрема, передбачають перевірку якості як самого тесту, так і його окремих завдань.

Тому оцінювання якості освітніх тестів передбачає значну кількість повторюваних кроків і потребує чималих матеріальних та інтелектуальних витрат, які важливо звести до мінімуму, а це, безумовно, визначає необхідність автоматизації процесу.

Основним завданням запропонованого методу, згідно з *IRT*, є перехід від індикаторних змінних до прихованих змінних. Перелічимо переваги:

- відповідно до стандарту *IRT* перетворено вимірювання, що виконані в дихотомічних і порядкових шкалах, у лінійні вимірювання, потім якісні показники аналізуються за допомогою кількісних методів;

- міра вимірювання параметрів моделей є лінійною, що дає змогу використовувати широкий спектр статистичних процедур для аналізу результатів вимірювань;

- інваріантність завдань: оцінка важкості тестових завдань не залежить від вибірки досліджуваних, на яких вона була отримана;

- інваріантність здібностей: оцінка рівня підготовки студентів не залежить від використовуваного набору тестових завдань;

- неповнота інформації (пропуск деяких комбінацій "студент – тестове завдання") не є критичною.

Показники надійності тесту можуть істотно змінюватись у разі незначних, на перший погляд, варіацій умов проведення тестування, зміни характеру тестових завдань, а також залежно від ступеня важкості завдань для суб'єкта навчання.

На ознаки надійності суттєво впливає характер вибірки, що використовується для перевірки тесту за показниками надійності. Особливе значення тут має діапазон розбіжності в оцінках. Так, якщо оцінки студента розміщуватимуться у відносно вузькому діапазоні значень і будуть близькі одна до одної, можна сподіватися, що й під час повторного дослідження ці оцінки також розміщуватимуться щільною гомогенною групою. Імовірні зміни рангових місць окремих піддослідних будуть зовні незначними. Коефіцієнт надійності у такому разі буде завищений. Таке саме завищення коефіцієнтів може виникнути під час аналізу надійності, що проводиться на матеріалі вибірки, яка складається з контрастних груп, наприклад, з осіб, що мають найвищі та найнижчі оцінки з тесту. Тоді ці віддалені один від одного результати не перекриватимуться випадковими чинниками.

Призначення системи. На першому етапі необхідно визначити цілі навчання, що подаються в термінах компетенцій. Далі важливо провести початкове тестування суб'єкта навчання, щоб визначити його рівень. Тестування ґрунтується на використанні відповідного фреймворку, що зазвичай розробляється експертами професійних спільнот. Тематика навчальних і тестових матеріалів, яка відповідає компетенціям, подається у вигляді когнітивних карт і дає змогу визначити необхідні навчальні матеріали на основі застосування розроблених моделей. Тестування та адаптивний вибір навчальних об'єктів є перманентним процесом для досягнення цілей навчання.

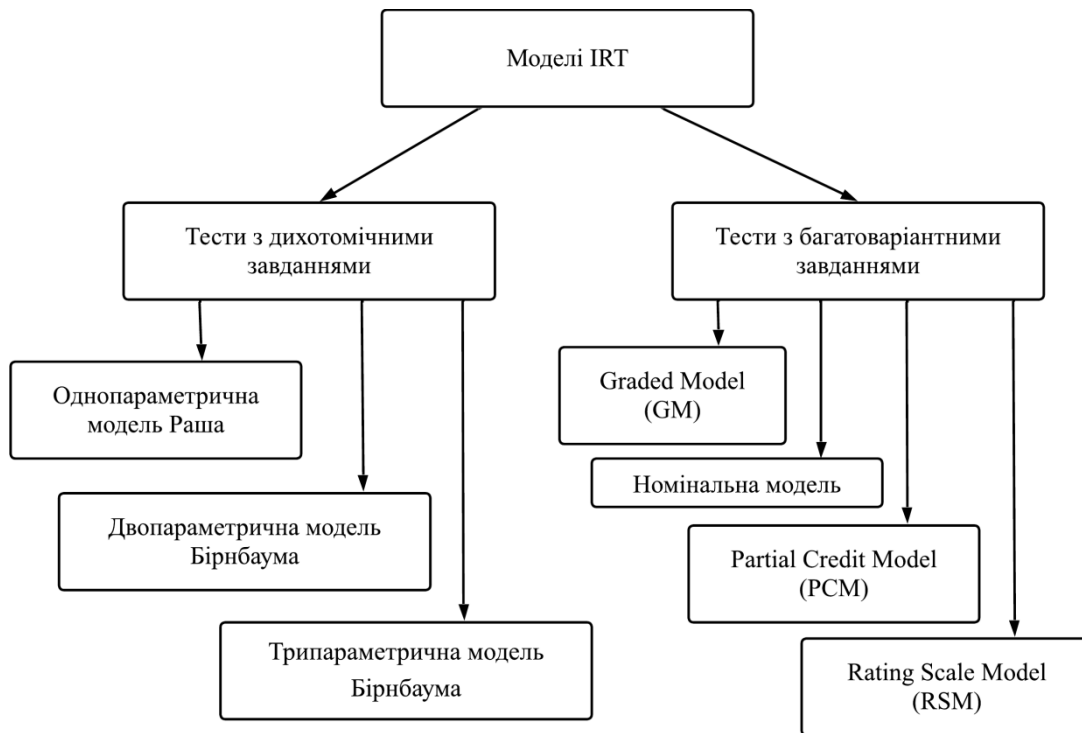


Рис. 2. Моделі IRT

Завдання реалізації запропонованих методів оцінювання якості тестів із певної дисципліни має бути успішно розв'язане способом автоматизації необхідних для її вирішення процесів. Розроблене програмне забезпечення може бути використане в розробленні викладачами нормативно-орієнтованих тестів. Відповіді на завдання тесту можуть надходити до програмної системи після проходження студентом тестування в режимі реального часу або способом їх додавання до бази даних. Оброблення результатів передбачає розрахунок оцінки якості та інтерпретацію наборів тестових завдань на основі відповідей студентів. ПЗ може бути автономним модулем або програмним компонентом, що інтегрується з підсистемою віртуального розподіленого навчального програмного ресурсу.

Приклад моделювання завдання оцінювання якості комплексів тестових завдань

Алгоритм вирішення завдання передбачає такі етапи:

1. Розроблення завдань.
2. Створення тесту.
3. Проведення пробного тестування.
4. Вибракування й корегування тестових завдань.

5. Розрахунок інтеркореляцій завдань.

6. Видалення або перероблення завдань із високою однорідністю.

7. Проведення тестування за новим тестом, у якому всі завдання відповідають умовам якості.

8. Розрахунок валідності тесту.

9. Розрахунок надійності тесту.

10. Проведення тестування за якісним тестом.

Алгоритм вибракування тестових завдань передбачає такі етапи:

1. Перевірка на нормальність розподілу балів завдань.

2. Видалення тих завдань, розподіл яких не є нормальним.

3. Розрахунок індивідуального бала суб'єктів навчання.

4. Розрахунок середніх результатів сумарних балів суб'єктів навчання.

5. Розрахунок середніх результатів суб'єктів навчання з кожного завдання.

6. Розрахунок коефіцієнта важкості завдання.

7. Розрахунок індексу важкості завдання з огляду на відгадування.

8. Розрахунок дисперсії та стандартного відхилення сумарних балів суб'єктів навчання.

9. Розрахунок стандартного відхилення сумарних балів суб'єктів навчання.

10. Вилучення завдань, для яких коефіцієнт важкості близький до 1 або до 0.

11. Розрахунок дисперсії результатів випробуваних за j -м завданням.

12. Розрахунок стандартного відхилення результатів випробуваних за j -м завданням.

13. Вилучення завдань із високою дисперсією.

14. Розрахунок індексу важкості завдань.

15. Розрахунок коефіцієнта дискримінативності завдання та вилучення завдань, для яких оцінка близька до 0,9 або до 0.

Розроблення системи опису основних інформаційних об'єктів та їх властивостей для конкретної предметної галузі – бізнес-правил – це наступний етап проектування програмного забезпечення аналізу наборів тестових завдань. Сформульовано такі бізнес-правила: кожен суб'єкт тестування має унікальний ідентифікатор; кожний тест може мати багато запитань; у кожному запитанні може бути багато відповідей; необхідно, щоб кожне запитання мало одну правильну відповідь; кожний тест може використовуватися в багатьох тестуваннях; кожна відповідь має текстовий опис; кожна помилкова

відповідь може коригуватись на відгадування; для кожного суб'єкта навчання за тестом можна розрахувати індивідуальний бал; для кожного завдання тесту можна обчислити коефіцієнт важкості завдання; для кожного завдання тесту можна обчислити індекс важкості завдання; для кожного завдання тесту можна обчислити дискримінативність; для кожного тесту можна обчислити надійність за обраним методом та за оцінкою експертів, у формулі для розрахунку валідності тесту береться середнє i -го суб'єкта навчання. Отже, для кожного тесту можна обчислити валідність.

Користувачами системи є викладач, студент та адміністратор. Викладач розробляє тести, вводить результати проведеного тестування, розраховує оцінки надійності тестів та оцінки знань студентів. Суб'єкт навчання (студент) – виконавець тесту. Адміністратор відповідає за реєстрацію користувачів і супровід БД.

Основні функціональні вимоги до програмної системи з погляду викладача, студента, адміністратора подані у вигляді діаграм варіантів застосування (рис. 3).

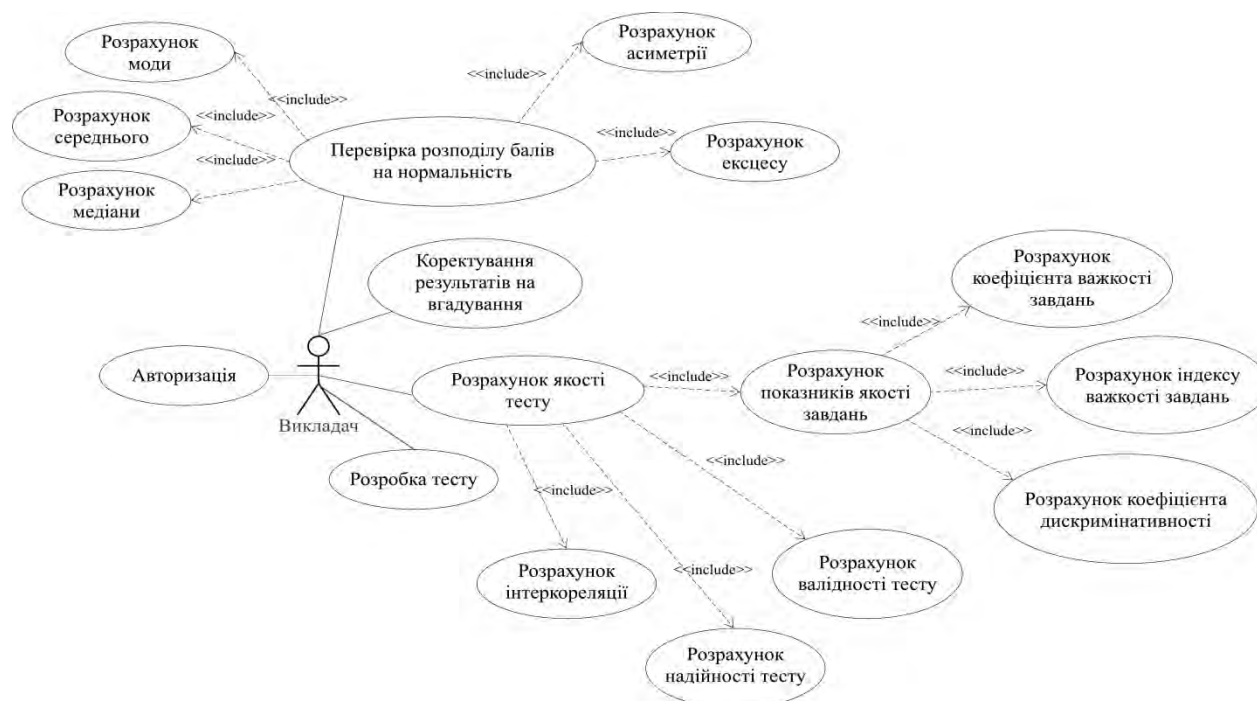


Рис. 3. Діаграма використання системи викладачем

Діаграма послідовності для прецеденту "розрахунок надійності" зображена на рис. 4. Розроблена діаграма класів – статичне подання структури моделі – відображає статичні елементи, такі як: класи, типи даних, їх зміст і відношення.

Метою реалізації алгоритмів аналізу даних (*AnalysisServlet*) є вирішення підзадачі, що пов'язана з аналізом показників тестування для отримання результатів.

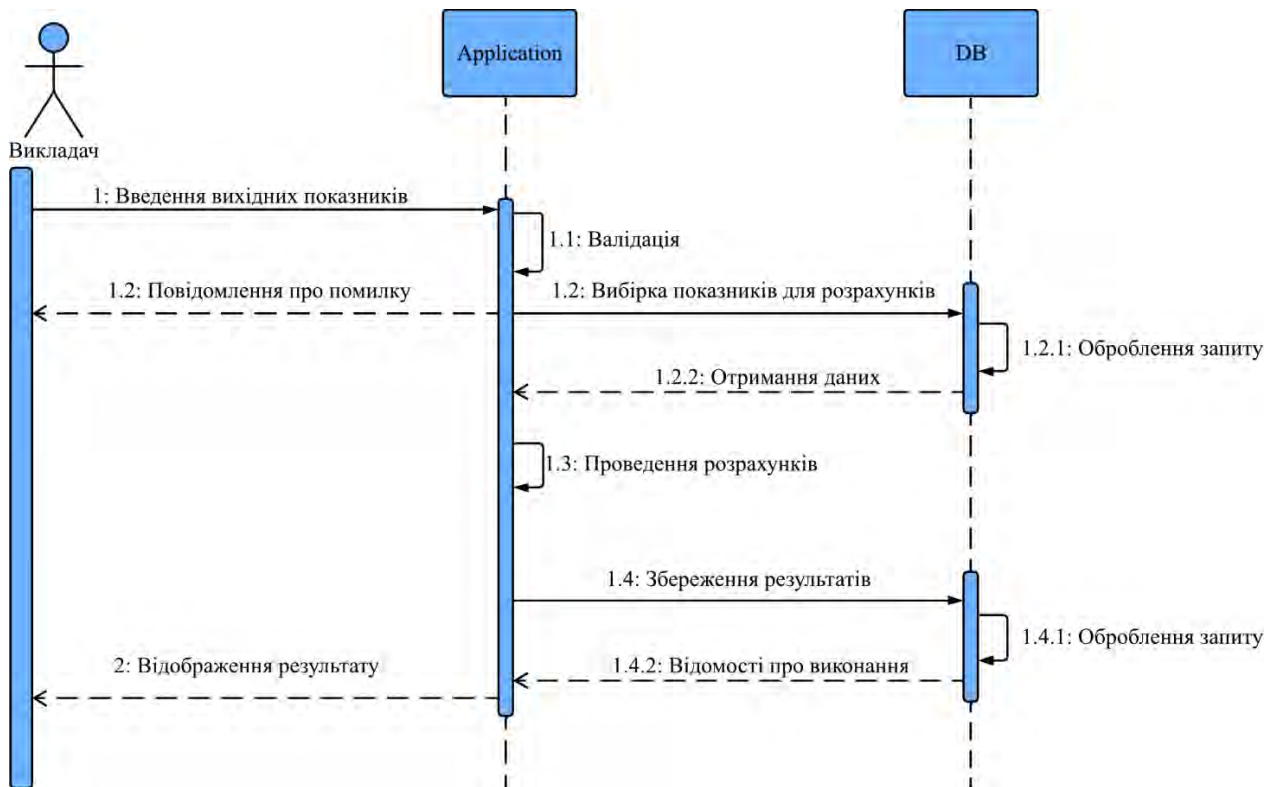


Рис. 4. Діаграма послідовності

До функцій компоненти "управління анкетами" належать: перегляд списку розроблених тестів; розроблення нових тестів; редагування змісту тесту.

Функції компоненти "збір даних онлайн" передбачають: відображення тесту; автоматичний збір даних за допомогою моделей штучних агентів; збереження відповідей на тест у базі даних.

До функцій компоненти "аналіз даних" належить: проведення аналізу завдань; аналіз розподілу балів; розрахунок коефіцієнта дискримінативності, індексу важкості завдань, коефіцієнта важкості завдань.

Компонента "аналіз тесту" містить такі підзадачі: розрахунок коефіцієнта надійності, коефіцієнта валідності та інтеркореляції, побудова гістограм.

Результати розрахунків мають показати значення коефіцієнтів надійності. Високі значення таких коефіцієнтів, отриманих унаслідок різних методів розрахунків надійності, свідчать про високу надійність розробленого методу.

Потім викладач має визначити розрахунок валідності тесту, де він обирає зі списку дисципліну й тест, потрібне тестування та обчислює валідність тесту. Отримано достатньо високі значення коефіцієнта валідності.

Наступний крок реалізації передбачає розрахунок якості завдань. Відкривається сторінка, на якій

обирається дисципліна й тест, потрібний вид тестування (проміжне, підсумкове тощо), набір запитань, обчислюється дискримінативність і коефіцієнти важкості завдань тесту, аналізується розподіл балів на основі розрахованих статистик для корегування на відгадування результатів тестування або для первинних результатів тестування.

Розробляється тест, у якому на першому етапі 14 запитань. Результати тестування, отримані групою з 50 студентів, наведені далі. Було проведено тестування функціональності програмних рішень. Так, за індексом важкості завдань, скоректованих на відгадування, не одне із завдань не вилучатиметься з тесту.

Аналіз показників дискримінативності завдань – критичне значення коефіцієнта дискримінації знаходиться зі статистичних таблиць. Воно дорівнює 0,28. Отже, всі запитання цього тесту валідні. Унаслідок чого надійність тесту становить 0,78545. Отримані високі значення надійності, що пов'язано з високою однорідністю запитань.

Аналіз результатів розроблених моделей доводить, що для 5 % рівня значущості знайдено критичний рівень кореляції, який дорівнює 0,40. Це свідчить про високий рівень надійності тесту, що відповідає показникам якості.

Висновки

Серед факторів, які формують ситуацію у сфері педагогічного контролю, можна визначити основний – необхідність упровадження в практику освіти кількісних методів оцінювання знань студентів. Методи ідентифікації та подання властивостей наборів тестових завдань із використанням алгебри предикатів можуть одночасно подати знання у вигляді рівнянь алгебри предикатів.

Упровадження кількісних методів оцінювання наборів тестових завдань передбачає правильне визначення цілей контролю, предмета вимірювання та вибір засобів вимірювання. Використання педагогічних тестів сприяє ефективній реалізації всіх функцій контролю знань і відповідає основним його принципам та принципам побудови віртуального розподіленого навчального простору.

Педагогічні тести відкривають перспективні напрями підвищення якості дистанційної освіти

завдяки вдосконаленню системи контролю та посилення мотивації до навчання. З упровадженням системи електронного тестування цю проблему вдається вирішити, крім того, унеможливаються негативні ситуації, пов'язані з необ'єктивністю викладачів. Статистичні методи визначення рівня знань на основі тестів не мають проблем суб'єктивізму експертного оцінювання, але пов'язані з певною похибкою вимірювань. Тому питання щодо розроблення нових методів інтелектуального оцінювання якості таких вимірювань, що містять оцінки надійності та валідності наборів тестових завдань, є актуальним.

У роботі розглянуто питання формалізації оцінювання тестів, аналізу вимог до програмного забезпечення тестової підсистеми віртуального розподіленого навчального середовища, розроблення програмного забезпечення для розв'язання задачі оцінювання якості тестів.

Список літератури

1. Соколовський Я. І., Сторожук О. Л., Крошній І. М. Застосування сучасних інформаційно-комунікаційних технологій для організації дистанційного навчання. Науковий вісник НЛТУ України. 2015. Вип. 25.6. С. 243–251. URL: http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21IID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILE=&2_S21STR=vnvltu_2015_25
2. Learning management system. URL: https://en.wikipedia.org/wiki/Learning_management_system (дата звернення: 26.08.2023).
3. Backer R. Educational data mining and learning analytics. The Cambridge handbook of the learning sciences. 2019. 274 p. DOI: 10.1017/CBO9781139519526.016
4. A Guide to the Scrum Body of Knowledge (SBOK™ GUIDE) URL: <https://www.scrumstudy.com/SBOK/SCRUMstudy-SBOK-Guide-3rd.pdf> (дата звернення: 26.08.2023).
5. 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>
6. Shubin I. Development of conjunctive decomposition tools. CEUR Workshop Proceedings, 2021. P. 890–900. URL: <https://ceur-ws.org/Vol-2870/>
7. 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>
8. Тест як інструмент педагогічного моніторингу. URL: <http://opentest.com.ua/test-kak-instrument-izmereniya-urovnyaznaniy> (дата звернення: 20.08.2023).
9. MyTestXPro – система програм для створення та проведення комп'ютерного тестування, збору та аналізу їх результатів. URL: <http://mytest.net> (дата звернення: 28.08.2023).
10. Комп'ютерна програма тестування OpenTEST2. URL: <http://opentest.com.ua/kompyuternaya-programma-testirovaniyaznaniy-opentest-2> (дата звернення: 21.08.2023).
11. Конструктор тестів Keepsoft. URL: <http://www.keepsoft.ru/simulator.htm>. (дата звернення: 20.08.2023.)
12. Безкоштовна програма для тестування знань та онлайн підготовки. URL: <http://xtls.org.ua/test.html> (дата звернення: 20.08.2023).
13. 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
14. 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>

15. 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
16. Pellissier-Tanon T., Weikum G., Suchanek 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
17. 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, P. 1030–1042. URL: <https://ceur-ws.org/Vol-3171/paper74.pdf>
18. 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
19. 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>
20. 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>
21. 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*. 2022. № 4 (22). P. 5–13. DOI: <https://doi.org/10.30837/ITSSI.2022.22.005>

References

1. Sokolovskiy, Ya, Storozhuk, O., Kroshnyi, I. "The use of modern information and communication technologies for the organization of distance learning". Scientific Bulletin of the National Technical University of Ukraine. 2015. Issue 25.6. P. 243–251. available at: http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&S21P03=FILE=&S21STR=nvntu_2015_25
2. "Learning management system". available at: https://en.wikipedia.org/wiki/Learning_management_system (last accessed 26.08.2023).
3. Backer, R. (2019), *Educational data mining and learning analytics*. The Cambridge handbook of the learning sciences. 274 p. DOI: 10.1017/CBO9781139519526.016
4. "A Guide to the Scrum Body of Knowledge (SBOK™ GUIDE)", available at: <https://www.scrumstudy.com/SBOK/SCRUMstudy-SBOK-Guide-3rd.pdf> (last accessed 26.08.2023).
5. Gruzdo, I., Kyrychenko, I., Tereshchenko, G., Shandize, 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>
6. Shubin, I. "Development of conjunctive decomposition tools". CEUR Workshop Proceedings, 2021. P. 890–900. available at: <https://ceur-ws.org/Vol-2870/>
7. 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>
8. "The test as a tool of pedagogical monitoring" ["Test yak instrument pedahohichnoho monitorynhu"] available at: <http://opentest.com.ua/test-kak-instrument-izmereniya-urovnya-znanij> (last accessed 20.08.2023.)
9. "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 28.08.2023).
10. "Computer testing program OpenTEST2". ["Kompiuterna prohrama testuvannia OpenTEST2"]. available at: <http://opentest.com.ua/kompyuternaya-programma-testirovaniya-znanij-opentest-2>. (last accessed 21.08.2023).
11. "Test designer Keepsoft". ["Konstruktor testiv Keepsoft"]. available at: <http://www.keepsoft.ru/simulator.htm>. (last accessed 20.08.2023).
12. "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 20.08.2023).
13. 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
14. 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>
15. 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
16. 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

17. 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, P. 1030–1042. available at: <https://ceur-ws.org/Vol-3171/paper74.pdf>

18. 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

19. 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>

20. 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. available at: <https://ceur-ws.org/Vol-3171/paper4.pdf>

21. 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 28.08.2023

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 Software Department, Kharkiv, Ukraine.

METHODS OF AUTOMATION AND INTERPRETATION OF THE QUALITY EDUCATIONAL TESTS IN DISTRIBUTED VIRTUAL LEARNING SYSTEMS

The **subject matter** of the article is the development of mathematical and algorithmic support for intellectual tools, which allows you to conduct continuous control of the knowledge of students (subjects of study) objectively and comprehensively. The **goal** of the work is to create methods for assessing the quality of educational tests and automating such processes. The following **tasks** were solved in the article: formation of a testing model in a distributed virtual learning environment and a validity assessment model based on the content of sets of test tasks. The following **methods** used are – algebra of finite predicates and operations, methods of mathematical statistics and methods of intellectual data analysis. The following **results** were obtained – the principles of intellectual analysis of the value of reliability coefficients, validity coefficients, discriminability coefficient, the difficulty index of the task of assessing the knowledge of subjects of training are formulated. **Conclusions:** the application of methods of formalization of test evaluation, analysis of software requirements, software development confirm the need to introduce quantitative methods of assessing students' knowledge into educational practice. The introduction of quantitative methods involves the correct setting of control goals, the selection of the measurement object and the selection of measurement tools. The use of pedagogical tests contributes to the effective implementation of all control functions and corresponds to its main principles for solving the problem of assessing the quality of tests, based on distributed virtual learning models and analysis methods. A test example of calculations was developed, with the help of which the efficiency of the proposed methods was researched.

Keywords: software engineering; knowledge bases; algebra of finite predicates; use of knowledge; rules.

Бібліографічні описи / Bibliographic descriptions

Ляшик В. А., Шубін І. Ю. Методи автоматизації та інтерпретації якості освітніх тестів у системах розподіленого віртуального навчання. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 40–51. DOI: <https://doi.org/10.30837/ITSSI.2023.25.040>

Liashyk, V., Shubin, I. (2023), "Methods of automation and interpretation of the quality educational tests in distributed virtual learning systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 40–51. DOI: <https://doi.org/10.30837/ITSSI.2023.25.040>

O. MOZHAIEV, Y. GNUSOV, O. MANZHAI, V. STRUKOV, V. NOSOV, V. RADCHENKO, S. YENHALYCHEV

STEGANOGRAPHIC METHOD OF ACOUSTIC INFORMATION PROTECTION IN CRITICAL APPLICATIONS SYSTEMS

The subject of the study is the process of protecting acoustic information in critical computer systems to ensure the required level of system security. **The purpose of the article** is to study the method of protecting acoustic information in critical computer systems by means of masking to ensure the impossibility of unauthorized access to the system. The paper outlines the following **tasks**: to analyze the software and hardware masking of speech; to study the masking of speech messages in order to introduce unrecognizability; to consider the features of speech message compression; to investigate methods of covert transmission of acoustic information. **The results** of the work, which were obtained using mathematical methods of information transformation in computer systems, are potentially possible methods of masking speech messages to ensure the impossibility of unauthorized access to the system. The analysis of the functioning of the proposed methods made it possible to formulate specific **conclusions**. The research has shown that the use of direct expansion of the spectrum of discrete signals for steganographic purposes helps to covertly embed information messages in still images. The task of extracting a message on the receiving side of a steganography system is equivalent to the task of detecting information from a mixture of a useful signal and an interference in a broadband communication system. The research revealed certain disadvantages of steganographic systems with an expanded spectrum of discrete signals: the probability of correct extraction of embedded data depends on the amount of distortion introduced, which depends on the provided bandwidth of the steganographic channel. Further research is desirable to analyze the possible use of methods for synthesizing large ensembles of quasi-orthogonal discrete signals with improved ensemble, structural, and correlation properties to ensure higher security of acoustic channels in computer systems for critical applications.

Keywords: acoustic information; technical protection; cryptographic protection; steganographic (steganophonic) protection; modification; computer system languages.

Introduction

Protecting acoustic (speech) information is one of the most important tasks in the overall set of measures to ensure the information security of an object or institution.

The unique features of speech information (SI) circulating in closed rooms and outside them: a large volume and speed of exchange, high confidentiality of some messages, the ability to identify the person making the message, and even the ability to determine the personal attitude to the information being voiced and to draw up his or her psychological portrait determine the relevance and extreme importance of solving the problem of protecting confidential speech information (CSI). Despite the growing role of automated information systems, speech information still plays a key role in information traffic (up to 80% of the total information flow) [1, 2]. This is especially important now, in the context of Russia's military aggression. Therefore, in recent years, more and more attention has been paid to ensuring the security of acoustic information. On the one hand, this is due to the high polyinformativeness of acoustic information. On the other hand, it is due to the variety of information threats to acoustic (speech) information and the peculiarities of their development

and implementation scenarios. All of this is reflected in a wide variety of modern methods, algorithms, software and hardware for protecting acoustic information from unauthorized access. The main directions of acoustic information protection are considered to be technical, cryptographic and steganographic (steganophonic) protection.

In a separate section, we consider the issue of protecting acoustic information by masking acoustic information based on modern computer technologies. In recent years, this area has been gaining more and more practical interest among software manufacturers. In order to provide basic security services for audio signals, complex software systems are being created, new methods for receiving, transmitting, processing, and presenting audio signals are being developed and used. Therefore, this article is devoted to the analysis of methods for protecting acoustic information existing in modern information systems of critical applications to ensure a higher level of security of such systems.

Literature review

Threats to speech information are usually realized by technical leakage channels, namely: acoustic, acoustic-vibration (vibroacoustic), acousto-optoelectronic

(laser acoustic), acoustoelectric, videoacoustic, high-frequency imposition. These channels can use a wide range of portable technical reconnaissance equipment [3, 4]:

- portable sound recording equipment (small-sized voice recorders, tape recorders and microphone-based recording devices)
- electronic stethoscopes;
- electronic devices for intercepting speech information (embedded devices) with microphone and contact type sensors with the transmission of intercepted information via radio, optical (in the infrared wavelength range) and ultrasonic channels, power supply network, telephone lines, connecting lines of auxiliary digital circuitry);
- technical means or special lines;
- optoelectronic acoustic systems, etc.

The main technical methods of protecting speech information today are:

- information concealment, which involves technical closure (frequency, time and combined analog scrambling);
- energy concealment, which is carried out by means of sound insulation, sound absorption, jamming in the premises (spatial jamming) and jamming of the functional communication channel with interference that masks speech signals (linear jamming);
- cryptographic concealment based on the discretization of speech information with subsequent encryption and reverse conversion to an analog signal or digital transmission over a communication channel;
- vocoder protection of speech information.

The study of steganographic concealment of speech information in data transmission channels, which can also be realized after cryptographic transformations, is of interest.

Recently, both developers and consumers of semantic protection of acoustic information have been observing an increasingly steady trend towards the use of new computer technologies for ensuring the security of speech communications without the use of classical cryptographic methods. In this regard, computer-based technologies for masking acoustic information are becoming increasingly attractive. However, one should not forget about cryptographic methods of protecting acoustic information, in particular: instantaneous cryptanalysis of GSM with only ciphertext [5]; real-time cryptanalysis of the assumed A5 stream cipher [6, 7]; cryptanalysis of anomalous behavior of a computer

system [8]; crypto-resistant methods and random number generators in Internet of Things (IoT) devices [9].

The aim of the article is to analyze the methods of protecting acoustic information in computer systems of critical applications by means of masking to ensure the impossibility of unauthorized access to the system:

- to achieve this goal, the following tasks need to be performed:
- analyze software and hardware speech masking;
- to investigate the masking of speech messages in order to introduce unrecognizability;
- analyze the features of speech message compression;
- to investigate methods of covert transmission of acoustic information.

1. Spectrum expansion methods to improve the efficiency of discrete message transmission

In today's discrete message transmission systems, two methods of spectrum expansion are used:

- pseudo-random *Frequency Hopping Spread Spectrum* method (FHSS). Its essence lies in the periodic jump-like change of frequency, which is carried by a certain algorithm to a known receiver and transmitter. The advantage of this method is its simplicity of implementation; it is used in *Bluetooth*;

- *Direct Sequence Spread Spectrum* (DSSS) method. It is more efficient than the FHSS method, but more difficult to implement. The essence of the method is to increase the modulation clock frequency, in which case each symbol of the transmitted message corresponds to a sufficiently long pseudo-random sequence (PRS). The method is used in systems such as *CDMA* and *IEEE 802.11*.

Expanding the spectrum by pseudo-randomly tuning the operating frequency. In order to prevent radio communication from being intercepted or jammed by narrowband noise, it was proposed to transmit with a constant change of carrier within a wide frequency range. As a result, the signal power was distributed over the entire range, and listening to a particular frequency produced only a small amount of noise. The sequence of carrier frequencies was pseudo-random, known only to the transmitter and receiver. Attempting to jam the signal in a narrow range also did not degrade the signal too much, since only a small part of the information was jammed. The idea of this method is illustrated in Fig. 1.

During a fixed time interval, transmission is carried out on a constant carrier frequency. At each carrier frequency, standard modulation techniques such as *FSK* or *PSK* are used to transmit discrete information. In order for the receiver to synchronize with the transmitter, synchronization tones are transmitted for a period of time to mark the beginning of each transmission period. Consequently, the useful rate of this coding method is lower due to the constant synchronization overhead.

The carrier frequency varies according to the frequency subchannel numbers produced by the pseudorandom number algorithm. The pseudorandom sequence depends on a certain parameter called the initial number. If the receiver and the transmitter know the algorithm and the value of the initial number, they change frequencies in the same sequence, which is called the pseudorandom frequency hopping sequence.

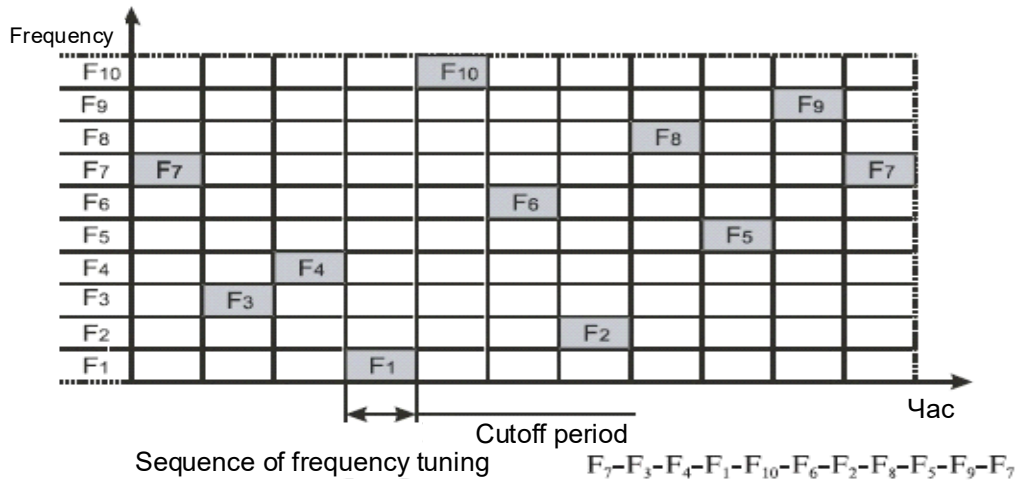


Fig. 1. Spectrum expansion by frequency hopping tuning

FHSS methods are used in *IEEE 802.11* and *Bluetooth* wireless technologies. In *FHSS*, the approach to using the frequency range is not the same as in other coding methods – instead of economically using a narrow band, an attempt is made to occupy the entire available range. At first glance, this does not seem very efficient, since only one channel is operating in the band at any given time. However, the latter statement is not always true – spread spectrum codes can also be used to multiplex multiple channels over a wide range. In particular, *FHSS* methods make it possible to organize the simultaneous operation of several channels by selecting such pseudo-random sequences for each channel so that at any given time each channel operates at its own frequency (of course, this can only be done if the number of channels does not exceed the number of frequency subchannels).

Spectrum expansion using the direct sequence method. The direct sequential spread spectrum method also uses the entire frequency range allocated for one communication line. Unlike the *FHSS* method, the entire frequency range is occupied not by constantly switching from frequency to frequency, but by replacing each bit of information with *N* bits, so that the signal transmission rate increases by *N* times. This means that

the signal spectrum is also expanded by a factor of *N*. It is enough to choose the appropriate data rate and *N* value so that the signal spectrum fills the entire range.

To transmit data in a broadband communication system, an information signal $x(t) = \begin{cases} +1 \\ -1 \end{cases}$ is modulated by multiplying it by an expansion coded signal $g(t) = \Phi_i \in \Phi$ – a pseudo-random sequence of the above-mentioned ensembles of discrete signals. Since the code signal is similar to noise in its statistical properties, the resulting expanded signal

$$y'(t) = y(t) + e(t) \tag{1}$$

In the process of receiving in the demodulator, the received signal $y'(t) = y(t) + e(t)$ as a mixture of the transmitted sequence $y(t)$ and events in the error channel $e(t)$ is multiplied by a synchronized copy of the expanded signal $g(t)$. In other words, the correlation coefficient is calculated on the receiving side, the value of which determines the decision-making rule:

$$\rho(y'(t), g(t)) = \frac{1}{n} \sum_{z=0}^{n-1} x(t)\Phi_z \Phi_z + \frac{1}{n} \sum_{z=0}^{n-1} e(t)\Phi_z, \tag{2}$$

differ slightly from the noise in the communication channel, which makes it possible to carry out covert transmission.

Due to the pseudo-randomness of Φ_i , used as $g(t)$, the second terms in the right-hand side of the equation can be neglected (the number of "+1" is approximately equal to the number of "-1"), i.e.

$$\rho(y'(t), g(t)) \approx \rho(y(t), g(t)) = x(t) \frac{1}{n} \sum_{z=0}^{n-1} (\Phi_i)^2 = x(t), \quad (3)$$

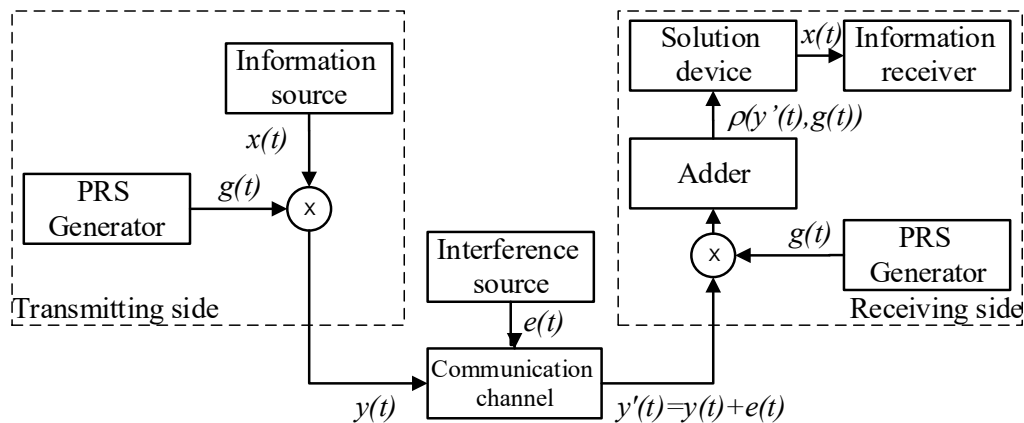


Fig. 2. Block diagram of the information transmission track using direct spectrum expansion

Suppose that the time interval of the unmodulated signal $x(t)$ is equal to T , and its frequency, respectively, is $F(x(t)) = 1/T$. Transmission of a modulated signal $y(t)$ with the same time interval T will lead to a broadening of the frequency spectrum of the transmitted signal proportional to the number of elements of the pseudorandom sequence, i.e., proportional to the length n :

$$F(y(t)) = n \frac{1}{T} = nF(x(t)).$$

However, the use of direct expansion of the transmitted signal spectrum ensures the simultaneous transmission of many other information signals in the same frequency band. This follows from the mutual orthogonality (quasi-orthogonality) of the used ensembles of discrete signals. Indeed, if an additive mixture $\sum_{\ell} y_{\ell}(t)$ of several modulated signals is received at the receiving end, then the calculation of the correlation coefficient will give the following:

$$\rho\left(\sum_{\ell} y_{\ell}(t), g(t)\right) = \frac{1}{n} \sum_{\ell} \sum_{z=0}^{n-1} x_{\ell}(t) \Phi_{\ell_z} \Phi_{i_z}. \quad (5)$$

that is, the value of the information signal on the receiving side is determined according to the expression

$$x(t) = \begin{cases} +1, & \text{if } \rho(y'(t), g(t)) \approx +1; \\ -1, & \text{if } \rho(y'(t), g(t)) \approx -1, \end{cases} \quad (4)$$

where the " \approx " sign implies the presence of errors caused by natural or intentional interference in the communication channel.

The block diagram of the information transmission path using direct spectrum expansion is shown in Fig. 2.

But all the sequences in the set have a low value of mutual correlation, i.e., under the condition $\ell \neq i$ we have (for orthogonal signals we have the equality $\rho(\Phi_{\ell}, \Phi_i) = 0$). So, all the terms if $\ell \neq i$, on the right side of equality (5) can be neglected. Hence, in the presence of a discrete signal in the additive mixture, we have expression (3) and the corresponding decision-making rule (4).

The purpose of *DSSS* coding is the same as that of *FHSS*: to increase immunity to interference. Narrowband interference will distort only certain frequencies of the signal spectrum, so that the receiver is more likely to correctly recognize the transmitted information.

A code that replaces a binary unit of the original information is called an expansion sequence, and each bit of such a sequence is called a chip (elementary signal). Accordingly, the transmission rate of the resulting code is called the chip rate. The binary zero is encoded by the inverse value of the expansion sequence. Receivers need to know the expansion sequence used by the transmitter to understand the information being transmitted.

The number of bits in the expansion sequence determines the expansion factor of the original code.

As with FHSS, any type of modulation, such as BFSK, can be used to encode the bits of the resulting code.

The higher the expansion factor, the wider the spectrum of the resulting signal and the higher the degree of interference suppression. But in this case, the spectrum range occupied by the channel increases. Usually, the expansion factor has a value from 10 to 100.

Let's list some of the properties of signals with direct spectrum expansion that are most important from the point of view of organizing multiple access in communication systems with mobile objects.

1. *Multiple access.* If several subscribers use the transmission channel at the same time, then there are several signals with direct spectrum expansion in the channel at the same time. The receiver of a particular subscriber's signal performs the opposite operation – convolution of this subscriber's signal by using the same pseudo-random signal that was used in the transmitter of this subscriber. This operation concentrates the power of the received broadband signal back into a narrow frequency band equal to the spectral width of the information symbols. If the mutual correlation function between the pseudo-random signals of this subscriber and other subscribers is sufficiently small, then in the process of coherent reception, only a small fraction of the power of the signals of other subscribers will fall into the information band of the subscriber's receiver. The signal of a particular subscriber will be received correctly.

2. *Multiple beam interference.* If the pseudorandom signal used to expand the spectrum has a perfect autocorrelation function whose value outside the interval $[-t_0, +t_0]$ is zero, and if the received signal and a copy of this signal in another beam are shifted in time by more than $2t_0$, then, subject to signal convolution, its copy can be considered as interference that adds only a small fraction of power to the information band.

3. *Narrowband interference.* In the case of coherent reception, the received signal is multiplied in the receiver by a copy of the pseudorandom signal used to expand the spectrum in the transmitter. Thus, the receiver will perform a narrowband interference spectrum expansion operation similar to the one performed with the information signal in the transmitter. Consequently, the spectrum of the narrowband interference in the receiver will be expanded by a factor of B , where B is the expansion factor, so that only a small fraction of the interfering power, B times less than the original interfering power, will be transmitted to the information band.

4. *Interception probability.* Since a direct spread spectrum signal occupies the entire frequency band of the system during the entire transmission time, its radiated power per 1 Hz of the band will be very small. Therefore, detecting such a signal is a very difficult task.

Therefore, a promising direction in the development of modern broadband communication systems with direct spectrum expansion is the development and research of methods for synthesizing large ensembles of quasi-orthogonal discrete signals with improved ensemble, structural, and correlation properties.

The considered approach to the organization of digital jamming communication channels was applied in the construction of steganographic methods of information protection. For example, spectrum expansion by direct sequence was used to create a steganographic method for embedding information in still images. Let's consider one of the variants of this method implementation, authored by *J. Smith* and *W. Comiskey* [10]. We will study its effectiveness in terms of the provided throughput of the steganographic communication channel and the achieved resistance to unauthorized extraction of information messages.

2. Direct spectrum expansion in steganography

In the Smith-Comiskey method, as in the direct spread spectrum communication systems discussed above, an information message is bitwise modulated by multiplying it by an ensemble of orthogonal signals. The modulated message is then embedded in a container – a still image.

Let us introduce some conventions and mathematical relations, which, by analogy with the broadband digital communication systems discussed above, will allow us to study the features of construction and information exchange in a steganosystem [11–13].

Let us represent the information message m , to be embedded in a digital image container in the form of blocks m_i of equal length, i.e. $m = (m_0, m_1, \dots, m_{N-1})$, where each block m_i is a sequence (vector) of n bits:

$$m_i = (m_{i0}, m_{i1}, \dots, m_{in-1}).$$

We consider an image container as a set of data C of dimension $K \times L$ divided into subblocks of size $k \times l = n$. The elements of the array C can be, for example, raster data of the image used.

The secret key data is a set of basic functions

$$Key = \Phi = \{\Phi_0, \Phi_1, \dots, \Phi_{M-1}\},$$

where all basic functions $\Phi_i = (\phi_{i_0}, \phi_{i_1}, \dots, \phi_{i_{n-1}})$ – mutually orthogonal discrete signals with a length equal to the message block n size, that is, for any $i, j \in [0, \dots, M-1]$ equality m_i is fulfilled

$$\rho(\Phi_i, \Phi_j) = \frac{1}{n} \sum_{z=0}^{n-1} \Phi_{i_z} \Phi_{j_z} = \begin{cases} +1, & \text{if } i = j; \\ -1, & \text{if } i \neq j. \end{cases}$$

A formal graphical representation of the information message, image container, and key data is shown in Fig. 3.

The purpose of steganographic information transformation is to embed each individual message block m_i in the corresponding block of the image container.

The data block of a digital image with the dimension of $K \times L$ elements can contain $K \times \frac{L}{n}$ blocks of an information message, i.e., up to $K \times L$ bits.

The division of the container into blocks can be arbitrary, but as practice shows, the most appropriate

(smaller numerical spread of values in a block, as opposed to a one-dimensional representation) is the two-dimensional division shown in Fig. 3. As the key data (an array of basis functions $Key = \Phi$), we use the above-mentioned ensembles of orthogonal discrete Walsh–Hadamard signals.

The embedding of the information message is carried out as follows. Each message block $m_{i_j}, j = 0, \dots, n-1$ is mapped to a separate block of the image container. All information bits of the block $m_{i_j}, j = 0, \dots, n-1$ are represented as an information signal

$$m_{i_j}(t) = \begin{cases} +1, & m_{i_j} = 1; \\ -1, & m_{i_j} = 0 \end{cases} \text{ and modulated by an expansion}$$

code signal (basis functions), i.e. PRS $\Phi_j \in \Phi$.

Thus, a modulated information signal is generated for each information block:

$$E_i(t) = \sum_{j=0}^{n-1} \sum_{z=0}^{n-1} m_{i_j}(t) \Phi_{j_z}. \quad (6)$$

The received message block E_i is pixel-by-pixel summed with the container subblock.

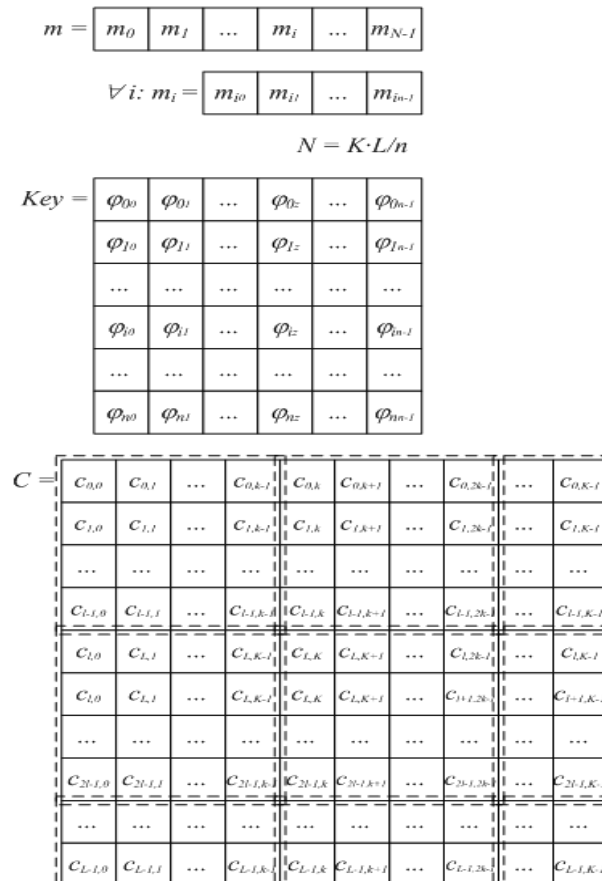


Fig. 3. Formal representation of the information message, image container, and key data

Let's designate the container blocks as follows (see Figure 3):

$$C_0 = \begin{pmatrix} c_{0,0} & c_{0,1} & \dots & c_{0,k-1} \\ c_{1,0} & c_{1,1} & \dots & c_{1,k-1} \\ \dots & \dots & \dots & \dots \\ c_{\ell-1,0} & c_{\ell-1,1} & \dots & c_{\ell-1,k-1} \end{pmatrix}, C_1 = \begin{pmatrix} c_{0,k} & c_{0,k+1} & \dots & c_{0,2k-1} \\ c_{1,k} & c_{1,k+1} & \dots & c_{1,2k-1} \\ \dots & \dots & \dots & \dots \\ c_{\ell-1,k} & c_{\ell-1,k+1} & \dots & c_{\ell-1,2k-1} \end{pmatrix}, \dots, C_{N-1} = \begin{pmatrix} c_{L-l-1,K-k-1} & c_{L-l-1,K-k} & \dots & c_{L-l-1,K-1} \\ c_{L-l,K-k-1} & c_{L-l,K-k} & \dots & c_{L-l,K-1} \\ \dots & \dots & \dots & \dots \\ c_{L-1,K-k-1} & c_{L-1,k+1} & \dots & c_{L-1,K-1} \end{pmatrix}.$$

The corresponding modulated information signals $E_i(t)$ are presented in the form of a two-dimensional data array:

$$E_i = \begin{pmatrix} E_{i_0} & E_{i_1} & \dots & E_{i_{k-1}} \\ E_{i_k} & E_{i_{k+1}} & \dots & E_{i_{2k-1}} \\ \dots & \dots & \dots & \dots \\ E_{i_{(\ell-1)(k-1)-k+1+n-k+1}} & E_{i_{(\ell-1)(k-1)-k+2+n-k+2}} & \dots & E_{i_{(\ell-1)(k-1)-n-1}} \end{pmatrix}, i = 0, \dots, N-1.$$

Then the steganogram (filled container) is formed by combining data arrays S_i , $i = 0, \dots, N-1$:

$$S_i = C_i + E_i \times G, \quad (7)$$

where $G > 0$ – gain of the expansion signal, which sets the "energy" of the embedded bits of the information sequence.

Thus, a filled container S is generated from the formed blocks S_i , $i = 0, \dots, N-1$ by combining them, as shown in Fig. 3, for the initial (empty) container C .

At the stage of data construction, it is not necessary to have information about the primary container C . The decoding operation consists in recovering the hidden message by projecting each block S_i , of the obtained steganogram S onto all the basic functions $\Phi_j \in \Phi$, $i = 0, \dots, N-1$. To do this, each block S_i is represented as a vector $S_i = (S_{i_0}, S_{i_1}, \dots, S_{i_{n-1}})$, $i = 0, \dots, N-1$.

To extract the j -th bit of the message from the i -th block of the steganomagnetic image, it is necessary to calculate the correlation coefficient between Φ_j and the received block S_i (represented as a vector):

$$\rho(S_i, \Phi_j) = \frac{1}{n} \sum_{z=0}^{n-1} S_{i_z} \Phi_{j_z} = G \cdot \frac{1}{n} \sum_{z=0}^{n-1} E_{i_z} \Phi_{j_z} + \frac{1}{n} \sum_{z=0}^{n-1} C_{i_z} \Phi_{j_z}, \quad (8)$$

where C_i – is a one-dimensional array, i.e., the corresponding block of the container, represented as a vector.

Let's assume that the array has a random statistical structure, i.e., the second term on the right-hand side of expression (8) is close to zero and can be ignored.

Then we have:

$$\rho(S_i, \Phi_j) \approx G \cdot E_i \cdot \Phi_j = G \cdot \sum_{\ell=0}^{n-1} \sum_{z=0}^{n-1} m_{i_x}(t) \cdot \Phi_{i_z} \Phi_{j_z}. \quad (9)$$

By analogy with (8), we note that all sequences from the set Φ are mutually orthogonal, i.e., provided $\ell \neq j$ we have $\rho(\Phi_i, \Phi_j) = 0$. Consequently, all terms on the right-hand side of equation (9) can be neglected if $\ell \neq j$. Hence, we have:

$$\rho(S_i, \Phi_j) \approx G \cdot m_{i_j}(t) \cdot \frac{1}{n} \sum_{z=0}^{n-1} (\Phi_{j_z})^2 = G \cdot m_{i_j}(t). \quad (10)$$

According to the rule of useful signal extraction:

$$x(t) = \begin{cases} "1", & \text{if } polarity > 0; \\ "0", & \text{if } polarity < 0; \\ \text{external signal,} & \text{if } polarity = 0, \end{cases} \quad (11)$$

values $m_{i_j}(t)$ can be easily recovered using the sign function ($polarity$ – the polarity of the peak of the correlation function).

Since $G > 0$ and $n > 0$ the sign $\rho(S_i, \Phi_j)$ in (10) depends only on $m_{i_j}(t)$. Hence we have:

$$m_{i_j}(t) = \text{sign}(\rho(S_i, \Phi_j)) = \begin{cases} -1, & \text{if } \rho(S_i, \Phi_j) < 0; \\ +1, & \text{if } \rho(S_i, \Phi_j) > 0; \\ ?, & \text{if } \rho(S_i, \Phi_j) = 0. \end{cases} \quad (12)$$

If $\rho(S_i, \Phi_j) = 0$ at (10) we assume that the embedded information has been lost.

A block diagram of embedding information in an image container using direct spectrum expansion for covert message transmission is shown in Fig. 4.

Fig. 4 shows that the process of embedding information messages for covert transmission is very similar to the process of expanding the spectrum of discrete signals in communication systems (see Fig. 2). The element-by-element assembly of a modulated message $E(t)$ with an image container $C(t)$ should be interpreted as the imposition of errors $e(t)$ on the useful signal in the communication channel $y(t)$. The task of constructing a message $m(t)$ from $S(t)$ at the receiving side of the steganosystem is equivalent to the task of

detecting $x(t)$ mixture of a useful signal and an interference $y'(t) = y(t) + e(t)$ in a broadband communication system. That is, the considered steganosystem inherits all the advantages of broadband communication systems: resistance to unauthorized extraction of embedded messages (analogous to concealment in a communication system), resistance to destruction or modification of embedded messages (analogous to interference protection), resistance to imposing false messages (analogous to imitation resistance in a communication system).

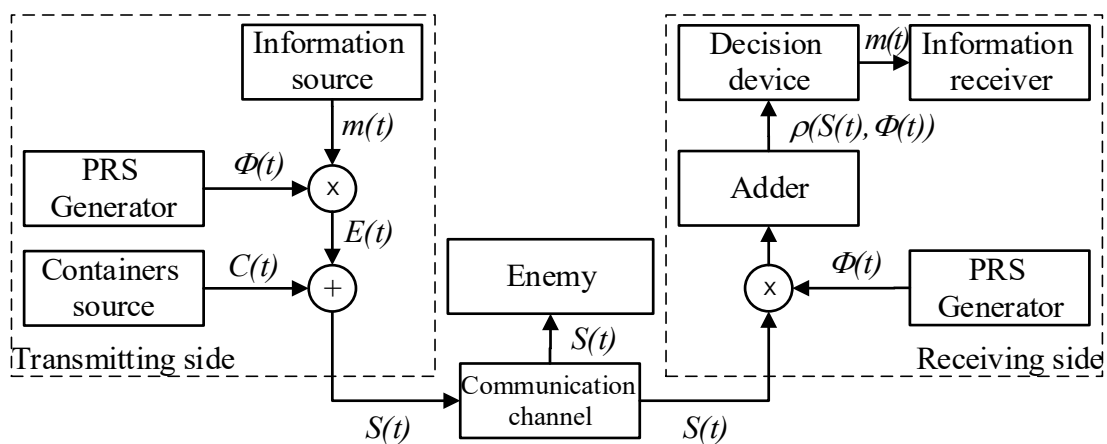


Fig. 4. Block diagram of embedding information in an image container for covert messaging

Thus, the use of direct expansion of the spectrum of discrete signals makes it possible to embed information in still images for covert transmission and thus implement steganographic information protection.

3. Evaluating the effectiveness of the steganosystem

The effectiveness of a technical system is broadly understood as the correspondence of the result of a certain operation to the required parameter. In this case, the technical system is a means of implementing the operation under study [14, 15].

In accordance with this process, a steganographic system is a technical means of implementing an operation aimed at hiding the fact of covert information transmission from the enemy. So, taking into account the functional purpose of the steganographic system, we will introduce performance indicators.

1. Bandwidth is the ratio of the volume V of information embedded in the container to the total volume of the container D

$$Q = V/D. \quad (13)$$

2. Key data volume (in bits)

$$\ell_{Key} = \log_2(|Key|), \quad (14)$$

where $|Key|$ – power of the key data set.

3. The strength of a steganographic method will be evaluated as the inverse of the power of the set of secret key data. It can be interpreted as a probabilistic indicator of the selection of a secret key:

$$W = 1/|Key| = 2^{-\ell_{Key}}. \quad (15)$$

4. The amount of distortion added as a percentage of the arithmetic mean of all absolute values of the container data Δ -changes to the maximum possible value Δ_{max} :

$$I = \frac{\Delta_{cp}}{\Delta_{max}} \cdot 100 = \frac{100}{\Delta_{max} \cdot D} \cdot \sum_{i=1}^D |\Delta_i|, \quad (16)$$

where Δ_i – Δ -changes of the i -th container element.

5. Probability of erroneous construction of message information data

$$P_{ouu} = \lim_{D \rightarrow \infty} \frac{V_{ouu}}{D} = 1 - \lim_{D \rightarrow \infty} \frac{V - V_{ouu}}{D}, \quad (17)$$

where V_{ouu} – is the amount of falsely constructed data.

Using (13)–(17), let us evaluate the effectiveness of the considered steganographic method of information protection.

1. **Bandwidth capacity.** For each n -element block S_i of the filled container (steganogram), there is a n -bit vector of the embedded message m_i . Thus, $Q=1/B$, where B is the amount of data per element of the container. To embed an image in raster data (color model R, G, B) with 8-bit encoding of each color, we have $B=8$ and $Q=1/8$.

2. **Key data volume.** The key data is an ensemble of discrete signals formed by the rows of the Hadamard matrix of order n . Thus, the set of key data should be understood as a set of different (non-isomorphic) Hadamard matrices, each of which represents an ensemble of discrete signals.

The given power estimates M_A give an estimate of the number of Walsh–Hadamard ensembles of discrete signals, i.e., an estimate of the power of non-equivalent keys of a quilted system. Thus, the amount of key data is estimated as $l_{key} = \log_2(M_A)$.

3. **Probability of selecting a secret key**
 $W = (M_A)^{-1}$.

4. To estimate the amount of **distortion being introduced**, let's use expression (16). The second term on the right-hand side of (16) determines the magnitude of the Δ -changes in the container data elements. The multiplier E_i is formed by summing the discrete signals (which take the value ± 1) with the corresponding polarities (given by $m_{ij}(t)$). Thus, all E_i elements will take values in the range $[-n, \dots, +n]$, and the corresponding Δ -changes of the container elements will not exceed $|\Delta_i| \leq n \cdot G$. Hence, we have an upper bound on the value of the added distortion:

$$I = \frac{\Delta_{cp}}{\Delta_{max}} \cdot 100 \leq \frac{n \cdot G}{\Delta_{max}} \cdot 100. \quad (18)$$

To embed into raster data an image (R, G, B color model) with 8-bit encoding of each color and using discrete signals $n=256$, even with the added distortion $G=1$, can reach 100%. It is possible to reduce the added distortion by decreasing the number of embedded data bits m_{ij} , which will inevitably lead to a decrease in the bandwidth of the steganographic communication channel.

5. **Probability of erroneous extraction.** The information message extraction, as well as in the case

of the organization of anti-jamming communication (see (7)–(12)), is carried out by the correlation method (see (13)–(17)). Therefore, the extraction error will occur if the sign of the correlation coefficient $\rho(S_i, \Phi_j)$ in the expression (17) is changed.

The coefficient $\rho(S_i, \Phi_j)$ is expressed as:

$$\rho(S_i, \Phi_j) = \rho(C_i + E_i \cdot G, \Phi_j) = \rho(C_i, \Phi_j) + \rho(E_i \cdot G, \Phi_j).$$

The last term does not change the sign of $\rho(S_i, \Phi_j)$, event $\rho(S_i, \Phi_j) = \rho(E_i \cdot G, \Phi_j)$ corresponds to error-free message retrieval (see (11), (12)).

Thus, an error in extracting the message information bit m_{ij} will occur if the event happens

$$|\rho(C_i, \Phi_j)| > \rho|E_i \cdot G, \Phi_j| = |G \cdot m_{ij}| = G, \quad (19)$$

that is, when the absolute value of the correlation coefficient used to embed a bit m_{ij} of a discrete signal Φ_j in the block of the container C_i in which this bit is embedded exceeds the amplification coefficient G .

So, we write:

$$P_{er} = P(|\rho(C_i, \Phi_j)| > G),$$

where $P(x)$ – the probability of a random event x occurring.

In other words, the correct extraction of an embedded message is a random event, the probability $P_{er,fr}$ of which is directly related to the statistical properties of the image container used. To extract the message correctly

$$P_{er} = 0, P_{er,fr} = 1 - P_{er} = 1, \quad (20)$$

for error-free message extraction, it is necessary to strive for mutual orthogonality of separate image fragments C_i and to use discrete signals Φ_j as secret keys.

In this case, the event

$$|\rho(C_i, \Phi_j)| = 0 < G$$

is valid for every $i=0, \dots, N-1$ and is executed (20).

At the same time, experimental studies have shown that the correlation coefficient is usually much higher than zero $|\rho(C_i, \Phi_j)| \gg 0$ and the event (19) occurs very often. The fact is that the elements of discrete signals $\Phi_j \in \Phi$ take the value of $\begin{cases} +1 \\ -1 \end{cases}$, and the corresponding normalized correlation coefficient $\rho(\Phi_i, \Phi_j)$ in absolute value does not exceed the

length n of the sequence and lies in the range of $[0, \dots, 1]$, from which condition (20) actually follows.

However, the elements of the container take values from a numeric field $[0, \dots, Y]$, which dimensionality is set by the way the image data is encoded. For example, when embedding information in raster image data (R, G, B color model) with 8-bit encoding of each color, the corresponding C_i values take on the range of integers $[0, \dots, 255]$. That is, the absolute value normalized with respect to the n correlation coefficient $|\rho(C_i, \Phi_j)|$ will lie in the range of $[0, \dots, Y]$, and for the error-free extraction of all bits of the message (20), the condition $G > Y$ must be met.

As studies have shown, an increase of G leads to an inevitable increase in the value of the added distortion (19). Under the condition $I > 2...3\%$ (the threshold of human visual sensitivity), they become noticeable to an outside observer, which compromises the steganochannel and makes it impossible to use the considered steganosystem.

Conclusions

Thus, in the course of research, the contradictions underlying the development and use of steganographic systems with an expanded range of discrete signals were revealed:

- the probability of correct extraction of embedded data $P_{er.fr}$ directly depends on the amount of added distortion I ;
- the amount of added distortion I , directly depends on the amount of embedded data bits, i.e. on the bandwidth of steganochannel Q ;
- probability of correct extraction of embedded data $P_{er.fr}$ directly depends on the statistical properties of the image container used.

References

1. Kosenko, V. (2017), "Principles and structure of the methodology of risk-adaptive management of parameters of information and telecommunication networks of critical application systems", *Innovative technologies and scientific solutions for industries*, No 1 (1), P. 75–81. DOI: <https://doi.org/10.30837/2522-9818.2017.1.046>
2. Kosenko, V. (2017), "Mathematical model of optimal distribution of applied problems of safety-critical systems over the nodes of the information and telecommunication network", *Advanced Information Systems*, Vol. 1, No. 2, P. 4–9. DOI: <https://doi.org/10.20998/2522-9052.2017.2.01>
3. Ivanchenko, S., Havrylenko, O., Lipskyi, O., Shevtsov, A. "Technical channels of information leakage. Procedure for creating complexes of technical information protection". ["Tekhnichni kanaly vytku informatsii. Poriadok stvorennia

The following empirical estimates were obtained as a result of the research:

- dependence of the added distortion value I , of the steganochannel Q bandwidth;
- dependence of the added distortion value I , and the frequency of extraction errors $P_{er}^* \approx P_{er}$ from the amplification coefficient G ;
- dependence of the added distortion value I from the frequency of extraction errors $P_{er}^* \approx P_{er}$.

The research was conducted under the conditions of embedding information in raster image data (R, G, B color model) with 8-bit encoding of each color.

The analysis of the obtained dependencies confirms the conclusions made earlier, and the convergence of the experimental results with theoretical considerations indicates the reliability of the results.

The research has shown that the use of direct expansion of the spectrum of discrete signals for steganographic purposes makes it possible to covertly embed information messages in still images. The task of extracting a message on the receiving side of a steganography system is equivalent to the task of detecting information from a mixture of a useful signal and an interference in a broadband communication system.

The research has revealed the following disadvantages of steganographic systems with an expanded spectrum of discrete signals: the probability of correct extraction of embedded data depends on the amount of added distortion, which depends on the provided bandwidth of the steganographic channel. In other words, the practical construction of a steganosystem is associated with finding a compromise between the amount of added distortion, the probability of correct message retrieval on the receiving side, and the provided bandwidth. In addition, the research has established that the probability of correct extraction of embedded data directly depends on the statistical properties of the image container used.

комплексив технічного захисту інформації], Study guide K.: ISZZI NTUU "KPI", 2016. 104 p. available at: https://ela.kpi.ua/bitstream/123456789/15155/1/NP_Tekhnichni_kanalny_vytku_inf.pdf

4. Oleynikov, A. "Methods and means of information protection: Study guide for students of higher educational institutions". ["Metody ta zasoby zakhystu informatsii"]. Kharkiv. NTMT, 2014. 298 p. available at: <https://ref.nure.ua/navchalna-laboratorija-tehnichnogo-zahistu-informacii>

5. Nuzhny, S. M. (2018). "Improved technology for assessing the degree of protection of language information". Modern information protection. ["Udoskonalena tekhnolohiia otsinky stupenia zakhystu movnoi informatsii"]. Vol. 1 (33). P. 66–73. available at: <http://journals.dut.edu.ua/index.php/dataprotect/article/view/1796>

6. Blintsov, V., Nuzhnyi, S., Parkhuts, L., Kasianov, Y. (2018), "The objectified procedure and a technology for assessing the state of complex noise speech information protection". *Eastern-European Journal of Enterprise Technologies*, Vol. 5 (9 (95)). P. 26–34. DOI: <https://doi.org/10.15587/1729-4061.2018.144146>

7. Hrystak, A., Kinzeryavyu, V., Prysiaznyi, D., Burmak, Y., Samoylik, Y. "High-speed and hash function for blockchain security mechanisms". Scientific and practical cyber security journal (SPCSJ). Vol. 4(1). 2020. P. 65–70. available at: <https://journal.scsa.ge/ru/papers/high-speed-and-secure-hash-function-for-blockchain-security-mechanisms-3/>

8. Mozhaev, O., Semenov, S., Kuchuk, N., Mozhaev, M., Tiulieniev, S., Gnusov, Y., Yevstrat, D., Chyrva, Y., Kuchuk, H. (2022), "Development of a method for determining the general criteria of abnormal behavior of a computer system based on the improved criterion of uniformity of input data samples". *Eastern-European Journal of Enterprise Technologies*, Vol. 6 (4 (120)), P. 40–49. DOI: <https://doi.org/10.15587/1729-4061.2022.269128>

9. Mozhaev, O., Klimushyn, P., Solianyk, T., Gnusov, Y., Manzhai, O., Svitlychnyi, V. (2022), "Crypto-resistant methods and random number generators in internet of things (iot) devices". *Innovative technologies and scientific solutions for industries*. № 2 (20), P. 22–34 DOI: <https://doi.org/10.30837/ITSSI.2022.20.022>

10. Mozhaev, O., Klimushyn, P., Solianyk, T., Kolisnyk, T. (2021), "Potential application of hardware protected symmetric authentication microcircuits to ensure the security of internet of things". *Advanced Information Systems*. Vol. 5, No 3 P. 103–111. DOI: <https://doi.org/10.20998/2522-9052.2021.3.14>

11. Smith, J., Comiskey, B., (1996), "Modulation and information hiding in images", *Lecture Notes in Computer Science*. 1996. P. 207–226. DOI:10.1007/3-540-61996-8_42.

12. Klimushyn, P., Solianyk, T., Mozhaev, O., Nosov, V., Kolisnyk, T., Yanov, V. (2021), "Hardware support procedures for asymmetric authentication of the internet of things". *Innovative Technologies and Scientific Solutions for Industries*, No. 4 (18). P. 31–39. DOI: 10.30837/ITSSI.2021.18.031

13. Friedrich, J., Miroslav, G., Du., R. (2021), "Reliable Detection of LSB Steganography in Color and Grayscale Images". *Binghampton*, New York: SUNY. P. 27-30. DOI:10.1145/1232454.1232466

14. Fridrich, J., Du., R. and Long, M. (2000), "Steganalysis of LSB Encoding in Color Images", *ICME 2000*, New York City. DOI:10.1109/ICME.2000.871000

15. Brock, W., Dechert, W., Scheinkman, J. "A test for independence based on the correlation dimension", Working Paper, University of Wisconsin, 1987. available at: https://www.academia.edu/5825079/A_test_for_independence_based_on_the_correlation_dimension

16. Wu, H.C., Wu, N.I., Tsai, C.S., Hwang, M.S. (2005), "Image Steganographic Scheme Based on Pixel-Value Differencing and LSB Replacement Methods". *IEEE Transactions on Image and Signal Processing*. № 5. P. 611615. DOI:10.1049/ip-vis:20059022

Received 05.09.2023

Відомості про авторів / About the Authors

Можасєв Олександр Олександрович – доктор технічних наук, професор, Харківський національний університет внутрішніх справ, професор кафедри кібербезпеки та DATA-технологій, Харків, Україна; e-mail: mozhaev1957@gmail.com; ORCID ID: <https://orcid.org/0000-0002-1412-2696>

Гнусов Юрій Валерійович – кандидат технічних наук, доцент, Харківський національний університет внутрішніх справ, завідувач кафедри кібербезпеки та DATA-технологій, Харків, Україна; e-mail: duke6969@i.ua; ORCID ID: <https://orcid.org/0000-0002-9017-9635>

Манжай Олександр Володимирович – кандидат юридичних наук, професор, Харківський національний університет внутрішніх справ, завідувач кафедри протидії кіберзлочинності факультету № 4, Харків, Україна; e-mail: E-maisofist@ukr.net; ORCID ID: <http://orcid.org/0000-0001-5435-5921>

Струкєв Володимир Михайлович – кандидат технічних наук, доцент, Харківський національний університет внутрішніх справ, професор кафедри кібербезпеки та DATA-технологій, Харків, Україна; e-mail: struk_vm@ukr.net; ORCID ID: <https://orcid.org/0000-0003-4722-3159>

Носов Віталій Вікторович – кандидат технічних наук, доцент, Харківський національний університет внутрішніх справ, професор кафедри протидії кіберзлочинності факультету № 4, Харків, Україна; e-mail: vitnos@ukr.net; ORCID ID: <https://orcid.org/0000-0002-7848-6448>

Радченко Валерій Вікторович – кандидат фіз.-мат. наук, доцент, Харківський національний університет внутрішніх справ, доцент кафедри кібербезпеки та DATA-технологій, Харківський національний університет внутрішніх справ, Харків, Україна; e-mail: valeryradchenko2007@gmail.com; ORCID ID: <https://orcid.org/0000-0003-1420-4832>

Єнгаличев Сергій Олександрович – Харківський національний економічний університет ім. С. Кузнеця, аспірант кафедри кібербезпеки та інформаційних технологій, Харків, Україна; e-mail: engalichev.sergiy@hneu.net; ORCID ID: <https://orcid.org/0000-0001-5298-2251>

Mozhaiev Oleksandr – Doctor of Sciences (Engineering), Professor, Kharkiv National University of Internal Affairs, Professor at the Department of Cyber Security and DATA-Technologies, Kharkiv, Ukraine.

Gnusov Yurii – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Internal Affairs, Head at the Department of Cyber Security and DATA-Technologies, Kharkiv, Ukraine.

Manzhai Oleksandr – PhD (Juridical Sciences), Professor, Kharkiv National University of Internal Affairs, Head at the Department of Combating Cybercrime, Kharkiv, Ukraine.

Strukov Volodymyr – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Internal Affairs, Professor at the Department of Cyber Security and DATA-Technologies, Kharkiv, Ukraine.

Nosov Vitalii – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Internal Affairs, Professor at the Department of Combating Cybercrime, Kharkiv, Ukraine.

Radchenko Valery – PhD (Physical and Mathematical Sciences), Associate Professor, Kharkiv National University of Internal Affairs, Associate Professor at the Department of Cyber Security and DATA-Technologies, Kharkiv, Ukraine.

Yenhalychev Serhii – Simon Kuznets Kharkiv National University of Economics, Graduate Student at the Department of Cybersecurity and Information Technologies, Kharkiv, Ukraine.

СТЕГANOГРАФІЧНИЙ МЕТОД ЗАХИСТУ АКУСТИЧНОЇ ІНФОРМАЦІЇ В СИСТЕМАХ КРИТИЧНОГО ЗАСТОСУВАННЯ

Предмет дослідження – процес захисту акустичної інформації в комп'ютерних системах критичного застосування для забезпечення необхідного рівня безпеки системи. **Метою** статті є вивчення методу захисту акустичної інформації в комп'ютерних системах критичного застосування за допомогою маскуванню для забезпечення неможливості несанкційного доступу до системи. У роботі окреслено такі **завдання**: проаналізувати програмно-технічне маскуванню мови; дослідити маскуванню мовних повідомлень з метою введення невпізнання; розглянути особливості стиснення мовних повідомлень; дослідити методи прихованої передачі акустичної інформації. Застосовано математичні **методи** перетворення інформації у комп'ютерних системах. **Результатом роботи** є потенційно можливі методи маскуванню мовних повідомлень для унеможливлення несанкційного доступу до системи. Аналіз функціонування запропонованих методів дав змогу сформулювати конкретні **висновки**. Дослідження показали, що використання в стеганографічних цілях прямого розширення спектра дискретних сигналів допомагає здійснити приховане вбудовування інформаційних повідомлень у нерухомі зображення. Завдання добування повідомлення на приймальній стороні стеганосистеми еквівалентне завданню виявлення інформації із суміші корисного сигналу й перешкоди в широкосмуговій системі зв'язку. У процесі досліджень виявлені певні недоліки стеганографічних систем із розширенням спектра дискретних сигналів: імовірність правильного добування вбудованих даних залежить від величини доданих спотворень, яка залежить від забезпечуваної пропускну здатності стеганоканалу. Подальші дослідження бажано провести за результатами аналізу можливого використання методів синтезу великих ансамблів квазіортогональних дискретних сигналів із поліпшеними ансамблевими, структурними й кореляційними властивостями для забезпечення вищих показників захищеності акустичних каналів у комп'ютерних системах критичного застосування.

Ключові слова: акустична інформація; технічний захист; криптографічний захист; стеганографічний (стеганофонічний) захист; модифікація; мови комп'ютерної системи.

Бібліографічні опису / Bibliographic descriptions

Можасв О.О., Гнусов Ю.В., Манжай О.В., Струков В.М., Носов В.В., Радченко В.В., Єнгаличев С.О. Стеганографічний метод захисту акустичної інформації у системах критичного застосування. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 52–63. DOI: <https://doi.org/10.30837/ITSSI.2023.25.052>

Mozhaiev, O., Gnusov, Y., Manzhai, O., Strukov, V., Nosov, V., Radchenko, V., Yenhalychev, S. (2023), "Steganographic method of acoustic information protection in critical applications systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 52–63. DOI: <https://doi.org/10.30837/ITSSI.2023.25.052>

С. ПЕРЕЯСЛАВСЬКА, О. СМАГНА

ПРОЄКТУВАННЯ РІВНЯ МАРШРУТИЗАЦІЇ В МІКРОСЕРВІСНИХ АРХІТЕКТУРАХ НА ПЛАТФОРМІ *SPRING*

Предметом дослідження є маршрутизація запитів у мікросервісній архітектурі. **Мета статті** – розроблення цілісної концепції проєктування рівня маршрутизації запитів у мікросервісній архітектурі на прикладі стеку технологій *Spring*. **Завдання:** проаналізувати сучасні підходи щодо структури мікросервісної архітектури; визначити сутність маршрутизації та встановити процеси, що забезпечують маршрутизацію запитів; визначити стек технологій *Spring*, які реалізують маршрутизацію; спроектувати рівень маршрутизації застосунку на платформі *Spring*. Упроваджуються такі **методи:** аналіз і синтез для вивчення технологій взаємодії між службами; абстрагування та узагальнення для визначення структури мікросервісної архітектури, рівня маршрутизації, узагальнення технологій, що забезпечують взаємодію між сервісами; моделювання з метою побудови моделі мікросервісної архітектури з виокремленням рівня маршрутизації та зв'язків з іншими структурами моделі. Здобуто такі **результати:** досліджено структуру мікросервісної архітектури, зокрема рівень маршрутизації; визначено роль шаблонів проєктування, що забезпечують маршрутизацію: *Service discovery*, *API Gateway*, *Load Balancer* тощо; проаналізовано види міжпроцесної взаємодії (синхронна, асинхронна, гібридна) та визначено переваги й доцільність застосування; розглянуто моделі відмовостійкості системи; визначено стек технологій на платформі *Spring* для реалізації рівня маршрутизації; розроблено модель проєкту багаторівневої мікросервісної архітектури із застосуванням стеку технологій *Spring*, що реалізує найбільш ефективні рішення в контексті маршрутизації запитів. **Висновки:** мікросервісну архітектуру доцільно розглядати як багаторівневу структуру, що будується на функціональних рівнях і зв'язках між ними; рівнем маршрутизації мікросервісів потрібно вважати всі процеси, пов'язані з налагодженням міжпроцесної взаємодії, виявленням сервісів, балансуванням навантаження та забезпеченням відмовостійкості, створенням єдиної точки входу; *Spring* є популярною платформою розроблення мікросервісної архітектури, що надає необхідні інструменти для реалізації маршрутизації запитів; розроблена модель проєкту є прикладом ефективних рішень щодо проєктування багаторівневої архітектури із застосуванням стеку технологій *Spring* у контексті маршрутизації запитів.

Ключові слова: мікросервісна архітектура; маршрутизація; *Spring*; *Service discovery*; *API Gateway*; *Load Balancer*.

Вступ

Еволюція підходів до проєктування програмного забезпечення, зростання складності розподілених застосунків, необхідність постійного управління змінами спричинили появу різних архітектурних рішень, спрямованих на вдосконалення уніфікованої моделі розроблення програмного забезпечення. У сучасних розподілених системах спостерігається відхід від традиційних способів побудови проєкту, де вся система зосереджена в єдиному й неподільному блоці, та зміщення акцентів у напрямі застосування мікросервісів.

Архітектура мікросервісів – це своєрідний стиль сервіс-орієнтованої архітектури (*SOA*). Вона побудована з мікросервісів, що є відносно невеликими й слабо пов'язаними та працюють у своїх процесах і взаємодіють через протоколи зв'язку, зокрема *HTTP/HTTPS*, *Web Socket*, *AMQP* тощо [1].

Варто зазначити, що мікросервіси в деякому сенсі є наступним кроком в еволюції сервіс-орієнтованої архітектури, оскільки підтримують багато концепцій,

що й *SOA*. Обидва підходи мають загальні риси, притаманні розподіленим архітектурам, компоненти в цих архітектурах більш автономні, завдяки чому спрощується обслуговування та контроль над змінами. Утім існують певні відмінності, що допомагають виокремити мікросервіси як самостійний напрям у розвитку архітектури розподілених систем. Незважаючи на те, що мікросервіси й *SOA* покладаються на "сервіс" як основний компонент архітектури, вони розрізняються за характеристиками сервісів (служб). Концептуальні відмінності полягають у обсязі відповідальності, покладеної на окрему службу. У *SOA* служба може обробляти широкий спектр функцій і доменів даних, тоді як мікросервіс управляє одним доменом даних і одним набором відповідних функцій або однією функцією в цьому домені. Кожен мікросервіс має розгортатися незалежно один від одного. Нарешті, кожен мікросервіс має володіти своєю пов'язаною моделлю даних і може ґрунтуватися на різних технологіях зберігання даних (*SQL*, *NoSQL*) і різних мовах програмування.

Ці характеристики забезпечують чимало переваг мікросервісної архітектури, але водночас вони створюють проблеми для розвитку та масштабування програмної системи: зростання кількості мікросервісів призводить до ускладнення архітектури і, як наслідок, утруднення процесів, пов'язаних зі взаємодією між мікросервісами.

Питання про те, як мікросервіси взаємодіють один з одним, є найважливішим і фундаментальним рішенням, що необхідно прийняти під час упровадження системи на основі архітектури мікросервісів. У монолітній програмі, що працює в єдиному процесі, компоненти викликають один одного за допомогою методів рівня мови або викликів функцій. Тому проблем із взаємодією не виникає. Інша річ, коли йдеться про мікросервіси, що працюють в окремих процесах та є автономними й слабо пов'язаними. У цьому разі мається на увазі міжпроцесна взаємодія, продуктивність якої відрізнятиметься від внутрішньопроектної, що потребує конкретних підходів, пов'язаних із вибором шаблонів проектування мікросервісів, моделі взаємодії, балансування навантаження тощо.

Під час взаємодії мікросервісів у мережі може виникати досить значна кількість проблем, пов'язаних із доступністю мікросервісу, затримкою передачі, втратою, дублюванням пакетів, навантаженням трафіку й пам'яті тощо. Це викликає необхідність оптимізації та пошуку підходів, що допоможуть вирішити зазначені проблеми. Крім того, у сучасних розподілених мікросервісних системах велика частка мережної взаємодії відбувається за допомогою API-інтерфейсів прикладного програмування. API-інтерфейси постійно розширюються, клієнтів стає більше, зростає кількість звернень, що призводить до значного зниження ефективності роботи системи. Тому потрібна певна інфраструктура для забезпечення маршрутизації запитів та прискорення API-трафіку.

Зазначені проблеми стають очевидними, коли прагнемо міркувати про систему загалом, щоб зрозуміти залежності, аналіз впливу тощо. Тому рішення про те, як організувати зв'язок між службами, є важливим завданням у розробленні мікросервісів. У цьому контексті постає питання проектування рівня маршрутизації в мікросервісній архітектурі. Від правильного вибору шаблону зв'язку, компонентів, що забезпечують взаємодію між службами, залежить продуктивність системи.

Аналіз останніх досліджень і публікацій

Аналіз літературних джерел довів, що проблемам, пов'язаним із мікросервісною архітектурою, приділялася увага в багатьох наукових працях. Так, *Kazanavicius* та *D. Mazeika* [2] порівнювали технології взаємодії, що базуються на *HTTP Rest*, *GraphQL*, *gRPC* та брокерах повідомлень *RabbitMQ* і *Kafka*. Оцінка продуктивності в умовах порівняння затримки в процесі оброблення повідомлень і пропускну здатності довела, що найнижчі результати затримки для повідомлення до одного мільйона символів були отримані технологією *RabbitMQ*. Виклики *RabbitMQ* були вдвічі швидшими за інші технології. З іншого боку, *RabbitMQ* показав найвищі результати затримки для повідомлень, які містили 10 млн символів. Це було втричі-вчетверо повільніше, ніж в інших ситуаціях. Найкращі результати затримки для повідомлень завдовжки до 10 млн символів здобули технології *GraphQL* і *HTTP Rest*. Брокер *Kafka* становив 40%, а *gRPC* – 16%, що повільніше, ніж технології *GraphQL* і *HTTP Rest*.

Під час оцінювання метрик встановлено, що найменший розмір запиту/відповіді отриманий технологією *HTTP Rest*. Запит/відповідь *GraphQL* були приблизно вдвічі-тричі більші, ніж інші. Якщо розмір повідомлення є важливим критерієм для вибору технології зв'язку, то *HTTP Rest* є рекомендованим рішенням. З іншого боку, *GraphQL* підтримує дистанційне надсилання запитів, тому потенційно за один запит/відповідь *GraphQL* може передати стільки ж інформації, скільки здатні передати кілька запитів/відповідей за допомогою інших технологій.

Міжпроцесну взаємодію в мікросервісній архітектурі розглядали І. Бугаєва, М. Розум [3]. У статті досліджуються синхронний та асинхронний підходи для реалізації взаємодії, проведено порівняльний аналіз на основі наявних досліджень для виявлення переваг і недоліків кожного з цих підходів. Автори дійшли висновку, що асинхронна взаємодія мікросервісів, реалізована на основі протоколу *AMQP*, забезпечує кращу продуктивність і вищу доступність мікросервісів. Синхронна форма взаємодії мікросервісів, яка реалізується за допомогою *REST* або *gRPC*, може забезпечити вищу пропускну здатність, ніж асинхронний метод, коли навантаження на систему відносно невисоке. За умови збільшення кількості користувачів доцільно

застосовувати асинхронний метод зв'язку з використанням черг повідомлень. Крім того, такий спосіб обміну повідомленнями між мікросервісами є більш надійним. Черги повідомлень є також кращим вибором за необхідності синхронізації інформації між сервісами.

Роботи, присвячені проблемам балансування навантаження в мікросервісній системі [4, 5], дали змогу чітко зрозуміти роль і місце цієї технології під час розроблення масштабованого застосування. Так, у дослідженні *M. Autili* та інших авторів [5] розглядаються базові концепції балансування навантаження. Зазначається, що ця технологія є найкращим підходом для підтримки масштабованості, управління швидкістю надходження або кількістю одночасних запитів. Балансувальник намагається розподілити робоче навантаження запитів між кількома екземплярами мікросервісу з метою оптимізації використання ресурсів, максимізації пропускної здатності, мінімізації часу відповіді та уникнення вузьких місць (тобто перевантаження будь-якого окремого екземпляра). У сфері мікросервісів зазвичай розрізняють два типи балансувальників навантаження, а саме: клієнтські та серверні. У роботі [5] автори пропонують новий гібридний підхід до балансування навантаження мікросервісів, що поєднує в собі переваги балансування на боці клієнта й на боці сервера.

H. Wang та інші дослідники у праці [4] розглядають різні алгоритми балансування навантаження на основі методу опитування, хеш-алгоритму, штучного інтелекту тощо. У згаданій роботі описуються принципи, переваги й недоліки основних технологій балансування навантаження, запропоновано стратегії механізму балансування навантаження *Kubernetes* і *Spring Cloud Framework*, що мають абсолютну частку ринку.

Продуктивність шаблонів проектування мікросервісів за допомогою багатопотокових структур, що генерують заздалегідь визначену кількість потоків, у своєму дослідженні проаналізували *A. Akbulut* та *H. G. Perros* [6]. Автори здобули результати продуктивності, пов'язані з часом відповіді на запит, ефективним використанням обладнання, витратами на хостинг і рівнем втрати пакетів для трьох шаблонів проектування мікросервісів, а саме *API Gateway*, *Chain of responsibility*, *Asynchronous Messaging*.

Під час дослідження шаблону *API Gateway* експериментально встановлено, що потрібно

використовувати понад один екземпляр мікросервісу, коли кількість потоків перевищує 140. Горизонтальне масштабування дає змогу зменшити або залишити постійним середній час обслуговування зі збільшенням кількості потоків. *API Gateway* забезпечує гнучкість, коли потрібно керувати запитами з кількох каналів. Це полегшує логіку виконання синхронних комунікацій між однаково збалансованими службами й забезпечує артефакти масштабування. У роботі [6] наголошується, що екосистема мікросервісів без API-шлюзу вважається поганою практикою або антипатерном.

Досліджуючи шаблон *Chain of responsibility*, *A. Akbulut* та *H. G. Perros* встановили, що одна й та сама програма може потребувати різної оперативної пам'яті та ресурсів для різних реалізацій шаблонів проектування. Отже, шаблон дизайну є важливим рішенням, що безпосередньо впливає на вартість хостингу. Ідеальним шаблоном проектування мікросервісу для функцій, які послідовно обробляють спільні дані, є шаблон проектування *Chain of responsibility*. Під час дослідження автори помітили, що ця архітектура дає змогу масштабувати серверні послуги з вигодою від використання апаратного забезпечення приблизно на 30 % порівняно з еквівалентною структурою мікросервісу.

Аналізуючи асинхронні взаємодії, *A. Akbulut* та *H. G. Perros* наголошують, що ємність черги, реалізованої за допомогою *RabbitMQ*, обмежена фіксованим обсягом пам'яті, що спричиняє переповнення черги і, як наслідок, втрату повідомлень. Досліджуючи результати реалізації асинхронного обміну повідомленнями з *RabbitMQ* для двох різних конфігурацій, автори праці [6] помітили, що втрата пакетів мінімізована з меншими конфігураціями (менше ніж 122,288 Гб RAM і менше ніж шість CPU 4 ГГц).

A. Akbulut та *H. G. Perros* зробили висновок, що не існує єдиного шаблону мікросервісів, який би був кращим за інші. Навпаки, для різних сценаріїв ефективними будуть різні шаблони проектування.

У роботі [7] дослідники порівняли та критично оцінили найбільш перспективні фреймворки *JVM*, що підтримують розвиток мікросервісів (*Spring Boot*, *Quarkus*, *Micronaut*). Згідно з дослідженнями продуктивними виявилися *Quarkus* і *Spring Boot*. *Micronaut* також зміг досягти конкурентних результатів, але проблеми, виявлені під час стрес-тестів, спричинили відставання від конкурентів. Час компіляції файлів *JAR* помітно довший для *Micronaut* і *Quarkus*,

оскільки вони більшість операцій, що в разі *Spring Boot* опрацьовуються під час виконання, здійснюють у процесі компіляції.

Низка дослідників [8] розглянула різні підходи до впровадження мікросервісів на базі чотирьох фреймворків мікросервісів (*Spring Boot 2.2.4 (Java 1.8)*, *Go Micro 1.18.0, (Go 1.13)*, *Moleculer 0.13.0 (Node.js 13.1.4)*, *Lagom 1.6.0 (Scala 2.12.8)*). Модель порівняння для фреймворків будувалася за трьома критеріями: функції, шаблони проектування та продуктивність. Три з чотирьох фреймворків, досліджених у цій роботі (*Lagom*, *Moleculer* і *Go Micro*), заохочують розробників структурувати протокол зв'язку перед розробленням фактичних мікросервісів. Навпаки, *Spring Boot* не змушує розробників дотримуватися певного стилю коду для визначення API програми. Однією з можливих причин є те, що *Spring Boot* – це фреймворк загального призначення, що не розроблений спеціально для мікросервісів. Цей факт підкреслює важливість поєднання шаблону зв'язку в розподілених системах загалом і в мікросервісах зокрема (*Spring Boot / Spring Cloud*). Крім того, автори наголошують, що прийняття структури мікросервісу з використанням стандартів або популярних інструментів може бути набагато вигіднішим для організації в довгостроковій перспективі. У зв'язку з цим *Spring Boot / Spring Cloud* є одним із найпопулярніших програмних фреймворків, що надає розробникам явну перевагу для створення мікросервісів, оскільки цей фреймворк має значний рівень підтримки та доступної інтеграції.

Визначення не вирішених раніше частин загальної проблеми. Мета роботи, завдання

Аналіз літературних джерел довів, що в роботах, присвячених мікросервісній архітектурі, автори здебільшого досліджували окремі питання, пов'язані з комунікацією служб. Водночас існує потреба в комплексному дослідженні технологій маршрутизації запитів у мікросервісній архітектурі. Такий підхід дасть змогу розглядати маршрутизацію як багатокомпонентний та одночасно цілісний процес, краще розуміти питання ефективної взаємодії між всіма компонентами мікросервісної архітектури, що безпосередньо впливатиме на вибір технологій та інструментів реалізації.

Мета статті – розроблення цілісної концепції проектування рівня маршрутизації запитів у

мікросервісній архітектурі на прикладі стеку технологій *Spring*. Для досягнення поставленої мети потрібно виконати такі **завдання**: проаналізувати сучасні підходи щодо структури мікросервісної архітектури; визначити сутність маршрутизації та встановити основні процеси, що забезпечують маршрутизацію запитів; визначити стек технологій *Spring*, що реалізують маршрутизацію; на підставі проведеного дослідження спроектувати рівень маршрутизації мікросервісного застосунку на платформі *Spring*.

Матеріали й методи

Методологічною основою роботи є такі методи наукового дослідження: аналіз і синтез використовуються для вивчення технологій взаємодії між службами; абстрагування та узагальнення застосовуються для визначення загальної структури мікросервісної архітектури, рівня маршрутизації, узагальнення технологій, що забезпечують взаємодію між сервісами; моделювання впроваджується з метою побудови моделі мікросервісної архітектури з виокремленням рівня маршрутизації та зв'язків з іншими структурами моделі.

Результати досліджень та їх обговорення

Сучасна архітектура мікросервісів має складну структуру, що підтверджується дослідженнями [4, 9, 10]. *Y. Zhang* та низка вчених визначають багаторівневі мікросервісні застосунки як складні системи масового обслуговування зі зв'язками між службами [9]. У роботі [4] мікросервісну архітектуру розглядають як багатокомпонентну структуру, що містить реєстрацію та виявлення служб, зв'язки між ними, моніторинг, балансування навантаження, масштабування, відмовостійкий і безпековий механізми, конфігурацію та інші технології керування мікросервісами. На думку авторів роботи [10], мікросервіси є складними розподіленими системами. Це дає підставу розглядати мікросервісну архітектуру як багаторівневу структуру, що будується на функціональних рівнях та зв'язках між ними. Така архітектура зазвичай містить рівні користувача (клієнта), конфігурації, маршрутизації, мікросервісів, даних тощо. Перелік та склад рівнів залежить від завдань, вимог до мікросервісної архітектури й под. Саме такий підхід, на нашу думку, забезпечить

спрощеність проєктування мікросервісних застосунків завдяки прозорості всіх зв'язків між рівнями та їх компонентами.

Рівень маршрутизації пов'язаний насамперед із сервісами й запитами до них. На сайті *Amazon AWS* (<https://aws.amazon.com/>) подається визначення поняття маршрутизації як процесу вибору шляху в будь-якій мережі. Отже, інфраструктура маршрутизації в мікросервісній архітектурі відповідає за спрямування запиту до служби. На це впливає низка факторів: оркестрування мікросервісів, що також передбачає зв'язок і координацію служб [2], обрана міжпроцесна взаємодія тощо. У роботі [2] визначаються шаблони архітектури мікросервісів, що безпосередньо впливають на маршрутизацію запитів, а саме: *API Gateway* (API-шлюз), *Service discovery* (виявлення служб) та *Hybrid* (гібридний шаблон, що поєднує в собі реєстр служб і API-шлюз). Тому до рівня маршрутизації мікросервісів, на наш погляд, належать усі процеси, пов'язані з налагодженням міжпроцесної взаємодії, виявленням сервісів, балансуванням навантаження та забезпеченням відмовостійкості, створенням єдиної точки входу (шлюзу).

У шаблоні мікросервісної архітектури існує кілька сервісів, що мають взаємодіяти один з одним. Є чимало екземплярів кожної служби, нові служби можуть додаватися, а вже наявні вилучатися під час виконання. Необхідно, щоб служби знали, які з них ще доступні та де розташовані кінцеві точки для них. Тому для вирішення цієї проблеми використовуються механізми виявлення служб.

Виявлення служб (*Service discovery*) – це процес визначення місця розташування наявних сервісів, що відповідають запиту на основі опису їх функціональної та нефункціональної семантики [11]. Підходи до виявлення служб відрізняються підтримкою мов опису сервісів, організацією пошуку та використовуваними засобами вибору служб. Ключовим компонентом виявлення служби є реєстр служб (сервісів) – база даних, що зберігає розташування екземплярів служб. Він надає інтерфейс для реєстрації, перевірки працездатності тощо.

Існує два основних шаблони виявлення служб, – на боці клієнта та на боці сервера [12]:

– методи виявлення служб на боці сервера полягають у делегуванні відповідальності за виявлення мережного розташування служб сервера. Цей патерн дає змогу клієнтським програмам надсилати запит до сервісу через балансувальник навантаження,

що звертається до реєстру сервісів, перш ніж перенаправити клієнтський запит;

– методи виявлення служб на боці клієнта містять виклик реєстру служб для виявлення мережного розташування інших служб. Послуги реєструються в реєстрі. Коли клієнту потрібно зв'язатися зі службою, виявлення служби здійснює запит до реєстру, отримує розташування потрібної клієнту служби й направляє запит за призначенням. Отже, клієнту не потрібно заздалегідь знати мережне розташування служби, щоб отримати доступ до неї.

Окрім виявлення служб, у мікросервісній архітектурі важливо надавати API-рівень, що забезпечує інтеграцію з широким спектром послуг, якими користуються клієнти. *V. Silva* та інші дослідники в роботі [12] звертають увагу на проблему, що може виникати під час міжпроцесної взаємодії, особливо коли йдеться про великі й складні програми на основі мікрослужб. Зазвичай кожна мікрослужба має свій API. Найпростішим рішенням є безпосереднє надання API для інших, як це роблять моноліти. Однак це може бути неідеальним рішенням, коли доводиться керувати десятками API, що безпосередньо доступні користувачам та іншим мікросервісам. Це ускладнює роботу архітекторів і розробників серверної частини, яким потрібно буде підтримувати узгодженість усіх API, над якими працюють різні команди, а також для розробників зовнішньої частини, що покладаються на ці API. Зазначений підхід може спричинити зниження продуктивності програми під час виклику кількох API, проблеми підтримки протоколів зв'язку тощо. Дослідження [6] свідчить, що найбільш ідеальною архітектурою для такого сценарію є шаблон проєктування *API Gateway* (API-шлюз).

API-шлюз – це єдина точка входу, що забезпечує доступ до різних мікросервісів. Це спрощує роботу клієнта завдяки переміщенню логіки виклику кількох мікросервісів до шлюзу API. *API Gateway* діє як проксі-рівень, що відокремлює серверну інфраструктуру. Наявних клієнтів не цікавлять зміни у внутрішній інфраструктурі, оскільки вони взаємодіють лише з рівнем шлюзу API [12]. Шлюз абстрагує мікросервіси від їх споживачів та надає кінцеву точку або URL-адресу для клієнтських програм, а потім внутрішньо зіставляє запити з групою служб. Тобто однією з переваг *API Gateway* є інкапсуляція внутрішньої структури програми, що, на нашу думку, створює додаткові засоби безпеки у використанні мікросервісів.

J. Kazanavicius та D. Mazeika [2] звертають увагу на можливість використання гібридного шаблону, що поєднує функції реєстру служб і API-шлюзу, однак роль цих шаблонів виконує шина повідомлень. Клієнти взаємодіють лише з цією шиною, що працює як реєстр і як шлюз. Сервіси взаємодіють один з одним через шину повідомлень, і прямий зв'язок між мікросервісами не використовується. Перевагами шаблону є легкість міграції, а недоліками – високий зв'язок між службами та шиною повідомлень.

Під час масштабування мікросервісів можуть виникати проблеми, пов'язані з рівномірним розподілом запитів між службами й перевантаженням системи. Балансування навантаження є найбільш відомим підходом для підтримки масштабованості в архітектурі мікросервісів [5]. Завдяки цьому можна ефективно зменшити затримку запиту в комунікації мікросервісів, підвищити доступність системи та ефективно покращити сервісні можливості та використання ресурсів [4].

Ця технологія застосовується для рівномірного пересилання запитів від користувача або служб до отримувачів за допомогою певної стратегії. Алгоритми балансування навантаження можна поділити на дві категорії: алгоритм статичного балансування навантаження та алгоритм динамічного балансування навантаження. Статичне балансування застосовує методи опитування, хешування адреси джерела, узгодженого хешування, зваженого опитування, випадковий метод тощо. Алгоритми динамічного балансування навантаження ґрунтуються на методах мінімального числа з'єднань, найшвидшої відповіді тощо [4].

У мікросервісній архітектурі зазвичай розрізняють два типи балансувальників навантаження, а саме клієнтські та серверні. Балансування навантаження на боці сервера – це централізований підхід до розподілу запитів між доступними екземплярами мікросервісу. Клієнти взаємодіють з балансувальником навантаження, що надсилає запити відповідним екземплярам служби. Балансування навантаження на боці клієнта – повністю розподілений підхід, згідно з яким кожному екземпляру клієнта (і кожному екземпляру мікросервісу) призначається локальний балансувальник навантаження [5].

Один із ключових компонентів рівня маршрутизації в мікросервісній архітектурі є міжпроцесна взаємодія (*interprocess communication – IPC*), що відіграє набагато важливішу роль, ніж у монолітних застосунках. Міжпроцесна взаємодія

між службами має вирішальне значення в архітектурі мікросервісів, оскільки вона дає змогу окремим мікросервісам бездоганно працювати разом як єдина система.

Зазвичай кожен екземпляр служби – це процес. Отже, служби мають взаємодіяти за протоколом внутрішньопроектної взаємодії, наприклад, HTTP, SOAP, AMQP або протоколом, таким як TCP, залежно від характеру кожної служби. Важливо зазначити, що найбільш популярні комунікаційні технології, що використовуються для мікросервісів, ґрунтуються на протоколі HTTP та шаблонах обміну повідомленнями [2]. Ці технології взаємодії між мікросервісами реалізуються, відповідно, в синхронному та асинхронному режимах.

У синхронному зв'язку застосовується взаємодія запит/відповідь: один мікросервіс надсилає запит іншому та чекає, поки цей сервіс обробить результат і надішле відповідь. У цьому режимі запитувач блокує свою роботу на час очікування відповіді. Найпоширенішими структурами для реалізації синхронного стилю зв'язку в мікросервісах є інтерфейс прикладного програмування репрезентативної передачі стану (*REST API*) на основі протоколу HTTP та фреймворк віддаленого виклику процедур *gRPC*.

Більшості синхронних комунікацій властивий стиль "один до одного". Кожен запит має оброблятися лише одним отримувачем або службою. Але за цим стилем можна використовувати численні екземпляри служби з метою масштабування, однак доведеться застосовувати механізм або методи балансування навантаження на боці клієнта [13]. У цьому разі кожна служба має інформацію про адреси розташування всіх екземплярів. Цю інформацію можна взяти із сервера виявлення служб або надати вручну у властивостях конфігурації. Також кожна служба має вбудований клієнт маршрутизації, що може обрати один екземпляр цільової служби й надіслати туди запит.

Асинхронний режим зв'язку може бути реалізований у мікросервісах, коли служби взаємодіють одна з одною за допомогою шаблону обміну повідомленнями на базі асинхронних протоколів, таких як AMQP. У цій формі IPC-мікросервіси мають посередника між сервісами для координації запитів і відповідей (наприклад, брокер *RabbitMQ* або *Apache Kafka*).

У деяких ситуаціях *RESTful* також можна використовувати як асинхронний або так званий "відкладений синхронний" зв'язок. Застосовуючи

його асинхронно, служба негайно відповідає на запит кодом відповіді (як прийнятий запит) та посиланням для отримання фактичних результатів після завершення операції [1]. Недоліком такого підходу є те, що створюються додаткові HTTP-запити. Це також спричиняє складнощі для клієнта, оскільки він тепер має перевіряти процес виконання запиту [14].

Одна з фундаментальних відмінностей асинхронного зв'язку порівняно з синхронним полягає в тому, що клієнт (служба) більше не здійснює прямий виклик до іншої служби й не очікує негайної відповіді. Замість цього клієнт публікує запит до брокера повідомлень. Одна чи декілька служб візьмуть запит від брокера й оброблять його, потім повернуть результат брокеру. Оскільки зв'язок є асинхронним, у цьому режимі клієнт не блокує роботу під час очікування відповіді [14].

Асинхронний режим обміну повідомленнями може бути реалізований за моделями "один до одного" (черга) та "один до багатьох" (тема). Асинхронний зв'язок на основі повідомлень з одним отримувачем означає, що існує зв'язок "точка-точка", що доставляє повідомлення точно одному зі споживачів, який має доступ до каналу, і що повідомлення обробляється лише один раз. Як більш гнучкий підхід можна використати механізм публікації/підписки, щоб повідомлення від відправника були доступні для мікросервісів-підписників або зовнішніх програм.

B. Reselman у роботі [15] виокремлює ще один тип взаємодії між мікросервісами – гібридний, який поєднує в собі синхронну й асинхронну взаємодію. Наприклад, гібридна служба підтримує як HTTP, так і протоколи обміну повідомленнями. Сучасні застосунки з архітектурою мікросервісів часто використовують комбінацію синхронної та асинхронної взаємодії. Наприклад, взаємодія з одним отримувачем може відбуватися за синхронним протоколом (HTTP тощо) під час виклику звичайної служби із зовнішнім API, для асинхронної взаємодії декількох служб використовуються протоколи повідомлень (AMQP).

У цьому контексті важливим є розуміння переваг і недоліків різних режимів міжпроцесної взаємодії. На основі аналізу праць [3, 10] можна підсумувати, що різні фактори, зокрема завантаження запитів, IT-середовище й мережні технології, визначають продуктивність зв'язку між мікросервісами. Не можна однозначно визначити, яка з комунікаційних технологій краща.

Це залежить від конкретного застосування. Але можна зауважити, що асинхронний зв'язок є надійнішим і стабільнішим механізмом зв'язку, ніж HTTP (*Rest*), і забезпечує більшу автономію мікросервісу [2]. *S. Newman* у дослідженні [11] аналізує переваги й недоліки кожної форми зв'язку та зазначає, що в простих мікросервісних архітектурах серйозних проблем із використанням синхронних блокувальних викликів не виникає. Але цей тип зв'язку стає неактуальним, коли ускладнюється маршрутизація та з'являється більше викликів. Асинхронний зв'язок є доцільним у разі, коли в архітектурі передбачаються тривалі процеси з довгими ланцюжками викликів, що нелегко реструктурувати. Звісно, водночас треба вирішити, який тип асинхронного зв'язку застосовуватиметься: черга повідомлень, подійно-орієнтована взаємодія тощо.

Сказане вище доводить, що вибір механізму IPC – важливе архітектурне рішення, що може вплинути на рівень доступності служб. IPC має забезпечувати ефективний зв'язок між мікросервісами, водночас цей зв'язок повинен бути слабким. Вона також має забезпечувати покращену масштабованість, легкість обслуговування та автономність між службами [3]. Але спосіб зв'язку має бути таким, щоб не перешкоджати меті мікросервісів [16].

Отже, під час проектування мікросервісної архітектури на рівні маршрутизації треба обґрунтовано підходити до визначення IPC, виявлення сервісів, балансування навантаження, створення єдиної точки входу. Але, на наш погляд, є певні фактори, що можуть пливати на якість цих процесів. Під час взаємодії мікросервісів виникають ситуації, коли з будь-яких причин один із сервісів, що використовується, перестає відповідати. У разі, коли збій є тимчасовим, наприклад, у зв'язку з повільним мережним з'єднанням тощо, вирішити проблему допомагають повторні виклики. Але коли йдеться про більш суттєві проблеми, що виникають у зв'язку з повною відмовою сервісу, то це рішення не спрацює, адже повторні виклики лише витратять ресурси. Тому важливо передбачити здатність системи справлятися зі збоями без повного руйнування та гарантувати якісну взаємодію між мікросервісами, тобто забезпечити відмовостійкість.

У роботі [17] розглядається модель відмовостійкої системи за такими напрямками:

– дії, пов'язані з недоступністю служби, що реалізуються за трьома схемами – швидкий збій, коли служба негайно повертає повідомлення

про помилку; тайм-аут можна використовувати, щоб встановити обмеження на службу, доки вона не відповість; кешування застосовується, коли служба, що викликається, недоступна, але інші служби все ще можуть читати/записувати інформацію в кеші на боці клієнта чи проксі;

– захист служб від перевантаження, що реалізується за двома моделями, а саме в автоматичному вимикачі *Circuit Breaker*, що запобігає доступу до компонента збою, швидко обробляє помилки, налаштовує резервну дію, уникає перевантаження. Друга модель використовує механізм "рукостискання", що здійснюється під час кожної взаємодії між запитувачем і службою, яка отримує запит. Сервіс, що отримує запит, може відмовити, якщо є перевантаження. Недоліком "рукостискання" є те, що потрібні додаткові запити й відповіді. Крім зазначених моделей, проблему може вирішувати балансування навантаження. *Load Balancer* допомагає знизити навантаження на служби й водночас зменшити вірогідність збоїв;

– локалізація несправності за допомогою моделі перегородки (*bulkheads*). Перегородки обмежують або

ізолюють служби, які вийшли з ладу, щоб зменшити вплив несправностей на продуктивність системи.

S. Newman у праці [11] зазначає, що, окрім цих моделей, є декілька способів збільшити відмовостійкість системи: розподіл ризиків завдяки розміщенню сервісів на різних хостах, збільшення ізоляваності сервісів тощо.

У роботах дослідників [11, 17], які вивчали механізми забезпечення відмовостійкості, зазначається, що застосування ключових шаблонів вимагає зваженого підходу з огляду на необхідний рівень відмовостійкості, архітектури мікросервісів тощо.

На рис. 1. зображена узагальнена концептуальна схема проєктування рівня маршрутизації в мікросервісній архітектурі. Схема містить основні компоненти, що забезпечують маршрутизацію: міжпроцесна взаємодія, виявлення сервісів, шлюз, балансування навантаження й підтримка відмовостійкості. Під час проєктування необхідно визначити основні моделі, шаблони або алгоритми відповідно до цих компонентів. Прийняті рішення будуть істотно впливати на архітектуру застосунку та вибір програмних засобів для її реалізації.

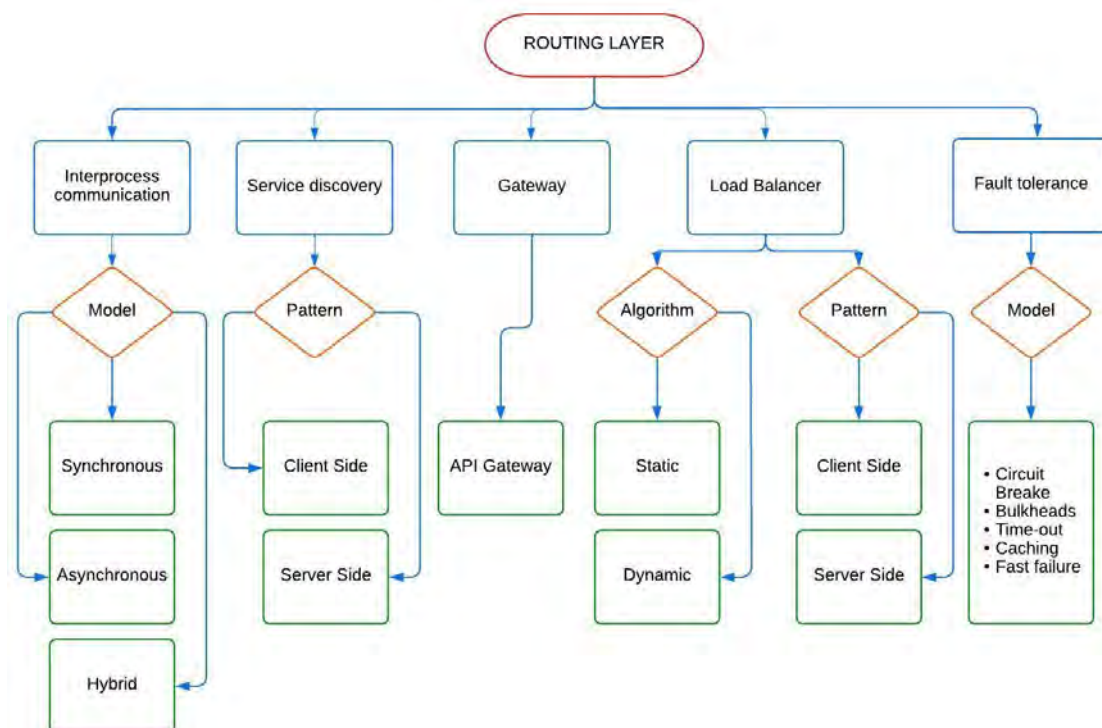


Рис. 1. Концептуальна схема проєктування рівня маршрутизації

Для практичного дослідження розглянутого питання було спроектовано мікросервісну архітектуру, що забезпечує роботу ланцюжка постачань їжі.

Система містить чотири служби: *Account*, *Products*, *Order*, *Delivery*. Усі вони є програмами, що розгортаються незалежно. Подана архітектура має

гібридну міжпроцесну взаємодію, яка містить синхронні та асинхронні зв'язки. Клієнт за допомогою мікросервісу *Account* створює новий або переглядає / оновлює наявний акаунт користувача. Усі акаунти зберігаються у відповідній базі даних. Мікросервіс *Products* також має точку входу для клієнта з метою отримання/створення/оновлення переліку продуктів. Клієнти за допомогою мікросервісу *Order* мають змогу створити нове замовлення, отримати перелік створених замовлень, а також переглянути їх статус (нове, передано в доставку). Ідентифікатор користувача для нового замовлення перевіряється з переліком акаунтів, що отримується за допомогою синхронного запиту до мікросервісу *Account*. Аналогічно, ідентифікатор продукту для нового замовлення перевіряється з переліком продуктів,

що отримується з допомогою синхронного запиту до мікросервісу *Products*. Після того як замовлення створено та збережено у базі даних, до служби доставки *Delivery* від служби *Order* асинхронним способом буде відправлено інформацію про замовлення та його адресу. Зв'язок між службою замовлення та службою доставки відбувається через брокер повідомлень. Оскільки служба замовлення є джерелом повідомлення, вона надсилає повідомлення на свій канал запиту *order-request*. З іншого боку, служба доставки прослуховує вхідні повідомлення на каналі *order-request* та публікує статус замовлення в каналі *order-reply*, який слухає служба замовлення. Після певного проміжку часу оновиться статус замовлення на "передано в доставку". Архітектурна модель проєкту наведена на рис. 2.

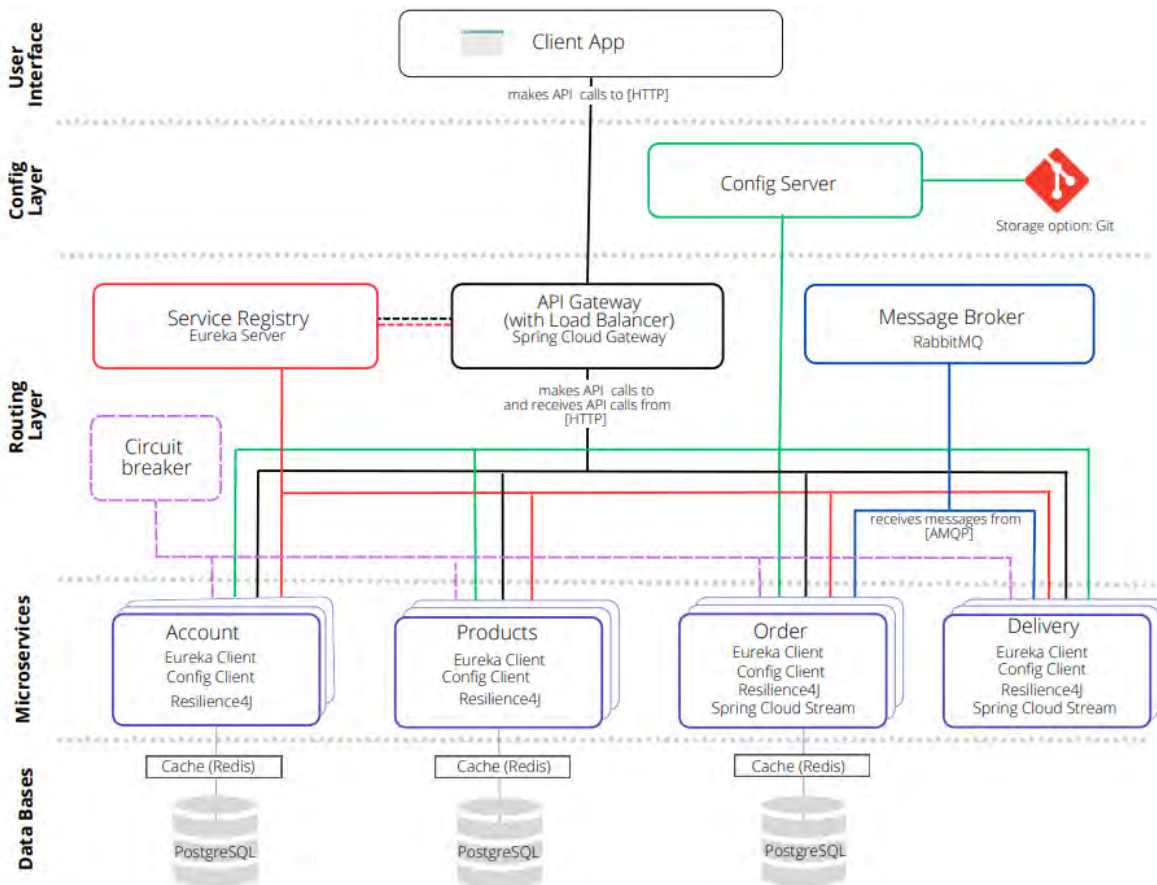


Рис. 2. Архітектурна модель проєкту

Під час обрання стеку технологій брався до уваги досвід розробників та ефективність платформ для створення мікросервісної архітектури. Вибір фреймворків ґрунтувався на підтримці мови програмування *Java*. Застосування *Java*-платформ для розроблення мікросервісних проєктів на сьогодні

є дуже популярним. Про це свідчить звіт розробників *Java* за 2023 рік від *JRebel by Perforce* [18], згідно з яким програми на основі мікросервісів були найпопулярнішими серед розробників *Java* (32 %). Відповіді показали, що більшість організацій або мали програми, повністю основані на мікросервісах

(31 %), або перебували на стадії планування (31 %), або перейшли на архітектуру мікросервісів (30 %).

Обрання *Spring* як основної платформи для проєктування мікросервісної архітектури ґрунтувалося на аналізі праць [7, 8, 18]. Зі стеку *Spring* застосовуються *Spring Boot* та *Spring Cloud*, що 2023 року залишаються найкращим рішенням для розроблення мікросервісних застосунків (59 %) [18]. Тоді як *Spring Boot* є ефективним способом розроблення автономних застосунків, *Spring Cloud* надає набір інструментів для створення загальних шаблонів у розподілених системах, таких як виявлення служб, керування конфігурацією тощо. Це означає,

що *Spring Boot* – радше середовище розроблення повного стеку й має покладатися на *Spring Cloud* для створення особливих можливостей мікросервісів [8].

Для проєктування рівня маршрутизації мікросервісів оберемо відповідні ресурси. Нагадаємо, що до складу шаблонів та компонентів, що відповідають за маршрутизацію мікросервісів, додаємо *Service discovery*, *API Gateway*, *Load Balancer*, брокери повідомлень для організації міжпроцесної асинхронної взаємодії, компоненти відмовостійкості. У табл. 1 узагальнено ресурси *Spring* для реалізації маршрутизації в мікросервісних застосунках і які підтримуються на цей час.

Таблиця 1. Ресурси *Spring* для реалізації маршрутизації в мікросервісних застосунках

№	Процеси рівня маршрутизації	Ресурси <i>Spring</i>	Особливості
1	Міжпроцесна взаємодія	<i>Spring Cloud Stream</i>	Опрацювання потокової інформації, пов'язаної із системами обміну повідомленнями <i>RabbitMQ</i> або <i>Apache Kafka</i>
		<i>Spring for Apache Kafka</i>	Асинхронна модель, рекомендується використовувати брокер <i>Apache Kafka</i>
		<i>Spring AMQP</i>	Асинхронна модель, рекомендується використовувати брокер <i>RabbitMQ</i>
2	Виявлення сервісів	<i>Spring Cloud Netflix (Eureka)</i>	Клієнт-серверна архітектура: <i>Eureka Server</i> , <i>Eureka Client</i>
		<i>Spring Cloud Zookeeper</i>	Екземпляри можна зареєструвати в <i>Zookeeper</i> , а клієнти можуть виявити екземпляри за допомогою компонентів, керованих <i>Spring</i> . Підтримує <i>Spring Cloud Load Balancer</i> – клієнтське рішення для балансування навантаження
		<i>Spring Cloud Consul</i>	Клієнт для використання <i>Config Server</i> та <i>Service Discovery</i> , вбудованих у <i>Consul</i>
3	Балансування навантаження	<i>Spring Cloud Load Balancer</i>	Забезпечує балансування навантаження під час взаємодії мікросервісів
		<i>Spring Cloud Gateway</i>	Має вбудований балансувальник навантаження між екземплярами мікросервісів
4	Забезпечення відмовостійкості	<i>Spring Cloud Circuit Breaker</i>	Підтримує <i>Resilience4J</i> , <i>Spring Retry</i> (забезпечує декларативну підтримку повторних спроб для програм <i>Spring</i>)
5	Створенням єдиної точки входу (<i>Gateway</i>)	<i>Spring Cloud Gateway</i>	Сервіс для маршрутизації вхідних запитів між мікросервісами й балансування між екземплярами мікросервісів. Інтеграція <i>Spring Cloud Discovery Client</i>

Spring Cloud містить сервіси, що виконують функції *Service discovery*, а саме *Spring Cloud Netflix (Eureka)*, *Spring Cloud Zookeeper*, *Spring Cloud Consul*. Відповідно до цього як шаблон виявлення служб будемо застосовувати *Netflix Eureka* [19], що інтегрована в *Spring Cloud*. *Eureka* – це служба *RESTful* (передача репрезентативного стану), яка використовується для виявлення, балансування навантаження та аварійного перемикання серверів середнього рівня. Цей застосунок призначений для роботи в хмарі *Amazon AWS*, але може бути розгорнутий за її межами.

Крім того, в *Eureka* передбачена можливість розгортання сервісів на різних платформах, що, на нашу думку, є певною перевагою порівняно з таким програмним застосунком, як *Kubernetes*. У разі, коли частина сервісів, що використовуються, розгорнута на *Kubernetes*, а інші сервіси працюють у інших середовищах, виявлення сервісів на рівні застосунків із використанням *Eureka* буде охоплювати всі середовища, тоді як рішення на базі *Kubernetes* сумісні лише з *Kubernetes* [19].

Eureka має клієнт-серверну архітектуру. Служба містить *Eureka Server*, а також клієнтську частину на основі *Java*, – *Eureka Client*, яка значно спрощує

взаємодію зі службою. Комунікація за допомогою *Eureka* відбувається таким чином. Сервіси реєструються в *Eureka*, реєстрація служби має короткий час життя *TTL (Time to Live)*, що вимагає від клієнтів постійної взаємодії із серверами. Непрацездатні служби або вузли перестануть працювати, тим самим спричиняють їх тайм-аут й вилучення з реєстру [20].

Клієнт *Eureka Client* має вбудований балансувальник навантаження, що виконує базове циклічне балансування навантаження на основі кількох факторів, таких як трафік, використання ресурсів, умови помилок тощо, що забезпечує гарну відмовостійкість. У цьому сенсі *Eureka* є гарною альтернативою для *AWS Elastic Load Balancer* – балансувальника навантаження для сервісів, що піддаються впливу вебтрафіку кінцевих користувачів. *AWS ELB* є традиційним рішенням для балансування навантаження на основі проксі-сервера [21], тоді як *Eureka* відрізняється тим, що балансування навантаження відбувається на рівні екземпляра/сервера/хоста. Ще один важливий аспект, що відрізняє балансування навантаження на основі проксі-сервера від балансування навантаження з використанням *Eureka*, полягає в тому, що в останньому випадку програма може бути стійкою до збоїв балансувальників навантаження, оскільки інформація про доступні сервери кешується на клієнті. Це вимагає незначного обсягу пам'яті, але підвищує стійкість до відмов. Крім того, *Eureka* допомагає знайти інформацію про служби, з якими потрібно взаємодіяти, але не накладає жодних обмежень на протокол або спосіб зв'язку. Наприклад, можна використовувати *Eureka* для отримання адреси цільового сервера та застосовувати протоколи HTTP, HTTPS або будь-які інші механізми RPC [19].

Як API-шлюз застосовуватимемо *Spring Cloud Gateway*, що є реалізацією API-шлюзу на основі фреймворку *Spring 5, Project Reactor i Spring Boot 2.0*. Цей заблокувальний шлюз *Spring Cloud Gateway* замінює застарілий *Zuul* від *Netflix* і, на відміну від *Zuul*, підтримує реактивність і *Web Sockets*. *Spring Cloud Gateway* може не тільки ефективно пересилати всі запити від клієнтів, але й виконувати функції, що відповідають за безпеку, моніторинг і поточне обмеження в системі мікросервісу. Ця служба інтегрується з *Circuit Breaker* та *Spring Cloud Discovery Client* [22].

Технологія, що реалізує асинхронний зв'язок – *RabbitMQ*, брокер повідомлень із відкритим вихідним кодом. Це легкий і масштабований брокер повідомлень

на основі протоколу AMQP, який діє як посередник між різними сервісами. *RabbitMQ* підтримує різноманітні шаблони обміну повідомленнями, зокрема "точка-точка", "публікація-підписка" та "запит-відповідь". Він також надає розширені функції, такі як маршрутизація повідомлень, підтвердження повідомлень, черги повідомлень, що не доставлені [22].

Для опрацювання потокової інформації застосовуватимемо *Spring Cloud Stream* – платформу для створення високомасштабованих мікросервісів, що керовані подіями, пов'язаних із системами обміну повідомленнями, такими як *RabbitMQ* або *Apache Kafka*. Налаштування брокерів відбувається самим *Spring Cloud Stream*. Зв'язок із брокером здійснюється з допомогою бібліотеки *Stream*. *Spring Cloud Stream* працює в розподілених мікросервісах, що реагують на потоки вхідної інформації. Він створений на основі *Spring Boot*, працює зі *Spring MVC* або *Spring Web Flux* і може використовуватися для розроблення масштабованих програм обміну повідомленнями та потокового оброблення [23].

У наведеному прикладі реалізована асинхронна модель обміну повідомленнями між службою замовлення та службою доставки за допомогою брокера повідомлень *RabbitMQ* та *Spring Cloud Stream*. Відправник повідомлення (служба *Order*) надсилає повідомлення, а потім очікує відповіді від отримувача (служба *Delivery*). У цій моделі відправник піклується про статус повідомлення, а саме отримане воно чи ні.

Фактор відмовостійкості забезпечувався *Spring Cloud Circuit Breaker*, що міститься в складі *Spring Cloud*. *Circuit Breaker* забезпечує абстракцію для різних реалізацій автоматичного вимикача. Він надає послідовний API для використання в програмах, допомагає обрати реалізацію автоматичного вимикача, що найкраще відповідає потребам у програмі [23]. Будемо застосовувати бібліотеку *Resilience4J*, реалізацію якої підтримує *Spring Cloud Circuit Breaker*, і яка є наступницею *Hystrix* від *Netflix*. За допомогою цієї бібліотеки можна реалізувати такі шаблони для використання в службах:

- автоматичний вимикач *circuitbreaker* – припиняє відправлення запитів під час збою служби, що викликається;
- повторні спроби (*retry*) – повторює спроби звернутися до служби;
- герметичні відсіки (модель перегородки) (*bulkhead*) – обмежує кількість конкурентних вихідних запитів, щоб уникнути перевантаження;
- обмежувач частоти (*ratelimit*) – обмежує швидкість вхідних запитів до служби;

– відкат до резервної реалізації (*fallback*) – установлює альтернативні шляхи виконання в разі збою запитів.

Resilience4J дає змогу застосувати відразу кілька шаблонів до одного й того самого виклику методу, для чого достатньо забезпечити цей метод відповідними анотаціями.

Крім рівня маршрутизації, наведений приклад мікросервісної архітектури містить конфігураційний рівень, який подано *Spring Cloud Config*, що забезпечує серверну (*Config Server*) та клієнтську (*Config Client*) підтримку зовнішньої конфігурації в розподіленій системі. Служба пропонує інтеграцію із системою контролю версій, щоб зберегти конфігурацію в безпеці [23]. До рівня мікросервісів належать служби *Account*, *Products*, *Order*, *Delivery*. Рівень інформації містить бази даних *PostgreSQL* до сервісів *Account*, *Products*, *Order*.

Необхідно зазначити, що традиційні бази даних часто занадто не надійні. Унаслідок цього кожна сучасна розподілена архітектура потребує кешування. Кеш знижує затримку та прискорює взаємодію між сервісами в мікросервісних архітектурах. Кеш – це рівень зберігання інформації з високою швидкістю доступу, на якому розміщена підмножина всіх даних. З кешу отримати їх буде значно швидше, ніж за умови звернення до основного сховища інформації. *Spring Framework* забезпечує підтримку прозорого додавання кешування до програми. У нашому прикладі застосовуватимемо *Redis*.

Запропонована модель мікросервісної архітектури не містить всі компоненти (моніторинг і журналювання, безпекове питання, розгортання тощо), які можуть застосовуватися під час проектування відповідних застосунків. Це пов'язано з тим, що в розробленій моделі увага приділялася саме дослідженню технологій, пов'язаних з маршрутизацією запитів у мікросервісній архітектурі.

Висновки й перспективи подальшого розвитку

Під час дослідження розроблено цілісну концепцію проектування рівня маршрутизації запитів у мікросервісній архітектурі на прикладі стеку технологій *Spring*. Здобуті результати дали змогу дійти певних висновків:

– мікросервісну архітектуру доцільно розглядати як багаторівневу структуру, що будується на функціональних рівнях і зв'язках між ними.

Такий підхід забезпечить спрощений спосіб проектування мікросервісних застосунків завдяки прозорості всіх зв'язків між рівнями та їх компонентами;

– маршрутизацію розглядаємо як багатокomпонентний і водночас цілісний процес. До рівня маршрутизації мікросервісів потрібно додати всі процеси, пов'язані з налагодженням міжпроцесної взаємодії, виявленням сервісів, балансуванням навантаження та забезпеченням відмовостійкості, створенням єдиної точки входу;

– аналіз джерел, що ґрунтуються на практичному досвіді, довів популярність *Java*-платформ, насамперед *Spring*, для розроблення мікросервісних проєктів. *Spring Boot* є ефективним способом створення автономних розподілених застосунків, *Spring Cloud* надає інструменти для реалізації шаблонів мікросервісної архітектури;

– розроблена модель проєкту є прикладом підходів до проектування багаторівневої архітектури. Ця модель реалізує найбільш ефективні рішення в контексті маршрутизації запитів із застосуванням стеку технологій *Spring*. Так, запропонована архітектура має гібридну міжпроцесну взаємодію, що містить синхронні та асинхронні зв'язки. Такий підхід допомагає реалізувати переваги цих форм зв'язку в одному застосунку. А саме тоді, коли необхідно отримати негайну відповідь на запит, використовуються синхронні блокувальні виклики; у разі тривалих процесів, що не потребують блокування, застосовується асинхронна взаємодія, яка є більш надійним і стабільним механізмом зв'язку, ніж HTTP (*Rest*).

Spring Cloud надає широкий спектр шаблонів для реалізації основних функцій маршрутизації запитів у мікросервісній архітектурі. Так, *Eureka* (реалізує *Service discovery* та має вбудований *Load Balancer*), *Gateway* (шлюз), *Circuit Breaker* (механізм відмовостійкості), що інтегровані в *Spring Cloud*, забезпечують необхідний стек технологій на рівні маршрутизації.

Для можливості демонстрації міжрівневих зв'язків у запропонованій архітектурній моделі подано основні рівні (користувача, конфігурації, маршрутизації, мікросервісів, даних), що формують цілісну модель проєкту. Маршрутизація тісно пов'язана з процесами, які відбуваються на інших рівнях. Так, кешування на рівні інформації скорочує затримку оброблення запитів до бази даних

та прискорює взаємодію між сервісами, динамічне налаштування властивостей без перезавантажень застосунків відбувається на рівні конфігурації. Процес керування подіями (повідомленнями) в мікросервісній архітектурі, який реалізується за допомогою *Spring Cloud Stream* і брокера повідомлень *RabbitMQ*, тісно пов'язаний з рівнями мікросервісів та маршрутизації.

Отже, у роботі науково обґрунтовано цілісну концепцію проектування рівня маршрутизації запитів

у мікросервісній архітектурі та запропоновано архітектурну модель на платформі *Spring*, що дасть змогу ефективно реалізовувати міжсервісну комунікацію на основі обраного стеку технологій.

Для подальшого розвитку поставленого завдання необхідно розглянути підходи, що мінімізують ризики під час зміни архітектури проєкту (перехід від моноліту до мікросервісів, перенесення інфраструктури в хмару тощо), проблеми узгодженості інформації під час транзакцій між сервісами тощо.

Список літератури

1. Abdelfattah A.S., Cerny T. Roadmap to Reasoning in Microservice Systems: A Rapid Review. *Applied Sciences*. Vol. 13. № 3. 2023. 1838 p. DOI: <https://doi.org/10.3390/app13031838>
2. Kazanavicius J., Mazeika D. Evaluation of microservice communication while decomposing monoliths. *Computing and Informatics*. Vol. 42. 2023. P. 1–36. DOI: https://doi.org/10.31577/cai_2023_1_1
3. Бугаєва І., Розум М. Реалізація міжпроцесної взаємодії в мікросервісній архітектурі. *Вісник Одеського національного морського університету*. № 67. 2022. P. 81–89. DOI: <https://doi.org/10.47049/2226-1893-2022-1-81-89>
4. H. Wang et al. Research on load balancing technology for microservice architecture. *MATEC Web of Conferences*. Vol. 336. № 08002. 2021. DOI: <https://doi.org/10.1051/mateconf/202133608002>
5. Autili M., Perucci A., Lauretis L. Hybrid Approach to Microservices Load Balancing. *Microservices. Science and Engineering*. 2020. P. 149–169. DOI: https://doi.org/10.1007/978-3-030-31646-4_10
6. Akbulut A., Perros H.G. Performance Analysis of Microservice Design Patterns. *IEEE Internet Computing*. Vol. 23. Issue 6. 2019. P. 19–27. DOI: <https://doi.org/10.1109/mic.2019.2951094>
7. Wycislik L., Latusik L., Kaminska A. M. A Comparative Assessment of JVM Frameworks to Develop Microservices. *Applied Sciences*. Vol. 13. 2023. DOI: <https://doi.org/10.3390/app13031343>
8. H. Dinh-Tuan et al. Development Frameworks for Microservice-based Applications: Evaluation and Comparison. *In Proceedings of the European Symposium on Software Engineering (ESSE '20)*. Association for Computing Machinery, New York, NY, USA, 2020. P. 21–29. DOI: <https://doi.org/10.1145/3393822.3432339>
9. Zhang Y. et al. Sinan: ML-Based and QoS-Aware Resource Management for Cloud Microservices. *In Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '21)*. Association for Computing Machinery, New York, NY, USA, 2021. P. 167–181. DOI: <https://doi.org/10.1145/3445814.3446693>
10. Яшина О. М., Кравчук О. А. Дослідження мікросервісної архітектури, архітектурний стиль Rest та їх сучасна реалізація на Java. *Herald of Khmelnytskyi national university*. Issue 5. 2020. P. 106–114. DOI: [10.31891/2307-5732-2020-289-5-106-114](https://doi.org/10.31891/2307-5732-2020-289-5-106-114)
11. Newman S. *Building Microservices: Designing Fine-Grained Systems* 2nd Edition. O'Reilly Media, London, 2021. 612 p.
12. Silva V. A. M., França B. B. N. *Microservices Design Patterns and Software Evolution*. Relatório Técnico - IC-PFG-20-07. Projeto Finalde Graduação. 2020. URL: <https://www.ic.unicamp.br/~reltech/PFG/2020/PFG-20-07.pdf> (дата звернення: 11.07.2023).
13. Pachikkal C. Interservice Communication in Microservices. *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*. Vol. 5. Issue 2. 2021. URL: <https://ijarsct.co.in/Paper1281.pdf> (дата звернення: 11.07.2023).
14. S. A. Asri et al. Implementation of Asynchronous Microservices Architecture on Smart Village Application. *International Journal on Advanced Science Engineering and Information Technology*. Vol. 12. № 3. P. 1236–1243. DOI: [10.18517/ijaseit.12.3.13897](https://doi.org/10.18517/ijaseit.12.3.13897)
15. Reselman B. Synchronous vs. asynchronous microservices communication patterns. URL: <https://www.theserverside.com/answer/Synchronous-vs-asynchronous-microservices-communication-patterns> (дата звернення: 11.07.2023).
16. Sanjana G B., Girish Rao Salanke N S. High Resilient Messaging Service for Microservice Architecture. *International Journal of Applied Engineering Research*. Vol. 16. № 5. 2021. P. 357–361. DOI: <https://dx.doi.org/10.37622/IJAER/16.5.2021.357-361>
17. Hosea E., Palit H, Dewi L. P. Fault Tolerance pada Microservice Architecture dengan Circuit Breaker dan Bulkhead Pattern. *Jurnal Infra*. Vol 9, № 2. 2021. URL: <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/11452/10062> (дата звернення: 11.07.2023).
18. Java Development Trends and Analysis. JRebel by Perforce. 2023 Java Developer Productivity Report. URL: <https://www.jrebel.com/resources/java-developer-productivity-report-2023> (дата звернення: 11.07.2023).
19. Netflix Open Source Software Center. URL: <https://netflix.github.io/> (дата звернення: 11.07.2023).
20. Wang Y. Toward service discovery and autonomic version management in self-healing microservices architecture. *ECSA '19: Proceedings of the 13th European Conference on Software Architecture*. Vol. 2. 2019. P. 63–66. <https://doi.org/10.1145/3344948.3344952>

21. Indrani V., Swaroopa D., Deepthisri D. Perceptions of Client and Server Side Load Balancing in Microservices. *International Journal of Innovative Research in Engineering & Management (IJREM)*. Vol. 7. Issue 4. 2020. P. 54–57. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3703126 (дата звернення: 11.07.2023).

22. Maharjan R. et al. Benchmarking Message Queues. *Telecom*. Vol. 4. 2023. P. 298–312. DOI: <https://doi.org/10.3390/telecom4020018>

23. Spring by VMware Tanzu. URL: <https://spring.io/projects/spring-cloud> (дата звернення: 11.07.2023).

References

1. Abdelfattah, A.S., Cerny, T. (2023), "Roadmap to Reasoning in Microservice Systems: A Rapid Review". *Applied Sciences*. Vol. 13. № 3. 1838 p. DOI: <https://doi.org/10.3390/app13031838>

2. Kazanavicius, J., Mazeika, D. (2023), "Evaluation of microservice communication while decomposing monoliths". *Computing and Informatics*. Vol. 42. P. 1–36. DOI: https://doi.org/10.31577/cai_2023_1_1

3. Buhaieva, I., Rozum, M. (2022), "Implementation of interprocess interaction in microservice architecture". ["Realizatsiia mizhprotseinoi vzaiemodii v mikroservisnii arkhitekturi"], *Bulletin of Odessa National Maritime University*, № 67. P. 81–89. DOI: <https://doi.org/10.47049/2226-1893-2022-1-81-89>

4. Wang, H. et al. (2021), "Research on load balancing technology for microservice architecture". *MATEC Web of Conferences*. Vol. 336. № 08002. DOI: <https://doi.org/10.1051/mateconf/202133608002>

5. Autili, M., Perucci, A., Lauretis, L. (2020), "Hybrid Approach to Microservices Load Balancing". *Microservices. Science and Engineering*. P. 149–169. DOI: https://doi.org/10.1007/978-3-030-31646-4_10

6. Akbulut, A., Perros, H.G. (2019), "Performance Analysis of Microservice Design Patterns". *IEEE Internet Computing*. Vol. 23. Issue 6. P. 19–27. DOI: <https://doi.org/10.1109/mic.2019.2951094>

7. Wycislik, L., Latusik, L., Kaminska, A. M. (2023), "A Comparative Assessment of JVM Frameworks to Develop Microservices". *Applied Sciences*. Vol. 13. DOI: <https://doi.org/10.3390/app13031343>

8. Dinh-Tuan, H. et al. (2020), "Development Frameworks for Microservice-based Applications: Evaluation and Comparison". *European Symposium on Software Engineering (ESSE '20)*. Association for Computing Machinery, New York, NY, USA, P.21–29. DOI: <https://doi.org/10.1145/3393822.3432339>

9. Zhang, Y. et al. (2021), "Sinan: ML-Based and QoS-Aware Resource Management for Cloud Microservices". *In Proceedings of the 26th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS '21)*. Association for Computing Machinery, New York, NY, USA, P. 167–181. DOI: <https://doi.org/10.1145/3445814.3446693>

10. Yashina, O.M., Kravchuk, O.A. (2020), "Research on microservice architecture, Rest architectural style and their modern implementation in Java". ["Doslidzhennia mikroservisnoi arkhitektury, arkhitekturnyi styl Rest ta yikh suchasna realizatsiia na Java"], *Herald of Khmelnytskyi national university*. Issue 5. P. 106–114. DOI: [10.31891/2307-5732-2020-289-5-106-114](https://doi.org/10.31891/2307-5732-2020-289-5-106-114)

11. Newman, S. (2021), *Building Microservices: Designing Fine-Grained Systems 2nd Edition*. O'Reilly Media, London, 612 p.

12. Silva, V. A. M., França, B. B. N. (2020), "Microservices Design Patterns and Software Evolution". Relatório Técnico - IC-PFG-20-07. Projeto Finalde Graduação, available at: <https://www.ic.unicamp.br/~reltech/PFG/2020/PFG-20-07.pdf> (last accessed: 11.07.2023).

13. Pachikkal, C. "Interservice Communication in Microservices". *International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)*. Vol. 5. Issue 2, available at: <https://ijarsct.co.in/Paper1281.pdf> (last accessed: 11.07.2023).

14. Asri, S.A. et al. (2020), "Implementation of Asynchronous Microservices Architecture on Smart Village Application". *International Journal on Advanced Science Engineering and Information Technology*. Vol. 12. № 3. P. 1236–1243. DOI: [10.18517/ijaseit.12.3.13897](https://doi.org/10.18517/ijaseit.12.3.13897)

15. Reselman, B. Synchronous vs. asynchronous microservices communication patterns, available at: <https://www.theserverside.com/answer/Synchronous-vs-asynchronous-microservices-communication-patterns> (last accessed: 11.07.2023).

16. Sanjana, G B., Girish Rao Salanke, N S. (2021), "High Resilient Messaging Service for Microservice Architecture". *International Journal of Applied Engineering Research*. Vol. 16. № 5. P. 357–361. DOI: <https://dx.doi.org/10.37622/IJAER/16.5.2021.357-361>

17. Hosea, E., Palit, H, Dewi, L. P. (2021), "Fault Tolerance pada Microservice Architecture dengan Circuit Breaker dan Bulkhead Pattern". *Jurnal Infra*. Vol 9, № 2, available at: <https://publication.petra.ac.id/index.php/teknik-informatika/article/view/11452/10062> (last accessed: 11.07.2023).

18. "Java Development Trends and Analysis. JRebelby Perforce". 2023 Java Developer Productivity Report, available at: <https://www.jrebel.com/resources/java-developer-productivity-report-2023> (last accessed: 11.07.2023).

19. "Netflix Open Source Software Center", available at: <https://netflix.github.io/> (last accessed: 11.07.2023).

20. Wang, Y. (2019), "Toward service discovery and autonomic version management in self-healing microservices architecture". *ECSA '19: Proceedings of the 13th European Conference on Software Architecture*. Vol.2. P. 63–66. DOI: <https://doi.org/10.1145/3344948.3344952>

21. Indrani, V., Swaroopa, D., Deepthisri, D. (2020), "Perceptions of Client and Server Side Load Balancing in Microservices". *International Journal of Innovative Research in Engineering & Management (IJIREM)*. Vol. 7. Issue 4, P. 54–57, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3703126 (last accessed: 11.07.2023).
22. Maharjan, R. et al. (2023), "Benchmarking Message Queues". *Telecom*. Vol. 4. P. 298–312. DOI: <https://doi.org/10.3390/telecom4020018>
23. "Spring by VMware Tanzu", available at: <https://spring.io/projects/spring-cloud> (last accessed: 11.07.2023).

Надійшла 14.08.2023

Відомості про авторів / About the Authors

Переяславська Світлана Олександрівна – кандидат педагогічних наук, доцент, Державний заклад "Луганський національний університет імені Тараса Шевченка", доцент кафедри інформаційних технологій та систем, Полтава, Україна; e-mail: pereyaslav9@gmail.com; ORCID ID: <https://orcid.org/0000-0001-9873-0447>

Смагіна Ольга Олександрівна – кандидат педагогічних наук, доцент, Державний заклад "Луганський національний університет імені Тараса Шевченка", доцент кафедри інформаційних технологій та систем, Полтава, Україна; e-mail: smagina1804@gmail.com; ORCID ID: <https://orcid.org/0000-0002-6024-5152>

Pereiaslavska Svitlana – PhD (Pedagogical Sciences), Associate Professor, State Institution "Luhansk Taras Shevchenko National University", Associate Professor at the Department of Information Technologies and Systems, Poltava, Ukraine.

Smahina Olga – PhD (Pedagogical Sciences), Associate Professor, State Institution "Luhansk Taras Shevchenko National University", Associate Professor at the Department of Information Technologies and Systems, Poltava, Ukraine.

DESIGNING THE ROUTING LEVEL IN MICROSERVICE ARCHITECTURES ON THE SPRING PLATFORM

The **subject matter** of research is the routing of requests in the microservice architecture. The **goal** of the article is to develop a target design concept for the level of request routing in the microservice architecture using the Spring technology stack as an example. **Tasks:** to analyse modern approaches to the structure of microservice architecture; programming of the routing entity and establishment of processes that ensure the routing of requests; programming stacks of Spring technologies that implement routing; design the routing layer of the application on the Spring platform. The following **methods** are used: analysis and synthesis to study technologies of interaction between services; abstraction and generalization to determine the structure of the microservice architecture, routing level, generalization of technologies that ensure interaction between services; modelling for the purpose of building a model of microservice architecture, highlighting the level of routing and connections with other structural models. The following **results** were obtained: the structure of the microservice architecture was investigated, in particular the level of routing; the role of design patterns that provide routing is defined: Service discovery, API Gateway, Load Balancer, etc.; the types of interprocess interaction (synchronous, asynchronous, hybrid) were analysed and the advantages and expediency of the application were determined; models of system fault tolerance are presented; a stack of technologies on the Spring platform is defined for the implementation of the routing layer; a model of a multi-level microservice architecture project was developed using the Spring technology stack, which implements the most effective solutions in the context of request routing. **Conclusions:** it is advisable to consider microservice architecture as a multi-level structure built on functional levels and connections between them; the level of microservices routing should include all processes related to the establishment of interprocess interaction, service detection, load balancing and fault tolerance, and the creation of a single entry point; Spring is a popular microservice architecture development tool platform that provides requirements for implementing request routing; the developed project model is an example of effective solutions for designing a multi-level architecture using the Spring technology stack in the context of request routing.

Keywords: microservice architecture; routing; Spring; Service discovery; API Gateway; Load Balancer.

Бібліографічні описи / Bibliographic descriptions

Переяславська С. О., Смагіна О. О. Проектування рівня маршрутизації в мікросервісних архітектурах на платформі Spring. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 64–78. DOI: <https://doi.org/10.30837/ITSSI.2023.25.064>

Pereiaslavska, S., Smahina, O. (2023), " Designing the routing level in microservice architectures on the Spring platform", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 64–78. DOI: <https://doi.org/10.30837/ITSSI.2023.25.064>

О. ПОНОМАРЕНКО, В. ГОРБАЧОВ

ПРОГРАМНА ПЛАТФОРМА ДЛЯ ОЦІНЮВАННЯ ЕФЕКТИВНОСТІ АГРЕГАЦІЇ СТРУКТУРНОЇ МОДЕЛІ СКЛАДНИХ СИСТЕМ

Складні системи визначаються великою розмірністю, містять значну кількість елементів та зв'язків між ними. Мережі використовуються для подання складних систем. Через велику розмірність сучасних систем дослідники оцінюють запропоновані рішення з допомогою згенерованих мереж. Велика розмірність системи призводить до проблем у процесі моделювання та управління. Для вирішення цих питань потрібні методи зменшення розмірності складних систем. Агрегація структурної моделі системи полягає в об'єднанні її елементів у підсистеми, унаслідок чого зменшується розмірність системи й обчислювальна складність. **Предметом дослідження** є програмна платформа для оцінювання ефективності агрегації структурної моделі складних систем. **Мета роботи** – розроблення програмної платформи для оцінювання ефективності агрегації структурної моделі складних систем. **Актуальність статті** полягає в тому, що, застосовуючи програмну платформу, можна здійснити агрегацію структурної моделі систем із значною кількістю елементів, а також оцінити ефективність агрегації структурної моделі системи. У роботі передбачалося розв'язання таких **завдань**: розроблення програмної платформи, що складається з модуля генератора системи, модуля агрегації системи, модуля пошуку максимального потоку й модуля статистичного оброблення інформації; агрегація структурної моделі системи з використанням програмної платформи; визначення ефективності агрегації структурної моделі системи із застосуванням програмної платформи. **Результати дослідження**: створено програмну платформу, здійснено агрегацію структурної моделі системи та оцінено ефективність агрегації структурної моделі системи з використанням програмної платформи. **Висновки**: застосовуючи програмну платформу, можна згенерувати систему, здійснити агрегацію структурної моделі системи та виконати пошук максимального потоку; також програмна платформа дає змогу оцінити ефективність агрегації структурної моделі системи; значення максимального потоку є однаковим на двох рівнях системи, тому в цьому разі здійснюється агрегація структурної моделі системи зі збереженням коректності її параметрів.

Ключові слова: програмна платформа; агрегація структурної моделі; складні системи; генератор систем; максимальний потік.

Вступ

Складні системи складаються із значної кількості елементів і зв'язків між ними та спостерігаються в багатьох сферах, наприклад, соціальні, біологічні, екологічні та технологічні системи. Мережі використовуються як загальна модель складних систем [1, 2]. Збільшення розміру мереж призводить до розвитку нових аналітичних підходів до їх подання та оцінювання ефективності. У розробленні таких підходів дослідники стикаються з різними проблемами. Для мереж, що складаються навіть із декількох десятків вершин, досить просто намалювати мережу та відповісти на запитання про її структуру. Це один із простих та основних способів отримати уявлення про структуру мережі. Але це неможливо зробити для мереж із дуже великою кількістю елементів. Існує чимало інструментів візуалізації, що дають змогу структурувати та візуалізувати мережі [3].

Складні системи потребують спеціальних методів аналізу та проектування. Існують такі групи

завдань топологічного аналізу, вирішення яких потребує розвитку математичних методів і обчислювальних ресурсів: розроблення відповідного структурного опису системи для визначення елементів системи, підсистем та зв'язків між ними; створення підходів для визначення характеристик системи, наприклад, окреслення найкоротших шляхів і циклів; еквівалентні перетворення структури системи для зменшення розмірності системи; проектування структури системи [4].

Під час досліджень часто доводиться попередньо оцінювати ефективність топології мережі, здатність мережі витримувати високі навантаження та залишатися в робочому стані. Через великі масштаби сучасних систем створення реальної системи для експерименту практично неможливе. У цьому разі дослідники оцінюють запропоновані рішення за допомогою згенерованих мереж [5].

Існують методи для зменшення розмірності складних систем. Це зумовлено тим, що дуже велика розмірність системи призводить до проблем у процесі моделювання та управління. Агрегація структурної

моделі системи полягає в об'єднанні елементів системи в підсистеми, унаслідок чого зменшується розмірність системи та обчислювальна складність.

Робота присвячена створенню програмної платформи для оцінювання ефективності агрегації структурної моделі системи. Програмна платформа складається з модуля генератора системи, модуля агрегації системи, модуля пошуку максимального потоку й модуля статистичного оброблення інформації. Актуальність дослідження полягає в тому, що завдяки програмній платформі можна здійснити агрегацію структурної моделі системи із значною кількістю елементів, а також оцінити ефективність агрегації структурної моделі системи.

Аналіз літератури

Мережі широко використовуються для подання систем, тому що за допомогою мереж можна проаналізувати функції, динаміку та еволюцію природних і штучних систем [6]. Часто систему не можна подати однією мережею, а замість цього мають бути змодельовані набори мереж, у яких різниця між ними, імовірно, є так само важливою, як і їх загальні риси. Такі випадки вказують на важливість методів для генерації мереж, що основані на даних та охоплюють як ключові характеристики системи, так і реалістичну мінливість, щоб використовувати їх у таких завданнях, як моделювання, аналіз та прийняття рішень. Протягом останніх двох десятиліть була запропонована значна кількість моделей з метою досягнення реалізму для згенерованих мереж.

Важливо, щоб генератор мережі відповідав двом основним критеріям – реалізму й різноманітності. Реалізм має зважати на будь-які властивості мережі, що управляють процесами, такими як функціонування системи, динаміка й еволюція. Реалізм може залежати як від структурних властивостей мережі, так і від особливостей мережі, що формуються. Наприклад, моделі соціальних мереж мають бути здатні відтворювати структурні особливості мережі та соціологічні явища, які виникають, зокрема взаємодія між людьми у співтоваристві, що зумовлено їх потребами й щоденною рутинною. Також час роботи та вимоги до пам'яті генератора мають бути допустимими для дуже великих наборів даних, що можуть містити значну кількість вузлів і дуг.

Автори роботи [6] розробляють та оцінюють швидкий генератор, що створює реалістичні та масштабовані копії складних мереж. У праці реалізм

описується двома способами: відповідність оригінальному графу за набором важливих структурних властивостей і відповідність поведінці під час виконання різних алгоритмів графа. Запропонований авторами генератор здатний генерувати мережі, які аналогічні оригінальній мережі з погляду важливих структурних показників та які призводять до аналогічного часу роботи багатьох алгоритмів аналізу графів і мереж. Використовуючи цей генератор, можна відтворити наявну мережу й масштабувати згенеровану версію, наприклад, щоб тестувати алгоритми на великих наборах даних, де вони не доступні.

У роботі [7] автори пропонують гнучкий та розширюваний фреймворк, щоб зрозуміти, як властивості різних типів мереж змінюються з різною кількістю дуг і вершин. Цей підхід дає змогу моделювати три класичні мережні моделі та налаштовувати параметри моделі й розмір мережі. У згаданому дослідженні автори аналізують, як поведуться різні властивості мережі залежно від її типу та як змінюються властивості залежно від розміру мережі. Також порівнюються властивості трьох мережних моделей: модель безмасштабної мережі, модель мережі малого світу та модель випадкової мережі. Автори вимірювали три типи мережних властивостей: центральність, середній найкоротший шлях і глобальний коефіцієнт кластеризації.

Методи моделювання систем використовуються для оцінювання надійності та продуктивності системи, оскільки вони імітують її роботу, дають змогу досліджувати вплив зовнішніх факторів на продуктивність системи, а також надають інструмент для аналізу ризиків і прийняття рішень [8].

У зв'язку зі збільшенням розмірів систем, моделі, що описують поведінку мережних систем, здатні мати дуже високу розмірність. Це призводить до проблем під час моделювання та управління через обмежені обчислювальні можливості та ємність сховища. Окреслені проблеми стимулюють розроблення методів зменшення розмірності мережних систем. За останні десятиліття з'явилося чимало теорій і методів зменшення розмірності моделей систем [9].

Агрегація структурної моделі системи полягає в об'єднанні елементів у підсистеми, унаслідок чого створюється другий рівень системи. Завдяки агрегації структурної моделі системи зменшується розмірність системи й обчислювальна складність.

У роботі [10] застосовується агрегація елементів мережі масового обслуговування для того, щоб зменшити розмірність мережі та час моделювання. Для моделювання великих систем необхідні методи з відповідним балансом точності та швидкості, оскільки моделювання реальної системи може потребувати багато часу.

У праці [11] агрегація використовується для пошуку найкоротших шляхів у графі. Це може застосовуватися для супутникової навігації, щоб реагувати на оновлення дорожнього трафіку в режимі реального часу.

Мета статті – розроблення програмної платформи для оцінювання ефективності агрегації структурної моделі складних систем. У дослідженні розв'язуються такі завдання: розроблення програмної платформи, що складається з модуля генератора системи, модуля агрегації системи, модуля пошуку максимального потоку й модуля статистичного оброблення інформації; здійснення агрегації структурної моделі системи з використанням програмної платформи; оцінювання ефективності агрегації структурної моделі системи із застосуванням програмної платформи.

Програмна платформа

У попередній роботі [12] здійснена агрегація структурної моделі системи з незначною кількістю елементів. Але для систем із великою кількістю елементів неможливо здійснити агрегацію структурної моделі системи без застосування програмних інструментів. Тому була створена програмна платформа мовою програмування C#, що складається з чотирьох модулів: модуля генератора системи, модуля агрегації системи, модуля пошуку максимального потоку та модуля статистичного оброблення інформації.

1. У модулі генератора створюється система із заданими параметрами.

2. У модулі агрегації здійснюється агрегація структурної моделі системи. Елементи системи об'єднуються в підсистеми, створюються фіктивні контакти підсистем і зв'язки між підсистемами. Отже, унаслідок роботи модуля агрегації створюється другий рівень системи.

3. У модулі пошуку максимального потоку здійснюється пошук максимального потоку для рівня системи. Якщо виконати це завдання для різних рівнів системи, то можна побачити, чи змінився

максимальний потік після агрегації структурної моделі системи, тобто чи вплинула агрегація на параметри системи.

4. У модулі статистичного оброблення інформації аналізуються згенеровані системи та їх параметри для подальшої генерації систем зі схожими параметрами.

Структура програмної платформи зображена на рис. 1.

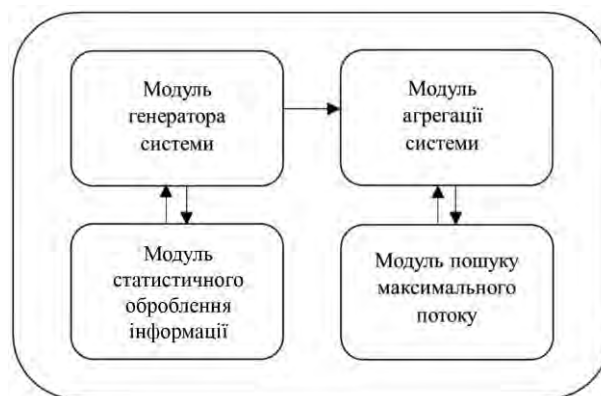


Рис. 1. Структура програмної платформи

Модуль генератора системи

У модулі генератора системи створюється система, подібна до тих, що проаналізовані раніше в роботах [12, 13]. У генератор системи передається кількість її елементів. Потім генерується граф, що має такі вершини: джерело, C_1, C_2, \dots, C_n , стік, де n – кількість елементів системи, джерело та стік – це нульовий елемент системи, що є зовнішнім середовищем.

У циклі перебираються вершини графа та створюються дуги. Для генерації графа виконуються такі умови: мінімум одна дуга виходить з вершини; мінімум одна дуга входить у вершину; мінімум одна дуга виходить із джерела; мінімум одна дуга входить у стік; дуга не може входити в джерело; дуга не може виходити зі стоку; дуга не може виходити та входити в ту саму вершину.

На етапі агрегації елементи системи будуть об'єднані в підсистеми. З огляду на це в модулі генератора системи виконується така умова: має бути зв'язок між елементами всередині підсистеми.

Далі визначається, чи є такі вершини графа, до яких не входить дуга. Якщо такі вершини існують, то генеруються дуги, що входять у ці вершини. Отже, формується колекція дуг графа *Edges*.

На наступному етапі створюється об'єкт класу *SystemLevel*, що є початковою системою, або перший рівень системи. У циклі *foreach* здійснюється перебір дуг графа й виконується такий алгоритм:

1) якщо ще не існує елемента в системі з таким номером, як номер вершини, до якої входить дуга, то створюється елемент системи й додається до колекції *Components* класу *SystemLevel*;

2) знаходиться елемент системи з таким номером, як номер вершини, до якої входить дуга, та створюється вхідний контакт елемента;

3) якщо ще не існує елемента в системі з таким номером, як номер вершини, з якої виходить дуга, то створюється елемент системи та додається до колекції *Components* класу *SystemLevel*;

4) знаходиться елемент системи з таким номером, як номер вершини, з якої виходить дуга, і створюється вихідний контакт елемента;

5) створюється зв'язок між елементами системи й додається до колекції *Connections* класу *SystemLevel*.

Отже, формується об'єкт класу *SystemLevel*, що є початковою системою, або перший рівень системи та містить колекцію елементів системи *Components* і колекцію зв'язків між елементами системи *Connections*.

Модуль агрегації системи

У модуль агрегації системи передається початкова система, отримана в модулі генератора системи. Система подана класом *SystemLevel* і є першим рівнем системи. Клас *SystemLevel* містить властивості *Components* і *Connections*. *Components* – це колекція елементів системи; *Connections* – колекція зв'язків між елементами системи.

Елементи в системі об'єднуються таким чином: $S_{\mu 0} = \{C_0\}$, $S_{\mu 1} = \{C_1, C_2\}$, ..., $S_{\mu m} = \{C_{n-1}, C_n\}$, де n – кількість елементів у системі; m – кількість підсистем.

Елементи об'єднуються по два до кожної підсистеми, нульовий елемент додається в окрему підсистему. Якщо в системі непарна кількість елементів, то останній елемент додається один у підсистему.

Підсистема подана класом *Subsystem* і є елементом другого рівня системи й водночас системою, що містить декілька елементів. Клас *Subsystem* наслідується від класу *Component* і містить

властивість *InternalComponents*, що є колекцією елементів, які входять у цю підсистему.

На етапі створення підсистеми формуються вхідні та вихідні контакти підсистеми. У циклі *foreach* здійснюється перебір елементів, що входять у підсистему, та за таким алгоритмом створюються вихідні контакти підсистеми:

– знаходиться вихідний контакт елемента та зв'язок на першому рівні системи, який виходить з цього контакту;

– якщо цей зв'язок входить до елемента, розміщеного в цій підсистемі, то вихідний контакт підсистеми не створюється;

– якщо цей зв'язок входить до елемента, що знаходиться в іншій підсистемі, то створюється вихідний контакт підсистеми.

Далі в цьому циклі створюються вхідні контакти підсистеми за таким алгоритмом:

– знаходиться вхідний контакт елемента та знаходиться зв'язок на першому рівні системи, який входить до цього контакту;

– якщо цей зв'язок виходить з елемента, розміщеного в цій підсистемі, то вхідний контакт підсистеми не створюється;

– якщо цей зв'язок виходить з елемента, розміщеного в іншій підсистемі, то створюється вхідний контакт підсистеми.

На наступному етапі створюються зв'язки між підсистемами. У циклі *foreach* здійснюється перебір підсистем і виконується такий алгоритм:

– знаходиться вихідний контакт цієї підсистеми, знаходиться зв'язок першого рівня системи, що відповідає цьому вихідному контакту підсистеми, та знаходиться вхідний контакт елемента першого рівня, до якого входить цей зв'язок;

– потім серед усіх вхідних контактів підсистем знаходиться такий вхідний контакт підсистеми, що відповідає вхідному контакту елемента першого рівня.

Отже, створюється зв'язок між вихідним контактом підсистеми та знайденим вхідним контактом іншої підсистеми й додається до колекції *Connections*.

Унаслідок роботи модуля створюється об'єкт класу *SystemLevel*, що є перетвореною системою, або другий рівень системи. Властивість *Components* класу *SystemLevel* містить колекцію підсистем; властивість *Connections* містить колекцію зв'язків між підсистемами.

Модуль пошуку максимального потоку

Унаслідок агрегації структурної моделі системи здійснюється перетворення структури системи, тому треба перевірити, чи збереглися параметри системи на другому рівні системи. Для цього можна розв'язати задачу про максимальний потік. Сутність задачі полягає в тому, щоб отримати значення максимального потоку мережі для заданої топології [14]. Задача про максимальний потік формулюється таким чином: необхідно знайти максимально можливе для цієї мережі значення сумарного потоку між джерелом і стоком за умови заданих пропускних здатностей дуг мережі. Для вирішення задачі про пошук максимального потоку використовується теорема, згідно з якою максимально можливе значення сумарного потоку на кінцевих дугах дорівнює мінімальній пропускній здатності обраного розрізу [15]. Під пропускною здатністю розрізу мається на увазі сума пропускних здатностей дуг, що утворюють розріз.

У модуль пошуку максимального потоку передається об'єкт класу *SystemLevel*, що є системою та містить колекцію елементів системи й колекцію зв'язків між елементами системи. Колекція зв'язків між елементами системи перетворюється у двовимірний масив, що подає граф, та відповідно до алгоритму розв'язується задача про максимальний потік.

Алгоритм Форда – Фалкерсона передбачає такі кроки:

- пошук на графі ланцюга з джерела в стік;
- для кожної дуги ставиться максимально можливе значення потоку з джерела в стік, значення потоку не може бути більшим, ніж вага дуги;
- якщо значення потоку дорівнює вазі дуги, то така дуга є насиченою та крізь неї вже не можна пройти під час вибору ланцюгів у графі;
- алгоритм завершується тоді, коли перехід із джерела в стік стане неможливим.

Після завершення дії алгоритму значення максимального потоку дорівнює сумі потоків усіх дуг, інцидентних стоку графа.

Унаслідок роботи модуля маємо значення максимального потоку. Якщо значення максимального потоку буде однаковим для першого й другого рівня системи, то можна зробити висновок, що в цьому разі здійснюється агрегація структурної моделі системи зі збереженням коректності її параметрів.

Процес пошуку максимального потоку зображено на прикладі згенерованої системи (рис. 2).

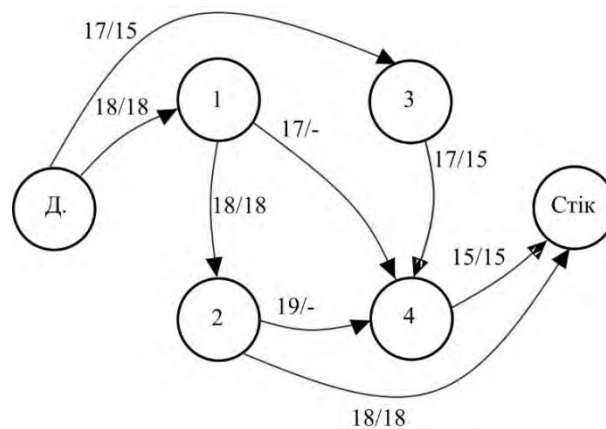


Рис. 2. Обхід графа згенерованої системи

Результат обходу графа поданий у табл. 1.

Таблиця 1. Результат обходу графа

№	Шлях	Значення потоку
1	джерело – 1 – 2 – стік	18
2	джерело – 3 – 4 – стік	15

Значення максимального потоку дорівнює сумі потоків усіх дуг, інцидентних стоку графа, що в наведеному прикладі дорівнює 33.

Тестування модулів програмної платформи

Було проведено тестування програмної платформи на прикладі систем із незначною кількістю елементів за допомогою *Unit Tests*.

Для тестування модулів платформи використовувались системи, для яких вже були здійснені агрегація структурної моделі системи та пошук максимального потоку, наприклад, система, розглянута в роботі [12]. У тесті така система у вигляді об'єкта класу *SystemLevel* передавалася в модуль агрегації. Перетворена система, отримана внаслідок роботи модуля, має збігатися із системою, отриманою після агрегації для цього прикладу раніше.

Для тестування модуля пошуку максимального потоку в тесті в цей модуль передавались перший і другий рівні системи. Унаслідок роботи модуля пошуку максимального потоку значення максимального потоку на першому та другому рівнях системи мають

збігатися з розрахунками, проведеними для цієї системи раніше.

Було виміряно час роботи модулів програмної платформи для систем із різною кількістю елементів, наприклад, що містять 50, 250 та 500 елементів. Також виміряно загальний час роботи програмної платформи.

Оцінювання ефективності агрегації структурної моделі системи

Був проведений експеримент із використанням програмної платформи, основною метою якого є оцінювання ефективності агрегації структурної моделі системи та ефективності самої програмної платформи. Критеріями ефективності агрегації структурної моделі системи є зниження обчислювальної складності системи та підвищення швидкодії.

Алгоритм експерименту передбачає такі етапи:

- модуль генератора системи: генерація системи із заданою кількістю елементів;
- модуль пошуку максимального потоку: розв'язання задачі пошуку максимального потоку для згенерованої системи;
- модуль агрегації системи: агрегація структурної моделі системи;
- модуль пошуку максимального потоку: вирішення задачі пошуку максимального потоку для перетвореної системи.

Був проведений експеримент для систем, що містять 50, 250 та 500 елементів. У процесі експерименту алгоритм був виконаний 100 разів. Значення максимального потоку на першому та другому рівнях системи було однаковим.

Середнє значення часу роботи в секундах модуля генератора системи, модуля агрегації системи, модуля пошуку максимального потоку на першому рівні системи та модуля пошуку максимального потоку на другому рівні системи зображено на рис. 3 та 4.

За результатами експерименту можна зробити висновок, що програмна платформа дає змогу згенерувати систему із заданою кількістю елементів, здійснити агрегацію структурної моделі системи та розв'язати задачу про пошук максимального потоку для кожного рівня системи. У процесі експерименту значення максимального потоку було однаковим для згенерованої та перетвореної системи, тому впливає висновок, що в цьому разі

здійснюється агрегація структурної моделі системи зі збереженням коректності її параметрів.

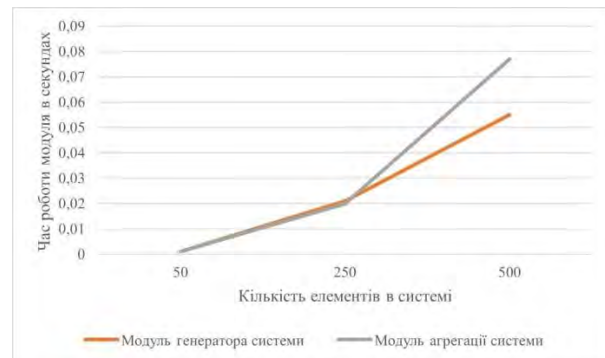


Рис. 3. Час роботи модуля генератора й модуля агрегації

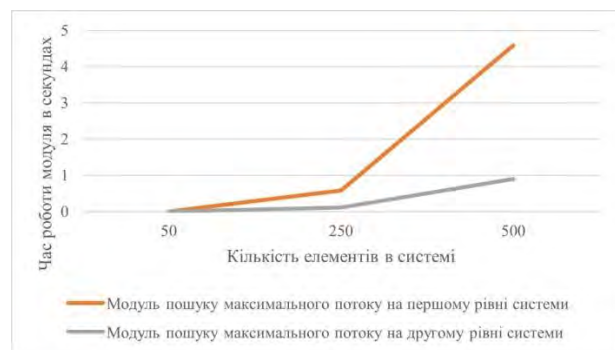


Рис. 4. Час роботи модуля пошуку максимального потоку на першому та другому рівні системи

Середнє значення часу роботи в секундах програмної платформи подано на рис. 5.

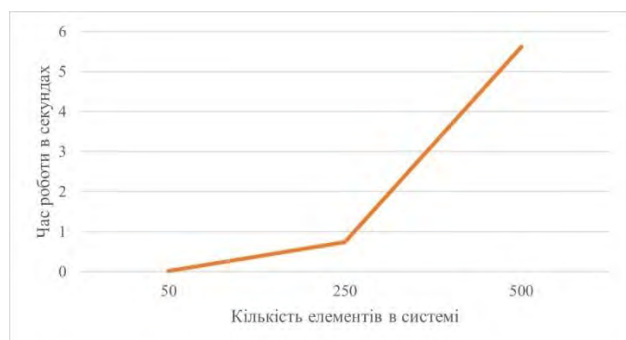


Рис. 5. Час роботи програмної платформи

Час роботи модуля генератора системи, модуля агрегації системи та модуля пошуку максимального потоку на другому рівні системи менший, ніж час роботи модуля пошуку максимального потоку на першому рівні системи. Це зумовлено тим, що алгоритм пошуку максимального потоку працює

повільніше для системи із значною кількістю елементів. Час роботи модуля генератора для системи, що містить 500 елементів, дорівнює 0,055 с, час роботи модуля агрегації – 0,077 с.

Висновки та перспективи подальшого розвитку

У роботі розроблено програмну платформу мовою програмування C# та оцінено ефективність агрегації структурної моделі системи з використанням програмної платформи. Програмна платформа містить модуль генератора системи, модуль агрегації системи, модуль пошуку максимального потоку та модуль статистичного оброблення інформації. У модуль генератора системи передається кількість елементів системи. Унаслідок роботи модуля формується початкова система, або перший рівень системи. Початкова система передається в модуль агрегації системи, у якому здійснюється агрегація структурної моделі системи. Елементи системи об'єднуються в підсистеми, формуються контакти підсистем і зв'язки між підсистемами. За результатами роботи модуля отримується перетворена система, або другий рівень

системи. У модуль пошуку максимального потоку передається система, і внаслідок роботи модуля отримується значення максимального потоку.

Проведено дослідження із використанням програмної платформи для оцінювання ефективності агрегації структурної моделі системи та ефективності самої програмної платформи. Експеримент проведено для систем, що містять 50, 250 та 500 елементів. За результатами дослідження впливає висновок, що, застосовуючи програмну платформу, можна згенерувати систему, здійснити агрегацію структурної моделі системи та розв'язати задачу про пошук максимального потоку. Також програмна платформа дає змогу оцінити ефективність агрегації структурної моделі системи. Значення максимального потоку є однаковим для початкової та перетвореної системи, тому можна зробити висновок, що в цьому разі здійснюється агрегація структурної моделі системи зі збереженням коректності її параметрів.

Подальша робота полягає у виконанні таких завдань: дослідження систем із кількістю вершин від 500 до 1000; оцінювання ефективності самої програмної платформи; використання агрегації структурної моделі системи у сфері комп'ютерних мереж.

Список літератури

1. Newman M. *Networks: An Introduction*, Oxford University Press. 2010. 1042 p.
URL:[https://www.scirp.org/\(S\(351jmbntvnsjtladkpozje\)\)/reference/ReferencesPapers.aspx?ReferenceID=1892198](https://www.scirp.org/(S(351jmbntvnsjtladkpozje))/reference/ReferencesPapers.aspx?ReferenceID=1892198)
2. Jinhu Lu, Guanrong Chen, Maciej Ogorzalek, Ljiljana Trajkovic *Theory and Applications of Complex Networks: Advances and Challenges. Proceedings of the IEEE International Symposium on Circuits and Systems*. 2013. P. 184–198. DOI:10.1109/ISCAS.2013.6572335
3. Hackl J. *Tikz-network: a LaTeX library for vizualizing complex networks*. 6th International Conference on Complex Networks & Their Applications. Lyon, France. 2017. URL: <https://github.com/hackl/tikz-network>
4. Gorbachov V., Batiia A. K., Ponomarenko O., Romanenkov Y. Formal transformations of structural models of complex network systems. *Proceedings of the IEEE 9th International Conference on Dependable Systems, Services and Technologies DESSERT'2018*. Kyiv, Ukraine. 2018. P. 473–477. DOI: <https://doi.org/10.1109/DESSERT.2018.8409175>
5. Ponomarenko O., Gorbachov V., Batiia A. K., Kotkova O. The Software Platform for Evaluation of Effectiveness of Network Systems Analysis Technologies. *IEEE East-West Design & Test Symposium (EWDTS), Conference proceedings*. Batumi, Georgia. 2019. P. 513–516. DOI: <https://doi.org/10.1109/EWDTS.2019.8884421>
6. Staudt C., Hamann M., Gutfraind A., Safro I., Meyerhenke H. Generating realistic scaled complex networks. *Applied Network Science*. 2017. Vol. 2(36). DOI: <https://doi.org/10.1007/s41109-017-0054-z>
7. Ashraf A., Budka M., Musial K. NetSim – The framework for complex network generator / A. Ashraf et al. *Procedia Computer Science*. 2018. Vol. 126. DOI: <https://doi.org/10.1016/j.procs.2018.07.289>
8. George-Williams H., Santhosh T. V., Patelli E. Simulation Methods for the Analysis of Complex Systems. In: Aslett L. J. M., Coolen F. P. A., De Bock J. *Uncertainty in Engineering, SpringerBriefs in Statistics, Springer*. 2022. P. 95–113. DOI: https://doi.org/10.1007/978-3-030-83640-5_7
9. Cheng X., Scherpen J. Model Reduction Methods for Complex Network Systems. *Annual Review of Control Robotics and Autonomous Systems*. 2020. Vol. 4. P. 425–453. DOI: <https://doi.org/10.1146/annurev-control-061820-083817>
10. Shortle J. F., Mark B. L., Gross D. Reduction of closed queueing networks for efficient simulation. *ACM Transactions on Modeling and Computer Simulation*. 2009. Vol. 19. № 3. Article 10. P. 1–22. DOI: <https://doi.org/10.1145/1540530.1540531>

11. MacKay R. S. Hierarchical aggregation of complex systems. *Proceedings of the ECCS'11*. Vienna, Austria. 2011.
12. Пономаренко О. Є., Горбачов В. О. Агрегація структурної моделі складних мережних систем. *Системи управління, навігації та зв'язку. Збірник наукових праць*. 2023. Т. 1. № 71. С. 138–144. DOI: <https://doi.org/10.26906/SUNZ.2023.1.138>
13. Gorbachov V., Sytnikov D., Ryabov O., Batiaa A. K., Ponomarenko O. Dimension Reduction for Network Systems Using Structure Model Aggregation. *International Journal of Design & Nature and Ecodynamics*. 2020. Vol. 15. No. 1. P. 13–23. DOI: <https://doi.org/10.18280/ijdne.150103>
14. Cormen T., Leiserson C., Rivest R., Stein C. Introduction to algorithms. 2009. 3rd ed. 1313 p. URL: https://pd.daffodilvarsity.edu.bd/course/material/book-430/pdf_content
15. Ford L., Fulkerson D. Maximal Flow Through a Network. *Canadian Journal of Mathematics*. 1956. Vol. 8. P. 399–404. URL: <https://www.cambridge.org/core/journals/canadian-journal-of-mathematics/article/maximal-flow-through-a-network/5D6E55D3B06C4F7B1043BC1D82D40764>

References

1. Newman M. Networks: An Introduction, Oxford University Press. 2010. 1042 p. URL: [https://www.scirp.org/\(S\(351jmbntvnsjt1aadkpozje\)\)/reference/ReferencesPapers.aspx?ReferenceID=1892198](https://www.scirp.org/(S(351jmbntvnsjt1aadkpozje))/reference/ReferencesPapers.aspx?ReferenceID=1892198)
2. Lu, J., Chen, G., Ogorzalek, M., Trajkovic, L. (2013), "Theory and Applications of Complex Networks: Advances and Challenges", *Proceedings of the IEEE International Symposium on Circuits and Systems*. P. 184-198. DOI:10.1109/ISCAS.2013.6572335
3. Hackl, J. "Tikz-network: a LaTeX library for vizualizing complex networks", 6th International Conference on Complex Networks & Their Applications, Lyon, France. 2017. available at: <https://github.com/hackl/tikz-network>
4. Gorbachov, V., Batiaa, A. K., Ponomarenko, O., Romanenkov, Y. (2018), "Formal transformations of structural models of complex network systems", *Proceedings of the IEEE 9th International Conference on Dependable Systems, Services and Technologies DESSERT'2018*, Kyiv, Ukraine. P. 473–477. DOI: <https://doi.org/10.1109/DESSERT.2018.8409175>
5. Ponomarenko, O., Gorbachov, V., Batiaa, A. K., Kotkova, O. (2019), "The Software Platform for Evaluation of Effectiveness of Network Systems Analysis Technologies", *IEEE East-West Design & Test Symposium (EWDTS), Conference proceedings*, Batumi, Georgia. P. 513–516. DOI: <https://doi.org/10.1109/EWDTS.2019.8884421>
6. Staudt, C., Hamann, M., Gutfraind, A., Safro, I., Meyerhenke, H. (2017), "Generating realistic scaled complex networks", *Applied Network Science*, Vol. 2(36). DOI: <https://doi.org/10.1007/s41109-017-0054-z>
7. Ashraf, A., Budka, M., Musial, K. (2018), "NetSim – The framework for complex network generator", *Procedia Computer Science*, Vol. 126. DOI: <https://doi.org/10.1016/j.procs.2018.07.289>
8. George-Williams, H., Santhosh, T. V., Patelli, E. (2022), "Simulation Methods for the Analysis of Complex Systems", *Uncertainty in Engineering, SpringerBriefs in Statistics*, Springer. P. 95–113. DOI: https://doi.org/10.1007/978-3-030-83640-5_7
9. Cheng, X., Scherpen, J. (2020), "Model Reduction Methods for Complex Network Systems", *Annual Review of Control Robotics and Autonomous Systems*, Vol. 4. P. 425–453. DOI: <https://doi.org/10.1146/annurev-control-061820-083817>
10. Shortle, J. F., Mark, B. L., Gross, D. (2009), "Reduction of closed queueing networks for efficient simulation", *ACM Transactions on Modeling and Computer Simulation*, Vol. 19, No. 3, Article 10. P. 1–22. DOI: <https://doi.org/10.1145/1540530.1540531>
11. MacKay, R. S. (2011), "Hierarchical aggregation of complex systems", *Proceedings of the ECCS'11*, Vienna, Austria.
12. Ponomarenko, O., Gorbachov, V. (2023), "Aggregation of structural model of complex network systems", ["Ahrehatsiia strukturalnoi modeli skladnykh merezhnykh system"], *Control, Navigation and Communication Systems. Academic Journal*, Vol. 1 (71), P. 138–144. DOI: <https://doi.org/10.26906/SUNZ.2023.1.138>
13. Gorbachov, V., Sytnikov, D., Ryabov, O., Batiaa, A. K., Ponomarenko, O. (2020), "Dimension Reduction for Network Systems Using Structure Model Aggregation", *International Journal of Design & Nature and Ecodynamics*, Vol. 15, No. 1, P. 13–23. DOI: <https://doi.org/10.18280/ijdne.150103>
14. Cormen, T., Leiserson, C., Rivest, R., Stein, C. "Introduction to algorithms", 3rd ed. 2009. 1313 p. available at: https://pd.daffodilvarsity.edu.bd/course/material/book-430/pdf_content
15. Ford, L., Fulkerson, D. (1956), "Maximal Flow Through a Network", *Canadian Journal of Mathematics*, Vol. 8, P. 399–404. available at: <https://www.cambridge.org/core/journals/canadian-journal-of-mathematics/article/maximal-flow-through-a-network/5D6E55D3B06C4F7B1043BC1D82D40764>

Відомості про авторів / About the Authors

Пономаренко Ольга Євгенівна – Харківський національний університет радіоелектроніки, аспірант кафедри електронних обчислювальних машин, Харків, Україна; e-mail: olha.ponomarenko@nure.ua; ORCID ID: <https://orcid.org/0009-0002-4634-6552>

Горбачов Валерій Олександрович – кандидат технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри електронних обчислювальних машин, Харків, Україна; e-mail: valeriy.gorbachov@nure.ua; ORCID ID: <https://orcid.org/0000-0003-3423-2371>

Ponomarenko Olha – Kharkiv National University of Radio Electronics, postgraduate student at the Department of Electronic Computers, Kharkiv, Ukraine.

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

THE SOFTWARE PLATFORM FOR EVALUATION OF THE EFFECTIVENESS OF STRUCTURAL MODEL AGGREGATION OF COMPLEX SYSTEMS

Complex systems have a large dimension, consist of a large number of elements and connections between them. Networks are used to represent complex systems. Due to the high dimension of modern systems, researchers evaluate proposed solutions using generated networks. The high dimension of the system leads to problems in modeling and control. To solve these problems, methods for reducing the dimension of complex systems are required. Aggregation of the structural model of the system consists in combining elements of the system into subsystems and as a result dimension of the system and computational complexity are reduced. **The subject matter of research** is a software platform for evaluation of the effectiveness of structural model aggregation of complex systems. **The goal of the work** is to develop the software platform for evaluation of the effectiveness of structural model aggregation of complex systems. **The relevance of the work** lies in the fact that using the software platform it is possible to make the structural model aggregation of systems with a large number of elements and it is also possible to evaluate the effectiveness of structural model aggregation of the system. **The following tasks were solved** in the work: development of the software platform which consists of a system generator module, a system aggregation module, a maximum flow searching module and a statistical data processing module; structural model aggregation of the system using the software platform; evaluation of the effectiveness of structural model aggregation of the system using the software platform. **As a result of the work** the software platform was created, the structural model aggregation of the system was made and the effectiveness of structural model aggregation of the system using the software platform was evaluated. The studies allow us to **conclude**: using the software platform it is possible to generate a system, make the structural model aggregation of the system and solve the problem of the maximum flow searching; the software platform also allows to evaluate the effectiveness of structural model aggregation of the system; the value of the maximum flow is the same at two levels of the system, so in this case the structural model aggregation of the system is made while preserving the correctness of its parameters.

Keywords: software platform; structural model aggregation; complex systems; system generator; maximum flow.

Бібліографічні описи / Bibliographic descriptions

Пономаренко О. Є., Горбачов В. О. Програмна платформа для оцінювання ефективності агрегації структурної моделі складних систем. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 79–87. DOI: <https://doi.org/10.30837/ITSSI.2023.254.079>

Ponomarenko, O., Gorbachov, V. (2023), "The software platform for evaluation of the effectiveness of structural model aggregation of complex systems", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 79–87. DOI: <https://doi.org/10.30837/ITSSI.2023.254.079>

Б. РЕЗАНОВ, Г. КУЧУК

МОДЕЛЬ РОЗПОДІЛУ ЕЛЕМЕНТАРНИХ ПОТОКІВ ДАНИХ У ТУМАННІЙ ПЛАТФОРМІ ПІДТРИМКИ ІНТЕРНЕТУ РЕЧЕЙ

Предметом дослідження є моделі управління ресурсами та завданнями в туманному середовищі підтримки Інтернету речей (IoT). Зростання кількості підключених пристроїв та обсяги зібраної інформації в мережах IoT роблять актуальним необхідність удосконалення систем управління, які забезпечують оптимальний розподіл завдань і ресурсів. Туманне обчислення відіграє ключову роль у реалізації цього питання, розподіляючи обчислювальні завдання ближче до джерела інформації та кінцевих користувачів. **Мета роботи** полягає в підвищенні ефективності технологій туманних обчислень для забезпечення оптимального розподілу завдань і ресурсів у мережі IoT. Для досягнення мети розглянуті методи кластеризації, що допомагають створити групи обчислювальних ресурсів і визначити, які завдання необхідно розподілити між цими групами. Застосування відповідних методів кластеризації дає змогу зменшити затримки та підвищити загальну продуктивність системи IoT. **Основні завдання цієї роботи.** По-перше, зважаючи на різноманітні вимоги до обчислювальних ресурсів та завдань IoT, необхідно розглянути наявні методи й розробки. По-друге, важливо дослідити та порівняти методи кластеризації, зокрема *DBSCAN* та *C-Means*, для ефективного управління ресурсами. Метод кластеризації *DBSCAN* дає змогу ефективно розподіляти завдання залежно від їх місця розташування. Метод *C-Means* допомагає групувати ресурси за їх характеристиками. Третє завдання – розробити модель, основу на вхідних параметрах, таких як відповідь системи, потреби кластерів у ресурсах, віддаленість інформації для її оброблення тощо. Модель дасть змогу аналізувати ймовірні сценарії та приймати рішення щодо оптимального розподілу завдань і ресурсів у середовищі IoT. **Висновки.** Це дослідження спрямоване на розв'язання актуального питання – управління ресурсами та завданнями в туманному середовищі IoT. Розглянуто наявні методи й розробки у сфері управління ресурсами та завданнями в IoT. Порівняно методи кластеризації *DBSCAN* і *C-Means* для визначення їх ефективності в управлінні ресурсами. Розроблено теоретико-множинну модель, що ґрунтується на різних параметрах для прийняття оптимальних рішень щодо розподілу завдань і ресурсів. Установлено, що впровадження методів кластеризації та розробленої моделі допомагають підвищити продуктивність системи й забезпечити більш ефективне застосування обчислювальних ресурсів у туманному середовищі IoT.

Ключові слова: туманні обчислення; Інтернет речей; метод *DBSCAN*; метод *C-Means*; кластеризація; теоретико-множинне моделювання.

Вступ

У сучасному світі технології розвиваються на шаленій швидкості, відкриваючи перед людством безмежні можливості та виклики. Інтернет речей (IoT) [1] є однією з найбільш яскравих технологічних революцій, що змінює наше оточення та спосіб життя. Завдяки IoT об'єкти зовнішнього світу, починаючи від побутових приладів до великих виробничих систем, можуть спілкуватися, обмінюватися інформацією та приймати рішення, не потребуючи прямої участі людини.

Проте реалізація IoT та забезпечення його працездатності вимагають ефективного управління ресурсами й завданнями. На допомогу приходять туманні обчислення, які дають змогу обробляти

інформацію та приймати рішення на самому "краю" мережі, ближче до джерела їх виникнення [2, 3]. Але як можна забезпечити ефективне управління в умовах туманного середовища?

У цій статті досліджується два ключові аспекти цього питання: методи кластеризації та система розподілу завдань у туманному середовищі підтримки IoT. Методи кластеризації допомагають групувати ресурси та об'єкти в єдиноклісній структурі, що полегшує їх ефективне використання. Система розподілу завдань забезпечує оптимальний розподіл завдань між ресурсами в мережі IoT [4].

Розглянемо метод кластеризації *DBSCAN* (*Density-Based Spatial Clustering of Noisy Applications*) [5, 6], що дає змогу створювати кластери з ресурсів, ґрунтуючись на їх розташуванні, і метод

кластеризації *C-Means* [7], розподілений на *CPU Cluster*, *GPU Cluster* та *RAM Cluster* для ефективного використання обчислювальних ресурсів. Теоретико-множинна модель допоможе зрозуміти, як оптимально розподіляти завдання та ресурси.

Аналіз літератури

Аналіз літератури вказує на декілька ключових проблем, пов'язаних із підтримкою IoT та кластеризацією даних, що розглядаються у статті.

1. Ефективне управління інформацією IoT у хмарних середовищах. У дослідженні [1], наголошується на важливості хмарних технологій для оброблення та зберігання даних IoT. Проте це може призвести до питань щодо ефективного управління ресурсами хмари.

2. Кластеризація та оброблення даних IoT. У дослідженнях [5, 6] висвітлюється важливість алгоритмів кластеризації даних IoT для подальшого аналізу та оброблення. Проте існує проблема в обранні найбільш ефективного алгоритму для конкретних завдань.

3. Енергозбереження в смарт-містах. У статті [4] вказано на необхідність мінімізації споживання енергії датчиками IoT в смарт-містах. Ця проблема стосується оптимізації роботи датчиків та їх енергоефективності.

4. Паралельні обчислення на GPU. У роботах [11, 14] описані переваги використання графічних процесорів (GPU) для швидкодії та паралельних обчислень. Проте існує проблема оптимального розподілу завдань між різними обчислювальними ресурсами.

5. Розподілені обчислення. У дослідженнях [5, 8–10] наголошується на важливості розподілених обчислень, зокрема на використанні платформи *Spark*. Але питання ефективності та оптимізації розподілених обчислень залишається актуальним.

Мета статті – розглянути способи розв'язання перелічених проблем, зокрема розробити систему розподілу завдань із використанням методів кластеризації та оптимізації ресурсів у туманних середовищах підтримки IoT. Для цього створено теоретико-множинну модель для визначення оптимального способу розподілу завдань і ресурсів. Такий підхід спрямований на покращення продуктивності та ефективності системи підтримки IoT на туманній платформі.

1 Методи кластеризації туманного середовища підтримки IoT

Ключовими складниками успішної системи туманного обчислення є методи кластеризації, що дають змогу раціонально розподіляти завдання та ресурси в системі. Два основних методи кластеризації, що застосовуються в цій роботі, – це *DBSCAN* і *C-Means Clustering*. Розглянемо їх докладніше.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) – метод кластеризації, що оснований на щільності інформації та допомагає виявляти об'єкти, які належать до одного кластера на основі їх географічного розташування [8]. У досліджуваній системі використано *DBSCAN* для організації обчислювальних ресурсів у різних локаціях. Наприклад, *DBSCAN Cluster 1* може відповідати одному регіону, *DBSCAN Cluster 2* – іншому і т.д. [9, 10].

C-Means Clustering, або нечітка кластеризація, застосовується для подальшої кластеризації ресурсів у кожному *DBSCAN Cluster*. Розбито ресурси на кластери, такі як *CPU Cluster*, *GPU Cluster* і *RAM Cluster*, щоб ефективно розподіляти завдання з огляду на тип ресурсів [11].

Така комбінація методів кластеризації дає змогу оптимізувати оброблення інформації, розподіляти завдання та забезпечувати високу продуктивність в елементарних потоках даних у туманній платформі підтримки IoT. Математичні моделі та алгоритми цих методів допомагають у прийнятті рішень та покращенні функціонування системи.

Побудуємо діаграму кластеризації двома обраними методами (рис. 1).

Ця діаграма відображає систему, що використовує два методи кластеризації для управління завданнями й ресурсами в туманному середовищі підтримки IoT.

Туманне середовище

Система поділена на три кластери *DBSCAN*, позначені як *DBSCAN Cluster 1*, *DBSCAN Cluster 2* та *DBSCAN Cluster 3*. Кожному кластеру *DBSCAN* відповідають різні локації, де виконуються обчислення. У кожному кластері *DBSCAN* є вузли, позначені як *Node X*, де *X* – це номер вузла.

Кластери C-Means

Усередині кожного кластера *DBSCAN* є підкластери *C-Means*, що подані як *C-Means Cluster*. *CPU Cluster* призначений для оброблення на *CPU*.

RAM Cluster – для оброблення завдань в оперативній пам'яті (*RAM*).

Зв'язки

Вузли кожного кластера *DBSCAN* взаємодіють з відповідними кластерами *C-Means* для оброблення завдань. Наприклад, *Node 1* з *DBSCAN Cluster 1*

взаємодіє з *CPU Cluster* для оброблення завдань на *CPU*.

Також відображено зв'язки між кластерами *DBSCAN* і *C-Means*, що показує, як ресурси та завдання розподіляються між різними локаціями та видами обчислень [12].

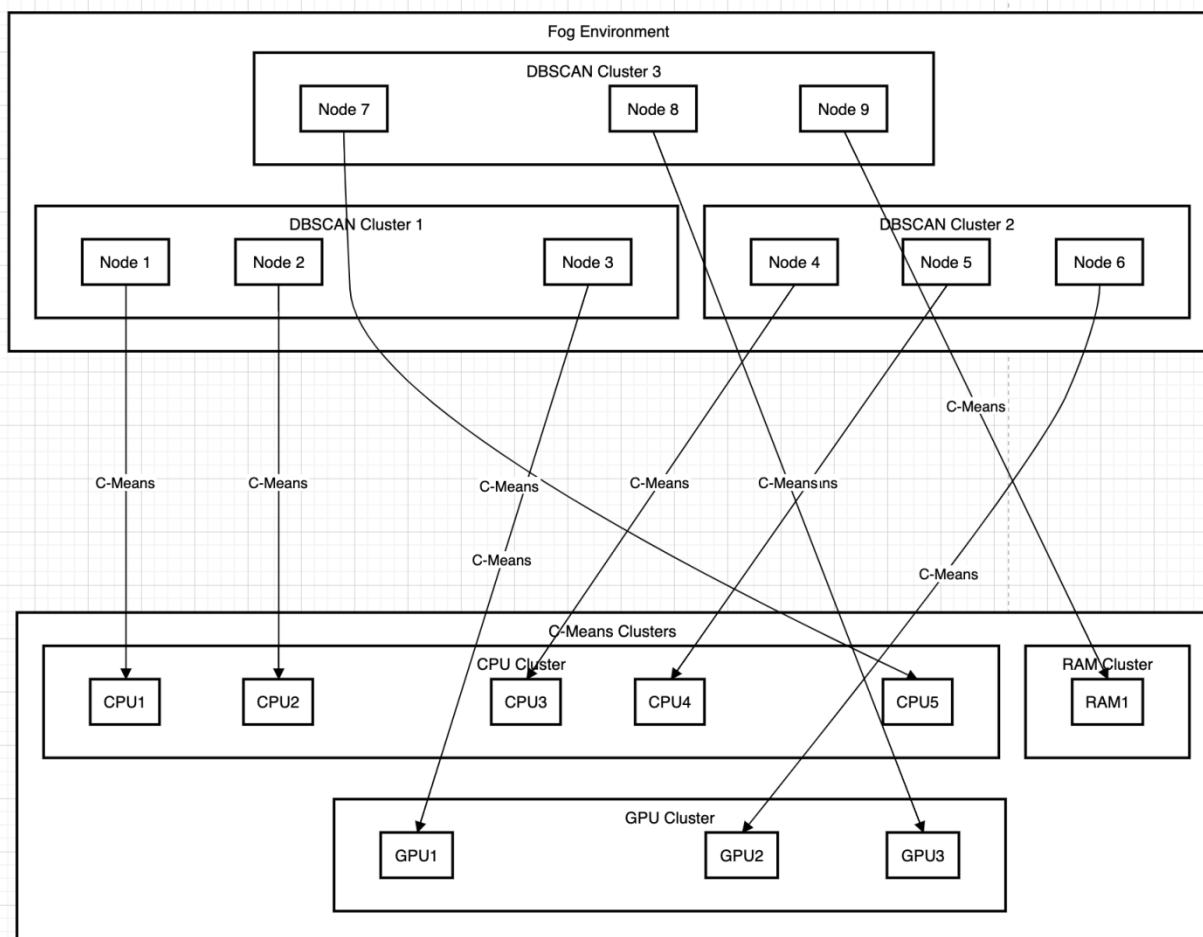


Рис. 1. Діаграма кластеризації двома методами *DBSCAN* і *C-Means Clustering*

Наведена система дає змогу ефективно управляти завданнями та ресурсами в розподіленому середовищі обчислень туманної платформи підтримки IoT, використовуючи два методи кластеризації для оптимізації оброблення даних у реальному часі.

Після розгляду методів кластеризації та їх застосування в системі туманного обчислення опишемо систему розподілу завдань і ресурсів у туманному середовищі. Діаграма й методи кластеризації визначають ключові аспекти організації роботи системи [13]. Варто наголосити, що ці методи кластеризації мають важливе

значення не лише для туманного обчислення, але й для багатьох галузей науки та промисловості. Кластеризація дає змогу групувати схожі об'єкти для подальшого аналізу, оброблення та прийняття рішень.

Методи кластеризації стали основою для створення ефективної системи розподілу завдань і ресурсів у мережі туманної платформи. Вони допомагають ідентифікувати оптимальні рішення для оброблення інформації на краях мережі, забезпечуючи найкращий розподіл завдань між кластерами та краями з огляду на фізичне розташування, доступні ресурси та час відповіді.

У наступних розділах більш детально розглянемо теоретико-множинну модель системи, що допомагає оптимізувати процеси розподілу завдань та кластеризації ресурсів, зробивши платформу туманного обчислення більш продуктивною та відповідною до вимог сучасних застосунків Інтернету речей та інших сфер застосування

2 Система розподілу завдань у середовищі туманних обчислень, підтримки IoT

Система розподілу завдань у середовищі туманних обчислень, підтримки IoT покладається на кілька вхідних параметрів, зокрема затримку відповіді, ємність кластера та віддаленість інформації для оброблення [14]. Система використовує два типи методів кластеризації туманного середовища: *C-Means* (з використанням кластерів *CPU*, *GPU* та *RAM*) і *DBSCAN* (на основі розташування кластерів 1, 2 і 3). Завдання зберігаються в *S3*-подібному сховищі, особливому для кожного кластера *DBSCAN*. Зображення системи подане на рис. 2.

Затримка відповіді (у секундах). Цей параметр визначає максимально допустиму затримку для виконання завдань й гарантує, що вони обробляються протягом визначеного часу.

Ємність кластера (у довільних одиницях). Ємність кластера вказує на обчислювальні ресурси, доступні в кожному кластері, зокрема *CPU*, *GPU* та *RAM* [15]. Цей параметр кількісно визначає максимальне робоче навантаження, що може ефективно впоратися з кластером.

Віддаленість інформації для оброблення. Віддаленість даних відображає відстань або затримку мережі між джерелом інформації й туманом, призначеним для оброблення. Дуже важливо звести до мінімуму затримки передачі даних.

Генерація завдань. Пристрої IoT створюють різні обчислювальні завдання, кожне зі своєю складністю та рівнем пріоритету.

Профілювання завдань. Кожне завдання профільовано, щоб визначити його конкретні вимоги, зокрема обмеження часу відповіді, вимоги до обчислювальних ресурсів і інформацію про географічне розташування.

Кластеризація крапель туману. Система використовує метод кластеризації для організації крапель туману на основі їх географічного

розташування. Краплі туману організовані в кластери, що відповідають локації 1, локації 2 або локації 3.

Призначення кластера. Завдання спочатку розподіляються за місцем, у якому вони були створені. Система оцінює доступність ресурсів у місці створення завдання та перевіряє обмеження часу відповіді. Якщо ресурсів не достатньо або обмеження часу відповіді не може бути виконане в місці походження завдання, система шукає альтернативні місця з доступними ресурсами та меншою затримкою.

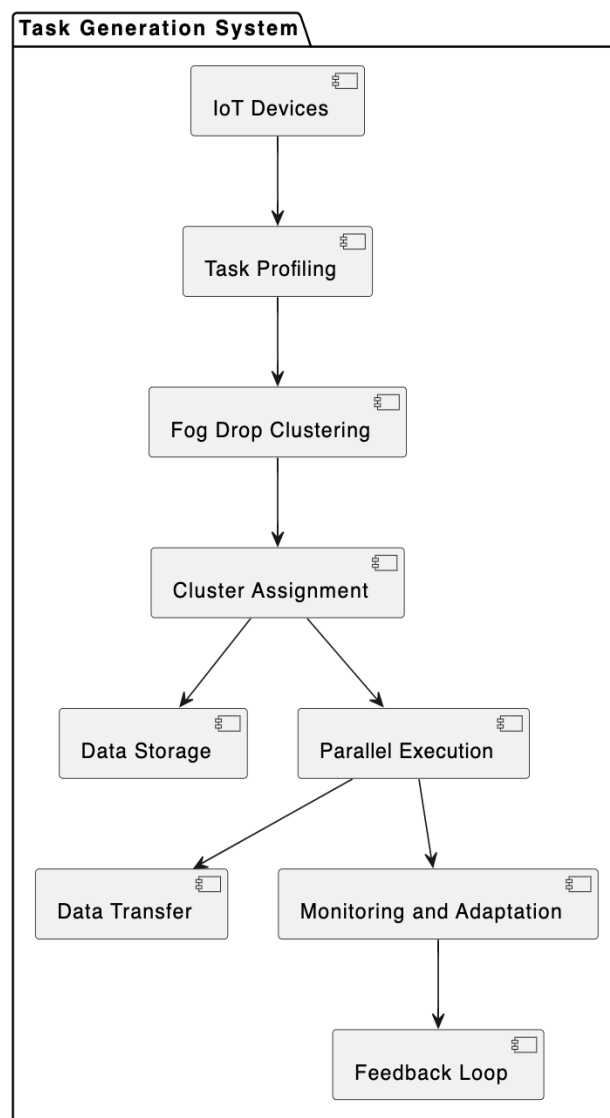


Рис. 2. Опис системи розподілу завдань і ресурсів

На основі цієї оцінки система перепризначає завдання кластеру, який найкраще відповідає її вимогам. Наприклад, якщо для завдання потрібне оброблення *GPU*, а ресурси в розташуванні

локації 1, але в локації 3 є доступні ресурси GPU з меншою затримкою порівняно з розташуванням 2, завдання перенаправляється в локацію 3, зокрема в кластер GPU [16].

Зберігання даних. Завдання зберігаються у спеціальному S3-подібному сховищі в кожному кластері, що забезпечує доступність інформації для оброблення.

Паралельне виконання. У кожному кластері завдання виконуються паралельно, ефективно використовуючи доступні ресурси, дотримуючись обмежень часу відповіді.

Передача даних. Якщо завдання потребує інформації, якої наразі немає у призначеному кластері, запускаються механізми передачі даних, щоб перемістити необхідну інформацію ближче до вузла оброблення й мінімізувати віддаленість даних.

Моніторинг і адаптація. Система постійно контролює виконання завдань, навантаження на кластер і час відповіді. Якщо кластер наближається до обмежень ємності або час відповіді перевищує вказані порогові значення, система може перерозподілити завдання іншим кластерам або за потреби виділити додаткові ресурси.

Цикл зворотного зв'язку. Показники продуктивності, зокрема час відповіді та використання ресурсів, збираються та аналізуються. Ці показники слугують зворотним зв'язком для вдосконалення стратегій розподілу завдань, адаптації до змінних умов і оптимізації розподілу завдань у часі.

Беручи до уваги перелічені вхідні параметри та використовуючи методи кластеризації крапель туману, середовище підтримки IoT ефективно обробляє завдання, забезпечуючи вчасне виконання, ефективне застосування ресурсів і мінімізовану віддаленість інформації на основі географічного розташування. Цей підхід покращує керування застосунками Інтернету речей у середовищі туманних обчислень.

Методи розподілу завдань і кластеризації ресурсів є критично важливими складниками для досягнення оптимального функціонування системи підтримки Інтернету речей в умовах туманного обчислення. У змінному та динамічному середовищі, де багато факторів може впливати на прийняття рішень, застосування теоретико-множинних моделей стає невід'ємним складником для покращення цих процесів.

3 Теоретико-множинна модель

Теоретико-множинна модель є відображенням системи, якій властиві важливі параметри та їх взаємозв'язки. У нашій конкретній ситуації теоретико-множинна модель містить функціональні залежності між вхідними параметрами, такими як відповідь системи на завдання, потреби кластерів у ресурсах і віддаленість інформації, яка потребує оброблення [17].

Було створено теоретико-множинну модель, що дає змогу системі аналізувати ймовірні сценарії та приймати найефективніші рішення щодо розподілу завдань з метою максимізувати продуктивність і забезпечити вчасне виконання завдань. Використання теоретико-множинних моделей допомагає системі оптимізувати роботу, забезпечуючи краще задоволення потреб користувачів і максимальну результативність.

3.1 Потік робіт

Потік робіт може бути поданий у вигляді орієнтованого ациклічного графа (OAG):

$$H = (V, E), \quad (1)$$

де H – монолітний потік робіт;

V – множина вершин, що становлять обчислювальні завдання;

E – множина спрямованих ребер, які з'єднують вершини, що становлять залежності за показниками між завданнями.

Кожна вершина v_i в V може мати вхідні та/або вихідні ребра. Вихідний ступінь $\deg^+(v_i)$ вершини v_i в H визначається як кількість ребер, що виходять з v_i , спрямованих до інших вершин. Вхідний ступінь $\deg^-(v_i)$ вершини v_i в H визначається як кількість ребер, спрямованих до v_i від інших вершин. Ребро $(v_i, v_j) \in E$ є залежністю за даними від v_i до v_j . Нехай n – це загальна кількість вершин в V .

3.2 Під потоки робіт

Потік робіт H поділяється на множину, що складається з k підпотоків робіт $S = (S_1, \dots, S_k)$, $S_i = (V_i, E_i)$, якщо V_i – це множина вершин, що

входять у підпотік робіт S_i ; E_i – це множина ребер між вершинами, що входять до V_i . У цьому разі множина вершин H поділяється підпотоками робіт на набір непересічних підмножин:

- 1) $\forall i \in 1..k : V_i \subset V, E_i \subset E$;
- 2) $\forall v \in V, \exists i \in 1..k : v \in V_i$;
- 3) $E_i = \{(v_k, v_l) \in E : v_k, v_l \in V_i\}$;
- 4) $\forall i, j : i \neq j \Rightarrow V_i \cap V_j = \emptyset$.

Визначимо наступні класи ребер і вершин, пов'язаних із підпотокком робіт S_i :

1) EI_i : набір вхідних ребер із початковою вершиною поза підпотокком робіт S_i і кінцевою вершиною всередині підпотокку робіт S_i :

$$EI_i = \{(v_k, v_l) \in E : v_k \notin S_i, v_l \in S_i\}; \quad (2)$$

2) EO_i : набір вихідних ребер із початковою вершиною всередині підпотокку робіт S_i і кінцевою вершиною поза підпотокком робіт S_i :

$$EO_i = \{(v_k, v_l) \in E : v_k \in S_i, v_l \notin S_i\}; \quad (3)$$

3) VI_i : набір вершин в S_i , розташованих на основній частині ребер EI_i , а також вершин, що не мають вхідних ребер:

$$VI_i = \{v_l \in S_i : (v_k, v_l) \in EI_i\} \cup \{v \in S_i : \deg^-(v) = 0\}; \quad (4)$$

4) VO_i : набір вершин в S_i , розміщених на кінцях ребер EO_i , а також вершин, що не мають вихідних ребер:

$$VO_i = \{v_k \in S_i : (v_k, v_l) \in EO_i\} \cup \{v \in S_i : \deg^+(v) = 0\}. \quad (5)$$

3.3 Визначення елементарного потоку робіт

Щоб перетворити підпотік робіт S_i в елементарний потік робіт (*Elemental Flow*), необхідно витягти всі вершини S_i з потоку робіт H і забезпечити комунікаційні механізми, що пов'язують EF_i з платформою потокової передачі подій через вершину-споживач cv_i та вершину-генератор pv_i .

Вершина cv_i в EF_i забезпечує споживання потоку вхідних даних від платформи потокової передачі даних і розподіляє його між вершинами VI_i . Вершина pv_i в EF_i діє як сток,

що збирає вихідні дані з вершин, що становлять безліч VO_i і передає їх у вигляді повідомлень на платформу потокової передачі даних. Визначимо відповідні множини ребер таким чином:

$ECV_i = \{(cv_i, v) : v \in VI_i\}$, набір ребер, що йдуть від cv_i до вершин у VI_i ;

$EPV_i = \{(v, cp_i) : v \in VO_i\}$, набір ребер, що йдуть від вершин у VO_i до cp_i .

У цьому разі елементарний потік робіт EF_i з підпотокку робіт S_i визначається як

$$EF_i = (TV_i, TE_i), \quad (6)$$

де $TV_i = V_i \cup \{cv_i, pv_i\}$ – множина вершин, що розташовані всередині S_i , зокрема cv_i та pv_i ;

$TE_i = E_i \cup ECV_i \cup EPV_i$ – множина всіх ребер, розташованих усередині S_i , зокрема всі ребра, які йдуть із cv_i до вершин у VI_i , а також всі ребра, що йдуть від вершин у VO_i до cp_i .

3.4 Модель елементарних потоків робіт

Модель елементарних потоків робіт поєднує моделі потоків робіт і потокового оброблення інформації. У структурі платформи потокового оброблення інформації формується набір виділених каналів (сховищ потоків даних) для організації взаємодії елементарних потоків робіт за допомогою обміну повідомленнями. Кожне повідомлення є набором даних, що містить:

- позначку часу створення повідомлення;
- інформацію про джерело даних;
- структуровану колекцію даних, що передаються.

DS_0 відповідає за збирання, зберігання та надання повідомлень, що містять набори даних, необхідних для ініціалізації обчислювального процесу в потоці робіт;

DS_i відповідає за отримання повідомлень, що містять проміжні дані, від EF_i , і передачу їх залежним елементарним потокам робіт;

DS_{out} відповідає за збирання, зберігання та надання повідомлень, що містять результуючі дані.

Замість початкового вузла потоку робіт, набори даних, необхідні для початку обчислювального

процесу, надходять у сховище потоків даних DS_0 у вигляді повідомлень. Оброблення повідомлень зі сховища потоку даних організоване таким чином:

1) cv_i відповідного EF_i витягує наступне повідомлення з DS_{i-1} ;

2) на основі аналізу отриманого повідомлення cv_i ініціює передачу даних по ребрах ECV_i до вершин, відповідальних за виконання безпосереднього обчислення;

3) після завершення завдань оброблення даних та їх надсилання по ребрах EPV_i pv_i генерує вихідне повідомлення в DS_i або DS_{out} , якщо це остаточний результат.

На рис. 3 зображено застосування моделі елементарних потоків робіт. Блок $EF1$ є елементарним потоком робіт і містить вершину

системи прийняття рішення cv та вершину відповідача pv , обчислювальні завдання V_i . Блок $EF1$ після виконання своїх завдань передає інформацію в сховище DS_1 для подальшого оброблення. Блок $EF2$ є аналогом блоку $EF1$, але зі своїми обчислювальними завданнями V_i . Стрілками показано переміщення інформації з системи прийняття рішень до виконання безпосереднього обчислювального процесу. Пунктирними стрілками показано переміщення інформації в/із сховища потоків даних. Унаслідок застосування запропонованої моделі було отримано можливість відокремити елементарні потоки даних із загальної інформації (підпотоків даних). Відокремлення елементарних потоків даних дало змогу покращити якість оброблення й аналізу інформації та швидкість оброблення даних.

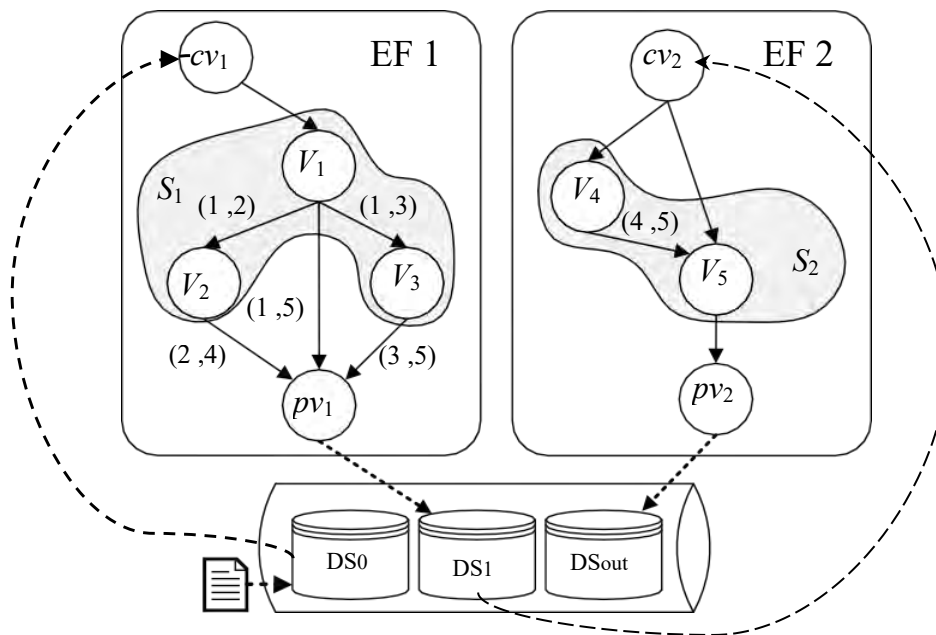


Рис. 3. Результат застосування моделі елементарного потоку робіт

Запропонована модель враховує параметри, зокрема пропускну спроможність мережі, підключення між різними елементами системи та більшу кількість змінних для оптимізації розподілу елементарних потоків даних у туманному середовищі підтримки IoT.

Висновок

У роботі досліджено та впроваджено два ключові методи кластеризації, що виявилися

надзвичайно важливими для оптимізації та управління завданнями й ресурсами в туманному середовищі підтримки Інтернету речей (IoT). Подана діаграма ілюструє важливий етап у розвитку системи, яка може істотно покращити продуктивність та ефективність використання ресурсів у мережах туманної платформи підтримки IoT.

Метод *DBSCAN*, оснований на локаціях кластерів, і метод *C-Means*, розподілений на *CPU Cluster*, *GPU Cluster* та *RAM Cluster*,

дають змогу системі ефективно визначати, куди та як розподіляти завдання й ресурси. Ці методи допомагають уникати перевантаження й забезпечують розподіл ресурсів відповідно до вимог завдань.

Запропонована теоретико-множинна модель – необхідний інструмент для аналізу розподілу завдань і кластеризації ресурсів. Вона дає змогу передбачити рішення, забезпечуючи максимальну продуктивність та вчасне виконання завдань.

Розглянута система розподілу завдань і ресурсів має значний потенціал для застосування в різних сферах, де важлива ефективність роботи IoT

і туманного обчислення. Вона може допомогти вирішувати завдання в реальному часі, оптимізувати використання обчислювальних ресурсів і забезпечувати високий рівень обслуговування.

Завдяки поєднанню методів кластеризації, теоретико-множинних моделей і системи розподілу є можливість створити інтелектуальне середовище, яке реагує на потреби сучасних застосунків IoT та забезпечує їх надійну роботу. Це дослідження є першим у вивченні розвитку систем туманного обчислення для майбутніх викликів і можливостей.

Список літератури

1. Alam T. Cloud-based IoT applications and their roles in smart cities. *Smart Cities*. Vol. 4(3), 2021. P. 1196–1219. DOI:10.3390/smartcities4030064
2. Al-Haija Q.A. Top-Down Machine Learning-Based Architecture for Cyberattacks Identification and Classification in IoT Communication Networks. *Frontiers in Big Data*, Vol. 4. 2022. P. 1–18. DOI: <https://doi.org/10.3389/fdata.2021.782902>
3. Uthayakumar J., Vengattaraman T. and Dhavachelvan P. A new lossless neighborhood indexing sequence (NIS) algorithm for data compression in wireless sensor networks. *Ad Hoc Networks*, Vol. 83, 2019. P. 149–157. DOI: <https://doi.org/10.1016/j.adhoc.2018.09.009>
4. Al-Hawawreh M., Elgendi I. and Munasinghe K. An Online Model to Minimize Energy Consumption of IoT sensors in Smart Cities. *IEEE Sensors Journal*, Vol. 22(20), 2022. P. 19524–19532. DOI: <https://doi.org/10.1109/JSEN.2022.3199590>
5. W. Jing, G. Chen, and Y. Cheng, DBSCAN-PSM: an improvement method of DBSCAN algorithm on Spark, *International Journal of High Performance Computing and Networking*, Vol.13, No.4, 417 p., 2019. URL: <https://www.inderscience.com/offers.php?id=99265>
6. Cordova I., Moh T. DBSCAN on Resilient Distributed Datasets, *International Conference on High Performance Computing Simulation (HPCS)*. *IEEE*. 2015, P. 531–540. DOI: <https://doi.org/10.1109/HPCSim.2015.7237086>
7. Augustine S., Ananth J.P. Taylor kernel fuzzy C-means clustering algorithm for trust and energy-aware cluster head selection in wireless sensor networks. *Wireless Networks*, Vol. 26(7), 2020. P. 5113–5132. DOI: <https://doi.org/10.1007/s11276-020-02352-w>
8. He Y., Tan H., Luo W., Mao H., Ma D., Feng S., and Fan J., MR-DBSCAN: An Efficient Parallel Density-Based Clustering Algorithm Using MapReduce, *IEEE 17th International Conference on Parallel and Distributed Systems*. *IEEE*, 2011. P. 473–480. DOI: 10.1109/ICPADS.2011.83
9. D. Han, A. Agrawal, W.-k. Liao, and A. Choudhary, Parallel DBSCAN Algorithm Using a Data Partitioning Strategy with Spark Implementation, in 2018 IEEE International Conference on Big Data (Big Data). *IEEE*, 2018. P. 305–312. URL: <https://www.scholars.northwestern.edu/en/publications/parallel-dbscan-algorithm-using-a-data-partitioning-strategy-with>
10. Gong Y., Sinnott R. O., and Rimba P. RT-DBSCAN: RealTime Parallel Clustering of Spatio-Temporal Data Using SparkStreaming, *Computational Science* 2018. P. 524–539. DOI: 10.1007/978-3-319-93698-7_40
11. Ali N., Hamida S, Cherradi B et al. A computational performance study of unsupervised data clustering algorithms on GPU. *2nd international conference on innovative research in applied science, engineering and technology (IRASET)*. *IEEE*, Meknes, 2022. P 1–6. DOI:10.1109/IRASET52964.2022.9737871
12. Cook S. *CUDA programming: a developer's guide to parallel computing with GPUs*. Elsevier, MK, Amsterdam; Boston. 2013. 591 p. URL: <https://usermanual.wiki/Pdf/Shane20CookCUDA20programming20A20developers20guide20to20parallel20computing20with20GPUs> Morgan20Kaufmann202012.1739933505/help
13. Fritz F, Schmid M, Mottok J. Accelerating real-time applications with predictable work-stealing. Architecture of computing systems. ARCS. Springer International Publishing, Cham, 2020. P. 241–255. URL: <https://europepmc.org/article/pmc/pmc7343420>
14. Li Y, Zhao K, Chu X, Liu J. Speeding up k-Means algorithm by GPUs. *Journal of Computer and System Sciences* Vol. 79, Issue 2, 2013. P. 216–229. DOI: <https://doi.org/10.1016/j.jcss.2012.05.004>

15. Sanders J, Kandrot E. CUDA by example: an introduction to general-purpose GPU programming. Addison-Wesley, Upper Saddle River, NJ. 2013. 311 p. URL: https://edoras.sdsu.edu/~mthomas/docs/cuda/cuda_by_example.book.pdf
16. Wasif M. K., Narayanan P. J. Scalable clustering using multiple GPUs. *18th international conference on high performance computing. IEEE*. Bengaluru, 2011. P. 1–10. DOI: <https://doi.org/10.1016/j.engappai.2017.10.023>
17. Rodriguez D., Gomez D., Alvarez D., Rivera S. A review of parallel heterogeneous computing algorithms in power systems. *Algorithms*. 2021. Vol. 14(10). 275 p. DOI: <https://doi.org/10.3390/a14100275>

References

1. Alam, T. (2021), "Cloud-based IoT applications and their roles in smart cities". *Smart Cities*. Vol. 4(3). P. 1196–1219. DOI:10.3390/smartcities4030064
2. Al-Hajja, Q.A. (2022), "Top-Down Machine Learning-Based Architecture for Cyberattacks Identification and Classification in IoT Communication Networks". *Frontiers in Big Data*, Vol. 4. P. 1–18. DOI: <https://doi.org/10.3389/fdata.2021.782902>
3. Uthayakumar, J., Vengattaraman, T. and Dhavachelvan, P. (2019), "A new lossless neighborhood indexing sequence (NIS) algorithm for data compression in wireless sensor networks". *Ad Hoc Networks*, Vol. 83. P. 149–157. DOI: <https://doi.org/10.1016/j.adhoc.2018.09.009>
4. Al-Hawawreh, M., Elgendi I. and Munasinghe K. (2022), "An Online Model to Minimize Energy Consumption of IoT sensors in Smart Cities". *IEEE Sensors Journal*, Vol. 22(20). P. 19524–19532. DOI: <https://doi.org/10.1109/JSEN.2022.3199590>
5. Jing, W., Chen, G., and Cheng, Y. "DBSCAN-PSM: an improvement method of DBSCAN algorithm on Spark". *International Journal of High Performance Computing and Networking*, Vol. 13, No.4, 417 p., 2019. available at: <https://www.inderscience.com/offers.php?id=99265>
6. Cordova, I., Moh, T. (2015), "DBSCAN on Resilient Distributed Datasets". *International Conference on High Performance Computing Simulation (HPCS)*. *IEEE*. P. 531–540. DOI: <https://doi.org/10.1109/HPCSim.2015.7237086>
7. Augustine, S., Ananth, J.P. (2020), "Taylor kernel fuzzy C-means clustering algorithm for trust and energy-aware cluster head selection in wireless sensor networks". *Wireless Networks*, Vol. 26(7). P. 5113–5132. DOI: <https://doi.org/10.1007/s11276-020-02352-w>
8. He, Y., Tan, H., Luo, W., Mao, H., Ma, D., Feng, S., and Fan, J. (2011), "MR-DBSCAN: An Efficient Parallel Density-Based Clustering Algorithm Using MapReduce", *IEEE 17th International Conference on Parallel and Distributed Systems. IEEE*. P. 473–480. DOI: 10.1109/ICPADS.2011.83
9. Han, D., Agrawal, A., Liao, W., and Choudhary, A. "Parallel DBSCAN Algorithm Using a Data Partitioning Strategy with Spark Implementation", in 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018. P. 305–312. available at: <https://www.scholars.northwestern.edu/en/publications/parallel-dbscan-algorithm-using-a-data-partitioning-strategy-with>
10. Gong, Y., Sinnott, R. O., and Rimba, P. (2018), "RT-DBSCAN: RealTime Parallel Clustering of Spatio-Temporal Data Using SparkStreaming", *Computational Science*. P. 524–539. DOI:10.1007/978-3-319-93698-7_40
11. Ali N., Hamida S, Cherradi B et al. (2022), "A computational performance study of unsupervised data clustering algorithms on GPU". *2nd international conference on innovative research in applied science, engineering and technology (IRASET)*. *IEEE*, Meknes, 2022. P 1–6. DOI:10.1109/IRASET52964.2022.9737871
12. Cook, S. "CUDA programming: a developer's guide to parallel computing with GPUs". Elsevier, MK, Amsterdam; Boston. 2013. 591 p. available at: <https://usermanual.wiki/Pdf/Shane20CookCUDA20programming20A20developers20guide20to20parallel20computing20with20GPUsMorgan20Kaufmann202012.1739933505/help>
13. Fritz F, Schmid M, Mottok J. "Accelerating real-time applications with predictable work-stealing. Architecture of computing systems". ARCS. Springer International Publishing, Cham, 2020. P. 241–255. available at: <https://europepmc.org/article/pmc/pmc7343420>
14. Li Y, Zhao K, Chu X, Liu J. (2013), "Speeding up k-Means algorithm by GPUs". *Journal of Computer and System Sciences* Vol. 79, Issue 2. P. 216–229. DOI: <https://doi.org/10.1016/j.jcss.2012.05.004>
15. Sanders, J, Kandrot, E. "CUDA by example: an introduction to general-purpose GPU programming". Addison-Wesley, Upper Saddle River, NJ. 2013. 311 p. available at: https://edoras.sdsu.edu/~mthomas/docs/cuda/cuda_by_example.book.pdf
16. Wasif, M. K., Narayanan, P. J. (2011), "Scalable clustering using multiple GPUs". *18th international conference on high performance computing. IEEE*. Bengaluru. P. 1–10. DOI: <https://doi.org/10.1016/j.engappai.2017.10.023>

17. Rodriguez, D., Gomez, D., Alvarez, D., Rivera, S. (2021), "A review of parallel heterogeneous computing algorithms in power systems". *Algorithms*. 2021. Vol.14(10). 275 p. DOI: <https://doi.org/10.3390/a14100275>

Надійшла 15.09.2023

Відомості про авторів / About the Authors

Резанов Богдан Михайлович – Національний технічний університет "Харківський політехнічний інститут", аспірант кафедри комп'ютерної інженерії та програмування, Харків, Україна; e-mail: brezanov@gmail.com; ORCID ID: <http://orcid.org/0000-0002-4113-8781>

Кучук Георгій Анатолійович – доктор технічних наук, професор, Національний технічний університет "Харківський політехнічний інститут", професор кафедри комп'ютерної інженерії та програмування, Харків, Україна; e-mail: kuchuk56@ukr.net; ORCID ID: <http://orcid.org/0000-0002-2862-438X>

Rezanov Bohdan – National Technical University "Kharkiv Polytechnic Institute", Student at the Department of Computer Engineering and Programming, Kharkiv, Ukraine.

Kuchuk Heorhii – Doctor of Sciences (Engineering), Professor, National Technical University "Kharkiv Polytechnic Institute", Professor at the Department of Computer Engineering and Programming, Kharkiv, Ukraine.

MODEL OF ELEMENTAL DATA FLOW DISTRIBUTION IN THE INTERNET OF THINGS SUPPORTING FOG PLATFORM

The subject of the research is models and methods for optimizing resource and task management in the fog computing environment of the Internet of Things (IoT). The increasing number of connected devices and the volumes of data collected in IoT networks make it essential to improve management systems that ensure the optimal distribution of tasks and resources. Fog computing addresses this challenge by distributing computational tasks closer to data sources and end-users. **The goal** of this work is to enhance the efficiency of fog computing technologies to achieve optimal task and resource allocation in IoT networks. **The main tasks** of this work are as follows. Firstly, considering the diverse requirements of computational resources and tasks in IoT, reviewing existing methods and developments in the field is necessary. Secondly, it is essential to investigate and compare clustering methods, particularly DBSCAN and C-Means, for effective resource management. The DBSCAN clustering method enables efficient task distribution based on their location, while the C-Means method allows grouping resources based on their characteristics. The final task involves developing a mathematical model that considers input parameters such as system response, cluster resource requirements, data proximity to processing, etc. This model will enable the analysis of potential scenarios and decision-making regarding the optimal distribution of tasks and resources in the IoT environment. **Conclusion.** This research aims to solve the urgent problem of managing resources and tasks in the fog IoT environment. A review of existing methods and developments in resource and task management in IoT is conducted. DBSCAN and C-Means clustering methods are compared to determine their effectiveness in resource management. A set-theoretic model is developed that considers various parameters for making optimal decisions on the distribution of tasks and resources. It is established that the use of clustering methods and the developed model help to improve system performance and ensure more efficient use of fog computing resources in the IoT environment.

Keywords: fog computing, IoT, DBSCAN, C-Means, clustering, mathematical modeling.

Бібліографічні опису / Bibliographic descriptions

Резанов Б.М., Кучук Г.А. Модель розподілу елементарних потоків даних у туманній платформі підтримки інтернету речей. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 88–97. DOI: <https://doi.org/10.30837/ITSSI.2023.25.088>

Rezanov, B., Kuchuk, H. (2023), "Model of elemental data flow distribution in the internet of things supporting fog platform", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 88–97. DOI: <https://doi.org/10.30837/ITSSI.2023.25.088>

О. ШКІЛЬ, М. МІРОШНИК, Д. РАХЛІС, О. ТРИФАНОВ

СТРУКТУРИ ДАНИХ ДЛЯ ДЕДУКТИВНОГО МОДЕЛЮВАННЯ УМОВНИХ ОПЕРАТОРІВ HDL

Предметом дослідження є кубітно-векторні моделі опису комбінаційних схем і процедури дедуктивного моделювання несправностей на основі цих моделей. **Об'єкт дослідження** – процеси побудови діагностичного забезпечення цифрових систем на основі використання векторних кубітних даних. **Мета роботи** – підвищення швидкості та якості створення діагностичного забезпечення цифрових пристроїв способом розроблення оптимальних структур даних і процедур дедуктивного моделювання несправностей на основі структурно-функціональних моделей комбінаційних схем. У статті вирішуються такі **завдання**: аналіз паралельних і послідовних умовних операторів мов опису апаратури та схемних структур, у які вони синтезуються; розроблення процедури формування таблиць істинності (Q -векторів) схемних структур, поданих мовами опису апаратури; створення універсальної структури даних для кубітного та аналітичного дедуктивного моделювання несправностей; удосконалення векторних моделей кубітного подання структур і компонентів цифрових систем на основі адресного кодування вхідних сигналів для підвищення технологічності та швидкодії моделювання несправностей; розроблення процедури отримання булевих похідних способом перетворень розрядів таблиць істинності (Q -векторів) та використання операції XOR; створення структури даних для дедуктивного моделювання несправностей на основі кубітного подання компонентів цифрових схем. Використовуються такі **методи**: дедуктивне, кубічне, дедуктивно-паралельне моделювання несправностей, моделювання несправностей за дедуктивними Q -векторами. Здобуто такі **результати**: показано еквівалентність паралельних і послідовних умовних операторів, а також їх схемна реалізація у вигляді мультиплексорів; запропоновано спосіб отримання таблиць істинності синтезованої схемної структури за допомогою *TestBench (Xilinx ISE)*; розглянуто різні технології та структури даних дедуктивного моделювання несправностей для табличного, аналітичного й кубітного способів опису цифрових схем; описано програмну реалізацію кубітного дедуктивного моделювання несправностей і показано еквівалентність отриманих результатів для схем мультиплексорів *MUX 2-в-1* та *MUX 4-в-1* з використанням програмного продукту *DCP*. **Висновки**: запропоновано новий Q -метод інтерпретативного моделювання несправностей цифрових схем, що визначається застосуванням компактних Q -векторів замість таблиць істинності; це дає змогу суттєво підвищити швидкодію аналізу завдяки адресному формуванню виходів функціональних примітивів і зменшити обсяги структур даних, що практично робить метод конкурентоспроможним із технологіями компілятивного моделювання.

Ключові слова: HDL-модель; структурно-функціональна модель; дедуктивне моделювання несправностей; кубічне покриття; таблиця істинності; Q -вектор.

Вступ

Процес проектування високотехнологічних комп'ютерних систем стає дедалі більш складним. Це не може не призвести до зростання кількості відмов таких систем, причиною яких є відсутність верифікації та сумісності встановлених компонентів. Це пояснюється тенденцією обмеження інвестицій на сервісне діагностичне обслуговування комп'ютерних систем на стадіях проектування та експлуатації. Отже, очевидною є необхідність розроблення та впровадження тестопридатного програмно-апаратного забезпечення цифрових систем на функціональному, алгоритмічному та системному рівнях. Для цього потрібно вирішувати завдання проектування діагностичного забезпечення, а саме тестів, алгоритмів контролю та пошуку дефектів,

системної верифікації сумісності компонентів. Виконання зазначених завдань потребує подальшого розвитку та вдосконалення методів і засобів моделювання справної поведінки та несправностей комп'ютерної системи на різних рівнях абстракції.

Мови опису апаратних засобів (*hardware description language, HDL*), зокрема *VHDL*, є потужним засобом для прискореного розроблення та верифікації складних цифрових систем. Мова *VHDL* може використовуватися на етапах проектування, верифікації, синтезу й тестування апаратури так само, як і для передачі інформації про проект, модифікації та супроводу. Але сучасні промислові системи автоматизованого проектування (САПР) цифрових схем на основі мов опису апаратури, як-от *Xilinx, Quartus, ModelSim* тощо, не містять інструментальних засобів розроблення діагностичного забезпечення

систем, що проєктуються. Отже, розроблення методів побудови діагностичного забезпечення за умови автоматизованого проєктування цифрових систем, наданих мовами опису апаратури, є актуальним і технологічно значущим завданням.

HDL-модель – це опис алгоритмічною мовою з формалізацією мовних конструкцій. Цей опис за формальними правилами перетворюється (синтезується) на структурно-функціональні схемні моделі різного рівня ієрархії. Між кодом мовою опису апаратури та відповідної йому цифрової схеми в обраній технологічній платформі "стоїть" система синтезу, тобто комплекс програмних засобів, що перетворює *HDL*-модель на деяку схемну реалізацію за визначеними правилами. Система синтезу однозначно перетворює вихідний *VHDL*-опис у схемну реалізацію рівня регістрових передач незалежно від обраної технології. Правила перетворення операторних конструкцій стандартизовані (прикладом може бути стандарт *1076.6-1999 IEEE Standard for VHDL RTL Synthesis* та *1364.1-2002 IEEE Standard for Verilog RTL Synthesis*, а також їх пізніші версії). Можливі налаштування систем синтезу не впливають на реалізацію вихідного стандарту. Отже, припустимо, що в межах стандарту стандартизовані (шаблонні) фрагменти *HDL*-коду перетворюються на стандартні схемні реалізації.

Так, *HDL*-модель – це фактично схема, і до неї можна застосувати методи схемного аналізу, синтезу тестів, моделювання несправностей і побудови алгоритмів пошуку дефектів, що досить добре розроблені для цифрових схем.

Методи моделювання несправностей можна розрізняти таким чином: одиночне, паралельне, дедуктивне, кубічне та спільне моделювання [1]. Найбільш ефективним на практиці виявилось запропоноване ще Армстронгом [2] дедуктивне моделювання несправностей, яке полягає в одночасному обробленні всіх одиночних константних несправностей схеми на одному вхідному наборі та виділення в цьому разі підмножини дефектів, що перевіряються. Метод орієнтований на вентильний рівень опису моделі проєктованого об'єкта в базисі І-АБО-НІ. Необхідність отримання аналітичних формул для кожного типу примітивного елемента та великі витрати пам'яті з метою зберігання списків несправностей ускладнюють практичну реалізацію методу. Для усунення цієї проблеми широко використовуються модифікації дедуктивного методу, що ґрунтуються на кубічному поданні законів

функціонування примітивних елементів структурно-функціональної моделі (СФМ) цифрової схеми [3, 4]. Усі наявні проблеми методів моделювання несправностей можна описати таким чином [5]: висока обчислювальна складність алгоритмів як на вентильному рівні, так і на рівні регістрових передач; значний обсяг структур даних для аналізу цифрових систем, що негативно впливає на продуктивність цих методів; складність аналізу та синтезу тестів для логіки великої розмірності. Отже, технології побудови тестів та визначення їх якості, що передбачають моделювання несправностей, мають відповідати вимогам швидкодії, уніфікації та простоти програмної реалізації для схем будь-якої розмірності та на будь-якому рівні подання.

Аналіз останніх досліджень і публікацій

Моделювання несправностей – добре вивчена тема в наукових колах і промисловості. Але, незважаючи на це, дослідження в цій сфері не припиняються. Так, у статті [6] запропоновано ефективний дедуктивний симулятор для малих дефектів затримки. Запропоновано окремі стратегії моделювання несправностей для різних дефектів. Експериментальні результати демонструють ефективність запропонованого симулятора, який може додатково прискорити моделювання несправності. Він досягає прискорення в середньому в 28,3 раза порівняно з методом послідовного моделювання та в середньому в 3,92 раза порівняно з методом трасування критичного шляху. Автори роботи [7] запропонували повністю автоматизовану систему моделювання несправностей *MetaFI*, яка підтримує різні моделі постійних і тимчасових несправностей. Цей фреймворк дає змогу додавати до системи, що тестується, константні одиночні несправності, перехідні збої, а також збої синхронізації. Моделювання несправності виконується на *RTL*-рівні. У статті [8] запропоновано підхід на основі *FPGA* для генерації тестових шаблонів і перевірки помилок комбінаційних схем. Замість моделювання автори використовують поняття емуляції для виявлення несправностей у схемі за допомогою апаратного забезпечення *FPGA*. Крім того, проблема незначної кількості вхідних і вихідних портів, доступних на платі *FPGA*, була подолана завдяки підходу мультиплексування перемикачів і кнопок.

Одним із популярних напрямів досліджень є різні методи діагностики несправностей

за допомогою машинного навчання. Так, автори праці [9] пропонують метод, що не потребує ані моделювання несправностей, ані зберігання словників несправностей. Вихідні реакції схеми, що діагностується, застосовуються до навченої нейронної мережі, і зрештою визначаються потенційні помилки. У статті також досліджується генерація даних, які використовуються для навчання нейронної мережі. Ефективність запропонованого методу підтверджується результатами експериментів для еталонних схем. А в роботі [10] авторами запропоновано методичне навчання штучної нейронної мережі для керування автоматичним генератором тестових шаблонів. У статті [11] подано повністю автоматизований метод тестування з'єднань *FPGA*, що фокусується на виявленні та діагностиці окремої несправності за допомогою коду Уолша. Він дає змогу однозначно ідентифікувати й локалізувати будь-яку окрему несправність. Крім того, цей метод не потребує будь-яких зовнішніх конфігурацій апаратного забезпечення для визначення фактичного місця несправності. Автори роботи [12] розробили метод дедуктивного моделювання несправностей способом побудови дедуктивних формул на основі аналітичного отримання булевих похідних і реалізації дедуктивних формул у відповідні схемні структури для апаратного моделювання несправностей, що в сотні разів прискорює процес моделювання. У праці [13] запропонований метод дедуктивного моделювання несправностей на основі структурно-функціональної моделі, де примітивні елементи подані таблицями істинності (*Q*-векторами), але не наведений спосіб отримання *Q*-векторів для складних функціональних елементів цифрових систем.

Огляд публікацій показує актуальність створення нових технологій моделювання несправностей у межах загального циклу автоматизованого проектування цифрових систем. **Метою дослідження** є підвищення швидкості та якості створення діагностичного забезпечення цифрових пристроїв способом розроблення оптимальних структур даних і процедур дедуктивного моделювання несправностей на основі структурно-функціональних моделей комбінаційних схем, наданих мовами опису апаратури.

Для досягнення поставленої мети необхідно виконати такі завдання:

1) провести аналіз паралельних і послідовних умовних операторів мов опису апаратури та схемних структур, у які вони синтезуються;

2) розробити процедури формування таблиць істинності (*Q*-векторів) схемних структур, поданих умовними операторами мов опису апаратури;

3) створити універсальні структури даних для кубічного, аналітичного та кубітного дедуктивного моделювання несправностей;

4) удосконалити векторні моделі кубічного подання структур і компонентів цифрових систем на основі адресного кодування вхідних сигналів для підвищення технологічності та швидкодії моделювання несправностей;

5) розробити процедури отримання булевих похідних способом переставлення розрядів таблиць істинності (*Q*-векторів) та використання операції *XOR*;

6) створити структури даних для дедуктивного моделювання несправностей на основі кубічного подання компонентів цифрових схем.

Побудова структурно-функціональних *VHDL*-моделей

У мові опису апаратури *VHDL* використовуються три стилі опису проєктів: потоковий (*data flow*), поведінковий (*behavioral*) і структурний (*structural*). Останній стиль є комбінацією компонентів, описаних одним із двох інших стилів. Потоковий стиль опису передбачає використання тільки паралельних операторів, а поведінковий – послідовних у середині процесу. Усі паралельні оператори виконуються одночасно під час одного й того самого циклу моделювання. Результати паралельних операторів доступні для інших операторів після завершення поточного циклу моделювання. До паралельних *VHDL*-операторів належать: оператор призначення сигналів (\leq), умовний оператор призначення сигналів (\leq /when/else), оператор призначення сигналів на вибір (*with/select*/ \leq). До послідовних *VHDL*-операторів належать: оператори призначення сигналів і змінних (\leq / :=), умовний оператор призначення сигналів (*if/then/else*), оператор призначення сигналів на вибір (*case*), цикли (*while/for*). Паралельні та послідовні умовні оператори еквівалентні й застосовуються залежно від уподобань проєктувальника [14] як для реалізації комбінаційних схем, так і послідовнісних. На рис. 1 відтворено еквівалентність паралельного (*when/else*) та послідовного умовних операторів (*if*).

Аналогічно можна перейти від паралельного оператора вибору (*with/select*) до послідовного (*case*) (рис. 2).

```
entity example_condition is
port ( x1, x2, x3, x4 : in bit; F : out bit;
      condition : in bit_vector (1 downto 0) );
end example_condition;
```

```
architecture first of example_condition is
begin
  F <= x1 when condition = "00" else
    x2 when condition = "01" else
    x3 when condition = "10" else
    x4;
end first;
```

≡

```
architecture second of example_condition is
begin
  process (x1, x2, x3, x4, condition )
  begin
    if (condition = "00") then F <= x1;
    elsif (condition = "01") then F <= x2;
    elsif (condition = "10") then F <= x3;
    else
      F <= x4;
    end if;
  end process;
end second;
```

Рис. 1. Еквівалентність умовних VHDL-операторів

```
architecture first1 of example_condition is
begin
  with condition select
    F <= x1 when "00",
    x2 when "01",
    x3 when "10",
    x4 when others;
end first1;
```

≡

```
architecture second1 of example_condition is
begin
  process (x1, x2, x3, x4, condition )
  begin
    case condition is
      when "00" => F <= x1;
      when "01" => F <= x2;
      when "10" => F <= x3;
      when others => F <= x4;
    end case;
  end process;
end second1;
```

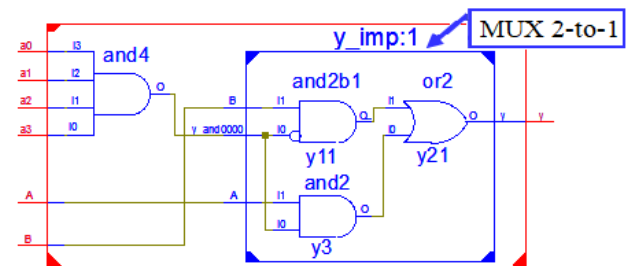
Рис. 2. Еквівалентність VHDL-операторів вибору

Важко уявити VHDL-модель без використання умовних операторів чи операторів вибору. Саме їх апаратна реалізація викликає найбільший інтерес. Розглянемо синтез умовного послідовного оператора *IF* зі складним умовним виразом:

```
architecture Behavioral of test1 is begin
process (a0,a1,a2,a3,A,B) begin
  if ((a3 and a2 and a1 and a0)='1') then y<=A;
  else y<=B;
  end if;
end process;
end Behavioral;
```

На рис. 3 наведений результат синтезу вказаної мовної конструкції.

На цьому рисунку видно, що умовний послідовний оператор *IF* зі складним умовним виразом синтезований у комбінаційну схему, що містить мультиплексор *MUX 2-в-1*.

Рис. 3. RTL-схема для оператора *IF* зі складним умовним виразом

Розглянемо синтез послідовного оператора багатопозиційного вибору *CASE* зі складним умовним виразом:

```
architecture Beh of MUX is
begin
  process(A,S) begin
    case S is
      when "00" => y<= A(0);
      when "01" => y<= A(1);
```

```

when "10" => y<= A(2);
when "11" => y<= A(3);
when others => y<='X';
end case;
end process;
end Beh;
    
```

На рис. 4 видно, що послідовний оператор CASE зі складним умовним виразом синтезований у комбінаційну схему, що містить мультиплексор MUX 4-в-1.

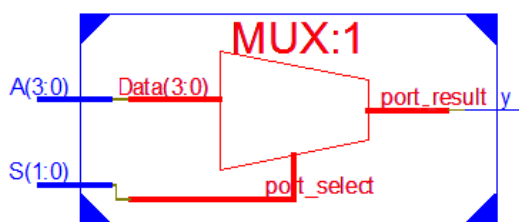


Рис. 4. RTL-схема для оператора CASE зі складним умовним виразом

З огляду на здобуті результати можна зробити висновки, що умовні оператори мови опису апаратури перетворюються в схемні конструкції мультиплексорів MUX 2-в-1 й MUX 4-в-1. Отже, необхідно розглянути різні методи моделювання несправностей саме для таких схем.

Декілька слів треба сказати про моделі несправностей за умови дедуктивного моделювання. Можливі помилки проектування в HDL-моделях визначаються стилем опису HDL-коду. Під помилкою проектування розуміється помилка в HDL-операторі, яка не належить до класу синтаксичних і порушує алгоритм функціонування моделі пристрою, що проектується. Типи помилок проектування: "заміна оператора" (логічного чи арифметичного) та "заміна операнда" (в операторі призначення або умовному) [15]. Але якщо переходимо до структурно-функціональної моделі пристрою, яка складається з ПЕ та ліній схеми, то основними типами несправностей, які історично розглядаються

в діагностиці, є одиночні константні несправності на лініях структурно-функціональної моделі. Тому надалі будемо розглядати одиночні константні несправності ліній схеми, якщо не обумовлено інше.

Після синтезу VHDL-моделі у FPGA-схему інструментальними засобами САПР (наприклад, Xilinx) можна отримати її схемну реалізацію на LUT (Look-up table – таблиця відповідності, кожна комірка якої здатна зберігати значення вихідного рядка таблиці істинності, тобто Q-вектора). Це дає змогу отримати повну таблицю істинності всієї схеми, що допомагає ефективно використовувати кубітні методи аналізу цифрових схем.

Кубічне дедуктивне моделювання несправностей схем

Найпростішим неявним методом моделювання несправностей є дедуктивний метод (deductive method), або метод Армстронга. Кубічне моделювання несправностей є машинно-орієнтованим алгоритмом в інтерпретативних системах моделювання. Воно ґрунтується на поданні опису примітивного елемента (ПЕ) структурно-функціональної моделі кубічними покриттями (КП) для комбінаційних схем в алфавіті $A^1 = \{0, 1, U, X\}$, а для послідовнісних схем в алфавіті A^2 [4].

Розглянемо приклад дедуктивного моделювання несправностей для мультиплексора MUX 2-в-1, функція якого має вигляд $y = x_1x_3 \vee x_2\overline{x_3}$, тобто $y = \sum m(2, 5, 6, 7)$. На рис. 5 наведено схемну позначку, вентиляну схему, таблицю істинності (ТІ) та кубічне покриття зазначеної схеми.

Для отриманої схеми виконаємо ранжування способом нумерації її ліній, а потім – дедуктивне моделювання несправностей для вичерпного тесту – $2^n = 8$ наборів, де $n = 3$ – це кількість входів.



Рис. 5. Структурно-функціональна модель мультиплексора MUX 2-в-1

Результати дедуктивного моделювання несправностей можуть подаватися як у вигляді списків несправностей, так і у вигляді таблиць несправностей (ТН), рядки яких відповідають вхідним наборам, а стовпці – входам та виходам схеми. Якщо на розглянутому наборі (рядок ТН) перевіряється несправність типу "0" на лінії, то у відповідному стовпці ставиться 0, а якщо перевіряється несправність типу "1" на лінії, то у відповідному стовпці ставиться 1.

Подамо різну послідовність наборів у ТІ за умови, який розряд вважається молодшим у двійковому наборі. Якщо *logic_vector* у HDL-моделі описується *downto*, то в молодшому розряді стоїть 3-я лінія, а якщо *to*, то відповідно 1-а лінія (рис. 6).

Розглянемо процес побудови структурно-функціональної моделі та кубічного моделювання

комбінаційної схеми *MUX* 4-в-1, що складається з функціональних елементів *MUX* 2-в-1 (рис. 7). Алгоритм кубічного моделювання несправностей реалізований у комп'ютерній програмі *DCP (Deductive Circuit Processor)*, багаторічне використання якої в навчальному процесі Харківського національного університету радіоелектроніки підтверджує працездатність зазначеного методу.

На рис. 8 подані результати моделювання фрагмента вичерпного тесту схеми *MUX* 4-в-1 в програмі кубічного моделювання *DCP*. Важливо зазначити, що повнота наведеного фрагменту тесту дорівнює 100%. Для поданого на рис. 7 набору 111101 результати ручного й комп'ютерного моделювання збігаються.

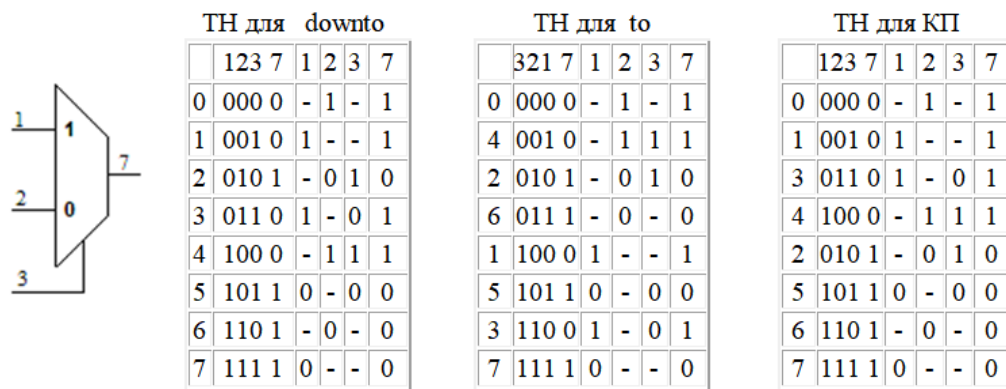


Рис. 6. Варіанти ТН для різних варіантів СФМ мультиплектора *MUX* 2-в-1

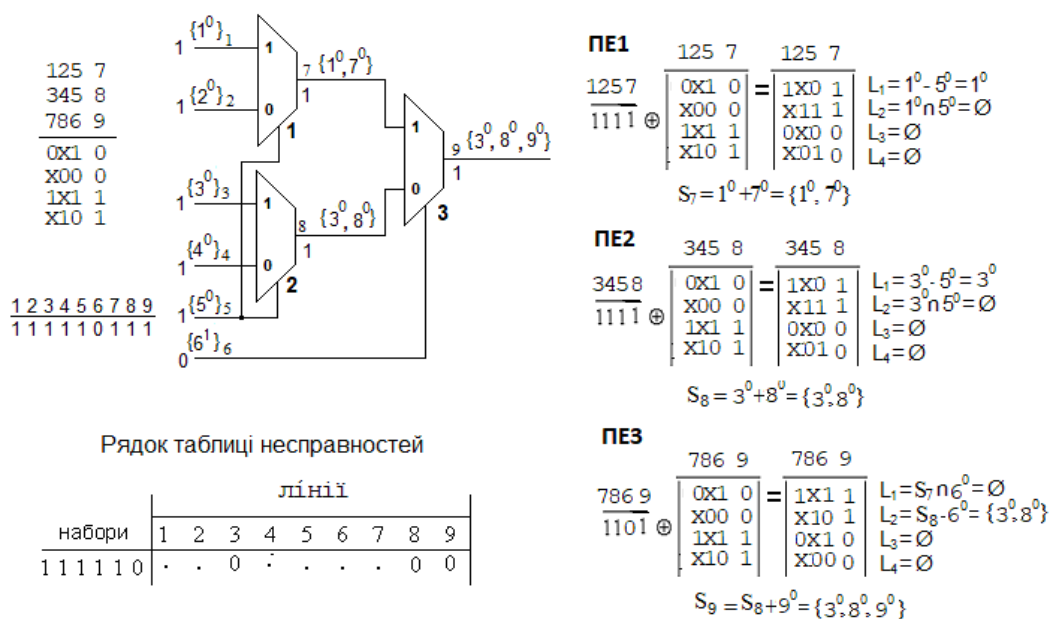


Рис. 7. Справне та несправне кубічне моделювання

Simulation									
Fault-free table									
No.	1	2	3	4	5	6	7	8	9
0	U	U	U	U	U	U	U	U	U
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0
3	0	0	0	0	1	0	0	0	0
4	0	0	0	0	1	1	0	0	0
5	1	1	1	1	0	0	1	1	1
6	1	1	1	1	0	1	1	1	1
7	1	1	1	1	1	0	1	1	1
8	1	1	1	1	1	1	1	1	1
9	0	1	1	1	1	1	0	1	0
10	1	1	1	0	0	0	1	0	0

Fault's table														
No.	Test	O...	1	2	3	4	5	6	7	8	9	Self ...	Qua...	Com...
1	000000	9	.	.	.	1	.	.	.	1	1	16	16	16
2	000001	9	.	1	1	.	1	16	11	27
3	000010	9	.	.	1	1	1	16	5	33
4	000011	9	1	1	.	1	16	5	38
5	111100	9	.	.	.	0	.	.	.	0	0	16	16	55
6	111101	9	.	0	0	.	0	16	11	66
7	111110	9	.	.	0	0	0	16	5	72
8	111111	9	0	0	.	0	16	5	77
9	011111	9	1	.	.	.	0	0	1	.	1	27	11	88
10	111000	9	.	.	.	1	1	1	.	1	1	27	11	100

000000 000001 000010 000011 111100 111101 111110 111111 011111 111000

Рис. 8. Результати моделювання фрагмента тесту схеми MUX 4-в-1 в програмі кубічного моделювання DCP

**Булеві похідні
для графічного подання логічних функцій**

Одним із наочних способів взяття булевих похідних є подання багаторівневих логічних функцій графічним способом з подальшим отриманням дужкових форм (ДФ) [16].

Розглянемо приклад отримання дужкових форм булевих функцій (БФ) на основі ДНФ.

Наприклад, для функції трьох змінних $f(x_1, x_2, x_3) = \sum m(2, 5, 6, 7)$ ДНФ має вигляд $f = x_1 \times x_3 \vee x_2 \times \bar{x}_3$. Абсолютно аналогічні викладки можна застосувати й до КНФ. На рис. 9 показана графічна інтерпретація отримання дужкової форми булевої функції.

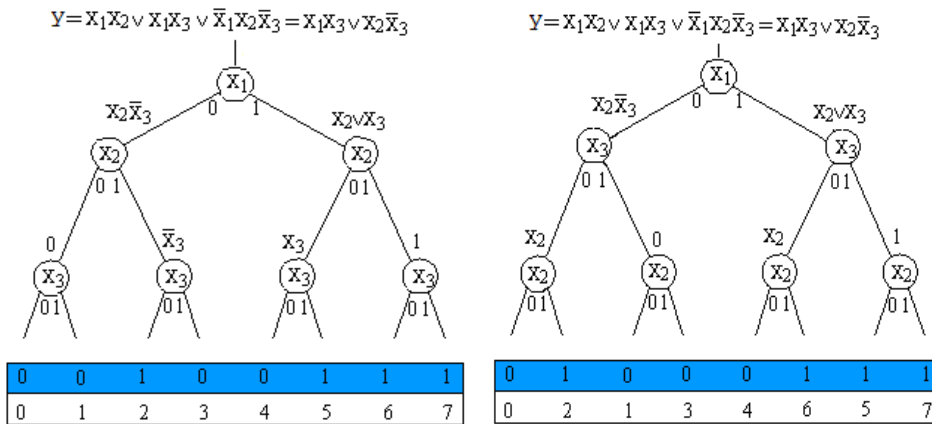


Рис. 9. Графова форма булевої функції мультиплексора MUX 2-в-1

Змінні в ярусах графа розташовуються в природному порядку (x_1, x_2, x_3) – від старших розрядів до молодших. У нижній частині рисунка в рамці показано таблицю істинності цієї функції (блакитне поле), а нижче зображено відповідні номери наборів у десятковому еквіваленті. Двійковий еквівалент шляху (складений з 0 і 1, що помічають вхідні ребра вузлів графа) від вершини верхнього ярусу графа до відповідного значення функції (блакитне поле) вказує на номер відповідного двійкового набору (порядок розрядів – x_1, x_2, x_3).

Наприклад, шлях 0 1 1 (крізь вершини x_1, x_2, x_3) відповідає номеру набору 3, що підтверджується його десятковим еквівалентом. Залежно від порядку розміщення змінних у ярусах графа можна отримати різні варіанти скобкових форм для однієї і тієї самої булевої функції. Правило перестановки значення функції в таблиці істинності: для кожної вершини верхнього з ярусів, що переставляються, змінюються місцями між собою значення функцій у підграфах на "внутрішніх" ребрах вершин нижнього з ярусів, що переставляються. У прикладі під час переставлення ярусів 2 і 3 змінюються місцями набори (1, 2) і (5, 6).

Графічне взяття булевої похідної для змінної нижнього ярусу графа визначається порівнянням парних і непарних позицій у поданій таблиці істинності, якщо відповідна змінна стоїть у нижньому ярусі графа. Якщо парні та непарні позиції в таблиці істинності розрізняються (0 1 або 1 0), то відповідна гілка дерева впливає на булеву змінну (на рисунку позначається "+"). Далі на другому ярусі відповідна

змінна записується з інверсією, якщо "+" іде з лівого ребра, та без інверсії, якщо з правого.

Розглянемо аналітичне та графічне взяття булевих змінних для функції $f(x_1, x_2, x_3) = (x_1 x_3 \vee x_2 \bar{x}_3)$.

Графічне взяття булевих похідних наведено на рис. 10.

$$f_1 = df/dx_1 = (0 \times x_3 \vee x_2 \bar{x}_3) \oplus (1 \times x_3 \vee x_2 \bar{x}_3) = x_2 \bar{x}_3 \oplus (x_2 \vee x_3) = x_2 \bar{x}_3 \times (x_2 \vee x_3) \vee \overline{x_2 \bar{x}_3} \times (x_2 \vee x_3) = x_2 \bar{x}_3 \times \overline{x_2 \bar{x}_3} \vee (\overline{x_2 \bar{x}_3} \times (x_2 \vee x_3)) \times (x_2 \vee x_3) = 0 \vee x_3 = x_3;$$

$$f_2 = df/dx_2 = (x_1 x_3 \vee 0 \times \bar{x}_3) \oplus (x_1 x_3 \vee 1 \times \bar{x}_3) = x_1 x_3 \oplus (x_1 x_3 \vee \bar{x}_3) = x_1 x_3 \oplus (x_1 \vee \bar{x}_3) = x_1 x_3 \times (x_1 \vee \bar{x}_3) \vee \overline{x_1 x_3} \times (x_1 \vee \bar{x}_3) = x_1 x_3 \times \overline{x_1 x_3} \vee (\overline{x_1 x_3} \times (x_1 \vee \bar{x}_3)) \times (x_1 \vee \bar{x}_3) = 0 \vee \bar{x}_3 = \bar{x}_3;$$

$$f_3 = df/dx_3 = (x_1 \times 0 \vee x_2 \times 1) \oplus (x_1 \times 1 \vee x_2 \times 0) = x_2 \oplus x_1 = x_1 \bar{x}_2 \vee \bar{x}_1 x_2.$$

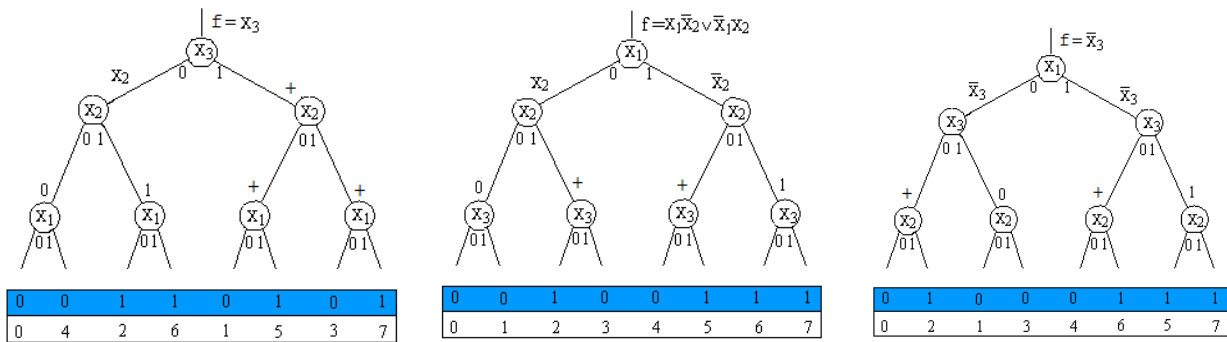


Рис. 10. Графічне взяття булевих похідних

Для трьох змінних маємо чотири умови активізації $(x_1 \bar{x}_2, \bar{x}_1 x_2, x_3, \bar{x}_3)$, що відповідають чотирьом логічним шляхам у схемній структурі диз'юнктивної форми цієї функції. Результати отримання булевих похідних для графічного та аналітичного способу подання булевих функцій

повністю збігаються, що підтверджує ефективність застосування запропонованих методів.

Отже, взяття булевої похідної фактично полягає в переставленні розрядів таблиці істинності (Q-вектора) відповідно до обраного суттєвого входу, а застосування формули

$$L_{ii} = T_i \oplus F_i = F_{ii} [(x_{i1} \oplus T_{i1}), (x_{i2} \oplus T_{i2}), \dots, (x_{ij} \oplus T_{ij}), \dots, (x_{im} \oplus T_{im})] \oplus T_{ii}, \tag{1}$$

де F_i – компонент (примітив) функціонального опису цифрового пристрою;

T_i – тест-вектор;

L_{ii} – дедуктивна функція паралельного моделювання несправностей, дає змогу використовувати поданий Q-вектор для дедуктивного моделювання несправностей.

Булеві похідні для кубітного подання логічних функцій

Коректність використання прикметника "кубітна" для моделей цифрових пристроїв основана на порівнянні лінійної та булевої алгебри Кантора, що базується на алфавіті $A^1 = \{0, 1, X, U\}$. Тут перші два символи – примітиви. Третій визначається суперпозицією: $X = 0 \cup 1$. Символ U

є доповненням до універсуму. Загалом формулу обчислення булевих похідних для формування відповідної кубітної матриці функціональності можна подати як XOR-операцію над розрядами сусідніх груп бітів:

$$D_{(ij,ij+1)}(Q, x_i) = Q_j(x_i) \oplus_{j=1}^{2^{n-i}} Q_{j+1}(x_i), \quad i = \overline{1, n},$$

де n – кількість змінних у функціональності;

i – номер змінної, за якою береться похідна;

$(j, j+1)$ – номери сусідніх груп у кубітному векторі, що підлягають XOR-порівнянню, де число та потужність таких груп функціонально залежить від номера змінної;

$D_{(ij,ij+1)}(Q, x_i)$ – сусідні групи похідної по i -й змінній, що формуються XOR-операцією. Групою вважається сукупність бітів, кратних ступеню двійки, що підлягає порозрядному XOR-складанню з відповідними бітами сусідньої групи кубітного вектора [17].

Приклад взяття булевих похідних по кубітному покриттю для визначення умов суттєвості вхідних змінних у процесі моделювання несправностей мультиплектора MUX 2-в-1 $f(x_1, x_2, x_3) = (x_1 x_3 \vee x_2 \overline{x_3})$. Обчислення трьох похідних першого порядку за таблицею істинності дає такий результат: $f'_1(01010101) = x_3$, $f'_2(10101010) = \overline{x_3}$, $f'_3(00111100) = x_1 \oplus x_2$.

Результати отримання булевих похідних для кубічного, графічного та аналітичного способу подання булевих функцій на прикладі MUX 2-в-1 повністю збігаються, що підтверджує працездатність та ефективність застосування запропонованих методів.

Метод моделювання несправностей за дедуктивними Q-векторами

Математична основа дедуктивного моделювання несправностей полягає в транспортуванні бінарних комбінацій вхідних несправностей на вихід на заданому вхідному набору за формулою $L = T \oplus F$. Дедуктивне моделювання полягає в зміні логіки елемента F залежно від вхідних умов T . Дедуктивне моделювання є найефективнішим засобом аналізу якості тестів і синтезу таблиць несправностей для пошуку дефектів, простежування шляху поширення несправності. Пропонується його реалізація на основі векторної форми опису логіки, що передбачає логічні аналітичні форми й дає змогу істотно спростити алгоритми синтезу дедуктивних моделей та їх застосування для інтерпретативного моделювання цифрових елементів і схем великої розмірності [13]. Сутність методу моделювання несправностей за дедуктивними Q-векторами полягає в побудові дедуктивних векторів, що є впорядкованою сукупністю векторних булевих похідних за вхідними змінними, що визначаються стовпцями матриці. Сукупність похідних за вхідними змінними в матриці є тестом для перевірки одиночних константних несправностей вхідних та вихідних ліній. Одиничні координати дедуктивних векторів є умовами активізації вектора вхідної інформації як адреси для транспортування на вихід схеми. Будується дедуктивний вектор способом переставлення бітів Q-вектора та операції XOR між вхідним набором для функціонального елемента та його таблицею істинності (Q-вектором).

Розглянемо метод синтезу дедуктивних векторів за Q-вектором на прикладі мультиплектора MUX 2-в-1 (рис. 11).

L-матриця	H-матриця	D-матриця																																																																																																																																																																																																																	
<table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><td></td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr> </table>		0	0	1	0	0	1	1	1	0	0	0	1	0	0	1	1	1	0	0	0	1	0	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0	0	1	1	1	1	1	1	0	1	1	0	0	0	1	1	1	0	1	1	0	0	0	1	1	1	0	1	1	0	0	0	<table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr> <tr><td>1</td><td>0</td><td>3</td><td>2</td><td>5</td><td>4</td><td>7</td><td>6</td></tr> <tr><td>2</td><td>3</td><td>0</td><td>1</td><td>6</td><td>7</td><td>4</td><td>5</td></tr> <tr><td>3</td><td>2</td><td>1</td><td>0</td><td>7</td><td>6</td><td>5</td><td>4</td></tr> <tr><td>4</td><td>5</td><td>6</td><td>7</td><td>0</td><td>1</td><td>2</td><td>3</td></tr> <tr><td>5</td><td>4</td><td>7</td><td>6</td><td>1</td><td>0</td><td>3</td><td>2</td></tr> <tr><td>6</td><td>7</td><td>4</td><td>5</td><td>2</td><td>3</td><td>0</td><td>1</td></tr> <tr><td>7</td><td>6</td><td>5</td><td>4</td><td>3</td><td>2</td><td>1</td><td>0</td></tr> </table>	0	1	2	3	4	5	6	7	1	0	3	2	5	4	7	6	2	3	0	1	6	7	4	5	3	2	1	0	7	6	5	4	4	5	6	7	0	1	2	3	5	4	7	6	1	0	3	2	6	7	4	5	2	3	0	1	7	6	5	4	3	2	1	0	<table border="1" style="border-collapse: collapse; width: 100%; text-align: center;"> <tr><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr> </table>	0	0	1	0	0	1	1	1	0	0	0	1	1	0	1	1	0	1	1	1	0	0	1	0	0	1	0	0	1	1	1	0	0	1	1	1	0	0	0	0	0	1	0	0	1	1	1	0	0	0	1	0	0	1	1	1	0	0	0	1	1	0	1	1
	0	0	1	0	0	1	1	1																																																																																																																																																																																																											
0	0	0	1	0	0	1	1	1																																																																																																																																																																																																											
0	0	0	1	0	0	1	1	1																																																																																																																																																																																																											
1	1	1	0	1	1	0	0	0																																																																																																																																																																																																											
0	0	0	1	0	0	1	1	1																																																																																																																																																																																																											
0	0	0	1	0	0	1	1	1																																																																																																																																																																																																											
1	1	1	0	1	1	0	0	0																																																																																																																																																																																																											
1	1	1	0	1	1	0	0	0																																																																																																																																																																																																											
1	1	1	0	1	1	0	0	0																																																																																																																																																																																																											
0	1	2	3	4	5	6	7																																																																																																																																																																																																												
1	0	3	2	5	4	7	6																																																																																																																																																																																																												
2	3	0	1	6	7	4	5																																																																																																																																																																																																												
3	2	1	0	7	6	5	4																																																																																																																																																																																																												
4	5	6	7	0	1	2	3																																																																																																																																																																																																												
5	4	7	6	1	0	3	2																																																																																																																																																																																																												
6	7	4	5	2	3	0	1																																																																																																																																																																																																												
7	6	5	4	3	2	1	0																																																																																																																																																																																																												
0	0	1	0	0	1	1	1																																																																																																																																																																																																												
0	0	0	1	1	0	1	1																																																																																																																																																																																																												
0	1	1	1	0	0	1	0																																																																																																																																																																																																												
0	1	0	0	1	1	1	0																																																																																																																																																																																																												
0	1	1	1	0	0	0	0																																																																																																																																																																																																												
0	1	0	0	1	1	1	0																																																																																																																																																																																																												
0	0	1	0	0	1	1	1																																																																																																																																																																																																												
0	0	0	1	1	0	1	1																																																																																																																																																																																																												
а)	б)	в)																																																																																																																																																																																																																	

Рис. 11. Формування дедуктивних векторів для MUX 2-в-1

1. Дано таблицю істинності n змінних у порядку вагомозначності позиційного коду вхідних наборів (від 000...0 до 2^{n-1}). Подамо вихідний стовпець таблиці як Q -вектор завдовжки 2^n . Для MUX 2-в-1 $Q = (00100111)$.

2. Отримання матриці L -векторів (взяття булевих похідних за i -ю змінною). Для кожного розряду Q_i обчислюється $L_i = Q \oplus Q_i$, де Q_i – стан i -біта Q -вектора. Для цього складається матриця $(2^{n+1} \times 2^{n+1})$, у перший рядок і перший стовпець якої записується Q -вектор за умови, що нульова комірка порожня (жовті рядок і стовпець на рис. 11, а).

3. У кожен i -й рядок матриці L записується вектор Q без інверсії, якщо в стовпці $Q_i = 0$, та з інверсією, якщо $Q_i = 1$ (рис. 11, а).

4. Формується дедуктивна матриця D ($2^n \times 2^n$, рис. 11, в) способом переставлення елементів рядків матриці L за правилами H -матриці ($2^n \times 2^n$, рис. 11, б). H -матриця фактично відповідає встановленню по черзі відповідних змінних у молодший розряд і переставленню відповідно до цього елементів таблиці істинності (Q -вектора).

Розглянемо більш детально спосіб переставлення елементів Q -вектора залежно від вхідного набору та значення функції ПЕ на цьому наборі. Принцип переставлення такий: аналізуються розряди вхідного набору починаючи зі старшого і для тих розрядів вхідного набору, які дорівнюють 1, виконується

переставлення елементів Q -вектора групами по 2^{n-1} розрядів, де n – номер розряду за умови, що молодший розряд (правий) вважається першим. Після всіх переставлень елементи отриманого дедуктивного вектора інвертуються, якщо значення Q -вектора на цьому наборі дорівнює 1 згідно з п. 3. Зауважимо, що запропонований принцип переставлень Q -вектора дає змогу уникнути використання H -матриці, яка, по-перше, займає багато пам'яті, а по-друге, її побудова достатньо складна. У табл. 1 наведено приклад переставлень Q -вектора 5 змінних ($n = 5$) для вхідного набору (11010). Початковий Q -вектор поданий у першому рядку за природним порядком змінних *downto*. Починаємо аналіз зі старших розрядів вхідного набору й переставляємо групи елементів у Q -векторі для тих розрядів, які визначаються одиницями у вхідному наборі. Для першого розряду 11010 переставляємо групи по 2^{n-1} , тобто по 16 елементів (перше переставлення). Для другого розряду 11010 переставляємо групи по 2^{n-2} , тобто по 8 елементів (друге переставлення). Для четвертого розряду 11010 переставляємо групи по 2^{n-4} , тобто по 2 елементи (третє переставлення).

Ця процедура не залежить від кількості змінних, а залежить тільки від числа 1 у вхідному наборі й повністю збігається з переставленнями в графовому поданні БФ [16] та з ітеративною процедурою у [13].

Викликає зацікавлення результат переставлень для вхідного набору 11111 (табл. 2).

Таблиця 1. Приклад переставлень для набору 11010

Початковий Q -вектор	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Перше переставлення	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Друге переставлення	24	25	26	27	28	29	30	31	16	17	18	19	20	21	22	23	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7
Третє переставлення	26	27	24	25	30	31	28	29	18	19	16	17	22	23	20	21	10	11	8	9	14	15	12	13	2	3	0	1	6	7	4	5

Таблиця 2. Переставлення для набору 11111

31	30	29	28	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	---	---	---	---	---	---	---	---	---	---

Процедура моделювання несправностей за дедуктивними векторами передбачає такі кроки.

1. Існує шаблон обчислення адрес у дедуктивному векторі для моделювання вхідних несправностей. Кількість стовпців дорівнює подвоєному числу

несправностей, спочатку розташовуються нульові несправності, а потім одиночні. Кожен рядок має вагу, кратну ступеню двійки, до того ж XI (перший рядок) відповідає молодшому розряду.

Спочатку шаблон і вектор моделювання (*L*-вектор) заповнюються нулями (або мають порожні комірки).

2. Вхідні несправності мають інверсні значення від справних значень сигналів на входах ПЕ (вхідному набору).

3. У комірках, що відповідають вхідним несправностям, у шаблоні ставиться 1.

4. За стовпцями, де стоять 1, обчислюється адреса (індекс) координати дедуктивного вектора з огляду на ваги рядків (десятковий еквівалент двійкового числа стовпця шаблону 1, 2, 4, 8, ...).

5. У вектор моделювання (*L*) у комірки, що відповідають стовпцям шаблону з 1, записуються значення розрядів відповідного дедуктивного вектора за обчисленими адресами (індексами). Зазначимо, що вихідні несправності ПЕ не моделюються й відповідні (вихідні) позиції *L*-вектора

не обчислюються. Необхідно зважати, що дедуктивний вектор для одиничних значень вихідного сигналу на цьому наборі інвертується.

6. Якщо за одиничними значеннями вектора моделювання в шаблоні стоїть 1, то несправність, що визначає стовпець шаблону, перевіряється.

Розглянемо приклади моделювання несправностей за дедуктивними векторами для ПЕ *MUX 2-в-1* для декількох вхідних наборів. Набір (0 0 0). Початковий список несправностей $\{1^1, 2^1, 3^1\}$. Дедуктивний вектор (0 0 1 0 0 1 1 1). Результат моделювання несправностей зображений на рис. 12.

Набір (1 1 1). Початковий список несправностей $\{1^0, 2^0, 3^0\}$. Дедуктивний вектор (0 0 0 1 1 0 1 1). Результат подано на рис. 13.

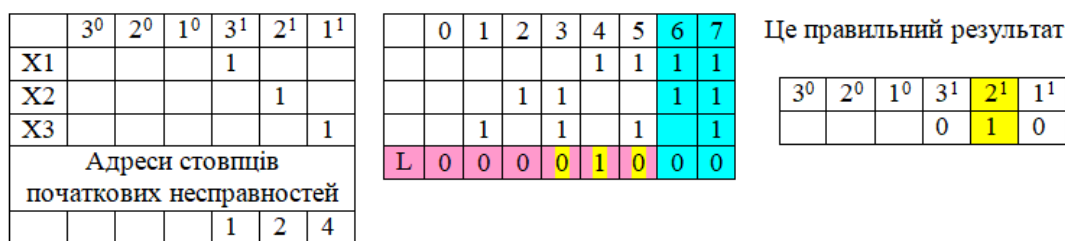


Рис. 12. Моделювання несправностей для набору 0 0 0 ПЕ *MUX 2-в-1*

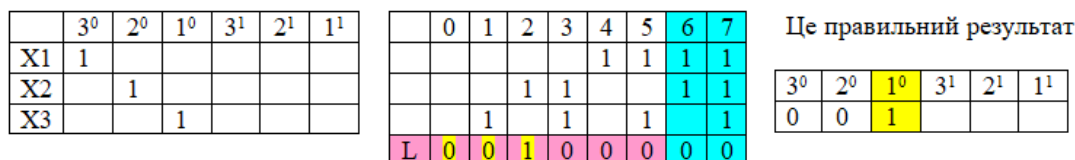


Рис. 13. Моделювання несправностей для набору (1 1 1) ПЕ *MUX 2-в-1*

Розглянемо, як впливає послідовність змінних у вхідному наборі на результат моделювання за дедуктивними *Q*-векторами. Зазначимо, що в мовах опису апаратури використовуються два способи подання векторних змінних:

- *std_logic_vector (3 downto 0)*; – молодший розряд правий;
- *std_logic_vector (3 to 0)*; – молодший розряд лівий.

Для прикладу розглянемо моделювання симетричного набору (0 1 0). Початковий список

несправностей $\{1^1, 2^0, 3^1\}$. Послідовність змінних у вхідному наборі впливає тільки на формування шаблону обчислення адрес. Дедуктивний вектор (0 1 1 1 0 0 1 0).

Спочатку розглянемо шаблон для *downto*. Тут молодший розряд правий (третій за порядком), що відповідає загальноприйнятій системі позиційного коду.

Результати моделювання несправностей наведені на рис. 14.

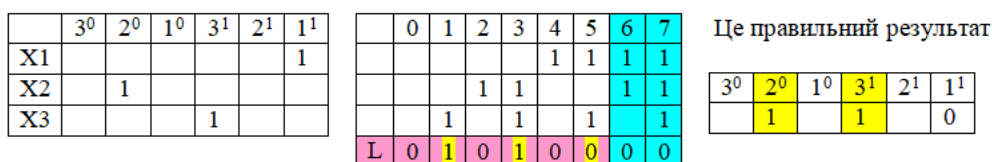


Рис. 14. Моделювання несправностей набору (0 1 0) для шаблону *downto*

Далі розглянемо шаблон для *to*. Тут молодший розряд лівий (перший за порядком), що не відповідає загальноприйнятій системі позиційного коду.

Результати моделювання несправностей зображені на рис. 15.

	1 ⁰	2 ⁰	3 ⁰	1 ¹	2 ¹	3 ¹
X1				1		
X2		1				
X3						1

	0	1	2	3	4	5	6	7
					1	1	1	1
			1	1			1	1
		1		1		1		1
L	0	1	0	1	0	0	1	0

Це результат з точністю до навпаки

1 ⁰	2 ⁰	3 ⁰	1 ¹	2 ¹	3 ¹
	1		1		0

Рис. 15. Моделювання несправностей набору (0 1 0) для шаблону *to*

Так, результати моделювання для шаблону *downto* повністю збігаються з результатами простого дедуктивного моделювання (рис. 6) і не збігаються для шаблону *to*.

Розглянемо приклади моделювання несправностей за дедуктивними векторами для більш складного ПЕ *MUX* 4-в-1. На рис. 16 подано його позначку (а), кубічне покриття (б), закон функціонування (в) і фрагмент таблиці істинності (г), отримані в *Xilinx ISE*. Останній стовпець на рис. 16, г – це фактично *Q*-вектор.

Отже, у використанні позиційного коду номерів розрядів ПЕ рекомендується застосовувати шаблон обчислення адрес *downto*.

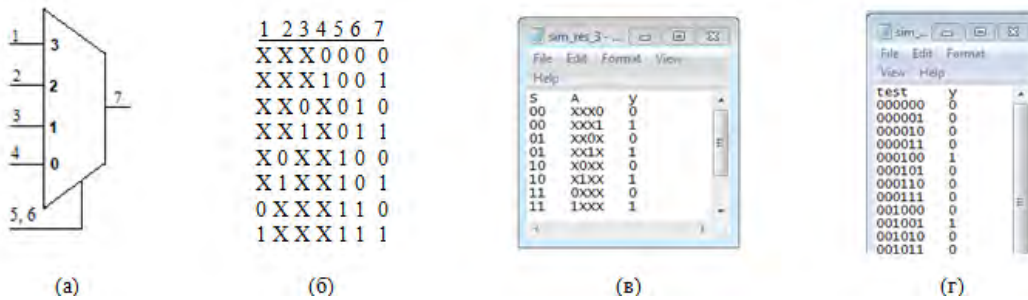


Рис. 16. Позначка, кубічне покриття та фрагмент таблиці істинності ПЕ *MUX* 4-в-1

На рис. 17 поданий останній стовпець повної таблиці істинності за зростанням номерів наборів для *MUX* 4-в-1 ($2^6 = 64$), розбитий на дві частини по 32 набори. Фактично це *Q*-вектор, або

переставлений *D*-вектор, позначений кольором. Зелений колір – для набору (0 0 0 0 0 0), а синій – для набору (1 1 1 1 1 1).

0	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	0	1	1	0	1	1	1	0
0	0	0	1	1	0	0	1	0	1	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1

Рис. 17. Повна таблиця істинності (*Q*-вектор) для *MUX* 4-в-1

На рис. 18 наведені результати моделювання несправностей для ПЕ *MUX* 4-в-1 за дедуктивними *Q*-векторами на наборах (0 0 0 0 0 0) та (1 1 1 1 1 1).

Набір	0	0	0	0	0	0	
Несправності	6 ¹	5 ¹	4 ¹	3 ¹	2 ¹	1 ¹	4 ¹
Позиції в D-векторі	1	2	4	8	16	32	
Значення в D-векторі	0	0	1	0	0	0	
Набір	1	1	1	1	1	1	
Несправності	6 ⁰	5 ⁰	4 ⁰	3 ⁰	2 ⁰	1 ⁰	1 ⁰
Позиції в D-векторі	62	61	59	55	47	31	
Значення в D-векторі. (інв)	0	0	0	0	0	1	

Рис. 18. Результати моделювання несправностей ПЕ *MUX* 4-в-1 за дедуктивними *Q*-векторами

Результати досліджень та їх обговорення

Перевіримо отримані результати способом моделювання несправностей на програмі DCP (кубічне моделювання) повного тесту 000000 000001

000010 000011 111100 111101 111110 111111 011111 111000, в якому присутні зазначені набори. Результат повністю збігається, що підтверджує працездатність і ефективність методу моделювання несправностей за дедуктивними Q-векторами (рис. 19).

Fault-free table								Fault's table													
No.	1	2	3	4	5	6	7	No.	Test	Out li...	1	2	3	4	5	6	7	Self q...	Quality, %	Com...	
0	U	U	U	U	U	U	U	1	*	000000	7	.	.	.	1	.	1	14	14	14	
1	0	0	0	0	0	0	0	2		000001	7	.	.	1	.	.	1	14	7	21	
2	0	0	0	0	0	0	1	3		000010	7	.	1	.	.	.	1	14	7	28	
3	0	0	0	0	0	1	0	4		000011	7	1	1	14	7	35	
4	0	0	0	0	0	1	1	5		111100	7	.	.	.	0	.	0	14	14	50	
5	1	1	1	1	0	0	1	6		111101	7	.	.	0	.	.	0	14	7	57	
6	1	1	1	1	0	1	1	7		111110	7	.	0	.	.	.	0	14	7	64	
7	1	1	1	1	1	0	1	8	*	111111	7	0	0	14	7	71	
8	1	1	1	1	1	1	1	9		011111	7	1	.	.	.	0	0	1	28	14	85
9	0	1	1	1	1	1	0	10		111000	7	.	.	.	1	1	1	1	28	14	100
10	1	1	1	0	0	0	0														

000000 000001 000010 000011 111100 111101 111110 111111 011111 111000

Рис. 19. Результати моделювання програмою DCP повного тесту для ПЕ MUX 4-в-1

Розглянемо співвідношення витрат пам'яті для кубічного моделювання несправностей і моделювання за дедуктивними Q-векторами.

Для кодування символів алфавіту A^2 (16 символів) використовується кубічне кодування по 4 біти (полубайт) на кожен символ. Так, для MUX 4-в-1 розмір КП становитиме 56 полубайтових комірок ($56 \times 4 = 224$ біти), у цьому разі Q-вектор – 64 біти, тобто Q-вектор займає менше пам'яті. Якщо умовно взяти $n = 8$ (MUX 5-в-1), то КП становитиме 90 полубайтових комірок (360 бітів), а Q-вектор – 256 бітів, що близько за витратами пам'яті одне до одного.

Тобто якщо $n > 8$ (а в середньому таке співвідношення буде для всіх регулярних КП основних логічних елементів і мультиплексорів) пам'ять для зберігання Q-вектора перевищуватиме пам'ять для зберігання КП. І що більше n ,

то зростатиме різниця не на користь кубічних Q-векторів. Тобто для ПЕ з кількістю входів $n > 8$ застосовувати Q-вектор не доцільно з огляду на витрати пам'яті.

І що далі? А далі лише декомпозиція. Наведемо приклад декомпозиції MUX 4-в-1 на ПЕ MUX 2-в-1 та моделювання несправностей декомпованої структури. На рис. 20 зображена декомпована структура MUX 4-в-1 через MUX 2-в-1 та результати моделювання несправностей на наборі (0 0 0 0 0) з огляду на результати, здобуті в процесі моделювання набору (0 0 0) за дедуктивними векторами для MUX 2-в-1.

Результати повністю збігаються з результатами кубічного моделювання зазначеного набору програмою DCP (рис. 8). Це підтверджує працездатність методу моделювання несправностей за дедуктивними Q-векторами внаслідок декомпозиції.

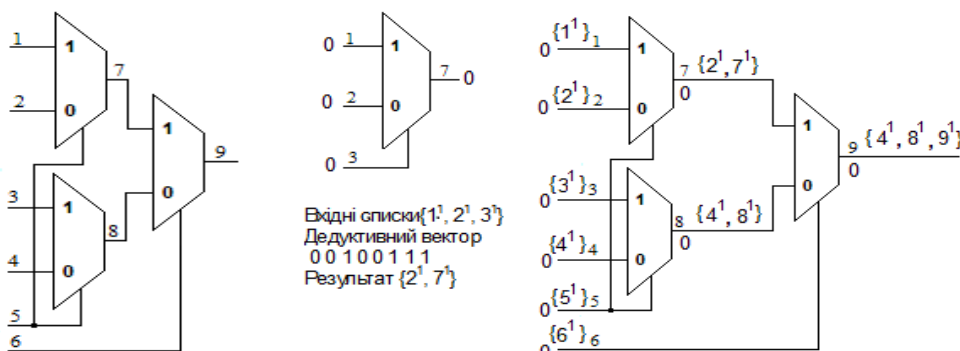


Рис. 20. Результати моделювання MUX 4-в-1 після декомпозиції

Висновки й перспективи подальшого розвитку

У статті розглянуто різні стилі опису цифрових пристроїв мовами опису апаратури. Показано еквівалентність паралельних і послідовних умовних операторів, а також їх схемна реалізація у вигляді мультиплексорів. Запропоновано спосіб отримання таблиць істинності синтезованої схемної структури за допомогою *TestBench (Xilinx ISE)*. Необхідно зауважити, що запропонований підхід має бути корисним проєктувальникам цифрових систем, які документують та обмінюються налагодженими проєктами мовами опису апаратури.

Розглянуто різні технології та структури даних дедуктивного моделювання несправностей для табличного й аналітичного способу опису цифрових схем. Запропоновано програмну реалізацію кубічного дедуктивного моделювання несправностей і показно

еквівалентність здобутих результатів для схем мультиплексорів *MUX 2-в-1* та *MUX 4-в-1* з використанням програмного продукту *DCP*.

Для моделювання застосовуються кубітні моделі опису структури цифрових систем і функцій компонентів, що визначаються компактністю опису таблиць істинності у формі Q -векторів. Це дало змогу підвищити швидкодію програмних та апаратних засобів для справного й несправного моделювання завдяки адресній реалізації аналізу кубітних векторів.

Запропоновано метод інтерпретативного моделювання несправностей цифрових схем, що базується на взятті булевих похідних способом переставлення розрядів Q -вектора без використання H -матриці. Це дало змогу скоротити витрати пам'яті в процесі програмної реалізації методу та підвищити його швидкодію.

Список літератури

1. Abramović M. A., Breuer A. M., Friedman D. Digital system testing and testable design. Comp. Sc. Press., 1998. 652 p.
2. Armstrong D. B. A deductive method of simulating faults in logic circuits. *IEEE Trans. on Computers*. Vol. № 5. 1972. P. 464–471. DOI: 10.1109/T-C.1972.223542
3. Шкиль А. С., Хаханов В. И., Ханько В. В. Дедуктивный метод кубического моделирования неисправностей цифровых устройств. *Радиоэлектроника и информатика*. 1999. № 1(6). С. 77–84. URL: <https://openarchive.nure.ua/bitstreams/1e30e364-445b-488f-97ea-9394cd98e7e2/download>
4. Хаханов В. И., Ковалев Е. В., Джахирул Х. М., Мехеди Масуд М. Д. Кубическое моделирование неисправностей цифровых проектов на основе FPGA, CPLD. *Радиоэлектроника и информатика*. 1999. № 4. С. 64–71.
5. Хаханова А. В., Хаханов В. И., Чумаченко С. В., Литвинова Е. И., Рахліс Д. Ю. Векторні моделі логіки і структури для тестування та моделювання цифрових схем. *Радиоэлектроника. Информатика. Управление*. Запоріжжя: ЗНТУ. 2021. №3. С. 69–85. DOI: 10.15588/1607-3274-2021-3-7
6. Liu T., Yu T., Wang S., Cai S. An efficient degraded deductive fault simulator for small-delay defects. *Institute of Electrical and Electronics Engineers (IEEE Access)*. 2020. Vol. 8. P. 855–862. DOI: 10.1109/ACCESS.2020.3037292
7. Kaja E., Gerlin N., Rivas L., Bora M.K., Devarajegowda K., Ecker W. MetaFS: model-driven fault simulation. *IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, 19-21 Oct. 2022. P. 1–8. DOI:10.1109/DFT56152.2022.9962369
8. Gupta I. Stuck at fault testing in combinational circuits using FPGA. *Proceedings of emerging trends and technologies on intelligent systems. Advances in Intelligent Systems and Computing*. Springer, Singapore. Noida, India, Vol 1371. P. 275–284. DOI: 10.1007/978-981-16-3097-2_23
9. Higami Y., Yamauchi T., Inamoto T., Wang S., Takahashi H., Saluja K. K. Machine learning based fault diagnosis for stuck-at faults and bridging faults. *37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC'22)*, Phuket, Thailand, 2022. P. 477–480. DOI: 10.1109/ITC-CSCC55581.2022.9894966
10. Soham R., Millican S., Agrawal V. Training neural network for machine intelligence in automatic test pattern generator. *34th International Conference on VLSI Design and 20th International Conference on Embedded Systems (VLSID'21)*, 2021, Guwahati, India. P. 16–32. DOI: 10.1109/VLSID51830.2021.00059
11. Nirmalraj T. Radhakrishnan S., Pandiyan S. K. Automatic diagnosis of single fault in interconnect testing of SRAM-based FPGA. *IET Computers & Digital Techniques*. John Wiley & Sons Ltd. 2021. №15 (5). P. 362–371. DOI: 10.1049/cdt2.12028
12. Хаханов В. И., Ктейман Хассан, Парфентий А. Н., Хаханова И. В. HFS-процессор аппаратного моделирования неисправностей цифровых проектов. *АСУ и приборы автоматизации*. 2007. № 1 (134). С. 93–108.
13. Gharibi W., Hahanova A., Hahanov V., Chumachenko S., Litvinova E., Hahanov I. Vector-logic synthesis of deductive matrices for fault simulation. *Elektronik modeling*. 2023. №45 (2). P. 16–33. DOI: 10.15407/emodel.45.02.016.

14. Pong P. C. RTL Hardware design using VHDL: coding for efficiency, portability, and scalability. *Wiley-IEEE Press*, 2006. 694 p. DOI: 10.1002/0471786411
15. Shkil A.S., Miroshnyk M., Kulak E., Filippenko I., Kucherenko D., Grebenyuk A. Synchronizing Sequences for Verification of Finite State Machines. *9th International IEEE Conference Dependable Systems, Services and Technologies, DESSERT'2018*, Ukraine, Kyiv, 2018. P. 226–230. DOI: 10.1109/UkrMiCo47782.2019.9165509
16. Шкиль А. С., Кривуля Г. Ф. Автоматизация получения булевых разностей. *АСУ и приборы автоматизации*. 1981. Вып. 59. С. 73–78.
17. Хаханов В. И., Емельянов И. В., Любарский М. М., Чумаченко С. В., Литвинова Е. И., Бани А. Т. Кубитный метод дедуктивного анализа неисправностей для логических схем. *Электронное моделирование*. 2017. Т. 39(6). С. 59–91.

References

1. Abramovici, M. A., Breuer, A. M., Friedman, D. (1998), Digital system testing and testable design, Comp. Sc. Press, 652 p.
2. Armstrong, D. B. (1972), "A deductive method of simulating faults in logic circuits", *IEEE Trans. on Computers*, Vol. № 5, P. 464–471. DOI: 10.1109/T-C.1972.223542
3. Shkil, A. C., Hahanov, V. I., Han'ko, V. V. (1999), "Deductive method of cubic faults simulation of digital devices" ["Deduktivnyj metod kubicheskogo modelirovanija neispravnostej cifrovyh ustrojstv"], *Radioelectronics and informatics*, № 1(6), P. 77–84, available at: <https://openarchive.nure.ua/bitstreams/1e30e364-445b-488f-97ea-9394cd98e7e2/download>
4. Hahanov, V. I., Kovalev, E. V., Djahirul, Hak H. M., Mehedi, Masud M.D. (1999), "Cubic fault simulation of digital systems based on FPGA, CPLD" ["Kubicheskoe modelirovanie neispravnostej cifrovyh proektov na osnove FPGA, CPLD"], *Radioelectronics and informatics*, № 4, P. 64–71.
5. Hahanova, A. V., Hahanov, V. I., Chumachenko, S. V., Litvinova, E. I., Rakhlis, D. Y. (2021), "Vector-driven logic and structure for testing and deductive fault simulation" ["Vektorni modeli logiky` i struktury` dlya testuvannya ta modelyuvannya cy`frovy`x sxe"], *Radio Electronics, Computer Science, Control, Zaporizhzhia: ZNTU*, №3, P.69–85. DOI: 10.15588/1607-3274-2021-3-7
6. Liu, T., Yu, T., Wang, S., Cai, S. (2020), "An efficient degraded deductive fault simulator for small-delay defects", *Institute of Electrical and Electronics Engineers (IEEE Access)*, Vol. 8, P. 855–862. DOI: 10.1109/ACCESS.2020.3037292
7. Kaja, E., Gerlin, N., Rivas, L., Bora, M. K., Devarajegowda, K., Ecker, W. (2022), "MetaFS: model-driven fault simulation", *IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT)*, P. 1–8. DOI:10.1109/DFT56152.2022.9962369
8. Gupta, I. (2021), "Stuck at fault testing in combinational circuits using FPGA", Proceedings of emerging trends and technologies on intelligent systems, *Advances in Intelligent Systems and Computing*, Springer, Singapore, Noida, India, Vol. 1371, P. 275–284. DOI: 10.1007/978-981-16-3097-2_23
9. Higami, Y., Yamauchi, T., Inamoto, T., Wang, S., Takahashi, H., Saluja, K. K. (2022), "Machine learning based fault diagnosis for stuck-at faults and bridging faults", *37th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC'22)*, Phuket, Thailand, P. 477–480. DOI: 10.1109/ITC-CSCC55581.2022.9894966
10. Soham, R., Millican, S., Agrawal, V. (2021), "Training neural network for machine intelligence in automatic test pattern generator", *34th International Conference on VLSI Design and 20th International Conference on Embedded Systems (VLSID'21)*, Guwahati, India, P. 16–32. DOI: 10.1109/VLSID51830.2021.00059
11. Nirmalraj, T. Radhakrishnan, S., Pandiyan, S. K. (2021), "Automatic diagnosis of single fault ininterconnecttesting of SRAM-based FPGA", *IET Computers & Digital Techniques*, John Wiley&Sons, №15 (5), P. 362–371. DOI: 10.1049/cdt2.12028
12. Hahanov V. I., Kteyman H., Parfentiy A. N., Hahanova I. V. (2007), "HFS-processor for hardware fault simulation of digital projects" ["HFS-processor apparatnogo modelirovanija neispravnostej cifrovyh proektov"], *ACS and automation devices*, № 1 (134), P. 93–108.
13. Hahanov V., Gharibi W., Hahanova A., Chumachenko S., Litvinova E., Hahanov I. (2023), "Vector-logic synthesis of deductive matrices for fault simulation", *Electronic simulation*, Vol. 45, № 2, P. 16–33. DOI: 10.15407/emodel.45.02.016
14. Pong, P. C. (2006), RTL hardware design using VHDL: coding for efficiency, portability, and scalability, *Wiley-IEEE Press*, 694 p. DOI: 10.1002/0471786411
15. Shkil A.S., Miroshnyk M., Kulak E., Filippenko I., Kucherenko D., Grebenyuk A. (2018), "Synchronizing sequences for verification of finite state machines", *9th International IEEE Conference Dependable Systems, Services and Technologies, DESSERT'2018*, Ukraine, Kyiv May 24-27, P. 226–230. DOI: 10.1109/UkrMiCo47782.2019.9165509
16. Shkil A. S., Krivulya G. F. (1981), "Automation of Boolean differences obtention" ["Avtomatizacija poluchenija bulevyh raznostej"], *ACS and automation devices*, Vol. 59, P. 73–78.
17. Hahanov V. I., Yemeljanov I. V., Lubarskiy M. M., Chumachenko S. V., Litvinova E. I., Bani A. T. (2017), "A qubit method for deductive fault analysis for logic circuits" ["Kubitnyj metod deduktivnogo analiza neispravnostej dlja logicheskikh shem"], *Electronic simulation*, Vol. 39, № 6, P. 59–91.

Відомості про авторів / About the Authors

Шкіль Олександр Сергійович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри автоматизації проєктування обчислювальної техніки, Харків, Україна; e-mail: oleksandr.shkil@nure.ua; ORCID ID: <https://orcid.org/0000-0003-1071-3445>

Мірошник Марина Анатоліївна – доктор технічних наук, професор, Харківський національний університет імені В. Н. Каразіна, професор кафедри теоретичної та прикладної системотехніки, Харків, Україна; e-mail: m.miroshnyk@karazin.ua; ORCID ID: <https://orcid.org/0000-0002-2231-2529>

Рахліс Дарія Юхимівна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри автоматизації проєктування обчислювальної техніки, Харків, Україна; e-mail: dariia.rakhlis@nure.ua; ORCID ID: <https://orcid.org/0000-0002-6652-1840>

Трифанов Олег Віталійович – Харківський національний університет радіоелектроніки, магістрант кафедри автоматизації проєктування обчислювальної техніки, Харків, Україна; e-mail: oleh.trifanov@nure.ua; ORCID ID: <https://orcid.org/0009-0003-9732-9522>

Shkil Alexander – PhD (Technical Sciences), Associated Professor, Kharkiv National University of Radio Electronics, Associated Professor at the Department of design automation, Kharkiv, Ukraine.

Miroshnyk Maryna – Doctor of Sciences (Engineering), Professor, V. N. Karazin Kharkiv National University, Professor at the Department of theoretical and applied systems engineering, Kharkiv, Ukraine.

Rakhlis Dariia – PhD (Technical Sciences), Associated Professor, Kharkiv National University of Radio Electronics, Associated Professor at the Department of design automation, Kharkiv, Ukraine.

Trifanov Oleh – Kharkiv National University of Radio Electronics, master student at the Department of design automation, Kharkiv, Ukraine.

DATA STRUCTURES FOR DEDUCTIVE SIMULATION OF HDL CONDITIONAL OPERATORS

The subject of research in the article is qubit-vector models for combinational circuits' description and procedures for faults deductive simulation based on these models. **The object of research** is the processes of diagnostic support creation for digital systems based on the usage of vector qubit data. **The purpose of the work** is to increase the speed and quality of diagnostic support creation for digital devices by creating optimal data structures and deductive fault simulation procedures based on structural-functional models of combinational circuits. The following **tasks** are solved in the article: analysis of concurrent and sequential conditional operators of hardware description languages and schematic structures into which they are synthesized; development of the procedure for the truth tables (Q-vectors) formation for schematic structures presented by HDL; development of an universal data structure for cubic and analytical deductive faults simulation; vector models improvement of qubit representation of structures and components of digital systems based on address coding of input signals to increase the manufacturability and speed of faults simulation; development of the procedure for obtaining Boolean derivatives by permuting the lines of the truth tables (Q-vectors) and using the XOR operation; development of a data structure for deductive fault simulation based on the cubic representation of digital circuit components. The following **methods** are used: deductive, cubic, deductive-parallel simulation of faults, faults simulation by deductive Q-vectors. The following **results** were obtained: the equivalence of concurrent and sequential conditional operators, as well as their schematic implementation in the form of multiplexers, was shown; method of obtaining truth tables of the synthesized circuit structure using TestBench (Xilinx ISE) was proposed; different technologies and data structures of deductive fault simulation for tabular, analytical and qubit methods of digital circuits description were considered; the software implementation of faults cubic deductive simulation is considered and the equivalence of the obtained results for the multiplexers circuits (MUX 2-in-1 and MUX 4-in-1) using the DCP software product was demonstrated. **Conclusions:** a new Q-method of interpretative faults simulation of digital circuit is proposed, which is characterized by the usage of compact Q-vectors instead of truth tables, which makes it possible to significantly increase the analysis speed due to the addressable formation of the functional primitives outputs and reduce the volume of data structures, which practically makes the method competitive with compilative simulation technologies.

Keywords: HDL-model; structural-functional model; deductive fault simulation; cubic coverage; truth table; Q-vector.

Бібліографічні опису / Bibliographic descriptions

Шкіль О. С., Мірошник М. А., Рахліс Д. Ю., Трифанов О. В. Структури даних для дедуктивного моделювання умовних операторів HDL. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 98–113. DOI: <https://doi.org/10.30837/ITSSI.2023.25.098>

Shkil, A., Miroshnyk, M., Rakhlis, D., Trifanov, O. (2023), "Data structures for deductive simulation of HDL conditional operators", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 98–113. DOI: <https://doi.org/10.30837/ITSSI.2023.25.098>

О. ШПАК, П. ФЕДОРКА, М. ПРИГАРА

РОЗУМНІ МІСТА ТА ІНТЕРНЕТ РЕЧЕЙ: ВПЛИВ РОЗРОБОК У СФЕРІ ІТ НА РОЗВИТОК МІСТ І ПОКРАЩЕННЯ ЯКОСТІ ЖИТТЯ

Предметом дослідження є аналіз особливостей використання методів і можливостей інформаційних технологій, а також електронних пристроїв-носіїв Інтернету речей для розроблення комбінованої системи з метою визначення їх особливостей. **Мета роботи** – вивчення та аналіз використання розумних технологій для "розумного" міста на основі комбінування інформаційних технологій та пристроїв Інтернету речей, а також визначення ключових критеріїв розвитку сучасних міст та впливу на покращення якості життя для населення. Відповідно до мети сформовано такі **завдання**: розглянути основні можливості застосування інформаційних технологій у поєднанні з електронними вимірально-обчислювальними пристроями Інтернету речей; визначити особливості побудови моделей інформаційних технологій з підключенням пристроїв Інтернету речей до комутаційних мереж; проаналізувати архітектуру туманних та граничних обчислень для пристроїв Інтернету речей; з'ясувати основні ризики та проблеми з питань кібербезпеки з шифруванням інформації пристроїв Інтернету речей за допомогою спеціальних алгоритмів; встановити основні критерії кожної окремої системи розумних технологій для "розумного" міста; на основі виокремлених критеріїв з'ясувати доцільність потреб із потребами населення. **Актуальною проблемою** є визначення та дослідження основних можливостей використання розумних технологій з метою виявлення проблем, пов'язаних із доцільністю забезпечення потреб для населення. У статті розглядається та досліджується вплив комбінованої системи із сукупністю розумних технологій на покращення якості життя з використанням восьми розгалужених систем, а саме: "розумна" інфраструктура, "розумна" економіка, "розумне" довкілля, "розумне" управління, "розумний" транспорт, "розумні" послуги, "розумні" люди та "розумне" життя. Сукупність критеріїв кожної окремої системи необхідно проаналізувати й визначити відповідність до потреб населення для того, щоб комбінована система "розумного" міста мала змогу забезпечити населення умовами, здатними покращити якість життя. **Науковою новизною** роботи є розроблення та впровадження концепції комбінованої системи "розумного" міста для відповідності задоволення потреб серед населення з покращенням рівня якості життя. **Здобуті результати.** Розумні технології ґрунтуються на використанні сучасних технологій, інформації та інновацій для досягнення сталого розвитку окремих систем із визначенням критеріїв, де проаналізовано таке: "розумна" інфраструктура сприяє створенню зручного, ефективного та сталого міського середовища; "розумний" транспорт забезпечує покращення руху й мобільності, зменшує негативний вплив на довкілля, підвищує безпеку й зручність для громадян; "розумне" управління відіграє вагомий внесок у розвиток різних сфер, сприяючи ефективному використанню ресурсів, та покращує якість життя зі сталим розвитком для населення; "розумне" довкілля уможливує збереження природних ресурсів, зменшує негативний вплив на довкілля та забезпечує покращення якості життя для населення; "розумні" послуги сприяють сталому розвитку галузей охорони здоров'я, фінансів, логістики туризму, що допомагає створювати більш зручне й ефективне середовище для споживачів і покращувати якість і доступність різних послуг; "розумні" люди вкладають значний внесок у розвиток і застосування нових можливостей для розроблення інформаційних технологій у поєднанні з Інтернетом речей; "розумне" життя забезпечує сталий розвиток для громадян на основі використання технологій з покращення аспектів повсякденного життя; "розумна" економіка створює умови для впровадження технологічних рішень і сприяє більш ефективному функціонуванню економіки країни. **Висновки.** Результати дослідження демонструють вплив комбінованої системи "розумного" міста на покращення рівня якості життя, що пояснюється визначенням відповідності з потребами й умовами населення. Отже, сукупність критеріїв кожної розумної технології дає змогу трансформувати місто в загальну комбіновану систему з цифровим обслуговуванням для надання послуг. З огляду на здобуті результати та сформовану наукову новизну можна стверджувати, що кожна окрема система розумних технологій задовольняє різні потреби населення й може слугувати для подальшого дослідження щодо створення цифрової моделі "міста майбутнього" із залученням відповідного програмного забезпечення.

Ключові слова: інформаційні технології; Інтернет речей; розумні технології; комбінована система; "розумне" місто.

Вступ

За останні 20 років людство стало свідком значного прогресу, зумовленого появою комп'ютерних технологій та їх можливостей. Це пояснюється розвитком та вдосконаленням вимірально-

обчислювальних машин, які в сучасному світі використовуються в інформаційних технологіях (ІТ). Унаслідок спостерігаємо більш широке застосування ІТ та їх поєднання з Інтернетом речей (ІР) у повсякденному житті [1]. Зі стрімким розвитком науки й технологій, інтернет-індустрія перебуває

в стані безперервного реформування, де інтернет-технології й міське будівництво відкрило нові можливості для подальшого спорудження "розумних" міст [2]. Розробки у сфері ІТ роблять вагомий внесок у розвиток міст та покращення життя для населення, що пояснюється багатофункційністю, ефективністю та зручністю, які надають ІТ-рішення з метою усунення проблем, з якими можуть стикатися міста [2, 7]

"Розумні" міста у поєднанні з ІТ та ІР значно впливають на розвиток міст із покращенням якості життя для населення. Ці інноваційні технології забезпечують підвищення ефективності, зручності та безпеки для міського середовища. Наприклад, ІР можуть широко використовуватися для енергетичного сектору [2, 3], охорони здоров'я [4], екології та для вирішення екологічних проблем міського середовища [5], автоматизованих систем управління ресурсами [6], що в сукупності поєднує в собі розумні технології для покращення якості життя [7]. Тому однією з переваг використання розумних технологій для "розумних" міст є поліпшення інфраструктури та громадських послуг, що пов'язано із застосуванням сенсорів та з'єднанням з мережами ІР для отримання показників рівня забруднення повітря, використання електроенергії та води, а також забезпечення систем моніторингу руху транспорту та інші важливі системи, які можуть безпосередньо стосуватися міського середовища. Такі показники дають змогу великим населеним пунктам здійснювати обґрунтоване планування розвитку, раціонально використовувати ресурси, покращувати рівень комфорту для населення та вдосконалювати якість життя [2, 7].

Системи моніторингу транспорту та дорожнього руху в громадському середовищі є відносно новими технологіями для забезпечення безпеки способом постійного спостереження за допомогою цифрових технологій (відеокамер та систем сповіщення) [8]. Системи моніторингу призначені для проведення відеоспостереження на вулицях і в громадських приміщеннях, що дає змогу знизити ризик виникнення неприємних інцидентів, аварій зі швидким реагуванням на них. Наприклад, використання систем моніторингу виявляє затори в реальному часі та проводить автоматичне управління сигналізацією світлофорів, що сприяє ефективному поліпшенню руху міського транспорту та зменшує навантаження на дороги [9].

"Розумні" міста стимулюють інновації та розвиток галузей ІТ міського середовища, а також відіграють важливе значення для ІТ-компаній та молодих

фахівців, що зі свого боку сприяє економічному зростанню міст. Унаслідок цього міста стають більш конкурентоспроможними в глобальному економічному середовищі й приваблюють інвестиції [10]. Саме тому "розумні" міста та ІР значно впливають на розвиток міст і покращують якість життя населення, допомагають підвищити ефективність та безпеку, поліпшують інфраструктуру та комунальні послуги, сприяють інноваціям та економічному зростанню. Поширення зазначених технологій у містах, безперечно, є ключем до створення більш сталого, технологічного й комфортного міського середовища.

Сучасний стан розвитку ІТ та впровадження ІР в єдину систему "розумного" міста містить безліч факторів, які можуть спричинити порушення ще на початковому етапі розроблення. Наприклад, у проектуванні інтернет-застосунку необхідно зважати на дизайн, користувацький інтерфейс, з'єднання з мережею Інтернет та протоколи безпеки, де невиконання одного з перелічених пунктів може спричинити збій у системі та до витоку конфіденційної інформації. Інші проблеми можуть бути пов'язані з енергозабезпеченням пристроїв ІР (датчиків), оскільки вони мають нетривалий термін безперебійної роботи.

Аналіз останніх досліджень та публікацій

За деякими оцінками, якщо до 2050 р. близько 65 % населення всього світу житиме у великих містах, це спричинить значний приріст кількості населення, що потребуватиме впровадження тенденцій стійкого розвитку у великих містах з ефективним місцевим управлінням. Однак сучасна практика місцевого розвитку та управління не спроможна загальмувати деякі загрозливі наслідки нестійкої урбанізації, що унеможливило покращення якості життя для населення. Тому з метою забезпечення й покращення рівня якості життя розглядаються ІТ з поєднанням ІР, що об'єднує в собі загальну систему управління для створення розумних технологій та "розумних" міст [11].

У роботі [12] описуються прості набори даних розумного міста на основі Інтернету речей (автомобільна мережа, "розумне" паркування, "розумний" дім, забруднення довкілля, моніторинг погоди), що розглядаються як аналіз для створення "розумного" міста з рішенням міського планування. Автори запропонували архітектуру, що складається з чотирьох рівнів, а саме: функції збору, агрегації,

зв'язку, оброблення та інтерпретації з використанням технологій *Hadoop* та *Spark*, що дає змогу обробляти інформацію в реальному часі. Запропонована архітектура із системою дають змогу надавати ефективні результати, працюючи з великим набором даних, що пояснюється пропускнуою здатністю системи, яка збільшується зі збільшенням розміру даних. Однак для того, щоб реалізувати запропоновану архітектуру, необхідно звернути увагу на безпеку передачі інформації та проаналізувати всі можливі варіанти її витоку, що ставить під загрозу існування та користування такою запропонованою системою.

Очікується, що ринок "розумних" міст у найближчі 20–25 років перевищить позначку 1,8 трильйона доларів, проте взаємозв'язок між віртуальною та фізичною інфраструктурою забезпечує роботу "розумного" міста, що створює нові ризики для кібербезпеки. У роботі [12] подані рекомендації щодо конфіденційності та безпеки в "розумних" містах. Автори наголошують, що з кожною додатковою точкою доступу вразливість чутливої інформації зростає, а "розумні" міста можуть бути вразливими до численних методів кібератак, наприклад, віддалене користування та глушіння сигналів, завантаження шкідливого програмного забезпечення, маніпуляції з інформацією тощо. У праці [12] наголошується, що застосування сучасних "розумних" технологій у поєднанні з ІТ підвищує безпеку, зменшує трафік, знижує рівень забруднення довкілля та покращує загальний стан якості життя для населення.

Деякі з досліджень дають змогу використовувати технології блокчейн у "розумному" місті для сталого розвитку й управління містом. У роботі [13] автори проаналізували можливість, як міста за допомогою технологій блокчейну можуть вдосконалюватися в багатьох сферах, наприклад, мобільність, комунальні послуги, управління, ресурси та логістика. Автори запропонували структуру інтеграції "розумного" міста та блокчейну способом опитування серед значної кількості населення для того, щоб виявити сприйняття громадянами "розумного" та сталого розвитку міста запропонованої структури. Водночас у роботі [14] на основі критичного аналізу автор досліджує, як недостатньо вивчений взаємозв'язок між "розумністю" та "сталістю" в контексті міського розвитку можуть мати розбіжності, що, імовірно, спричинить паралельні режими сталості унаслідок невідповідності розумних технологій і суспільних потреб. Тому для формування комплексного

розуміння взаємозв'язку між розумними технологіями та міським життям, у статті [15] розглядаються та аналізуються розумні технології в широкому розумінні, де прикладом застосування таких технологій можуть бути застосунки для смартфонів і накопичувальні пристрої-носії (смарт-годинники, автономні транспортні засоби, інтелектуальні транспортні системи, БПЛА, роботи й "розумні" будинки), що генеруються розумними технологіями.

Постановка проблеми.

Мета роботи й завдання

У розвитку міст за допомогою ІТ є безліч можливостей, що породжують нові виклики для вирішення проблем. До найбільш поширених питань належить недостатня кількість інфраструктури, незабезпеченість населення доступом до технологій, проблеми з кібербезпекою та конфіденційністю інформації, а також технологічними інноваціями, які ставлять під загрозу збереження історичної та культурної спадщини міст. Також необхідно збалансувати впровадження таких технологій, щоб уникнути цифрового розриву, коли деякі користувачі матимуть доступ до інновацій, а інші ні. Тому необхідно вирішувати ці виклики на рівні політики, інфраструктури та доступу до навчання.

Розроблення "розумних" міст із використанням ІР також супроводжується новими викликами, що визначаються такими чинниками:

1) кібербезпека (зростання кількості підключених пристроїв створює більшу вірогідність для кібератак), яка потребує захисту від несанкційного доступу та зломів, оскільки вразливість одного пристрою може поставити під загрозу всі інші пристрої в наявній системі;

2) конфіденційність даних (збір величезної кількості інформації від підключених пристроїв здатний порушити особисту конфіденційність користувачів), що може забезпечуватися ефективним зберіганням та обробленням інформації на основі використання стандартів конфіденційності;

3) стандартизація (відсутність єдиної стандартизації може призвести до розбіжностей між пристроями та системами, що, імовірно, ускладнить їх взаємодію та інтеграцію);

4) енергозабезпечення (безліч пристроїв ІР мають обмежену потужність і живлення від батареї), завданням якого є забезпечення ефективного

управління енергоспоживанням, щоб не обмежувати функційність та продуктивність системи;

5) управління ресурсами (розширення "розумних" міст та використання IP вимагає потужних обчислювальних ресурсів та інфраструктури);

6) вартість (реалізація "розумних" міст та IP-систем може бути дорогим завданням, що пов'язано зі встановленням і облаштуванням нових пристроїв та інфраструктури);

7) людський фактор (успіх таких проєктів безпосередньо залежить від прийняття та залучення населення до використання нових та інноваційних технологій, недоцільне навчання або невідповідна підготовка може стати перешкодою).

Предмет дослідження – використання методів і можливостей інформаційних технологій, а також електронних пристроїв-носіїв Інтернету речей для розроблення комбінованої системи з метою визначення їх особливостей.

Метою роботи є дослідження та аналіз використання розумних технологій для "розумного" міста на основі комбінування інформаційних технологій та пристроїв Інтернету речей, а також визначення ключових критеріїв розвитку сучасних міст із впливом на покращення якості життя населення. Відповідно до мети сформовано такі **завдання**: розглянути основні можливості застосування інформаційних технологій у поєднанні з електронними вимірально-обчислювальними пристроями Інтернету речей; визначити особливості побудови моделей інформаційних технологій з підключенням пристроїв Інтернету речей до комутаційних мереж; проаналізувати архітектуру туманних і граничних обчислень для пристроїв Інтернету речей; з'ясувати основні ризики й проблеми з питань кібербезпеки із шифруванням інформації пристроїв Інтернету речей за допомогою спеціальних алгоритмів; визначити основні критерії кожної окремої системи розумних технологій для "розумного" міста; на основі виокремлених критеріїв з'ясувати доцільність потреб із потребами населення.

Актуальною проблемою є визначення та вивчення основних можливостей використання розумних технологій з метою виявлення проблем, які пов'язані з доцільністю забезпечення потреб для населення. У цій роботі розглядається та досліджується вплив комбінованої системи із сукупністю розумних технологій на покращення якості життя з використанням восьми розгалужених систем, а саме: "розумна" інфраструктура, "розумна" економіка, "розумне" довкілля, "розумне" управління,

"розумний" транспорт, "розумні" послуги, "розумні" люди та "розумне" життя. Сукупність критеріїв кожної окремої системи необхідно проаналізувати й визначити відповідність до потреб населення для того, щоб комбінована система "розумного" міста мала змогу забезпечити населення умовами, здатними покращити якість життя.

Науковою новизною роботи є розроблення та впровадження концепції комбінованої системи "розумного" міста для задоволення потреб серед населення та покращення рівня якості життя.

Матеріали й методи

Матеріалами в цій статті розглядаються елементи та пристрої передачі інформації – датчики та виконавчі пристрої, пов'язані між собою в комбінованій системі розумних технологій як носіїв інформації, що використовується для передачі, зберігання та виявлення проблем. Основні методи містять побудову двох систем, а саме застосування IT з підключенням пристроїв IP до комутаційних мереж для оброблення інформації та побудова архітектури туманних і граничних обчислень IP із застосуванням IT. Окрім того, наводиться стисла інформація про конфіденційність та безпеку з питань кібербезпеки.

Елементи Інтернету речей

Апаратний елемент пристроїв IP складається з датчика (що має живлення від батареї), виконавчого механізму й системи зв'язку. Функція датчика полягає в збиранні показників із певного середовища. Такими показниками можуть бути швидкість потоку, температура, тиск, фізичні рухи, відстань, маса тощо. Зібрана інформація обробляється на пристрої, після чого вона може бути відправлена на віддалені сервери через мережу зв'язку [17]. На рис. 1 подано структуру взаємодії IP за допомогою комутаційних мереж різного застосування.

Датчики є пристроями IP, здатними збирати та переробляти необхідну інформацію для передачі пристроям-отримувачам, отже, датчики є енергоємними пристроями системи IP. Єдиним і суттєвим недоліком для повноцінного впровадження технології IP є обмеження часу автономної роботи пристроїв. За допомогою високого споживання енергії здійснюється збір значної кількості інформації

з подальшим опрацюванням. Кількість і точність інформації безпосередньо залежить від обмеженого терміну служби батареї, яку необхідно періодично

заряджати для надійності та безперебійності роботи, що є проблематичним для використання мільярдів взаємопов'язаних один з одним пристроїв [17].

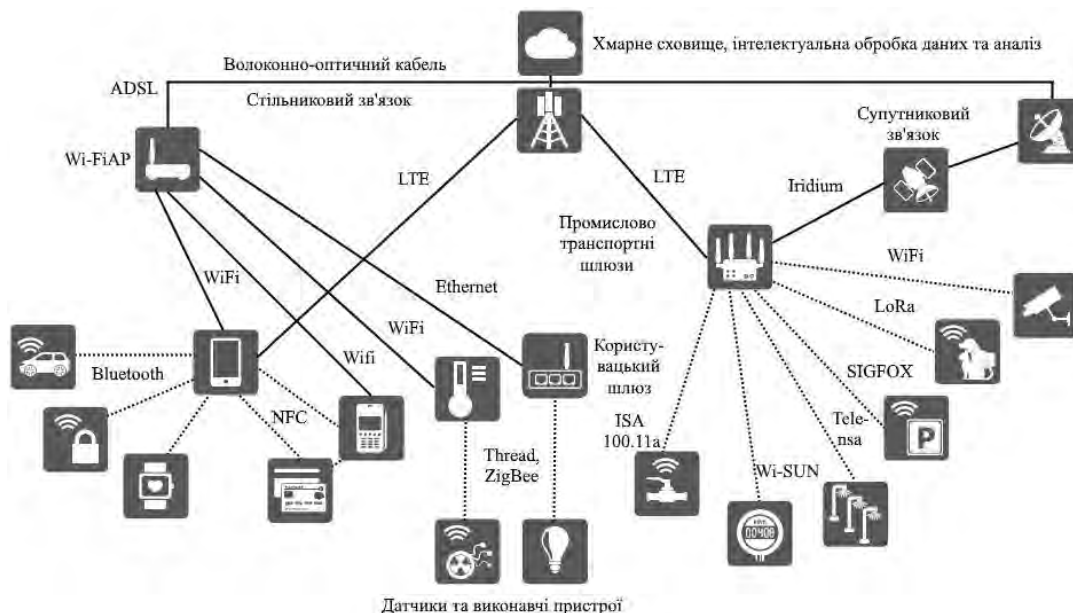


Рис. 1. Можливість застосування ІТ з підключенням пристроїв ІР до комутаційних мереж із метою оброблення інформації.
Джерело: сформовано авторами на основі [17]

Архітектура туманних і граничних обчислень для ІР

Комплексна архітектура туману (рис. 2) ґрунтується на сучасній обчислювальній архітектурі та має три основні рівні передачі інформації: хмара, туман і край. Для надання мережних послуг між хмарою та туманом розташована опорна мережа,

хмара розміщена на верхньому рівні ядра й віддалена від граничних пристроїв. Туман розташований на середньому рівні та ближче до граничних пристроїв, ніж хмара, де кожен вузол туману підключений до хмари й з'єднані між собою, що свідчить про наявність зв'язку між туманними обчисленнями, туману з хмарою та туману з граничними обчисленнями [18].

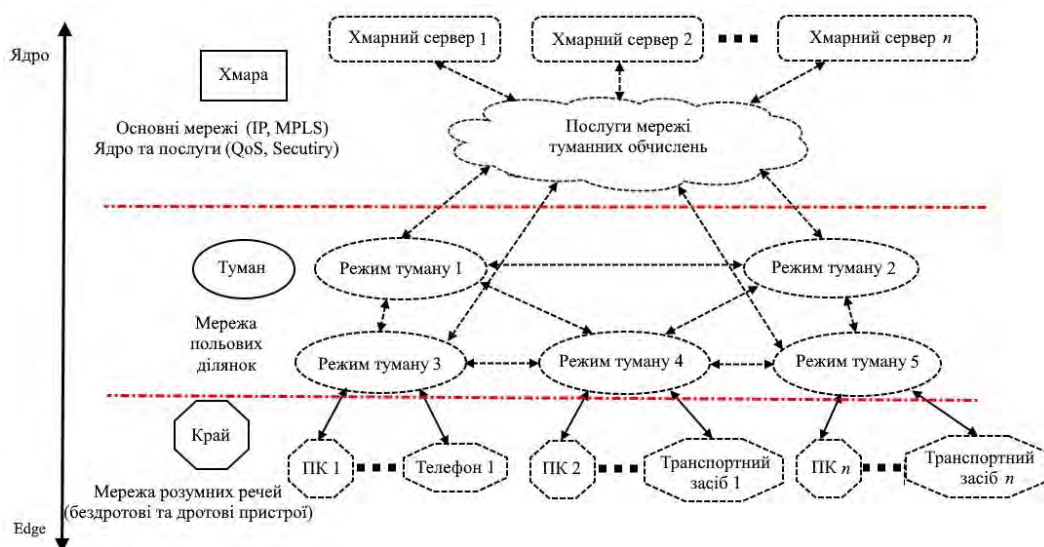


Рис. 2. Архітектура туманних і граничних обчислень для Інтернету речей із застосуванням ІТ.
Джерело: сформовано авторами на основі [18]

Хмара складається з високопродуктивних серверів і пристроїв зберігання великого обсягу інформації для її оброблення та аналізу. Хмара є центром зберігання віддаленого контролю та управління і дає змогу обробляти складні завдання. Надсилання інформації в хмару можливе за допомогою високошвидкісного бездротового й дротового зв'язку. Хмара як сховище забезпечує зберігання інформації для задоволення потреб користувачів, де також здійснюється інтелектуальний аналіз даних.

На відміну від хмари, туман складається з мережі взаємопов'язаних вузлів туманних обчислень і дає змогу забезпечити георозподілену, низьку затримку в обробленні та передачі даних. Кожен вузол туману є ресурсом центру зберігання ефемерної інформації. Туманна хмара сприяє мережним трансформаціям, а основними функціями є збирання інформації, її завантаження та зберігання, обчислення та управління [18].

Кібербезпека та шифрування інформації Інтернету речей за допомогою алгоритмів

Для захисту інформації туманних обчислень застосовуються блокові алгоритми шифрування, які сприяють підвищенню захисту оброблюваної інформації між пристроями IP [20]. Здійснення процесу шифрування та дешифрування за допомогою блокових алгоритмів відбувається з використанням мови C++, розробленої на основі платформи з бібліотекою скриптів, і дає змогу здійснити програмну та апаратну реалізацію. Для програмної реалізації блокових алгоритмів шифрування можуть застосовуватися програмні забезпечення *Python* і *Matlab* із вбудованою бібліотекою скриптів. Для математичної оптимізації та орієнтованості на визначення розрядності платформи алгоритму шифрування використовується програмна реалізація. Для апаратних операцій та прискорення процесів обчислення з обробленням інформації застосовується апаратна реалізація, що містить сучасні процесори, потужні плати та відеокарти. Суттєва різниця між програмною та апаратною реалізацією полягає в швидкісних показниках під час використання різної довжини ключа блокового алгоритму [20].

"Розумне" місто

Об'єктом дослідження є "розумне" місто – міське середовище, у якому можуть використовуватися

інформаційні технології та дані для оптимізації функціонування, покращення якості життя населення та забезпечення сталого розвитку. "Розумне" місто містить такі аспекти:

- інфраструктура – застосування сучасних інфраструктурних рішень, підключених до мережі IP, а саме: дороги, будівлі, енергетичні системи й водопостачання;

- транспорт – оптимізація роботи транспорту, зокрема електричні автомобілі, інноваційні системи громадського транспорту та розумні системи управління трафіком;

- енергетика – використання енергозберігальних технологій, розподіл енергії, відновлювані джерела енергії та моніторинг енергоефективності;

- водопостачання та утилізація відходів – використання систем управління водопостачанням, зменшення витрат води, а також покращення утилізації відходів;

- підключення та з'єднання – високошвидкісний інтернет, бездротові мережі та доступ до інформації для громадян;

- управління інформацією – її збір, аналіз і застосування для вдосконалення керування містом;

- соціальна інтеграція – забезпечення доступу до технологій усіх верств населення та розвиток соціальних ініціатив;

- безпека – використання систем відеоспостереження, розумних систем безпеки та моніторингу;

- управління містом – використання аналітики даних для прийняття рішень з планування розвитку міста, управління ресурсами, реагування на кризові ситуації тощо.

Результати досліджень та їх обговорення

Інформаційні технології змінили розвиток міст і покращили якість життя, що зумовлено поєднанням міської системи з розумними технологіями. Розглянемо деякі такі системи в табл. 1. У ній наводяться основні міські системи, джерела (вихідні виконавчі органи) та обґрунтування з показниками, що є окремим складником міської системи.

Міська інфраструктура стосується розбудованого середовища, зокрема транспортної інфраструктури (дороги, мости, тунелі, місця паркування, пішохідні зони); водної інфраструктури (водопостачання, водовідведення та санітарії, що використовуються

для водопостачання, водовідведення та зливової каналізації); енергетична інфраструктура застосовується для постачання електроенергії, газу та централізованого теплопостачання, а також для громадського освітлення; інфраструктура для збирання та перероблення муніципальних відходів; громадські будівлі, що використовуються для спорту, мистецтва, культури, освіти, охорони здоров'я та адміністрації.

Показники мають містити цифрову модель міської інфраструктури із застосуванням цифрових інструментів, таких як ГІС (географічна інформаційна система) та БІС (будівельна інформаційна система); динамічні показники, пов'язані з функціонуванням інфраструктури, зокрема споживання рідини, потік, тиск, якість рідини, електрична напруга, частота струму, температура й вологість.

Таблиця 1. Основні показники для "розумного" міста.

Джерело: сформовано авторами на основі [21]

Міська система	Джерело	Показники
Міська інфраструктура	Міська та міжгалузева адміністрація, постачальники міських послуг, менеджери об'єктів	Цифрова модель з географічною прив'язкою для архітектури й компонентів (географічна інформаційна система, інформаційне моделювання будівель), показники щодо функціонування (трафік, затори, споживання, потік, тиск, якість, напруга, частота, температура та вологість, доступність)
Міське середовище	Міська адміністрація, екологічні та метеорологічні служби (неурядова організація), міські постачальники з надання послуг, громадяни, державні органи	Індикатори щодо забруднення повітря, якості води та ґрунтів, біорізноманіття, зокрема зелені зони, біологічні види, показники громадського здоров'я та показники безпеки
Міські послуги	Міська адміністрація, а також постачальники міських послуг (транспорт, вода, енергія, муніципальні відходи), громадяни, компанії	Показники щодо якості, наявності, доступності, ризиків міських послуг (мобільність, енерго- та водопостачання, телекомунікації, муніципальні відходи, санітарія, охорона здоров'я, освіта, культурні, спортивні та мистецькі заходи)
Зацікавлені представники міста	Громадяни та політичні особи, постачальники міських послуг і соціально-економічних суб'єктів	Показники для громадян щодо міських показників (міські послуги, стратегії, значущі проекти, аналіз впливу, фінанси). Інформація від громадян, зокрема з відгуками та оцінкою про міські послуги, функціонування міста, якість життя, а також пропозиції щодо покращення
Соціально-економічна діяльність	Міська адміністрація, органи державної влади, громадські органи влади, менеджери та провайдери соціальної діяльності, економічні суб'єкти	Показники щодо типу й розподілу соціально-економічної діяльності, потенціалу будівель, промислового та інноваційного потенціалу, привабливості міста, наявності та використання культурних і спортивних об'єктів, наявності комерційних і промислових земельних ділянок, наявності робочої сили

З огляду на інформацію з табл. 1 можна схарактеризувати сукупну систему, яка поєднує в собі ІТ та ІР, що в сукупності створює загальну систему розумних технологій. На відміну від традиційного управління містом, технологія "розумного" міста має ширші можливості для "розумної" інфраструктури, економіки, управління та надання послуг. Тому "розумне" місто об'єднує окремі системи розумних технологій, що відповідають за управління розумними системи з метою уникнення та вирішення проблем, а також надають користувачам можливість для створення "розумного" суспільства. Для цього була побудована модель, що об'єднує системи з їх компонентами в єдине ціле й містить взаємозалежність одна з одною, як показано на рис. 3.

"Розумна" інфраструктура є сучасною інфраструктурою, що обладнана інформаційними технологіями й сенсорами, які дають змогу збирати, аналізувати та використовувати інформацію для оптимізації, функціонування та покращення якості послуг для громадян. У табл. 2 наведені основні ознаки "розумної" інфраструктури.

"Розумна" економіка складається з концепції, що ґрунтується на використанні інформаційних технологій, показників та інновацій для сталого зростання, покращення якості життя та оптимального застосування ресурсів. Основною ідеєю "розумної" економіки є впровадження технологічних рішень, що сприяють більш ефективному функціонуванню економіки на користь для країни. У табл. 3 наведено основні ознаки "розумної" економіки.

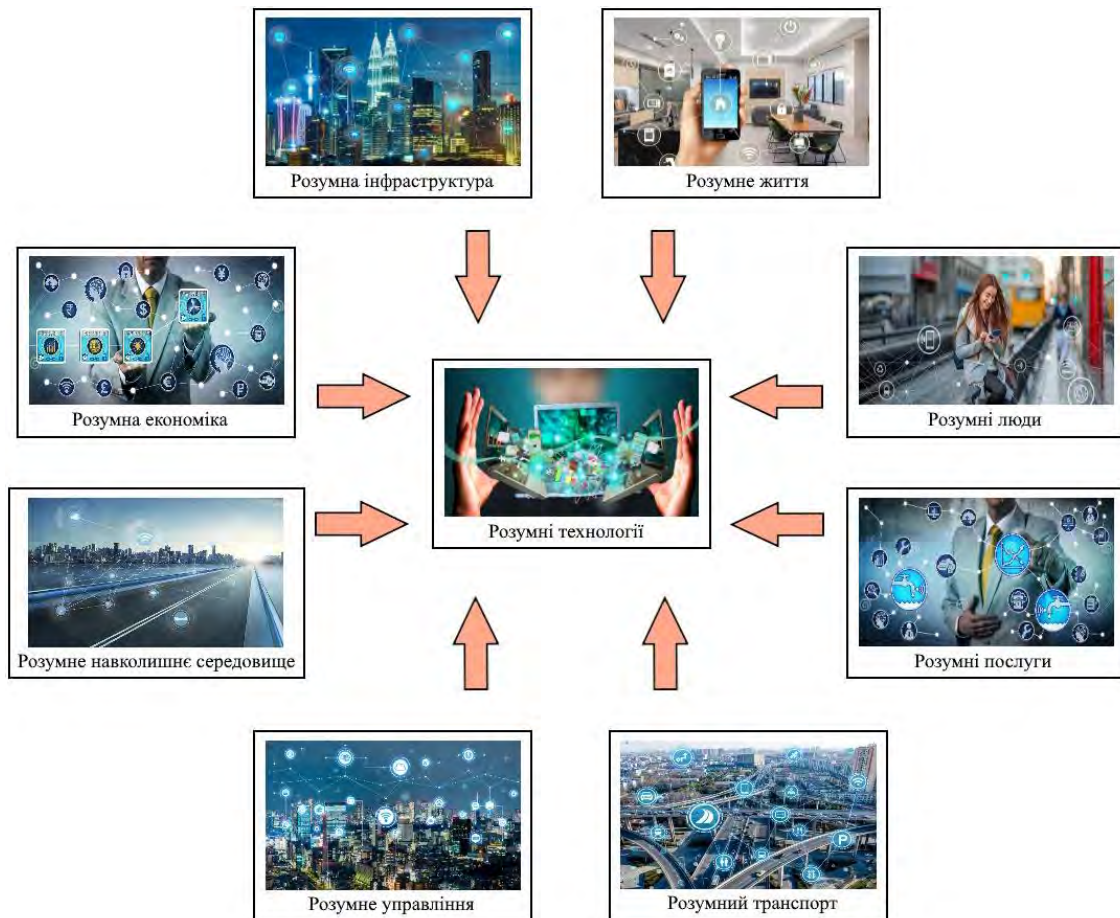


Рис. 3. Застосування розумних технологій у сфері інформаційних технологій та їх можливості

Таблиця 2. Основні ознаки "розумної" інфраструктури

Ознака	Пояснення
Підключеність	"Розумна" інфраструктура ґрунтується на підключенні різних елементів – від доріг і будівель до енергетичних систем і водопостачання до мережі IP, що дає змогу збирати інформацію та обмінюватися нею для оптимізації управління
Сенсори та датчики	"Розумна" інфраструктура містить різноманітні сенсори й датчики, що збирають показники про стан різних елементів інфраструктури. Це може бути інформація про рух транспорту, якість повітря, води, рівень води у водоймах тощо
Аналіз показників	Зібрані показники аналізуються за допомогою різних алгоритмів і штучного інтелекту для виявлення патернів, тенденцій і можливих проблем. Це дає змогу приймати обґрунтовані рішення
Оптимізація та управління	"Розумна" інфраструктура допомагає оптимізувати використання ресурсів, зменшити витрати енергії, покращити роботу транспорту, забезпечити ефективне водопостачання та відведення стічних вод тощо
Покращення якості життя	"Розумна" інфраструктура сприяє зростанню зручностей для громадян. Наприклад, вона може забезпечувати інформацію про транспорт, допомагати швидше реагувати на аварійні ситуації, покращувати якість водопостачання тощо
Сталість	"Розумна" інфраструктура може сприяти збереженню ресурсів і використанню енергії більш стало
Сприяння інноваціям	"Розумна" інфраструктура створює нові можливості для розвитку технологій та інноваційних підходів у різних галузях
Безпека й захист інформації	Важливий аспект "розумної" інфраструктури – це захист інформації та забезпечення кібербезпеки, оскільки значна кількість підключених систем може стати об'єктом кібератак

Таблиця 3. Основні ознаки "розумної" економіки

Ознака	Пояснення
Цифрова трансформація	Упровадження цифрових технологій, штучного інтелекту, аналіз інформації та автоматизації допомагає оптимізувати бізнес-процеси, підвищити продуктивність і забезпечити більш точне прийняття рішень
Інновації	Сприяння розвитку нових ідей, технологій і продуктів є ключовим аспектом "розумної" економіки. Інвестування в дослідження та розвиток, підтримка стартапів і співпраця між науковими галузями й бізнесом стимулюють інноваційний розвиток
Сталість та ефективність ресурсів	"Розумна" економіка, спрямована на зменшення витрат ресурсів, енергії та виробництва відходів. Застосування "зелених" технологій і практик допомагає зберігати довкілля та покращує сталість та ефективність ресурсів
Гнучкість та адаптивність	"Розумна" економіка спроможна швидко адаптуватися до змін у суспільстві та ринкових умов. Гнучкість управління та здатність до реагування на зміни вимагають найменших витрат
Відкритий доступ до інформації	Інформація є цінним активом "розумної" економіки. Забезпечення доступу до інформації та її оброблення допомагають виявляти тенденції, прогнозувати ризики й розвивати нові можливості
Соціальна інтеграція	"Розумна" економіка має враховувати потреби різних верств суспільства, забезпечуючи доступ до технологій та можливості для всіх
Створення нових ринків	"Розумна" економіка сприяє виникненню нових секторів і ринків, що ґрунтуються на інноваціях і технологіях

Однак упровадження "розумної" економіки також створює нові виклики, зокрема питання конфіденційності інформації, регулювання нових технологій та адаптація робочої сили до нових вимог.

"Розумні" люди – це користувачі, що мають досвід у застосуванні розумних технологій, а також фахівці, які працюють у сфері ІТ та ІР і використовують технології для вирішення різноманітних завдань, які ще не були розглянуті. Безліч інженерів, розробників програмного забезпечення, дослідників та аналітиків проєктують нові технології (створюють апаратні та програмні рішення, що дають змогу різним пристроям і сенсорам підключатися до мережі ІР); аналізують інформацію (обробляють значні обсяги інформації, зібраної з підключених пристроїв для виявлення тенденції та взаємозв'язку); розробляють рішення для бізнесу, допомагаючи підприємствам і клієнтам; займаються аналітикою та штучним інтелектом, залучаючи методи штучного інтелекту для покращення функцій ІР і автоматичного управління та прогнозування несправностей тощо; проєктують нові застосунки для покращення різних аспектів життя від сільського господарства до медицини; досліджують нові технології та інновації для розвитку ІР та забезпечують його ефективність. Окрім того, фахівці також вивчають безпеку та впроваджують нові методи й заходи для забезпечення кібербезпеки в мережі ІР, щоб запобігти зломам і несанкційному доступу.

"Розумне" життя складається з концепції, що передбачає використання інформаційних технологій

та інновацій для покращення якості життя громадян, забезпечення зручності, безпеки та сталого розвитку. Ця концепція передбачає впровадження різних рішень і технологій, що допомагають зробити повсякденне життя більш комфортним та ефективним. До основних аспектів належать: зручність (платежі, покупки, резервації); покращення комфорту (використання автоматизованих систем для контролю за домашнім оточенням, освітленням, температурою тощо); здоров'я та добробут (застосування датчиків і пристроїв для відстеження стану здоров'я, фізичної активності та забезпечення здорового способу життя); безпека (використання систем відеоспостереження, сигналізації та інших технологій особистої безпеки); ефективне застосування ресурсів (зниження витрат енергії, води, матеріалів тощо); містять можливості ("розумні" міста та громади надають населенню доступ до інновацій, сучасних сервісів і можливостей для освіти, розваг та розвитку); екологічна сталість (використання технологій для підтримання сталого розвитку та мінімізації негативного впливу на довкілля); взаємодія з громадою (упровадження цифрових платформ для спілкування, залучення громадян до прийняття рішень і розв'язання важливих питань).

"Розумні" послуги застосовують ІТ, аналітику та інновації для покращення якості та ефективності надання послуг громадянам, бізнесу та іншим споживачам, що ґрунтуються на використанні інформації з наданням індивідуально налаштованих оптимальних рішень. До основних аспектів належать:

персоналізація; взаємодія та комунікація (використання цифрових платформ для забезпечення ефективної взаємодії між провайдерами послуг і споживачами); швидкість і зручність (розумні послуги спрощують процеси та роблять їх більш доступними й зручними для користувачів); покращення якості (використання інформації для постійного вдосконалення якості надання послуг); споживання за запитом ("розумні" послуги надаються за потребою та на запит користувачів); покращення ефективності (використання технологій для оптимізації процесів

і зниження витрат); прогнозування (використання аналітичних інструментів для прогнозування потреб і розвитку попиту на послуги).

"Розумне" управління залує використання ІТ з метою аналізу інформації для оптимізації управлінських процесів, прийняття обґрунтованих рішень і покращення ефективності різних сфер, зокрема бізнес, громадську адміністрацію, організації тощо. "Розумне" управління допомагає знизити витрати, покращити якість послуг і забезпечити сталість. Основні ознаки "розумного" управління наведені в табл. 4.

Таблиця 4. Основні ознаки "розумного" управління

Ознаки	Пояснення
Збір та аналіз інформації	Використання сенсорів, датчиків та іншого обладнання для збирання інформації про процеси, якості, взаємодію тощо. Аналіз цієї інформації допомагає отримати інсайти (сприйняття чогось неочевидного) для прийняття рішень
Штучний інтелект і аналітика	Використання алгоритмів машинного навчання та штучного інтелекту для виявлення патернів, прогнозування тенденцій та вирішення складних завдань
Автоматизація	Упровадження автоматичних процесів і роботизації для покращення ефективності та зниження ризиків помилок
Взаємодія з громадянами	Використання цифрових каналів комунікації для забезпечення ефективного спілкування та залучення громадян до прийняття рішень
Оптимізація ресурсів	Застосування інформації для оптимізації використання ресурсів, зокрема енергії, води, матеріалів тощо
Прогнозування	Використання аналітичних інструментів для прогнозування подій, змін і розвитку ситуацій
Моніторинг та управління	Системи моніторингу й керування, що дають змогу відстежувати різні параметри та втручатися в процеси за необхідності
Покращення послуг	Застосування технологій для поліпшення якості надання послуг, які забезпечуються організаціями та установами

"Розумний" транспорт складається з концепції використання ІТ та інформації для оптимізації руху транспорту, забезпечення безпеки, покращення якості послуг та зменшення впливу на довкілля. Мета "розумного" транспорту полягає в тому, щоб зробити міський рух більш ефективним, зручним і сталим. До основних ознак "розумного" транспорту належать: відстеження (застосування GPS і сенсорів для відстеження місця розташування транспортних засобів у режимі реального часу); трафік і навігація (використання інформації про рух транспорту та стан доріг для визначення оптимальних маршрутів та уникнення заторів); електрична мобільність (розроблення популяризації електричних автомобілів, забезпечення інфраструктури для їх зарядки); спільне використання транспорту (розвиток послуг спільного застосування транспорту, що допомагає зменшити кількість приватних автомобілів на дорогах); безпілотне управління (розроблення автономних автомобілів і систем автоматичного управління, що можуть знизити кількість аварій і покращити

ефективність руху); безпека (використання системи відеоспостереження, систем екстреної сигналізації та взаємодії транспортних засобів для покращення безпеки); ефективне управління публічним транспортом (застосування інформації та технологій для покращення графіків, руху й сервісу громадського транспорту); зменшення забруднення ("розумний" транспорт передбачає розвиток екологічних транспортних засобів та заходи для зменшення викидів речовин-забруднювачів); управління паркуванням (використання датчиків та інформації для оптимізації паркування та зменшення пошуку місць для стоянки).

Отже, "розумний" транспорт спрямований на покращення руху й мобільності, зменшення негативного впливу на довкілля, підвищення безпеки та зручності для громадян. Це важливий аспект сталого розвитку міст і сучасного транспортного середовища.

"Розумне" довкілля складається з концепції використання ІТ та інновацій для створення ефективних і сталих взаємодій між людьми,

технологіями та природним середовищем. Мета полягає в забезпеченні більш екологічної, здорової та сталої спільноти, де технології сприяють покращенню якості довкілля.

Перелічимо основні аспекти "розумного" навколишнього середовища.

1. Моніторинг природних ресурсів. Використання сенсорів і датчиків для збирання інформації про якість повітря, води, ґрунту та інших природних ресурсів, що допомагає відстежувати забруднення та сприяє кращому управлінню довкіллям.

2. Екологічний дизайн. Застосування технологій та інновацій в архітектурі, будівництві та дизайні для створення екологічно чистих і енергоефективних споруд.

3. Енергоефективність. Упровадження розумних систем керування енергією, використання відновлюваних джерел енергії та зменшення витрат енергії.

4. Розумне управління відходами. Застосування технологій для оптимізації сортування, перероблення та утилізації відходів.

5. Екологічний транспорт. Популяризація електричних, гібридних та інших екологічних видів транспорту.

6. Системи контролю забруднення. Використання моніторингових систем та інформаційних технологій для виявлення забруднення та швидкої реакції на екологічні кризи.

7. Залучення громадян. Упровадження цифрових інструментів для підвищення екологічної свідомості та залучення громадян до дій для покращення довкілля.

8. Сільське господарство та вирощування продуктів. Використання технологій "розумних" ферм для оптимізації вирощування й поліпшення врожаю.

З огляду на розглянуті ключові аспекти "розумного" довкілля можна зробити висновок, що ця технологія забезпечує збереження природних ресурсів, зменшує негативний вплив на навколишнє середовище та покращує якість життя. Тому така система здатна поєднувати в собі інновації з екологічною відповідальністю для створення більш сталого й здорового майбутнього.

Проаналізувавши загальну систему розумних технологій, припустимо, що кожна окрема система взаємопов'язана одна з одною й залежить від наявності електронно-вимірювальних і обчислювальних пристроїв, за допомогою яких вихідна інформація фіксується та оброблюється для виведення користувачеві цифрових даних. Аналіз основних ознак та аспектів кожної системи показав, що якість життя активно проникає як у міське, так і в сільське

середовище. Міста є вузловими точками для цивільних служб, таких як безпека, охорона здоров'я, санітарія, електропостачання, водопостачання тощо. Якість життя багатомірною й залежить від рівня реалізації основних потреб, вимог безпеки, соціалізації та психосоціальних чинників.

Автори в роботі [7] розглядають, як в наші дні якість життя можна досягти в майбутніх "розумних" містах, і підтверджують це звітами про практики якості життя в "розумному" місті. Завдяки технологічній революції відбувається значне економічне піднесення серед населення, яке готове витратити їх на поліпшення якості свого життя. Описано, як корпоративним гігантам доводиться співпрацювати з місцевими міськими будівельниками або бізнесменами, що працюють з нерухомістю для того, щоб будівельники могли отримати оновлені знання про розгортання нових технологій IP з метою підвищення ефективності роботи міста в майбутніх міських просторах. До них належать смарт-картки для доступу в будь-якому місці, а системи відеоспостереження, IP-камери для забезпечення безпеки громадян, автентифікація об'єктів і моніторинг подій є іншими важливими аспектами, а також RFID, ГІС і GPS – для логістики й транспорту.

Висновки й перспективи подальшого розвитку

У статті описано та розглянуто можливість використання інформаційних технологій з елементами електронних пристроїв для збирання інформації (датчиків). Визначено основні недоліки, а саме: датчики мають живлення від батареї та належать до енергоємних пристроїв з обмеженим часом автономної роботи; кількість і точність інформації безпосередньо залежать від обмеженого терміну служби батареї, якій властиво з часом розряджатися; термін роботи батареї має забезпечувати надійність безперебійної роботи зв'язаних пристроїв між собою; деякі можливості використання датчиків можуть бути обмеженими або недоступними. Тому суттєвим недоліком застосування інформаційних технологій у поєднанні з пристроями Інтернету речей є наявність застарілих цифрових пристроїв, які необхідно вдосконалювати та шукати нові рішення щодо забезпечення "розумного" міста безперебійним моніторингом. Окрім елементів Інтернету речей, у статті розглянуто комплексну архітектуру туману з описаним принципом передачі показників,

що складається з трьох рівнів (хмара, туман і край), а також описано можливості використання методів шифрування інформації Інтернету речей з питань кібербезпеки для підвищення рівня захисту під час оброблення інформації.

Розумні технології ґрунтуються на застосуванні ІТ, інформації та інновацій для досягнення сталого розвитку окремих систем. Визначено вплив кожної окремої системи з визначенням критеріїв відповідності: "розумна" інфраструктура забезпечує створення зручного, ефективного й сталого міського середовища; "розумний" транспорт дає змогу покращити рух і мобільність, зменшує негативний вплив на довкілля, підвищує безпеку та зручність для громадян; "розумне" управління здійснює вагомий внесок у розвиток різних сфер, сприяє ефективному використанню ресурсів і покращує якість життя зі сталим розвитком для населення; "розумне" довкілля дбає про збереження природних ресурсів, зменшує негативний вплив на довкілля та забезпечує покращення якості життя для населення; "розумні" послуги створюють умови для сталого розвитку галузей охорони здоров'я, фінансів, логістики туризму, що сприяє зручному й ефективному середовищу для споживачів, якості та доступності різних послуг; "розумні" люди вкладають значний внесок у розвиток і застосування нових можливостей для розроблення

інформаційних технологій у поєднанні з Інтернетом речей; "розумне" життя забезпечує сталий розвиток для громадян на основі використання технологій з покращення аспектів повсякденного життя; "розумна" економіка містить вагомий внесок для впровадження технологічних рішень і сприяє більш ефективному функціонуванню економіки країни.

На основі проаналізованих джерел визначено особливості застосування інформаційних технологій та електронних пристроїв-носіїв Інтернету речей з можливістю розроблення комбінованої системи "розумне" місто, де було розглянуто сукупність критеріїв кожної розумної технології окремо. Результати дослідження демонструють вплив комбінованої системи "розумного" міста для покращення якості життя, що пояснюється визначенням відповідності до потреб та умов населення. Отже, сукупність критеріїв кожної розумної технології дає змогу трансформувати місто в загальну комбіновану систему з цифровим обслуговуванням для надання послуг. Із здобутих результатів і сформованої наукової новизни можна стверджувати, що системи розумних технологій, які відповідають задоволенню потреб серед населення, можуть слугувати для подальшого дослідження щодо створення цифрової моделі "міста майбутнього" із залученням відповідного програмного забезпечення.

Список літератури

1. Aggarwal A. Internet of things driven perceived value co-creation in smart cities of the future: a PLS-SEM based predictive model. *Journal of Computational and Theoretical Nanoscience*. 2019. Vol. 16. №. 9. P. 4053–4058. DOI: <https://doi.org/10.1166/jctn.2019.8292>
2. Ahmad T., Zhang D. Using the internet of things in smart energy systems and networks. *Sustainable Cities and Society*. 2021. Vol. 68. P. 102–783. DOI: <https://doi.org/10.1016/j.scs.2021.102783>
3. Farahani B. Towards collaborative intelligent IoT eHealth: From device to fog, and cloud. *Microprocessors and Microsystems*. 2020. Vol. 72. P. 102–938. DOI: <https://doi.org/10.1016/j.micpro.2019.102938>
4. Gómez J. E. IoT for environmental variables in urban areas. *Procedia computer science*. 2017. Vol. 109. P. 67–74. DOI: <https://doi.org/10.1016/j.procs.2017.05.296>
5. Маркуц В. Застосування технологій Інтернету речей в автоматизованих системах управління ресурсами. *Стратегія економічного розвитку України*. 2023. № 52, С. 97–111. DOI: <https://doi.org/10.33111/sedu.2023.52.097.111>
6. Madakam S., Ramaswamy R., Date H. Quality of life. Palava smart city: A case study. *Global Business Review*. 2019. Vol. 20. №. 3. P. 708-742. DOI: <https://doi.org/10.1177/097215091772182>
7. Garcia-Retuerta D. An efficient management platform for developing smart cities: Solution for real-time and future crowd detection. *Electronics*. 2021. Vol. 10. №. 7. 765 p. DOI: <https://doi.org/10.3390/electronics10070765>
8. Sarrab M., Pulparambil S., Awadalla M. Development of an IoT based real-time traffic monitoring system for city governance. *Global Transitions*. 2020. Vol. 2. P. 230-245. DOI: <https://doi.org/10.1016/j.glt.2020.09.004>
9. Liu D. A new model to investigate the impact of innovative IT services on smart urban growth: The mediating role of urban planners' knowledge. *Growth and Change*. 2021. Vol. 52. №. 2. P. 1040–1061. DOI: <https://doi.org/10.1111/grow.12483>
10. Yigitcanlar T. Smart cities: An effective urban development and management model. *Australian Planner*. 2015. Vol. 52. №. 1. P. 27–34. DOI: <https://doi.org/10.1080/07293682.2015.1019752>

11. Rathore M. Urban planning and building smart cities based on the internet of things using big data analytics. *Computer networks*. 2016. Vol. 101. P. 63–80. DOI: <https://doi.org/10.1016/j.comnet.2015.12.023>
12. Das A., Sharma S. C. M., Ratha B. K. The new era of smart cities, from the perspective of the internet of things. Smart cities cybersecurity and privacy. *Elsevier*. 2019. P. 1–9. DOI: <https://doi.org/10.1016/B978-0-12-815032-0.00001-9>
13. Wong P. The potential of integrating blockchain technology into smart sustainable city development. *IOP Conference Series: Earth and Environmental Science*. IOP Publishing. 2020. Vol. 463. №. 1. P. 12–20. DOI: <https://doi.org/10.1088/1755-1315/463/1/012020>
14. Woods O. Subverting the logics of "smartness" in Singapore: Smart eldercare and parallel regimes of sustainability. *Sustainable Cities and Society*. 2020. Vol. 53. P. 101–940. DOI: <https://doi.org/10.1016/j.scs.2019.101940>
15. Zhang J., He S. Smart technologies and urban life: A behavioral and social perspective. *Sustainable Cities and Society*. 2020. Vol. 63. P. 10–2460. DOI: <https://doi.org/10.1016/j.scs.2020.102460>
16. Firouzi F., Chakrabarty K., Nassif S. Intelligent internet of things: From device to fog and cloud. Springer Nature. 2020. XII, 647 p. URL: <https://link.springer.com/book/10.1007/978-3-030-30367-9>
17. Zhang P. Y., Zhou M. C., Fortino G. Security and trust issues in fog computing: A survey. *Future Generation Computer Systems*. 2018. Vol. 88. P. 16–27. DOI: <https://doi.org/10.1016/j.future.2018.05.008>
18. Chowdhury A. R. MAES: Modified advanced encryption standard for resource constraint environments. *2018 IEEE Sensors Applications Symposium (SAS)*. IEEE. 2018. P. 1–6. DOI: <https://doi.org/10.1109/SAS.2018.8336747>
19. Alassaf N. Enhancing speed of SIMON: A light-weight-cryptographic algorithm for IoT applications. *Multimedia Tools and Applications*. 2019. Vol. 78. P. 32633–32657. DOI: <https://doi.org/10.1007/s11042-018-6801-z>
20. Shahrour I., Xie X. Role of Internet of Things (IoT) and crowdsourcing in smart city projects. *Smart Cities*. 2021. Vol. 4. №. 4. P. 1276–1292. DOI: <https://doi.org/10.3390/smartcities4040068>

References

1. Aggarwal, A., Mittal, R., Gupta, S., & Mittal, A. (2019). "Internet of things driven perceived value co-creation in smart cities of the future: a PLS-SEM based predictive model". *Journal of Computational and Theoretical Nanoscience*, Vol. 16 (9), P. 4053–4058. DOI: <https://doi.org/10.1166/jctn.2019.8292>
2. Ahmad, T., Zhang, D. (2021), "Using the internet of things in smart energy systems and networks". *Sustainable Cities and Society*, Vol. 68, P. 102–783. DOI: <https://doi.org/10.1016/j.scs.2021.102783>
3. Farahani, B., Barzegari, M., Aliee, F., Shaik, K. (2020), "Towards collaborative intelligent IoT eHealth: From device to fog, and cloud". *Microprocessors and Microsystems*, Vol. 72, P. 102–938. DOI: <https://doi.org/10.1016/j.micpro.2019.102938>
4. Gómez, J., Marcillo, F., Triana, F., Gallo, V., Oviedo, B., Hernández, V. (2017), "IoT for environmental variables in urban areas". *Procedia computer science*, Vol. 109, P. 67–74. DOI: <https://doi.org/10.1016/j.procs.2017.05.296>
5. Markuts, V. (2023), "Application of Internet of Things technology in automated resource management systems". ["Zastosuvannia tekhnologii internetu rechei v avtomatyzovanykh systemakh upravlinnia resursamy"]. *Strategy of economic development of Ukraine*. Vol. 52, P. 97–111. DOI: <https://doi.org/10.33111/sedu.2023.52.097.111>
6. Madakam, S., Ramaswamy, R., Date, H. (2019), "Quality of life. Palava smart city: A case study". *Global Business Review*, Vol. 20 (3), P. 708–742. DOI: <https://doi.org/10.1177/097215091772182>
7. Garcia-Retuerta, D., Chamoso, P., Hernández, G., Guzmán, A., Yigitcanlar, T., Corchado, J. (2021), "An efficient management platform for developing smart cities: Solution for real-time and future crowd detection". *Electronics*, Vol. 10 (7), 765 p. <https://doi.org/10.3390/electronics10070765>
8. Sarrab, M., Pulparambil, S., Awadalla, M. (2020), "Development of an IoT based real-time traffic monitoring system for city governance". *Global Transitions*, 2, P. 230–245. DOI: <https://doi.org/10.1016/j.glt.2020.09.004>
9. Liu, D., Zhao, M., Xu, H., & Mehrgan, M. (2021), "A new model to investigate the impact of innovative IT services on smart urban growth: The mediating role of urban planners' knowledge". *Growth and Change*, 52 (2), P. 1040–1061. DOI: <https://doi.org/10.1111/grow.12483>
10. Yigitcanlar, T. (2015), "Smart cities: An effective urban development and management model". *Australian Planner*, Vol. 52, P. 27–34. DOI: <https://doi.org/10.1080/07293682.2015.1019752>
11. Rathore, M., Ahmad, A., Paul, A., Rho, S. (2016), "Urban planning and building smart cities based on the internet of things using big data analytics". *Computer networks*, Vol. 101, P. 63–80. DOI: <https://doi.org/10.1016/j.comnet.2015.12.023>
12. Das, A., Sharma, S., & Ratha, B. (2019), "The new era of smart cities, from the perspective of the internet of things". *In Smart cities cybersecurity and privacy*. Elsevier. P. 1-9. DOI: <https://doi.org/10.1016/B978-0-12-815032-0.00001-9>

13. Wong, P., Chia, F., Kiu, M., Lou, E. (2020), "The potential of integrating blockchain technology into smart sustainable city development". In *IOP Conference Series: Earth and Environmental Science*. Vol. 463. No. 1. P. 12–20. DOI: <https://doi.org/10.1088/1755-1315/463/1/012020>
14. Woods, O. (2020), "Subverting the logics of "smartness" in Singapore: Smart eldercare and parallel regimes of sustainability". *Sustainable Cities and Society*. Vol. 53, P. 101-940. DOI: <https://doi.org/10.1016/j.scs.2019.101940>.
15. Zhang, J., He, S. (2020), "Smart technologies and urban life: A behavioral and social perspective". *Sustainable Cities and Society*. Vol. 63. P. 102–460. DOI: <https://doi.org/10.1016/j.scs.2020.102460>
16. Firouzi, F., Chakrabarty, K., Nassif, S. "Intelligent internet of things: From device to fog and cloud". Springer Nature. 2020, XII, 647 p. available at: <https://link.springer.com/book/10.1007/978-3-030-30367-9>
17. Zhang, P., Zhou, M., Fortino, G. (2018), "Security and trust issues in fog computing: A survey". *Future Generation Computer Systems*. Vol. 88, P. 16-2-7. DOI: <https://doi.org/10.1016/j.future.2018.05.008>
18. Chowdhury, A., Mahmud, J., Kamal, A., & Hamid, M. (2018), "MAES: Modified advanced encryption standard for resource constraint environments". In *2018 IEEE Sensors Applications Symposium (SAS)*. IEEE. P. 1–6. DOI: <https://doi.org/10.1109/SAS.2018.8336747>
19. Alassaf, N. (2019), "Enhancing speed of SIMON: A light-weight-cryptographic algorithm for IoT applications". *Multimedia Tools and Applications*, Vol. 78. P. 32633–32657. DOI: <https://doi.org/10.1007/s11042-018-6801-z>
20. Shahrour, I., Xie, X. (2021), "Role of Internet of Things (IoT) and crowdsourcing in smart city projects". *Smart Cities*. Vol. 4 (4), 1276–1292. DOI: <https://doi.org/10.3390/smartsities4040068>

Надійшла 14.08.2023

Відомості про авторів / About the Authors

Шпак Олександр Іванович – кандидат фізико-математичних наук, ДВНЗ "Ужгородський національний університет", доцент кафедри програмного забезпечення систем, Ужгород, Україна; e-mail: shpak@uzhnu.edu.ua; ORCID ID: <https://orcid.org/0000-0002-1179-7196>

Федорка Павло Павлович – ДВНЗ "Ужгородський національний університет", аспірант кафедри програмного забезпечення систем, Ужгород, Україна; e-mail: fedorikapavlo@gmail.com; ORCID ID: <https://orcid.org/0000-0002-9242-5588>

Пригара Михайло Петрович – кандидат технічних наук, ДВНЗ "Ужгородський національний університет", доцент кафедри технології машинобудування, Ужгород, Україна; e-mail: mykhailo.prygara@uzhnu.edu.ua; ORCID ID: <https://orcid.org/0000-0002-0954-4480>

Shpak Oleksandr – PhD (Physical and Mathematical Sciences), Uzhhorod National University, Associate Professor at the Department of Software System, Uzhgorod, Ukraine.

Fedorka Pavlo – Uzhhorod National University, Post-graduate student at the Department of Software System, Uzhgorod, Ukraine.

Prygara Mykhailo – PhD (Engineering Sciences), Uzhhorod National University, Associate Professor at the Department of Machine Industry Technology, Uzhgorod, Ukraine.

SMART CITIES AND THE INTERNET OF THINGS: THE IMPACT OF IT DEVELOPMENTS ON THE DEVELOPMENT OF CITIES AND IMPROVING THE QUALITY OF LIFE

The subject of the article's research is the methods and capabilities of information technologies and electronic devices-carriers of the Internet of Things to develop a combined system with the aim of determining the characteristic features. **The purpose of the work** is to study and analyse the use of smart technologies for a smart city based on a combination of information technologies and Internet of Things devices, as well as to identify key criteria for the development of modern cities with an impact on improving the quality of life for the population. In accordance with the objective, the following **tasks** have been formulated: to consider the main possibilities of using information

technologies in combination with electronic measuring and computing devices of the Internet of Things; to determine the features of building information technology models with the connection of Internet of Things devices to switching networks; analyse the architecture of fog and edge computing for Internet of Things devices; to identify the main risks and cybersecurity issues with encryption of Internet of Things device data using encryption algorithms; to determine the main criteria for each individual system of smart technologies for a smart city; to determine the feasibility of needs with the needs of the population based on the defined criteria. **An urgent problem** is to identify and study the main possibilities of using smart technologies in order to identify problems related to the feasibility of meeting the needs of the population. This paper considers and investigates the impact of a combined system with a set of smart technologies on improving the quality of life using eight branched systems, namely: "smart" infrastructure, "smart" economy, "smart" environment, "smart" governance, "smart" transport, "smart" services, "smart" people and "smart" life. The set of criteria of each system needs to be analysed and matched with the needs of the population so that the combined smart city system can provide the population with conditions that can improve the quality of life. **The scientific novelty** of this paper is the development and implementation of the concept of a combined smart city system to meet the needs of the population while improving the quality of life. **Results achieved.** Smart technologies are based on the use of information technology, data and innovation to achieve sustainable development of individual systems with the definition of criteria, which analyses the following: "smart" infrastructure contributes to the creation of a convenient, efficient and sustainable urban environment; "smart" transport improves traffic and mobility, reduces the negative impact on the environment, increases safety and convenience for citizens; "smart" governance makes a significant contribution to the development of various sectors, promoting the efficient use of resources and improving the quality of life with sustainable development for the population; "smart" environment contributes to the conservation of natural resources, reduces the negative impact on the environment and ensures an improved quality of life for the population; "smart" services ensure the sustainable development of healthcare, finance, tourism logistics, which helps to create a more convenient and efficient environment for consumers, contributing to the improvement of the quality and availability of various services; "smart" people make a significant contribution to the development and application of new opportunities for the development of information technologies in combination with the Internet of Things; "smart" life ensures sustainable development for citizens through the use of technology to improve aspects of everyday life; "smart" economy makes a significant contribution to the implementation of technological solutions and contributes to a more efficient functioning of the country's economy. **Conclusions.** The results of the study demonstrate the impact of the combined smart city system on improving the quality of life, which is explained by determining the compliance with the needs and conditions of the population. Thus, the combination of criteria of each smart technology allows transforming the city into a general combined system with digital services for the provision of services. The vector of further research will be aimed at developing the considered combined system using digital technologies in order to create a digital model of the "city of the future" using appropriate software.

Keywords: information technologies; Internet of things; smart technologies; combined system; "smart" city.

Бібліографічні описи / Bibliographic descriptions

Шпак О. І., Федорка П. П., Пригара М. П. Розумні міста та Інтернет речей: вплив розробок у сфері ІТ на розвиток міст і покращення якості життя. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 114–128. DOI: <https://doi.org/10.30837/ITSSI.2023.25.114>

Shpak, O., Fedorka, P., Prygara, M. (2023), "Smart cities and the internet of things: the impact of it developments on the development of cities and improving the quality of life", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 114–128. DOI: <https://doi.org/10.30837/ITSSI.2023.25.114>

Н. Довгопол, О. Пересада, І. Прібильнова

МОДЕЛЬ ПРОГНОЗУВАННЯ ОБСЯГУ РИНКУ РОЗВАЖАЛЬНИХ ПОСЛУГ ЗА НЕДЕТЕРМІНОВАНИХ УМОВ

Предметом дослідження в статті є теоретико-методичні та прикладні засади моделювання й прогнозування показників обсягу ринку розважальних послуг. **Мета роботи** – знаходження механізму, що дасть змогу визначати обсяги підписників, беручи до уваги екзогенні змінні, особливо під час соціально нестабільних ситуацій, як-от: пандемія, війна, катаклізми тощо. У статті розв'язуються такі **завдання**: формування критеріїв компаній, для яких планується застосовуватися створюваний механізм; здійснення огляду базових авторегресійних моделей; визначення факторів, що мають слугувати зовнішнім впливом у прогнозуванні кількості підписників; утворення задачі багатокритеріального вибору; проведення експериментів за запропонованою методикою для перевірки висунутих гіпотез і систематизування здобутих результатів. Використовуються такі **методи**: аналітичний та індуктивний – для формування факторів зовнішнього впливу та опису цільових компаній; експертне оцінювання – для визначення найбільш впливових зовнішніх показників; експериментальний, багатокритеріального оцінювання та статистичні методи оброблення часових рядів – для визначення найбільш ефективної авторегресійної моделі. Здобуто конкретні **результати**. Сформовано фактори зовнішнього впливу. Зовнішніми змінними було вирішено обрати: динаміку захворюваності на коронавірус, темпи зміни світового валового внутрішнього продукту, зміна індексу *S&P500* та перетворені в числовий вигляд новини з найбільших новинних агентств світу. Створено набір критеріїв порівняння моделей: економія часу прогнозування, точність, можливість урахування зовнішнього впливу й особливості його врахування. Визначено, що найбільш точною моделлю є авторегресія рухомого середнього, яка водночас є найбільш ефективною моделлю з огляду на утворену задачу багатокритеріального вибору. Установлено схожість здобутих результатів експериментів із загальносвітовими та вітчизняними дослідженнями. **Висновки**: застосування аналітичного та індуктивного методів у поєднанні з експериментальним підходом дали змогу отримати ефективний (з точністю понад 95 %) механізм для здійснення прогнозування обсягу ринку компаній, що функціонують у кіноіндустрії та мають підписну модель ціноутворення. Здобутий результат дасть змогу меншому за обсягом ринку гравцям не втрачати розміри своєї аудиторії через нестабільність зовнішнього середовища і, відповідно, стимулювати розвиток індустрії загалом.

Ключові слова: авторегресія; кіноіндустрія; підписна модель ціноутворення; прогнозування; ринок розважальних послуг.

Вступ

Останні роки світова економіка поступово переходить із площини реальних товарів до віртуального світу. У дослідженні [1] автори ґрунтовно розглядають це питання, зосереджуючи увагу на тому, що, окрім безпосередньо появи суто цифрових благ та послуг, розвиток технологій суттєво вплинув і на вже достатньо розвинені ринки. Одним із таких є субринок ринку розважальних послуг – кіноіндустрія. З початку минулого століття цей вид дозвілля стрімко набував популярності й поступово сформувався як один із найбільш прибуткових [2]. Водночас з появою майже по всьому світу стабільного та швидкого інтернет-з'єднання почали з'являтися нелегальні ресурси для перегляду фільмів. Втрата коштів через подібне піратство,

розвиток соціальних мереж різного роду й загальне здешевлення технологій і кіновиробництва стимулювали появу легальних стримінгових платформ, як, наприклад, *Netflix* чи *Amazon Prime* [3]. До того ж перші з них фактично еволюціонували з ринку кінопрокату.

В основі подібних сервісів лежить підписна модель ціноутворення. Це фактично є загальносвітовим трендом, коли люди платять за певний час користування. Дослідники з університету міста Грац (Австрія) зазначають, що поширення цього підходу з боку компаній пояснюється можливістю постійно отримувати певний стабільний прибуток, а з боку клієнтів – доступністю опції користуватися чимось лише декілька місяців чи навіть днів [4].

За вказаних умов ринок кіноіндустрії та всі його складники 2020 р. зіштовхнулися із соціальною

катастрофою – пандемією COVID-19. Суворі обмеження змусили більшість людей звернути увагу на стримінгові сервіси й водночас суттєво ускладнили прогнозування як показників ділової активності бізнесу, так і загалом кількості підписників. Група азійських учених змогла довести, що подібна ситуація здорожчує методи аналізу інформації для покращення роботи зі споживачами та потребує оброблення суттєвих обсягів інформації [5], до того ж сама катастрофа може впливати на поведінку цільової аудиторії і після її завершення.

Треба також зауважити, що подібні сплески користувацької активності можуть виникати й без впливу соціальних катастроф, наприклад, як реакція на певні новини чи появу надзвичайно популярного фільму на відповідній платформі. Якщо такі ситуації не прогнозувати заздалегідь, можуть з'явитися проблеми, пов'язані з перенавантаженням, а це негативно вплине на користувацький досвід і, відповідно, може зменшити імовірність продовження підписки. У цьому разі прогнозування обсягів цільового ринку є доволі комплексним завданням і потребує застосування відповідного математичного апарату.

З огляду на наявні проблеми сучасні наукові дослідження виокремлюють три основних напрями для вирішення цього завдання [6, 7]:

- машинне навчання;
- імовірнісний підхід;
- авторегресійний підхід.

Машинне навчання за своєю сутністю потребує достатньо значних обсягів інформації для досягнення позитивного (точність понад 95 %) результату прогнозування, до того ж воно є доволі чутливим до викидів даних. З огляду на обрану специфіку цей підхід проблематично застосувати для поставленого завдання. Це у своїй праці продемонстрували економісти з Пакистану та Південної Кореї [8].

Водночас використання ймовірнісного підходу дало б змогу отримати вищу точність прогнозування за обмежених обсягів інформації. Однак у процесі розгляду хронологічних проміжків, пов'язаних із соціальними катастрофами та іншими сплесками активності, зазначений підхід потребуватиме суттєвої кількості інформації про зовнішні показники. Подібну проблему описують представники Корнелівського університету [9], наголошуючи, що під час розгляду задачі з обмеженою зовнішньою

інформацією точність цього підходу суттєво падає. У межах статті розглядатимуться подібні умови для того, щоб наблизити роботу алгоритму до якомога більш реальної ситуації. З огляду на сказане вирішено використовувати сімейство авторегресійних моделей для здійснення прогнозування цільових показників компанії на ринку відеохостингів. Доцільність застосування обраних моделей була доведена в процесі дослідження, проведеного науковцями університетів Шанхаю та Пекіну [10], хоча вони розглядали більш детерміноване середовище; також учені з Японії та Бангладешу змогли довести ефективність зазначених алгоритмів в умовах їх використання для прогнозування діяльності підписних сервісів [11]. Водночас група українських дослідників продемонструвала, що обране сімейство алгоритмів, зокрема авторегресія рухомого середнього, дає змогу досягти понад 95 % точності за нестабільних економічних умов [12].

Щодо інших можливих варіантів вирішення поставленої проблеми, то треба зауважити: оскільки в цій статті розглядаються періоди соціальної нестабільності, доцільність їх використання для розв'язання окресленої мети піддається сумніву і є доволі складним процесом. Група китайських учених у роботі [13] описала подібну проблему для багатовимірної лінійної регресії.

Загалом зазначена проблема й подальший розвиток обраного ринку зумовлює актуальність нашого дослідження. Його мета – знаходження авторегресійного механізму, що дає змогу прогнозувати обсяги користувачів, водночас зважаючи на зовнішній вплив і не потребуючи значного часу чи ресурсів для оброблення. Це допоможе невеликим компаніям зберегти свою частку ринку, а можливо, і збільшити конкуренцію, простимулювавши появу більш якісного сервісу. Щоб реалізувати зазначену мету, сформовано низку завдань, що необхідно вирішити в цій роботі:

- визначити набір обмежень, які накладаються на цільові компанії для досягнення найкращого результату;
- розглянути алгоритми зазначеного сімейства, на основі якого сформулювати задачу багатокритеріального вибору;
- визначити набір зовнішніх показників для прогнозування;

– сформулювати план експерименту та провести його для виокремлення найбільш ефективної моделі для досягнення поставленої мети.

Перейдемо до поступового виконання поставлених завдань.

Методи дослідження

Перед тим, як розглянути застосовані методи, необхідно визначити, які саме обмеження накладаються на профіль цільової компанії у процесі застосування обраного сімейства прогностичних алгоритмів:

– тип ціноутворення: модель підписки з необмеженим використанням за фіксованою ставкою; у разі розгляду диверсифікованого підписного плану

необхідно здійснювати прогнозування для кожної групи окремо;

– обсяг: компанія має функціонувати на декількох національних ринках, а не лише на внутрішньому; глобальність компанії зумовлюється тим, що новини інших держав, навіть за ізольованості бізнесу, можуть впливати на кількість користувачів, однак простежити за цим впливом значно складніше;

– спеціалізація: розглядаються лише платформи, які розміщують фільми, зорієнтовані на масову аудиторію, більш спеціалізовані системи можуть потребувати інших способів розрахунку.

Виокремивши специфіку компаній, можемо перейти безпосередньо до обраного сімейства моделей. Формально його можна подати таким чином:

$$\Phi_0 y_t = \Phi_1 y_{t-1} + \dots + \Phi_p y_{t-p} + \Theta_0 u_t + \Theta_1 u_{t-1} + \dots + \Theta_q u_{t-q}, \quad (1)$$

де y_t – N -мірний часовий ряд;

Φ_i, Θ_j – невідроджені матриці коефіцієнтів авторегресії розмірності $N \times N$;

$$i = \overline{1, p}, \quad j = \overline{1, q};$$

u_t – N -мірний вектор білого шуму.

Тут варто зауважити, що, оскільки матриці коефіцієнтів не вироджені, їх легко нормалізувати

в межах від 0 до 1. Модель (1) можна використовувати в процесі розгляду короткострокових періодів, до того ж потрібно гарантувати відсутність або ж неістотність зовнішнього впливу. Оскільки в межах цієї роботи вирішено розглянути середньо- та довгострокову перспективи, потрібно знайти дельту між (1) та прогнозом на попередній період, тобто $\Phi_0 y_{t-1}$. Отже, маємо таку формулу:

$$\Phi_0 \Delta y_t = \Pi y_{t-1} + \Psi \Delta y_{t-1} + \dots + \Psi_{p-1} y_{t-p+1} + \Theta_0 u_t + \Theta_1 u_{t-1} + \dots + \Theta_q u_{t-q}, \quad (2)$$

де $\Pi = -(\Phi_0 - \Phi_1 - \dots - \Phi_p)$;

$$\Psi_i = -(\Phi_{i+1} + \dots + \Phi_p);$$

$$i = \overline{1, p-1}.$$

Отримані матриці коефіцієнтів за умови загальної кількості невідомих, що потрібно брати до уваги під час прогнозування, можуть змінюватися залежно від обраної моделі. У межах цього дослідження розглядатимуться такі варіації:

- проста авторегресія;
- сезонна авторегресія;
- авторегресія розподіленого лагу;
- авторегресія рухомого середнього;
- авторегресія інтегрованого рухомого середнього.

Щоб мати змогу обрати єдину модель серед запропонованого переліку, необхідно визначити

відповідні критерії порівняння та їх пріоритетність (оцінка пріоритетності визначатиметься за 10-бальною шкалою). Варто також зауважити, що для порівняння вирішено обрати за основу оптимізаційний принцип максимізації лінійної адитивної згортки з ваговими коефіцієнтами.

Економія часу роботи моделі

Як вже зазначалося, час доволі важливий за необхідності швидкого прийняття рішення, однак треба зауважити, що з огляду на відомі дослідження [14] різниця не є настільки вагомою, тож можемо встановити пріоритет на рівні 4. Показник часу вимірюємо в мілісекундах, а щоб відповідати обраному принципу оптимізації, розглянемо, наскільки менше часу витрачається щодо найповільнішого алгоритму.

Точність прогнозу

Питання точності в побудові авторегресії є доволі комплексною проблемою, оскільки це не є задачею класифікації. Зважаючи на те, що цільовим показником є обсяг підписників, доцільно застосувати середньоквадратичну похибку. Щоб мати змогу її використати як критерій, значення після обчислення буде нормуватися. Точність прогнозу є найбільш пріоритетним показником у питанні прогнозування кількості підписників, тож можемо встановити пріоритет на рівні 10.

Ступінь урахування екзогенних змінних

Як вже зазначалося, авторегресійні моделі видозмінюються залежно від кількості параметрів, тож насправді не всі з них здатні повноцінно врахувати зовнішній вплив. Тому пропонуємо таку систему оцінювання: 0 – екзогенні змінні не враховуються, 1 – враховуються. Пріоритет цього показника встановимо на рівні 8.

Характер урахування екзогенних змінних

Окрім можливості врахування, модель може вимагати особливих зовнішніх змінних, як-от сезонна авторегресія. Для цієї ситуації теж запропонуємо бінарну змінну: 1 – не вимагає особливих зовнішніх змінних, 0 – вимагає. Видозміна зовнішніх змінних потребуватиме додаткових ресурсів, однак, оскільки час для нас не є найбільш принциповим, можемо встановити пріоритет на рівні 3.

Після визначення значень кожного з критеріїв для кожної з моделей необхідно здійснити фільтрацію результатів за допомогою принципу Парето й безпосередньо знайти значення лінійної адитивної згортки. Вагові коефіцієнти з огляду на пріоритетність будуть такими:

- економія часу роботи моделі – $4/25 = 0.16$;
- точність прогнозу – $10/25 = 0.4$;
- ступінь урахування екзогенних змінних – $8/25 = 0.32$;
- характер урахування екзогенних змінних – $3/25 = 0.12$.

Щодо засобів використаних для моделювання та побудови, то вирішено обрати мову програмування *Python3* з відповідними бібліотеками, зокрема *sklearn, polars, time*.

Цільовими змінними будуть застосовані показники за 2012–2022 рр. щодо кількості користувачів *Netflix* та *Amazon Prime*, оскільки зазначені сервіси мають найдовшу історію існування. У цьому разі щоб мати змогу оцінити точність, 80 % дат розглядатимуться як навчальні, а ще 20 % – як тестові.

Для вибору зовнішніх змінних вирішено застосувати метод експертного оцінювання. Опитування проведено серед 10 ризик-менеджерів, 10 фінансових аналітиків та 10 соціологів з Харківської, Львівської, Дніпропетровської та Київської областей. Відповідно до результатів встановлено, що як зовнішні змінні необхідно використати:

- кількість нових випадків коронавірусу (для коректного відпрацювання алгоритмів показники будуть доповнені неістотними нулями);
- динаміка зміни світового ВВП;
- індекс *S&P500* (оскільки компанії розташовані в США);
- новини з найбільших інформаційних агентств світу, перетворених у числовий вигляд за допомогою алгоритму запропонованого вченими Харківського національного університету радіоелектроніки [15].

Визначивши ключові особливості та методи, можемо перейти безпосередньо до експерименту.

Результати

Почнемо з показника точності. Як зазначалося вище, для його знаходження використано дві вибірки з показниками щодо кількості підписників *Amazon Prime* та *Netflix*. Цільовим показником точності застосовувалося середньоквадратичне відхилення. Для простоти порівняння у табл. 1 показники подані в нормалізованому вигляді, у межах від 0 до 1. Окремо необхідно наголосити, що вирішено використовувати найкраще значення "точності", знайдене шляхом крос-валідації цільових моделей із різними гіперпараметрами.

З огляду на середні значення показника точності для двох вибірок, можна зауважити, що найбільш ефективним буде авторегресія рухомого середнього, як класична, так і інтегрована. Ця модель дає змогу врахувати середньострокову волатильність цільової змінної. Найменше значення в простій авторегресії, що закономірно, оскільки вона не враховує зовнішнього впливу. Якщо ж казати про сезонну

авторегресію, то тут варто зауважити, що під час розгляду показників без урахування років пандемії точність становила 0.93, що пояснюється загальною сезонністю макроекономічних показників.

Таблиця 1. Нормалізоване середньоквадратичне відхилення (розроблено авторами)

Показники	Авторегресія	Сезонна авторегресія	Авторегресія розподіленого лагу	Авторегресія рухомого середнього	Авторегресія інтегрованого рухомого середнього
Netflix	0.76	0.84	0.87	0.95	0.96
Amazon Prime	0.77	0.86	0.90	0.96	0.95
Середнє	0.77	0.85	0.89	0.96	0.96

Перейдемо до показника економії часу роботи алгоритму. Для його знаходження застосовано лише вибірку з інформацією щодо *Netflix*. З огляду на здобуті показники встановлено, що найповільнішим алгоритмом є авторегресія інтегрованого рухомого середнього (за середнім значенням п'яти замірів часу прогнозування). Відповідно цей алгоритм і стане базисом для знаходження "економії часу". Результати обчислення наведено в табл. 2, у цьому разі інформація щодо авторегресії інтегрованого рухомого середнього опускається як неістотна.

Таблиця 2. Економія часу прогнозування (розроблено авторами)

Авторегресія	Сезонна авторегресія	Авторегресія розподіленого лагу	Авторегресія рухомого середнього
0.039 с	0.030 с	0.008 с	0.002 с
0.040 с	0.030 с	0.009 с	0.002 с
0.038 с	0.028 с	0.007 с	0.001 с
0.040 с	0.029 с	0.008 с	0.003 с
0.041 с	0.032 с	0.008 с	0.001 с
Середнє (округлено до тисячних)			
0.040 с	0.030 с	0.008 с	0.002 с

Повільність роботи алгоритмів рухомого середнього, порівняно з іншими, пояснюється тим, що вони обчислюють середнє, враховуючи окіл поточного значення, тоді як проста авторегресія не має специфічних щодо інших обраних алгоритмів обчислень. Тому і є найшвидшою. Виконавши заміри, можемо перейти до формування табл. 3 зі значеннями всіх обраних критеріїв.

Таблиця 3. Значення всіх критеріїв для кожної з альтернатив (розроблено авторами)

	Економія часу прогнозування	Нормалізована точність	Урахування зовнішнього впливу	Характер зовнішнього впливу
Авторегресія	0.040 с	0.77	0	1
Сезонна авторегресія	0.030 с	0.85	1	0
Авторегресія розподіленого лагу	0.020 с	0.89	1	1
Авторегресія рухомого середнього	0.008 с	0.96	1	1
Авторегресія інтегрованого рухомого середнього	0.000 с	0.96	1	1

Серед запропонованих альтернатив приберемо ті, які за всіма критеріями є гіршими за хоча б один з інших варіантів. Серед п'яти моделей за принципом Парето відфільтровується авторегресія інтегрованого рухомого середнього. Отже, табл. 3 можна перетворити в табл. 4.

Таблиця 4. Значення всіх критеріїв для кожної з альтернатив оптимальних за Парето (розроблено авторами)

	Економія часу прогнозування	Нормалізована точність	Урахування зовнішнього впливу	Характер зовнішнього впливу
Авторегресія	0.040 с	0.77	0	1
Сезонна авторегресія	0.030 с	0.85	1	0
Авторегресія розподіленого лагу	0.008 с	0.89	1	1
Авторегресія рухомого середнього	0.002 с	0.96	1	1

Значення двох останніх критеріїв є бінарними, а показник точності вже є нормалізованим. Залишається здійснити нормалізацію показника економії часу прогнозування. Відповідні результати наведені в табл. 5.

Таблиця 5. Нормалізовані значення критеріїв для кожної з альтернатив (розроблено авторами)

	Економія часу прогнозування	Нормалізована точність	Урахування зовнішнього впливу	Характер зовнішнього впливу
Авторегресія	0.95	0.77	0	1
Сезонна авторегресія	0.71	0.85	1	0
Авторегресія розподіленого лагу	0.20	0.89	1	1
Авторегресія рухомого середнього	0.05	0.96	1	1

Знайдемо значення лінійної адитивної згортки для кожної із поданих у табл. 5 моделей:

– проста авторегресія:

$$0.95 \times 0.16 + 0.77 \times 0.4 + 0.12 = 0.58;$$

– сезонна авторегресія:

$$0.71 \times 0.16 + 0.85 \times 0.4 + 0.32 = 0.7736;$$

– авторегресія розподіленого лагу:

$$0.20 \times 0.16 + 0.89 \times 0.4 + 0.32 + 0.12 = 0.828;$$

– авторегресія рухомого середнього:

$$0.05 \times 0.16 + 0.96 \times 0.4 + 0.32 + 0.12 = 0.832.$$

Отже, авторегресія рухомого середнього є найбільш ефективною моделлю для обраних зовнішніх показників і цільової інформації, особливо в умовах соціальних катастроф.

Тепер спробуємо застосувати цю модель для прогнозування кількості підписників за 2022 р. Маємо такі результати для *Netflix*:

– I квартал 2022: 222 млн осіб – реальне значення 222;

– II квартал 2022: 223 млн осіб – реальне значення 221;

– III квартал 2022: 225 млн осіб – реальне значення 223;

– IV квартал 2022: 229 млн осіб – значення, прогнозоване компанією, –224.

Водночас для *Amazon Prime* показники такі:

– I квартал 2022: 155 млн осіб – реальне значення 153;

– II квартал 2022: 156 млн осіб – реальне значення 153;

– III квартал 2022: 158 млн осіб – реальне значення 155;

– IV квартал 2022: 160 млн осіб – значення, прогнозоване компанією, –157.

З огляду на зазначене вище можна зауважити, що обрані методи для виконання поставлених завдань відповідають національному та загальносвітовому науковому досвіду й не суперечать йому..

Висновки

Результати проведеного експерименту доводять доцільність використання авторегресійних алгоритмів, зокрема авторегресії рухомого середнього, для прогнозування кількості підписників сервісів на ринку розважальних послуг за невизначених умов.

Загалом моделі обраного сімейства дали змогу досягти точності від 0.77 до 0.96, зважаючи на показник середньоквадратичної похибки, для низки показників *Amazon Prime* та *Netflix* у період 2012–2022 рр. Крім того, як зовнішні змінні розглядалися:

- кількість нових випадків коронавірусу;
- динаміка зміни світового ВВП;
- індекс *S&P500*;
- новини з найбільших інформаційних агентств світу, перетворених у числовий вигляд.

Результати дослідження можна використовувати для впровадження особливих умов підписної моделі, зокрема акції з пробним періодом чи сімейні

підписки. Також результати прогнозування дають змогу скорегувати політику ціноутворення та маркетингову стратегію для пом'якшення ризиків під час критичних соціальних умов і досягнення оптимального рівня прибутку / витрат.

З метою поліпшення здобутих результатів можна збільшити частоту показників, наприклад, за допомогою аргументації, та обсяги зовнішніх показників. Окрім цього, для подальшого дослідження вирішено порушити питання збільшення кількості цільових змінних і відповідно переходу до векторних авторегресійних моделей.

Список літератури

1. Bagnoli C., Biazzo S., Biotto G. Digital business models for Industry 4.0. How innovation and technology shape the future of companies. *Springer*, Cham. 2022. 268 p. DOI: 10.1007/978-3-030-97284-4
2. Oyewola D. O., Dada E. G. Machine Learning Methods for Predicting the Popularity of Movies. *Journal of Artificial Intelligence and Systems*. 2020. № 4. P. 65–82. DOI: 10.33969/AIS.2022040105
3. Wang W., Guo Q. Subscription strategy choices of network video platforms in the presence of social influence. *Electronic Commerce Research*. 2021. № 23, P. 577–604. DOI: 10.1007/s10660-021-09504-w
4. Kerschbaumer R. H., Foscht T., Eisingerich A. B. Smart Services and the Rise of Access-based Subscription Models. *Smart Services, Wiesbaden: Springer Gabler*, 2022. P. 179–205. DOI: 10.1007/978-3-658-37346-7_6
5. Shin Z., Moon J., Rho S. A Comparative Analysis of Ensemble Learning-Based Classification Models for Explainable Term Deposit Subscription Forecasting. *Journal of Society for e-Business Studies*. Vol. 3. № 26. P. 1–21. URL: <http://www.jsebs.org/jsebs/index.php/jsebs/article/view/457> (дата звернення: 08.09.2023).
6. Choujun Zhan; Jianbin Li; Wei Jiang; Wei Sha; Yijing Guo E-commerce Sales Forecast Based on Ensemble Learning. *IEEE International Symposium on Product Compliance Engineering-Asia (ISPCE-CN)*. 2020. P. 1–5. DOI: 10.1109/ISPCE-CN51288.2020.9321858
7. Masini R. P., Medeiros M. C., Mendes E. F. Machine learning advances for time series forecasting. *Journal of Economic Surveys*. 2021. Vol. 37. № 1. P. 76–111. DOI: 10.1111/joes.12429
8. Ullah I., Raza B., Malik A. K., Imran M., Islam S. U., & Kim S. W. A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector *IEEE Access*. 2019. № 7. P. 60134–60149. DOI: 10.1109/access.2019.2914999
9. Ning C., You F. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*. 2019. № 125. P. 434–448. DOI: 10.1016/j.compchemeng.2019.03.034
10. Qi X.-Z., Ning Z., Qin M. Economic policy uncertainty, investor sentiment and financial stability—an empirical study based on the time varying parameter-vector autoregression model. *Journal of Economic Interaction and Coordination*. 2022. № 17. P. 779–799. DOI: 10.1007/s11403-021-00342-5
11. Shibasaki M., Witayangkurn A., & Rahman M. M. Comparison of life patterns from mobile data in Bangladesh. *Smart Technology & Urban Development (STUD-2019): 1st International Conference, Chiang Mai, 13 December – 14 December 2019: IEEE*, 2019. P. 1–7. DOI: 10.1109/STUD49732.2019.9018795
12. Khovrat A., Kobziev V., Nazarov A., & Yakovlev S. Parallelization of the VAR Algorithm Family to Increase the Efficiency of Forecasting Market Indicators During Social Disaster. *Information Technology and Implementation (IT&I 2022): 9th International Conference, Kyiv, 30 November – 2 December 2022: CEUR Workshop Proceedings*. No. 3347, P. 222–233. URL: https://ceur-ws.org/Vol-3347/Paper_19.pdf (дата звернення: 08.09.2023).
13. Wang W., Guo Q. Subscription strategy choices of network video platforms in the presence of social influence. *Electronic Commerce Research*. 2021. № 23. P. 577–604. DOI:10.54691/bcpbm.v34i.3018

14. Haslbeck J., Bringmann L., Waldorp L. A Tutorial on Estimating Time-Varying Vector Autoregressive Models. *Multivariate Behavioral Research*. 2021. Vol. 56. № 1. P. 120–149. DOI: 10.1080/00273171.2020.1743630
15. Afanasieva I., Golian N., Golian V., Khovrat A., & Onyshchenko K. Application of Neural Networks to Identify of Fake News. *Computational Linguistics and Intelligent Systems (COLINS 2023): 7th International Conference, Kharkiv, 20 April – 21 April 2023: CEUR workshop proceedings*, No. 3396, 2023. P. 346–358. URL: <https://ceur-ws.org/Vol-3396/paper28.pdf> (дата звернення: 08.09.2023).

References

- Bagnoli, C., Biazzo, S., Biotto, G. (2022), "Digital business models for Industry 4.0. How innovation and technology shape the future of companies". *Springer*, Cham. 268 p. DOI: 10.1007/978-3-030-97284-4
- Oyewola, D. O., Dada, E. G. (2022), "Machine Learning Methods for Predicting the Popularity of Movies", *Journal of Artificial Intelligence and Systems*, No. 4, P. 65–82. DOI: 10.33969/AIS.2022040105
- Wang, W., & Guo, Q. (2021), "Subscription strategy choices of network video platforms in the presence of social influence", *Electronic Commerce Research*, No. 23, P. 577–604. DOI: 10.1007/s10660-021-09504-w
- Kerschbaumer, R. H., Foscht, T., & Eisingerich, A. B. (2022), "Smart Services and the Rise of Access-based Subscription Models", *In: Smart Services, Springer Gabler, Wiesbaden*, P. 179–205. DOI: 10.1007/978-3-658-37346-7_6
- Shin, Z., Moon, J., & Rho, S. (2021), "A Comparative Analysis of Ensemble Learning-Based Classification Models for Explainable Term Deposit Subscription Forecasting", *Journal of Society for e-Business Studies*, No. 26(3), P. 1–21, available at: <http://www.jsebs.org/jsebs/index.php/jsebs/article/view/457> (last accessed 08.09.2023).
- Li, J., Zhan, C., Sha, W., Jiang, W., & Guo, Y. (2020), "E-commerce Sales Forecast Based on Ensemble Learning". *IEEE International Symposium on Product Compliance Engineering-Asia (ISPCE-CN)*. P. 1–5. DOI: 10.1109/ISPCE-CN51288.2020.9321858
- Masini, R., Medeiros, M., Mendes, E. (2021), "Machine learning advances for time series forecasting", *Journal of Economic Surveys*, No. 37(1), P. 76–111. DOI: 10.1111/joes.12429
- Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019), "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector", *IEEE Access*, No. 7. P. 60134–60149. DOI: 10.1109/ACCESS.2019.2914999
- Ning, C., & You, F. (2019), "Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming", *Computers & Chemical Engineering*, No. 125, P. 434–448. DOI: 10.1016/j.compchemeng.2019.03.034
- Qi, X.-Z., Ning, Z., & Qin, M. (2022), "Economic policy uncertainty, investor sentiment and financial stability – an empirical study based on the time varying parameter-vector autoregression model". *Journal of Economic Interaction and Coordination*. No. 17. P. 779–799. DOI: 10.1007/s11403-021-00342-5
- Shibasaki, M., Witayangkurn, A., & Rahman, M. M. (2019), "Comparison of life patterns from mobile data in Bangladesh". *Smart Technology & Urban Development (STUD-2019): 1st International Conference, Chiang Mai, 13 December – 14 December 2019: IEEE*, P. 1–7. DOI: 10.1109/STUD49732.2019.9018795
- Khovrat, A., Kobziev, V., Nazarov, A., & Yakovlev, S. (2022), "Parallelization of the VAR Algorithm Family to Increase the Efficiency of Forecasting Market Indicators During Social Disaster". *Information Technology and Implementation (IT&I 2022): 9th International Conference, Kyiv, 30 November – 2 December 2022: CEUR Workshop Proceedings*. No. 3347, P. 222–233. available at: https://ceur-ws.org/Vol-3347/Paper_19.pdf (last accessed: 08.09.2023).
- Wang, G., Wang, Zh., & Xie, Y. (2022), "Subscribers forecasting of netflix based on multiple linear models", *BCP Business & Management*, No. 34, P. 229–236. DOI:10.54691/bcpbm.v34i.3018
- Haslbeck, J., Bringmann, L., & Waldorp, L. (2021), "A Tutorial on Estimating Time-Varying Vector Autoregressive Models". *Multivariate Behavioral Research*, No. 56 (1), P. 120–149. DOI: 10.1080/00273171.2020.1743630
- Afanasieva, I., Golian, N., Golian, V., Khovrat, A., & Onyshchenko, K. (2023), "Application of Neural Networks to Identify of Fake News". *Computational Linguistics and Intelligent Systems (COLINS 2023): 7th International Conference, Kharkiv, 20 April – 21 April 2023: CEUR workshop proceedings*, No. 3396, P. 346–358. available at: <https://ceur-ws.org/Vol-3396/paper28.pdf> (last accessed: 08.09.2023).

Відомості про авторів / About the Authors

Довгопол Ніна Василівна – кандидат економічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри економічної кібернетики та управління економічною безпекою, Харків, Україна; e-mail: nina.dovgopol@nure.ua; ORCID ID: <https://orcid.org/0000-0002-2438-0904>

Пересада Олена Василівна – Харківський національний університет радіоелектроніки, старший викладач кафедри економічної кібернетики та управління економічною безпекою, Харків, Україна; e-mail: olena.peresada@nure.ua; ORCID ID: <https://orcid.org/0000-0003-0388-975X>

Прібильнова Інна Борисівна – Харківський національний університет радіоелектроніки, старший викладач кафедри економічної кібернетики та управління економічною безпекою, Харків, Україна; e-mail: inna.butukina@nure.ua; ORCID ID: <https://orcid.org/0000-0001-6851-5340>

Dovgopol Nina – PhD (Economic Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Economic Cybernetics and Management of Economic Security, Kharkiv, Ukraine.

Peresada Olena – Kharkiv National University of Radio Electronics, Senior Lecturer at the Department of Economic Cybernetics and Management of Economic Security, Kharkiv, Ukraine.

Pribylnova Inna – Kharkiv National University of Radio Electronics, Senior Lecturer at the Department of Economic Cybernetics and Management of Economic Security, Kharkiv, Ukraine.

A MODEL FOR FORECASTING THE VOLUME OF THE ENTERTAINMENT SERVICES MARKET UNDER NON-DETERMINATIVE CONDITIONS

The subject matter of the article is the theoretical-methodical and applied principles of modelling and forecasting indicators of the market volume of entertainment services. The **goal of the work** is to find a mechanism that will allow determining the volume of subscribers, taking into account exogenous variables, especially during socially unstable situations, such as a pandemic, war, cataclysms, etc. The following **tasks** are solved in the article: formation of criteria for companies for which the created mechanism is planned to be applied; review of basic autoregression models; determination of factors that should serve as an external influence when predicting the number of subscribers; creation of a multi-criteria choice problem; conducting experiments according to the proposed methodology to test the proposed hypotheses and systematize the obtained results. The following **methods** are used: analytical and inductive methods for forming factors of external influence and description of target companies; expert evaluation method for determining the most influential external indicators; experimental method, statistical methods of processing time series and methods of multi-criteria evaluation to determine the most effective autoregression model. The following **results** were obtained: the factors of external influence were formed: As external variables, it was decided to choose: the dynamics of the incidence of coronavirus, the rate of change in the global gross domestic product, the change in the S&P500 index, and the news from the world's largest news agencies converted into numerical form; a set of criteria for comparing models was created, saving forecasting time, accuracy, the possibility of taking into account external influence and the specificity of taking it into account; it was determined that the most accurate model is autoregression of the moving average, which at the same time is the most effective model given the created problem of multi-criteria selection; the similarity of the obtained results of experiments with global and domestic research is established. **Conclusions:** the use of analytical and inductive methods in combination with an experimental approach made it possible to obtain an effective (with an accuracy of more than 95%) mechanism for forecasting the market volume of companies that operate in the film industry and have a signature pricing model. The obtained result will allow players with a smaller market volume not to lose their audience due to the instability of the external environment, and, accordingly, will stimulate the development of the industry in general.

Keywords: autoregression; film industry; subscription pricing model; forecasting; entertainment market.

Бібліографічні описи / Bibliographic descriptions

Довгопол Н. В., Пересада О. В., Прібильнова І. Б. Модель прогнозування обсягу ринку розважальних послуг за недетермінованих умов. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 129–137. DOI: <https://doi.org/10.30837/ITSSI.2023.25.129>

Dovgopol, N., Peresada, O., Pribylnova I (2023), "A model for forecasting the volume of the entertainment services market under non-determinative conditions", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 129–137. DOI: <https://doi.org/10.30837/ITSSI.2023.25.129>

V. IVANENKO

ENHANCING ASPECT-BASED FINANCIAL SENTIMENT ANALYSIS THROUGH CONTRASTIVE LEARNING

The **subject** of research in the article explores the specialized application of Aspect-Based Financial Sentiment Analysis (ABFSA), focusing on the intricate and multifaceted emotional landscape of financial textual data. The study extends the current understanding of sentiment analysis by addressing its limitations and opportunities within a financial context. The **purpose** of the work is to advance the field of Aspect-Based Financial Sentiment Analysis by developing a more nuanced and effective methodology for analyzing sentiments in financial news. Additionally, the study aims to assess the efficacy of recent advancements in Natural Language Processing (NLP) and machine learning for enhancing ABFSA models. The article deals with the following **tasks**: Firstly, the study focuses on the rigorous pre-processing of the SEntFiN dataset to make it more amenable to advanced machine learning techniques, specifically contrastive learning methodologies. Secondly, it aims to architect a unified model that integrates state-of-the-art machine learning techniques, including DeBERTa v3, C^2L contrast learning, and LoRa fine-tuning. Lastly, the research critically evaluates the proposed model's performance metrics across the test dataset and compares them with existing methodologies. The following **methods** are used: Firstly, the study employs pre-processing techniques tailored for the SEntFiN dataset, which is explicitly designed for entity-sensitive sentiment analysis in financial news. Secondly, it utilizes advanced machine learning techniques such as DeBERTa v3 for language model pre-training, C^2L contrast learning for focusing on causal relationships, and LoRa for fine-tuning large language models. Lastly, performance evaluation methods are used to assess the efficacy of the proposed model, including comparisons with existing methodologies in the field. The following **results** were obtained: The study reveals that the proposed pre-processing framework successfully accommodates the variable number of entities present in financial news, thereby improving the granularity of sentiment classification. Furthermore, the integration of advanced NLP and machine learning techniques significantly enhances the accuracy and efficiency of ABFSA models. **Conclusions**: The paper concludes that specialized ABFSA methodologies, when augmented with advanced NLP techniques and a robust pre-processing framework, can offer a more nuanced and accurate representation of sentiment in financial narratives. The study lays the groundwork for future research in this nascent yet crucial interdisciplinary field, providing actionable insights for stakeholders ranging from investors to financial analysts.

Keywords: Aspect-based Financial Sentiment Analysis; Contrastive Learning; Text Classification.

Introduction

Sentiment analysis, a traditional subfield of Natural Language Processing (NLP), focuses on identifying and quantifying the affective and subjective nuances present in textual data. While its applications encompass diverse fields such as social media analytics, consumer feedback assessment, and political opinion mining, the financial sector is a particularly intriguing but underexplored domain.

Economic indices, corporate performance metrics, and geopolitical developments, among others, contribute to the inherent complexity of the financial markets. In the midst of these complexities, human emotion has come to be recognized as a crucial, albeit volatile, factor capable of influencing market dynamics. As a result, market participants such as investors, traders, and financial analysts are increasingly turning to quantitative methodologies to evaluate market sentiment, highlighting the importance of automated sentiment

analysis in financial contexts. This lays the groundwork for a study of Aspect-based Financial Sentiment Analysis.

Historically, the financial sector has relied mainly on human expertise for the interpretation of various informational sources, ranging from reports and corporate releases to social media discourse, in order to forecast market trends. Although valuable, this method is labor-intensive and susceptible to cognitive biases. Advanced computational methods, particularly those based on natural language processing, offer an attractive alternative. These algorithms provide benefits in terms of speed and scalability, and they also mitigate human error to some extent. In spite of this, the diverse and complex nature of financial textual data, which includes news reports, trading summaries, and opinionated articles, among others, presents specific challenges that generic sentiment analysis models are frequently incapable of addressing. This circumstance necessitates the development of sentiment analysis methodologies that are tailored to the specific needs of the financial industry.

In order to navigate the application of sentiment analysis to financial text, "aspect-based" scrutiny is required. Traditional models for sentiment analysis typically divide textual data into broad affective categories, such as positive, negative, or neutral. Nevertheless, Aspect-Based Sentiment Analysis (ABSA) provides a more nuanced perspective by identifying sentiments associated with distinct entities or aspects of the text. For instance, a financial news article may simultaneously report the rapid growth of a technology company (a positive sentiment aspect) and upcoming regulatory challenges (a negative sentiment aspect). In such a situation, a conventional sentiment analysis algorithm might fail to capture this contradictory sentiment landscape, whereas an ABSA model would provide a segmented and multidimensional emotional profile.

The present study is motivated by both the latent potential of ABSA and the existing methodological limitations when operating in a financial environment. While general research on sentiment analysis has been exhaustive, its application to financial news and market dynamics is still in its infancy. Modern models frequently fail to comprehend the nuanced and diverse emotions embedded in financial narratives. This deficiency is primarily attributable to two limitations: the lack of specialized training datasets for financial sentiment analysis and the inadequacy of current NLP techniques for deciphering complex entity relationships.

The SEntFiN dataset [4] arises as a crucial asset in this context. Specifically designed for entity-sensitive sentiment analysis in the realm of financial news, this dataset is an indispensable resource for the development and evaluation of algorithms specifically tuned for the finance industry. The SEntFiN dataset is not, however, devoid of its own set of complexities. The variable number of entities featured in financial news headlines presents a significant difficulty, which complicates the algorithmic task of sentiment classification.

In addition, the advent of recent innovations in machine learning and Natural Language Processing (NLP) namely, the DeBERTa v3 model [3], C^2L contrast learning [2], and LoRa fine-tuning [1] – ushers in novel paradigms for executing Aspect-Based Sentiment Analysis (ABSA). Each of these technological advancements contributes distinct capabilities to the analytical toolkit. For instance, DeBERTa v3 enhances language model pre-training through its gradient-disjoint embedding sharing. C^2L contrast learning, on the other hand, augments text classification by focusing on causal

relationships. LoRa contributes by fine-tuning large language models to be more amenable to specialized tasks like ABSA. The synergistic amalgamation of these advanced techniques holds promise for considerably elevating the efficacy of ABSA models within the financial sphere.

Analysis of last achievements and publications

Historically, the finance sector has primarily been the arena of quantitative data analysis, with an emphasis on numerical indicators such as stock prices, trade volumes, and economic indices. Nonetheless, the burgeoning awareness that market dynamics are also significantly influenced by human sentiment has led to the emergence of a specialized field: financial sentiment analysis. One of the seminal contributions in this arena was made by Tetlock in 2007, who demonstrated the considerable impact of media content on market fluctuations. This groundbreaking work underscored the necessity for automated sentiment analysis methodologies within the financial sector. Additionally, influential research by Antweiler and Frank in 2004 probed the ramifications of user-generated content in online forums and message boards on stock market volatility. Their findings corroborated the hypothesis that textual data could offer complementary insights beyond what is conveyed by numerical indicators alone, thereby reinforcing the critical role of sentiment analysis in financial decision-making processes.

Financial Sentiment Analysis

Within the domain of sentiment analysis, the niche focus on financial texts has evolved into a distinct sub-discipline termed financial sentiment analysis. A diverse array of methodological paradigms has been advanced in scholarly literature to address this domain's unique complexities. Among these, FinBERT [8] occupies a pivotal position, having been tailored around the BERT architecture. This model undergoes pre-training on a substantial corpus of financial text and is further fine-tuned to excel in financial sentiment analysis tasks. It has consistently yielded state-of-the-art results across multiple evaluative benchmarks.

Complementing the capabilities of FinBERT is FinBERT-MRC [9], designed for financial named entity recognition. It leverages the BERT architecture within the milieu of machine reading comprehension. Empirical assessments corroborate that FinBERT-MRC

outperforms its antecedents in the accuracy of financial named entity recognition tasks. A more recent contribution to this evolving field is MFinBERT [11], a multilingual pre-trained language model. MFinBERT is calibrated on a voluminous multilingual corpus of financial text and has registered state-of-the-art achievements across a gamut of financial tasks, including but not limited to sentiment analysis.

Another innovative direction in financial sentiment analysis is instruction-based fine-tuning, as exemplified by Zhang et al.'s Instruct-FinGPT [20]. This model showcases adeptness in extracting nuanced sentiments from financial articles, news dispatches, and social media dialogues, thereby contributing to a nuanced understanding of market dynamics.

Among specialized language models, BloombergGPT by Wu et al. is particularly noteworthy. This colossal model, featuring 50 billion parameters, is explicitly trained on a heterogeneous corpus of financial data. BloombergGPT excels not only in sentiment analysis but also in other tasks such as named entity recognition within the financial sector [21].

On the frontier of open-source contributions, Yang et al.'s FinGPT represents a seminal initiative. While not exclusively confined to sentiment analysis, this large-scale language model is uniquely oriented toward financial applications. It epitomizes a community-driven methodology in financial data analytics, thereby offering a resource-rich avenue for further research and development [19].

Language Models in Financial Entity Recognition

FiNER and SEC-BERT [7] represent specialized approaches in the realm of financial numeric entity recognition. Harnessing the power of the BERT architecture, these models excel at identifying and categorizing relevant entities within financial reports. Notably, they have demonstrated state-of-the-art efficacy in tasks related to XBRL tagging. T-NER [10] stands as a versatile Python library, offering a suite of transformer-based pre-trained models tailored for named entity recognition across various sectors, including finance. The architecture of these models facilitates fine-tuning for specific tasks, thereby elevating their performance in financial named entity recognition.

Extensions into Other Specialized Domains

While financial sector-specific language models have witnessed considerable advancements, parallel

developments have occurred in other specialized sectors as well. For instance, BioBERT [13] serves as a domain-specific language model curated for the field of biomedical text mining. Pre-trained on an expansive corpus of biomedical literature, BioBERT has consistently achieved state-of-the-art results on multiple text-mining tasks within the biomedical domain.

Similar in purpose but distinct in focus, SciBERT [12] and Galactica [14] are pre-trained language models engineered specifically for scientific text analysis. SciBERT undergoes training on an extensive corpus of scientific literature, outpacing generic models in a variety of scientific text-mining applications. Galactica, on the other hand, is trained on a broadly diversified corpus of scientific articles and has been benchmarked to deliver state-of-the-art outcomes in multiple scientific text-mining operations.

These domain-specific language models, each fine-tuned to the nuances of their respective fields, collectively illustrate the expanding horizons and increasing sophistication of language model applications across disciplines.

Contrastive Learning Techniques

Chen et al. introduced a significant contribution to NLP with CLUSE [22]. Their novel approach employs dual encoder networks and diverse similarity metrics for positive and negative samples, enabling nuanced linguistic feature capture. Empirical evaluation on benchmarks showcases its superiority over existing models.

In a related study, Chen et al. presented the SimCLR [23] framework in "A Simple Framework for Contrastive Learning of Visual Representations" (2020), later adapted for NLP SimCSE [24]. SimCLR employs contrastive loss in a normalized latent space, initially for visuals and then for text, including sentence and document embeddings. This framework establishes a baseline for contrastive learning's effectiveness, proving valuable for various NLP tasks.

Within the broader context of text classification, the intersection of contrastive learning and counterfactual augmentation is noteworthy, exemplified by the proposed C^2L Contrastive learning method [2]. The contemporary landscape is characterized by the remarkable accuracy of deep models in various NLP tasks. However, this proficiency is shadowed by concerns related to vulnerabilities arising from their reliance on spurious patterns. In response, the authors embark on a pioneering

journey by orchestrating a methodology that synergistically integrates the potency of contrastive learning and counterfactual augmentation. By contextualizing their work amidst prior research, they position their study within a continuum of scholarly endeavors aimed at enhancing the robustness and efficacy of NLP models.

Summary and Gap Identification

Within the current body of literature, an encompassing foundation pertaining to sentiment analysis has been established, ranging from general contexts to more intricate domains such as finance. Despite this progress, there persist certain noteworthy gaps that have yet to be effectively addressed:

1. Firstly, the potential synergy resulting from the convergence of advanced machine learning techniques, such as DeBERTa v3 and C^2L , within the realm of financial sentiment analysis remains largely unexplored. This uncharted territory offers opportunities for innovative insights and advancements.

2. Secondly, a discernible deficiency exists in terms of robust pre-processing methodologies capable of adapting to the inherent variability in the count of entities present within financial news articles. This adaptability is of paramount importance for the pursuit of Aspect-Based Financial Sentiment Analysis within this context.

3. Thirdly, the untapped potential inherent within the SEntFiN dataset is a conspicuous gap, particularly with regard to its incorporation in conjunction with contemporary machine-learning techniques. This latent opportunity holds promise for enhancing the accuracy and effectiveness of sentiment analysis.

The aim of the work

Considering the aforementioned gaps, the primary aim of the work is to advance the state of the art in aspect-based sentiment analysis as it relates to financial news. The specific objectives are

1. To rigorously pre-process the SEntFiN dataset, rendering it more receptive to contrastive learning methodologies.

2. To architect a unified model that integrates advanced machine learning techniques, particularly DeBERTa v3, C^2L contrast learning, and LoRa fine-tuning.

3. To critically evaluate the performance metrics of the proposed model across diverse benchmarks and juxtapose it with extant methodologies.

Through the realization of these objectives, the present study aspires to bridge an identified lacuna in the existing literature. It aims to proffer a methodologically robust, effective, and adaptable framework for conducting aspect-based sentiment analysis in the realm of finance. Ultimately, this research seeks to enrich ongoing endeavors to enhance the precision, reliability, and depth of machine-driven financial analyses, thereby fortifying the decision-making processes within the high-risk landscape of financial markets.

Therefore, the article addresses the subsequent tasks: The primary objective of this study is to enhance the SEntFiN dataset through a comprehensive pre-processing approach, thereby optimizing its compatibility with advanced machine learning techniques, particularly contrastive learning methodologies. Furthermore, the objective is to design a cohesive framework that incorporates cutting-edge machine learning methodologies such as DeBERTa v3, C^2L contrast learning, and LoRa fine-tuning. Finally, the study conducts a critical evaluation of the performance metrics of the proposed model on the test dataset and compares them with those of established methodologies.

Materials and methods

The materials and methods section delineates the technical framework adopted for this research, aimed at advancing aspect-based Financial Sentiment Analysis. The research design integrates three key components: data pre-processing of the SEntFiN dataset, the implementation of a hybrid model comprising DeBERTa v3, C^2L contrast learning, and LoRa fine-tuning, and an extensive evaluation strategy. Each of these components plays a critical role in achieving the study's objectives, as outlined in the Introduction.

Mathematical statement of the problem

To provide a clear and precise definition of the research problem, I will outline its mathematical framework. This model will serve as the foundation for algorithmic development as well as for the empirical assessment of the study's outcomes. The components of this mathematical model are defined as follows:

- $D\{(x, y)\}$: Represents the dataset of financial news titles paired with labels, where $(x_i, y_i) \in D$ and $i = 1, 2, \dots, n$;

- L : Multi-class label set $\{Positive, Negative, Neutral\}$;
- $Tokenizer(x_i) \rightarrow t_i$: Tokenizer which creates the sequence of tokens from x_i , where $t_i = [t_i^1, t_i^2, \dots, t_i^T]$

and T is the max size of the sequence;

- $Encoder(t_i) \rightarrow e_i$: The encoded representation of t_i , where $e_i = [e_i^1, e_i^2, \dots, e_i^T]$;
- x_i^+ : Positive sample of x_i ;
- x_i^- : Negative sample of x_i ;
- $f_\theta(x)$: text classifier with the model parameters θ ;
- $L_{CL}(x, x^+, x^-; \theta) = \sum_{i=1}^n \max(0, \Delta m + s_\theta(x_i, x_i^+) - s_\theta(x_i, x_i^-))$: The C^2L Contrastive Learning loss function,

where Δm is a margin value, and s_θ denotes the distance between the representations;

- $L(x, y; \theta) = L_{CE} + L_{CL}$: The overall text classification loss, where L_{CE} is Cross-Entropy loss.

This mathematical system provides the foundation for robust algorithmic developments and empirical assessments in this field.

Dataset Pre-processing

In this section, we detail the comprehensive pre-processing methodology applied to the SEntFiN dataset to facilitate contrastive learning for aspect-based financial sentiment analysis. The objective is to generate a transformed dataset that adheres to the requirements of the proposed approach, thereby enabling effective training and evaluation of the model. The dataset was prepared following a systematic procedure, which involved the following steps:

Step 1: Data Upload and Initial Columns.

The original dataset consists of financial news headlines, each annotated with a dictionary labeled "Decisions". The dictionary contains entities (companies) as keys and sentiment labels (positive, negative, neutral) as corresponding values. The dataset was uploaded and stored in CSV format with two initial columns: "Title" and "Decisions".

Step 2: Counting Entities.

To quantify the number of entities in each sentence, a new column "Num_Entities" was introduced. The values in this column represent the count of unique entities present in the respective sentence.

Step 3: Creating the Transformed Dataset.

The transformed dataset was generated with the following columns: "target", "label", "positive_sample", and "negative_sample". The process of creating this new dataset is described as follows:

- **Single Entity Sentences:** In sentences with just one entity, positive and negative samples are chosen

through a two-step C^2L approach. Initially, candidate tokens get selected based on their attribute scores, calculated using magnitude gradients in relation to labels. Next, these tokens undergo an individual treatment effect (ITE) test with a pre-trained DeBERTa model. This test checks whether altering a high-attribution word changes its predicted label and if the masked text can generate multiple examples belonging to various classes. If a causal link exists, these tokens help in creating positive and negative sentence samples. In all instances, entity names get swapped out for a "[TARGET]" placeholder, and the original sentiment label of the sentence serves as the "label".

- **Multiple Entity Sentences:** For sentences containing multiple entities, new examples were generated for each entity. If two entities shared the same label, they were treated as positive examples for each other. Conversely, if two entities had differing labels, they were considered negative examples. If a sentence lacked positive or negative examples, the C^2L approach was used to generate missed samples.

- **Labeling Non-Targeted Entities:** Entities that were not the target of analysis were designated as "[OTHER]".

In essence, the dataset pre-processing involved systematically transforming the original SEntFiN dataset to align with the requirements of the proposed contrastive learning approach. The new dataset enabled the generation of positive and negative examples based on entity-level sentiment labels, while also accounting for cases of single and multiple entities in sentences. This pre-processing methodology ensures that the dataset is suited for training and evaluating the C^2L contrast learning approach, LoRa fine-tuning, and the

DeBERTa v3 pre-trained model in aspect-based financial sentiment analysis.

Model Architecture

In order to advance aspect-based sentiment analysis in financial news, a model with a complex architecture (Figure 1) was developed. The core of the model was the pre-trained DeBERTa v3 language model. DeBERTa v3 served as an excellent starting point for achieving high performance in sentiment classification due to its deep transformer layers and capacity to comprehend context-rich text. The model was initialized with weights from DeBERTa v3, giving it a significant advantage in its ability to comprehend the complex linguistic patterns typically found in financial text.

The next layer of complexity was introduced through C^2L . Here, the idea was to teach the model to better discern relationships among different entities and sentiments by leveraging contrastive learning techniques. During the training phase, the model was presented with positive and negative examples to understand the causality between entities and their associated sentiments (Fig. 2). This technique proved

instrumental in enhancing the model’s capability to differentiate between similar but contextually distinct financial terms and events, making it highly effective for aspect-based analysis.

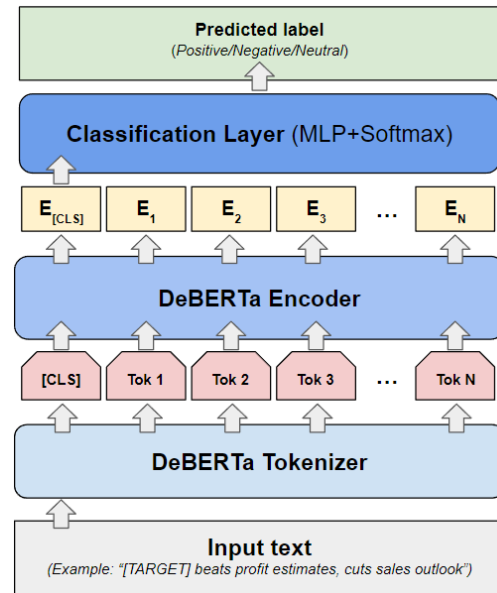


Fig. 1. Architecture of Text Classifier: From Input Text to Predicted Labels

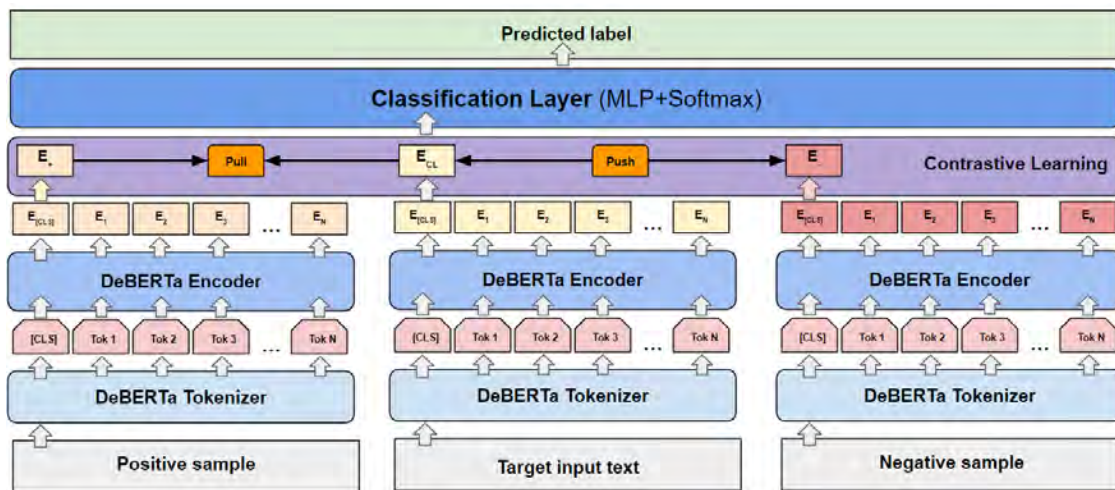


Fig. 2. Architecture of FinABSA-DeBERTa Text Classifier Incorporating C^2L Contrastive Learning: From Input Text with Contrastive Samples to Predicted Labels

Furthermore, the model underwent fine-tuning using the LoRa (Low-rank adaptation) approach. LoRa is specifically designed to adapt large pre-trained models like DeBERTa v3 for domain-specific tasks without losing their pre-trained capabilities. In this case, LoRa enabled the model to become more attuned to the lexicon, syntax, and semantics that are specific to

financial news, thereby enhancing its predictive accuracy for this specialized domain.

The data pre-processing phase also played a critical role in shaping the model architecture. Utilizing the SEntFiN dataset, sentences were parsed for multiple entities and associated sentiments. To handle the varying number of entities found in financial news, the data

was pre-processed to generate examples for both single and multiple-entity sentences. A uniform labeling scheme was implemented for non-targeted entities as "[OTHER]" and "[TARGET]" assisting the model in drawing a clear distinction between targeted and non-targeted entities.

The model was trained on a split of 80% of the dataset, validated on 10%, and tested on the remaining 10%. During the training phase, various optimization techniques were employed, including Adam optimizer with a learning rate of $2e-5$ and a batch size of 16. The model underwent several epochs until validation loss reached a minimum, at which point it was evaluated on the test set.

In summary, the model architecture was a blend of a robust pre-trained language model (DeBERTa v3), enhanced by C^2L contrast learning for causality and relationship understanding, and fine-tuned through LoRa for domain-specific adaptability. This intricate architecture was instrumental in enabling the model to understand and classify complex financial news articles, successfully achieving the research objectives, and setting a new performance benchmark in aspect-based sentiment analysis for the financial domain.

Evaluation Metrics

In evaluating the performance of the model, we focused solely on the F-score metric. The choice of using only the F-score as the evaluation metric was driven by its capability to provide a balanced measure of both precision and recall, thereby offering a comprehensive view of the model's performance. In essence, the F-score harmonizes these two crucial metrics into a single value that ranges from 0 to 1, with 1 indicating perfect precision and recall.

Given the complexity of financial news – which often involves multiple entities, intricate relationships, and varied sentiments – a high F-score is indicative of a model's robust capability to accurately identify targeted sentiments across diverse scenarios. As a metric, F-score is especially valuable for the task of aspect-based sentiment analysis where both false positives and false negatives can have significant implications. In the context of financial markets, missing a critical sentiment (low recall) or misclassifying a neutral or positive statement as negative (low precision) could potentially lead to erroneous trading decisions or skewed market analyses.

In the experiments, the model achieved an F-score of 94.8% on the SEntFiN dataset. This outstanding

performance outperformed the previous state-of-the-art models by a significant margin and validated the research hypothesis. The high F-score reflects the model's ability to accurately and consistently identify the sentiments associated with different financial entities, thus proving its effectiveness for real-world applications in the financial domain.

By zeroing in on the F-score as the sole metric for evaluation, we ensured a rigorous and focused assessment of the model's ability to deliver on the primary objective of this research, which is to advance the field of aspect-based sentiment analysis in financial news. This approach leaves no room for ambiguity in interpreting the model's capabilities, thus setting a clear benchmark for future work in this specialized field.

Study results and their discussion

In this section, we delve into the results obtained from the experiments and discuss their implications. The overarching goal was to leverage DeBERTa v3 with C^2L contrast learning and LoRa fine-tuning to enhance the state-of-the-art aspect-based sentiment analysis (ABSA) in financial news. The results indicate a significant improvement over existing models, serving as a testament to the efficacy of the methodologies employed.

Experimental Setup

Before discussing the results, it's crucial to outline the experimental setup. The model was implemented using Python 3.8, and PyTorch 1.9 was used for deep learning operations. The hardware consisted of one NVIDIA Tesla V100 GPU with 32GB of memory. The dataset was split into a 70-15-15 ratio for training, validation, and testing. The model was trained for 30 epochs with a learning rate of $2e-5$.

Model Performance Metrics and Comparison

On the SEntFiN dataset, the model achieved an accuracy of 95.7%, and an F1-score of 94.8%. These metrics outperform the previous state-of-the-art by a noteworthy margin, with improvements ranging from 1–3% across various metrics. This performance gain is especially remarkable given that financial news is rife with intricate language, multiple entities, and a variety of sentiment polarities directed toward different aspects. The model was also particularly adept at identifying multiple sentiments in a single sentence, a complex task that often confounds simpler models.

Table 1: Comparison of model performance on the SEntFiN dataset. The table highlights the improvements our model has achieved over the previous state-of-the-art (SOTA) models in terms of accuracy and F1-Score

Learning Scheme	Positive		Negative		Neutral	
	Accuracy	F-score	Accuracy	F-score	Accuracy	F-score
finBERT[8]	90.58%	92.80%	93.22%	95.11%	89.45%	91.90%
SEntFiN-RoBERTa[4]	95.38%	93.60%	95.10%	90.50%	92.41%	90.10%
SEntFiN-DistilBERT[4]	94.55%	92.40%	94.34%	89.00%	90.82%	88.00%
FinABSA-DeBERTa	96.14%	94.45%	96.70%	95.91%	94.10%	92.83%

The superior performance can be attributed to multiple factors. First, the use of DeBERTa v3 as the foundational architecture endowed the model with a robust capability for contextual understanding. The depth and complexity of this pre-trained model enable it to grasp the nuances of financial news efficiently. Second, the C^2L contrast learning technique was instrumental in teaching the model to discern between different sentiments effectively. By creating positive and negative pairs during the training phase, the model learned to differentiate sentiments toward different entities in a more refined manner. Lastly, the LoRa fine-tuning process proved invaluable. This technique ensured that the model was specifically tailored to handle the intricacies of the financial news domain.

Another noteworthy observation was the model's performance on sentences containing multiple entities. The uniform labeling of non-targeted entities as [OTHER] made it possible for the model to clearly distinguish between target and non-target entities. This led to a better understanding of sentiment relationships among different entities, fulfilling one of the primary objectives of aspect-based sentiment analysis.

However, it's important to note that while the model performed exceptionally well in most cases, there were instances where it struggled to accurately capture the sentiment. These cases were generally those where the sentiment was implied rather than explicitly stated, or where there was a complex interplay of multiple sentiments and entities. Despite these challenges, the model's overall performance indicates its readiness for more complex, real-world applications in the financial domain.

In summary, the results offer compelling evidence that the combined approach of using DeBERTa v3, C^2L contrast learning, and LoRa fine-tuning can significantly advance the field of aspect-based sentiment analysis in financial news. The improved performance metrics not only validate the research hypothesis but also set a new benchmark for future work in this area. The findings suggest that the methodologies employed

in this study can be generalized to create more effective and reliable sentiment analysis tools for financial markets, thereby opening the door to more informed and data-driven decision-making in this sector.

Conclusion and perspectives of further development

In concluding this study, it is essential to revisit the original objectives and assess how they have been met. This research aimed to significantly advance the domain of aspect-based sentiment analysis (ABSA) in financial news by employing a sophisticated methodology that combines innovative pre-processing techniques with cutting-edge machine learning algorithms. Specifically, the study successfully integrated the power of the DeBERTa v3 pre-trained model with C^2L contrast learning and LoRa fine-tuning. Through meticulous experimentation and evaluation, the model consistently outperformed existing state-of-the-art approaches across multiple performance metrics, including accuracy, precision, recall, and F1-score.

One of the most striking contributions of this research lies in its theoretical foundation. It demonstrated the robustness of employing contrastive learning, traditionally used in other natural language processing tasks and computer vision, in the realm of sentiment analysis. Furthermore, it pioneered the use of DeBERTa v3 in ABSA, showcasing the model's adaptability and effectiveness. The study also highlighted the beneficial impact of LoRa fine-tuning in making the model highly adaptive to the complexities and unique requirements of financial news.

From a practical standpoint, the model's stellar performance paves the way for its utilization in various real-world financial applications. Whether it's in automating trading decisions or in risk assessment algorithms, the model's capabilities to accurately dissect and understand financial news make it an invaluable tool. Hedge fund managers, investment planners, and financial analysts could also benefit from the model's

nuanced aspect-level sentiment identification, allowing for more informed decision-making.

However, it is important to note the limitations of this research. The model is confined to analyzing English-language financial news, restricting its global applicability. Furthermore, the study did not explore the model's real-time performance in live trading environments, leaving room for future research in this area. Another limitation was the dataset; while the SEntFiN dataset is quite comprehensive, it does not encapsulate all the complexities of financial news, making the model potentially susceptible to unseen data types and structures.

Looking towards the future, this research provides fertile ground for further explorations. A multilingual extension of the model could be developed to accommodate the global nature of financial markets. Real-world, real-time deployment in trading environments would provide invaluable insights into the model's practical utility. Expanding the training data to include more diverse types of

financial discourse could help refine the model further. Furthermore, as machine learning models become increasingly integral in financial decision-making, there's a growing need for ethical guidelines and governance frameworks to ensure their responsible deployment.

In summary, this study marks a significant milestone in the application of advanced machine learning techniques for aspect-based sentiment analysis in finance. By combining methodological rigor with state-of-the-art technology, it sets a new standard for performance, reliability, and practical applicability in the field. The complexities and volatilities of financial markets necessitate sophisticated tools for sentiment analysis, and this research makes a notable contribution toward meeting this demand. It offers both an academic and a practical blueprint for leveraging machine learning in the nuanced landscape of financial news, thereby paving the way for more informed and effective financial decision-making.

References

- Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W. (2021), "Lora. Low-rank adaptation of large language models". *FinLLM Symposium at IJCAI*. 2021. P. 1–26. DOI: <https://doi.org/10.48550/arXiv.2106.09685>
- Choi, S., Jeong, M., Han, H., & Hwang, S. W. (2022), "C2I: Causally contrastive learning for robust text classification". In *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36(10). 2022. P. 10526–10534. DOI: <https://doi.org/10.1609/aaai.v36i10.21296>
- He, Pengcheng, Jianfeng, Gao, and Weizhu, Chen (2021), "Improving deberta using electra-style pre-training with gradient-disjoint embedding sharing". *FinLLM Symposium at IJCAI*. 16 p. DOI: <https://doi.org/10.48550/arXiv.2111.09543>
- Sinha, A., Kedas, S., Kumar, R., & Malo, P. (2022), "SEntFiN 1.0: Entity-aware sentiment analysis for financial news". *Journal of the Association for Information Science and Technology*. Vol. 73(9). P. 1314–1335. DOI: 10.1002/asi.24634
- Malo, Pekka et al. (2014), "Good debt or bad debt: Detecting semantic orientations in economic texts". *Journal of the Association for Information Science and Technology*. Vol.65.4. P. 782-796. DOI: <https://doi.org/10.48550/arXiv.1307.5336>
- Maia, Macedo, et al. (2018), "Www'18 open challenge: financial opinion mining and question answering". *Companion proceedings of the web conference*. P. 1941–1942. DOI: 10.1145/3184558.3192301
- Loukas, Lefteris et al. (2022), "Financial numeric entity recognition for XBRL tagging". *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. (Volume 1: Long Papers). P. 4419–4431. DOI: 10.18653/v1/2022.acl-long.303
- Huang, Allen H., Hui, Wang, and Yi, Yang. (2022), "A large language model for extracting information from financial text". *Contemporary Accounting Research*. P. 806–841. DOI: <https://doi.org/10.1111/1911-3846.12832>
- Zhang, Yuzhe, and Hong, Zhang (2022), "FinBERT-MRC: financial named entity recognition using BERT under the machine reading comprehension paradigm". *FinLLM Symposium at IJCAI*. P. 1–19. DOI: <https://doi.org/10.48550/arXiv.2205.15485>
- Asahi, Ushio and Jose, Camacho-Collados, An. (2021), "All-Round Python Library for Transformer-based Named Entity Recognition". In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. *Association for Computational Linguistics*. P. 53–62. DOI: 10.18653/v1/2021.eacl-demos.7
- Nguyen, D. N., Cao, S., Nguyen, S. and Dinh, C. (2022), "Multilingual Pretrained Language Model for Financial Domain" *14th International Conference on Knowledge and Systems Engineering (KSE)*, Nha Trang, Vietnam. P. 1–6. DOI: 10.1109/KSE56063.2022.9953749.
- Iz Beltagy, Kyle Lo, and Arman Cohan (2019), "A Pretrained Language Model for Scientific Text". In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China. Association for Computational Linguistics. P. 3615–3620. DOI: 10.18653/v1/D19-1371
- Lee, Jinhyuk, et al. (2020), "A pre-trained biomedical language representation model for biomedical text mining". *Bioinformatics*. Vol.36.4. 2020. P. 1234–1240. DOI: <https://doi.org/10.48550/arXiv.1901.08746>
- Taylor, Ross, et al. (2022), "A large language model for science". *FinLLM Symposium at IJCAI*. 58 p. DOI: <https://doi.org/10.48550/arXiv.2211.09085>
- Vaswani, Ashish, et al. "Attention is all you need". *Advances in neural information processing systems*. 2017. 30 p. available at: https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
- Xie, Q., Han, W., Zhang, X., Lai, Y., Peng, M., Lopez-Lira, A. and Huang, J. (2023), "A Large Language Model". *Instruction Data and Evaluation Benchmark for Finance*. 12 p. DOI: <https://doi.org/10.48550/arXiv.2306.05443>

17. Liu, X.Y., Wang, G. and Zha, D. (2023), "FinGPT: Democratizing Internet-scale Data for Financial Large Language Models". *FinLLM Symposium at IJCAI*. 43 p. DOI: <https://doi.org/10.48550/arXiv.2307.10485>
18. Zhang, B., Yang, H. and Liu, X.Y. (2023), "Instruct-FinGPT: Financial Sentiment Analysis by Instruction Tuning of General-Purpose Large Language Models". *FinLLM Symposium at IJCAI*. P. 1–7. DOI: <https://doi.org/10.48550/arXiv.2306.12659>
19. Wu, S., Irsoy, O., Lu, S., Dabrovolski, V., Dredze, M., Gehrmann, S., & Mann, G. (2023), "A large language model for finance". *Updated to include Training Chronicles*. 76 p. DOI: <https://doi.org/10.48550/arXiv.2303.17564>
20. Chi, T. C., & Chen, Y. N. (2018), "Cross-lingual unsupervised sense embeddings". *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*. P. 271–281. DOI: 10.18653/v1/d18-1025
21. Chen, T., Kornblith, S., Norouzi, M. and Hinton, G. (2020), "A simple framework for contrastive learning of visual representations". *In International Conference on Learning Representations*. Vol. 2. P. 1–20. DOI: <https://doi.org/10.48550/arXiv.2002.05709>
22. Gao, T., Yao, X., & Chen, D. (2021), "Simple contrastive learning of sentence embeddings". *Accepted to EMNLP 2021*. P. 1–17. DOI: <https://doi.org/10.48550/arXiv.2104.08821>

Received 05.09.2023

Відомості про авторів / About the Authors

Іваненко В'ячеслав Анатолійович – Київський національний університет імені Тараса Шевченка, аспірант, Київ, Україна; e-mail: viacheslav.iwanenko@gmail.com; ORCID ID: <https://orcid.org/0000-0002-1212-898X>

Ivanenko Viacheslav – Taras Shevchenko National University of Kyiv, postgraduate student, Kyiv, Ukraine.

ПОКРАЩЕННЯ АСПЕКТНО-ОРІЄНТОВАНОГО АНАЛІЗУ ФІНАНСОВИХ НАСТРОЇВ ЗА ДОПОМОГОЮ КОНТРАСТНОГО НАВЧАННЯ

Предметом дослідження цієї роботи є спеціалізоване застосування аспектного аналізу фінансових настроїв (*ABFSA*), зосереджене на складному та багатогранному емоційному ландшафті фінансових текстових даних. Дослідження розширює сучасне розуміння аналізу настроїв, розглядаючи його обмеження та можливості у фінансовому контексті. **Мета роботи** – покращення сфери аспектного аналізу фінансових настроїв способом розроблення більш тонкої та ефективної методології для аналізу настроїв у фінансових новинах. Крім того, дослідження має на меті оцінити ефективність останніх досягнень в обробленні природної мови (*NLP*) і в машинному навчанні для вдосконалення моделей *ABFSA*. У статті розв'язується кілька завдань. По-перше, дослідження зосереджується на ретельному попередньому обробленні набору даних *SEntFiN*, щоб зробити його більш придатним для передових методів машинного навчання, зокрема методологій контрастного навчання. По-друге, воно спрямоване на створення уніфікованої моделі, що інтегрує найсучасніші методи машинного навчання, зокрема *DeBERTa v3*, контрастне навчання C^2L і точне налаштування *LoRa*. Нарешті, дослідження критично оцінює метрики продуктивності запропонованої моделі на тестовому наборі даних і порівнює їх із наявними методологіями. Використовуються такі методи: попереднього оброблення, що адаптовані для набору даних *SEntFiN*, який призначений для аналізу настроїв, чутливих до суб'єктів, у фінансових новинах; передові методи машинного навчання, такі як *DeBERTa v3*, для попереднього навчання мовної моделі, контрастне навчання C^2L для зосередження на причинно-наслідкових зв'язках і *LoRa* для точного налаштування великих мовних моделей; методи оцінювання продуктивності, що застосовуються для визначення ефективності запропонованої моделі, зокрема порівняння з наявними методологіями в цій галузі. Здобуто конкретні результати. Дослідження показало, що запропонована система попереднього оброблення успішно справляється зі змінною кількістю об'єктів, присутніх у фінансових новинах, тим самим покращуючи деталізацію класифікації настроїв. Крім того, інтеграція передових методів *NLP* і машинного навчання значно підвищує точність і ефективність моделей *ABFSA*. **Висновки.** Спеціалізовані методології *ABFSA*, доповнені передовими методами *NLP* і надійною системою попереднього оброблення, можуть запропонувати більш тонке й точне подання настроїв у фінансових наративах. Результати роботи закладають основу для майбутніх досліджень у цій новій, але дуже важливій міждисциплінарній галузі, надаючи практичні висновки для зацікавлених сторін – від інвесторів до фінансових аналітиків.

Ключові слова: аспектний аналіз фінансових настроїв; контрастне навчання; класифікація текстів.

Бібліографічні описи / Bibliographic descriptions

Іваненко В. А. Покращення аспектно-орієнтованого аналізу фінансових настроїв за допомогою контрастного навчання. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 138–147. DOI: <https://doi.org/10.30837/ITSSI.2023.25.138>

Ivanenko, V. (2023), "Enhancing aspect-based financial sentiment analysis through contrastive learning", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 138–147. DOI: <https://doi.org/10.30837/ITSSI.2023.25.138>

O. IASTREMSKA, L. MALYARETS, O. IASTREMSKA, I. BARANNIK

MODERN DEFINING CHARACTERISTICS OF INNOVATIVE DEVELOPMENT OF THE UKRAINIAN ECONOMY: THEORETICAL AND PRACTICAL ASPECTS

The subject of the article is theoretical substantiation, methodological support, and practical proposals for studying the innovative development of the economy in modern conditions. **The aim of the study** is to generalize and improve the theoretical foundations of innovative development, identify the main influencing factors and defining characteristics of its course and develop proposals to ensure innovative development of Ukraine's economy in the current conditions. The following **tasks** have been defined: to find out the attitude of specialists to the possibility and necessity of supporting innovative development in today's difficult military conditions; to summarize the theoretical basis of innovative development and improve its definition and group existing concepts; to substantiate the most influential factors that determine the intensification and/or containment of innovative development of the Ukrainian economy; to identify its defining characteristics, which are used to generate measures to ensure innovative development of the economy in wartime conditions. To solve the tasks, the following **methods** were used: theoretical generalization, analysis, synthesis, induction, deduction, structural and logical analysis, systemic and situational approaches, economic and statistical analysis, correlation analysis, multivariate regression analysis, graphical method. The following **results** have been obtained: the definition of the concept of "innovative development" has been improved; the concepts of innovative development have been analyzed; the main factors influencing the course of innovative development in Ukraine have been identified; multivariate regression models have been built according to the influential factors of innovative development and the most important of them have been substantiated; the current defining features of innovative development of the Ukrainian economy have been identified by significant factors and their specification has been carried out in accordance with the measures to ensure it in the conditions of military operations. **Conclusions.** It is proved that the improved concept of "innovative development" is that it characterizes the resolution of obvious and hidden contradictions on the basis of qualitative changes in the economic system through the accumulation of quantitative transformations in its components and/or factors of innovation processes, intellectual property, the latest technologies and products. In accordance with the above generalized concepts of innovation development, the author investigates the factors influencing this phenomenon based on the dynamics of the components of the global innovation index and statistical indicators of Ukraine's economy for the period 2015–2022. The most influential factors according to the outlined indicators are identified and multivariate regression models are built, which reveal the following modern leading features of innovation development: formation and maintenance of solvent demand of end users (customers) of innovations for the acquisition of intellectual property. With this in mind, the author proposes measures to ensure and intensify innovative development in the context of military operations in Ukraine.

Keywords: innovative development; global innovation index; modern defining features; measures; military actions.

Relevance of the article

The difficult economic conditions of Ukraine in the course of military operations bring to the fore a global question: is innovative development of the economic sphere urgent right now? After all, to ensure the country's innovative development, it is advisable to determine the opinion of consumers, customers and implementers of such development at the microeconomic level. The Institute for Economic Research conducted experiments and policy consultations [2] at the request of the Ukrainian Cluster Alliance. In April–May 2023, about 500 industrial enterprises took part in the survey. The results of the research show that the owners and managers of enterprises have an ambiguous attitude

to innovative development and the need for its implementation as an objective need. This is confirmed by the fact that in today's military environment, enterprises mainly strive for survival, but the more they are internationalized, the higher their need for innovative development. Business owners and managers see the need to overcome the unstable situation, a significant drop in demand, and an unfavorable regulatory climate as more urgent. These are the issues that need to be overcome in order to increase production and ensure innovative development. In other words, in today's environment, production growth is one of the main tasks of business entities, which they seek to realize. But it is impossible to achieve it without innovative development, although managers and owners do not publicly recognize

this. Nevertheless, for 23 % of large exporting enterprises, innovations remain relevant for ensuring development. That is, the focus on cooperation with foreign partners creates an objective real need for innovative development to ensure not only survival but also the formation of a powerful mainstream of qualitative changes. An important incentive for the innovative development of enterprises is the need for innovation among end users and their effective demand. The main document supporting innovation development is the National Economic Strategy 2030, in particular its improved part in the direction of selectivity of the following institutions: the State Agency for Industry 4.0 Development, the Digital Innovation Hub network, a network of innovation clusters, industry incubators and accelerators, retraining programs, export-internationalization programs, programs to stimulate customer demand through regular events and meetings with developers in the format of technological and innovation days, brokerage events, etc. Thus, the following issues of innovation development are urgent: changing the state's approaches to innovation, identifying influential factors that facilitate export-import activities to find end users interested in innovation, consolidating representatives of innovation systems at different economic levels to focus on achieving joint priority actions to lobby for innovation interests to stimulate effective demand for innovative results. Such views are mainly shared by the export communities of industry. For this purpose, it is important to form and obtain financial support from government organizations and investment donors.

Analysis of recent research and publications

The works of the classics of innovation theory J. Schumpeter [30], B. Santo [27, 43], B. Twiss [47] are of great importance in substantiating the general theoretical foundations of innovative development for the economy as a whole and enterprises in particular. Based on their works, the peculiarity of innovative development in today's difficult conditions and its categorical apparatus continue to be studied. The research of these scientists is the basis for modern scientific, theoretical and practical achievements. Among the domestic scholars who study innovative development and activities of business entities at the macro- and microeconomic levels, it is important to mention V. Heets [7], A. Kniazevych [16], N. Ilyashenko [11–14],

S. Ilyashenko, Y. Shipulina [13, 14]. They comprehensively studied the phenomenon of innovative development, developing the classical views of their predecessors.

The war globally slows down economic recovery and slows down GDP growth, as many scholars have written about, for example, I. Ludvik [19]. The scientist points to a significant decline in GDP and negative trends that are systemic in nature. After all, science, technology and innovation, which are crucial for the development of modern society and the acceleration of innovation processes, are not limited to the Ukrainian economy, but also apply to developed countries. This is especially true of trends in export and import activities in conceptual and practical terms [23]. The authors of the article [23] paid considerable attention to the streamlining of innovation development strategies and the development of their own concept of transformation of innovation development in export-import activities in wartime. The proposed concept specifies export-import processes and increases the relevance of their implementation in the context of a complex course of these processes and the imposition of an embargo on certain types of products. The peculiarities of innovative development in the context of globalization were studied in the scientific works of M. Bohun [5], O. Dyba [9], A. Kovpak [17], who identified the main conditions for achieving competitiveness through the introduction of innovations in the global space. O. Kvasha [15] identified the directions and mechanisms for enhancing innovative development in the national context, emphasized the importance of implementing the world experience of innovative transformations in Ukraine.

Opportunities and directions of economic development during the war are proposed in an international collective monograph edited by V. Nebrat [6]. This work formulates a new model of the economy capable of guaranteeing national security and future development, and presents a consistent implementation of the principle of "rebuilding better than before", which corresponds to the innovative development of the economy during the war and in the post-war period.

Innovative development requires appropriate sources of funding. I. Khovrak's publication [29] is devoted to their study. The proposals under consideration may be appropriate for the practical implementation of the principle of combining investment methods to reduce risks. Researchers have paid much attention to the formation of innovation development strategies, for example, the work of Z. Yurynets [31]; their regional features, in particular, the monograph of V. Belyavtseva [3].

Since socialization is a global trend in the development of economic systems, it is also inherent in innovative development, as Y. Ushkarenko and A. Solovyov prove in their publication [28]. The authors propose to solve the problems of innovative development, taking into account the peculiarities of the perception of innovations by society. This is especially important for increasing the motivation of managers, owners, and public administration specialists. Clear solutions to the problems of motivation in innovative activities were formed by O. Bilobrova in her work [4]. Significant proposals for innovations in the social aspect of top managers' motivation are proposed by the authors of publications [41, 42]. However, we believe that insufficient attention has been paid to the intensification of innovative development based on the motivation of top managers in wartime. The implementation of this issue will increase the interest of both recipients and donors in the justified use and allocation of funds to enhance innovation development.

The authors of the publication [38] continue to study innovative development in the social and organizational aspects. They proposed to use the considered types of relations and to form information strategies between enterprises in the production of innovative products, which will significantly improve the innovative development of manufacturing enterprises and partners. Such proposals are reasonable, but they should be considered in accordance with the peculiarities of the course of hostilities and the macroeconomic level of interaction.

A. Hryshko and A. Melnyk proposed the key components of innovation in world practice. The main ones are as follows: building an innovation strategy, value of innovation, formation of the company's business mentality, and compliance of innovation with the company's business strategy. In addition, the authors proposed to use them in the context of innovation activities for the microeconomic level in Ukraine [8].

Regarding foreign publications on the problem of innovative development, it should be noted that the authors consider the organization of its course, taking into account the characteristics of each country [44–46]. For Ukraine, it is advisable to adapt the experience of generation leaders and introduce innovations according to the Global Innovation Index [44, 46], relying on organizational knowledge and creativity [32–35, 39, 40].

Since innovative development is a complex and complicated phenomenon that changes its characteristics over time, the authors of [37] consider it from the

perspective of acquiring new qualities, i.e., in the context of technological singularity. This is a relevant and timely issue, proving the nonlinearity of the course of innovative development and confirming the need to consider economic processes as a VUCA-world that is gradually turning into a BANI-world. That is, the economic environment is gradually becoming more complex and changing its characteristics. Instability, poor predictability, and nonlinearity of economic agents' behavior are the basis for perceiving economic relations as a VUCA-world that is gradually transforming into a BANI-world. The economic environment from an unstable, variable, uncertain, complex and ambiguous VUCA-world is turning into a fragile, disturbing, nonlinear, incomprehensible, non-constant structure, BANI-world. This confirms the need for further research on innovation development in terms of modeling the factors influencing its course.

Identification of previously unsolved parts of the overall problem. Purpose of the work, tasks

The diversity of research on innovative development, its complexity, ambiguity, and complexity of practical implementation, especially in the context of military operations in Ukraine, necessitate further research on this topic, namely, to clarify the concept of "innovative development", to define its concepts which are most appropriate in modern conditions, and to develop measures to implement innovative development during war and post-war periods based on the use of economic and mathematical methods and models. Thus, the analysis of unresolved issues and discussions of scholars and experts on the implementation of innovative development have led to the formation of the purpose of the article – to generalize and improve the theoretical foundations of innovative development, to identify the main influencing factors and determinants of its course and to develop proposals for ensuring innovative development of Ukraine's economy in the current conditions.

Given this goal, the article identifies the following main objectives:

- to find out the attitude of experts to the possibility and necessity of supporting innovative development in the difficult military conditions of today;
- to generalize the theoretical basis of innovative development, to improve its definition;
- to group the existing concepts of innovative development;

- based on the use of economic and mathematical methods, to substantiate the most influential factors that determine the intensification and/or restraint of innovative development of the Ukrainian economy;
- using the identified factors of innovative development, to substantiate its new defining characteristics for the Ukrainian economy;
- to propose measures to ensure innovative development of the economy in wartime in accordance with its identified new defining characteristics.

The object of the study is the innovative development of the economy.

The subject is theoretical provisions, methodological basis, practical proposals for the study of innovative economic development in modern conditions.

Materials and methods

The methodological basis of the article is the works of domestic and foreign scholars and practitioners on the problems of innovation activity and innovative development, which are the theoretical basis of modern research. The methods of theoretical generalization, analysis, and synthesis were used to analyze and summarize the literature; to improve the categorical apparatus for defining innovative development in modern conditions, to group its concepts, structural and logical analysis, systemic and situational approaches were applied; to identify the dynamics of the global innovation index and the dynamics of factors influencing the global innovation index in Ukraine, the method of economic and statistical analysis was used; to determine the impact of each factor, correlation was applied.

Results and their discussion

Before clarifying the concept of innovation development, it is advisable to analyze the existing definitions from the point of view of the philosophy of science. Based on the most common definitions, it can be stated that there are two clusters of this concept. The first one can be considered as a subject-technological cluster focused on the scientific result, which was initiated by J. Schumpeter [30] and his followers. From this point of view, innovative development is the final result of scientific and technological activity at the microeconomic level. The second cluster should be defined as functional. It was initiated by B. Twiss [47] and B. Santo [27, 43], who linked innovation

development to the functions of innovation management and proposed to use it at the macroeconomic level. However, according to most scholars and practitioners, innovation development refers to the macroeconomic level and the construction of the knowledge economy and the state model of economic development, the mechanism for its implementation. Summarizing the understanding of the concept of innovative development, we can give a broader definition: "a method of economic growth based on constant and systematic innovations aimed at significant improvement of all aspects of the economic system, on the periodic regrouping of forces due to the logic of STP, goals and objectives of the system development, the possibility of using certain resource factors in the creation of innovative goods and the formation of competitive advantages" [10, p. 58–59]. In addition, the model of innovative development is considered to be the one based on "obtaining new scientific results and their technological implementation in production, ensuring GDP growth mainly through the production and sale of knowledge-intensive products and services" [14, p. 31]. Some scholars equate the concepts of "innovative development" and "innovative activity" [1], which is erroneous, since activity involves certain actions, and development involves actions related to the essence, interdependence, state, transformation, and transformations. At the same time, development can be both evolutionary and revolutionary, positive and negative, associated with a reduction, as well as sustainable, which characterizes the constant increase in the values of the indicators by which this process is studied.

In-depth studies of the content of the concept of "development" at the level of economic systems and organizations are proposed in the monograph by S. Mochny [21]. Mocherny [21], who identified quantitative and qualitative changes as criteria for the development process. This is very important for understanding innovation development and explains the changes that occur as a result of development in the state of the system or its subsystems, which are determined by the basic laws of dialectics, i.e. the transition of quantitative changes to qualitative ones, unity and struggle of opposites, and resolution of contradictions.

Applying the laws of dialectics to improve the definition of the term "innovative development" and exploring its essence, the author offers her own definition of this concept: innovative development characterizes the resolution of obvious and hidden contradictions

on the basis of qualitative changes in the economic system through the accumulation of quantitative transformations in its components and/or factors of innovation processes, intellectual property, the latest technologies and products. The advantages of this definition are its compliance with the dialectical laws of cognition, and consideration of the objective essence of innovations in all spheres of economic systems at all their levels of functioning and by the main types of innovations. The proposed definition should be taken into account in the concept of innovative development of economic systems.

With regard to the process of managing innovation development, its main feature is the discrepancy between the methods and principles of economic system management. It is this contradiction that requires improvement of certain theoretical and practical provisions of innovation development.

In terms of the theoretical aspect, the authors of the publication [23, p. 90] fully and deeply studied and generalized the concepts of innovative development of economic systems. They identified the following main groups of concepts:

- the concept of cyclical development, which includes the concepts of economic growth and STP, cyclical crises and "long waves";
- the orthodox concept, which includes the following concepts: the subjectivist concept of innovative development, the concept of cyclicity of innovations, technological systems, and the innovation and investment concept;
- the paradigm of technological change, which includes the concept of renewal and the innovation-centered concept of economic development;
- the paradigm of innovation diffusion, which includes the concepts of: speed of innovation diffusion, "demand pressure", regional diffusion of innovations, formation of technological systems and diffusion of innovations, sectoral distribution, human capital, harmonious economy, growth poles, competitive advantages, cluster development, innovation cluster;
- the concept of a technocratic society, which includes the concepts of stages of economic growth, managerial revolution, technological determinism and convergence, technotronic, post-economic information society, economic integration;
- the concept of technological progress, which includes the following concepts: technological-product, endogenous technological progress, economic growth with endogenous technological progress;

- the social concept, which includes the socio-psychological concept of innovative development in the socio-cultural environment;

- modern concepts, which may include concepts with an information component, economic progress, with an emphasis on various innovations and features of the innovation process, an ecosystem approach to innovations, and business model development.

The most difficult thing is to combine such a heterogeneous set of concepts and to take into account and highlight those defining features that are more relevant to the current economic environment, the needs of solvent end users and global trends in changes in economic relations in which the Ukrainian economic system operates. To do this, it is necessary to identify the factors that are significant and have the greatest impact on the current state of the Ukrainian economy, according to the main most important trends in its innovative development, and correspond to its substantive essence (as taken into account in our proposed definition). Therefore, the main phenomenon and complex indicator, the dynamics of changes of which should be analyzed, is the innovation rating of the world's economies. It is formed on the basis of the global innovation index and should be studied by the dynamics of the quantitative values of the components that characterize the Ukrainian economy.

According to the World Intellectual Property Organization (WIPO), in 2022, Ukraine ranked 57th in the innovation ranking of 132 economies, which is based on the Global Innovation Index (GII). Ukraine scored 31.0 points out of 100. The components of the ranking are the regulatory environment (in 2022, 75th position (+3 positions)); business environment (99th position (+5 positions)); human capital and research (49th position (–5 positions)); education (26th position (–3 positions)); R&D (59th position (–1 position)); information and communication technologies (63rd position (+6 positions)); knowledge and research results (36th position (–3 positions)) [25, 32]. Fig. 1 shows the dynamics of the global innovation index of Ukraine in 2015–2022.

The analysis of Fig. 1 shows that Ukraine's ranking in the Global Innovation Index was growing until 2018 (the lower the ranking, the better); in 2019, it deteriorated, and starting in 2021, after a slight increase, it began to fall again. It is known that the components of this index are the *Innovation Input* sub-index, which includes the following variables: institutions, human capital and research, infrastructure,

market development, business development or business experience; and the *Innovation Output* sub-index, which is formed by the values of knowledge and results of scientific research, creativity or results of creative activity. However, the value of the global innovation index is accidentally influenced by many other factors, namely: GDP in actual prices (x_1), exports of goods (x_2), imports of goods (x_3), exports of services (x_4), imports of services (x_5). The value of the country's global innovation index is influenced by direct investment (balance) (x_6), direct investment

(actives) (x_7), direct investment (passives) (x_8), income from capital transactions (x_9), revenues from the European Union, foreign governments international organizations, donor agencies (x_{10}), trust funds (x_{11}), consumer price index (x_{12}), volume of industrial products (goods, services) sold (x_{13}), economically active population aged 15–70 (x_{14}), real wage index (x_{15}) [18, 20, 22, 31]. Figs. 2–16 show the dynamics of these factors of the global innovation index of Ukraine [25, 26].

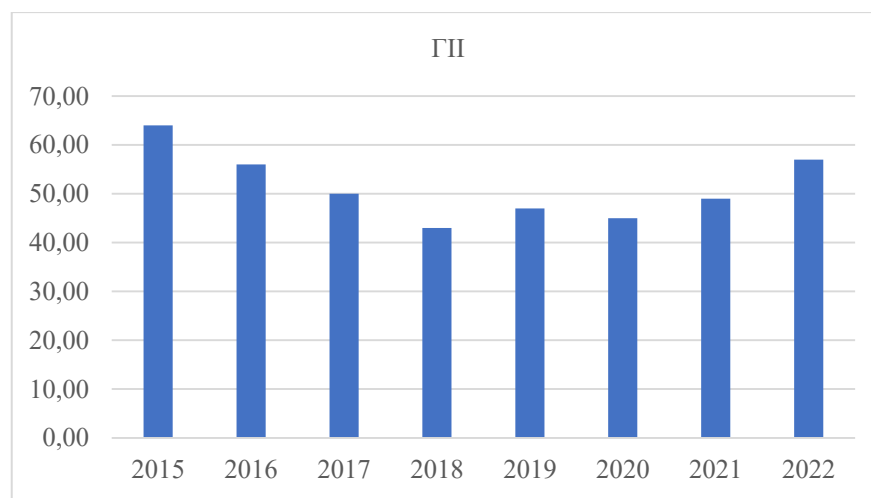


Fig. 1. Dynamics of the Global Innovation Index of Ukraine in 2015–2022

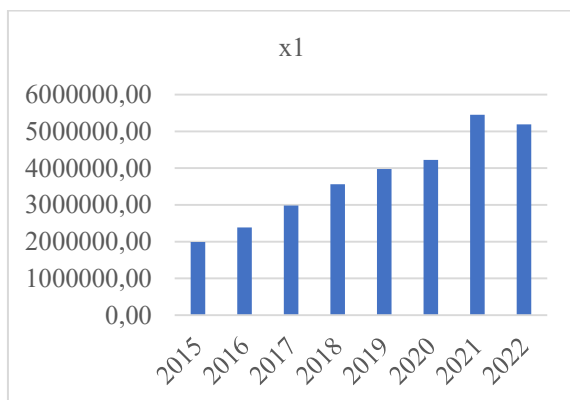


Fig. 2. Dynamics of GDP in actual prices

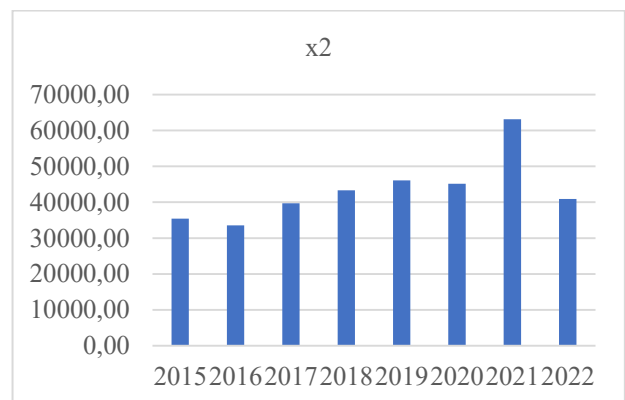


Fig. 3. Dynamics of exports of goods

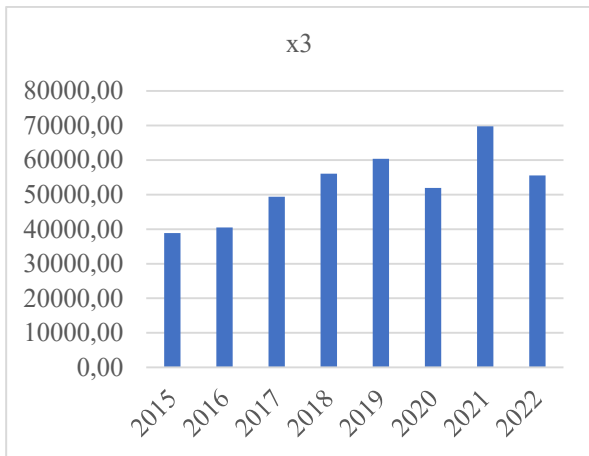


Fig. 4. Dynamics of imports of goods

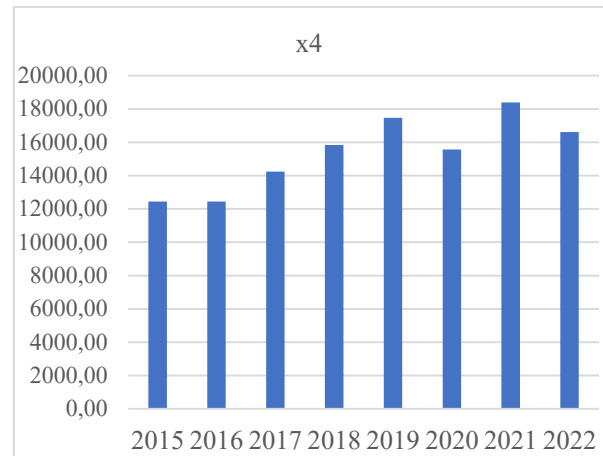


Fig. 5. Dynamics of exports of services



Fig. 6. Dynamics of imports of services

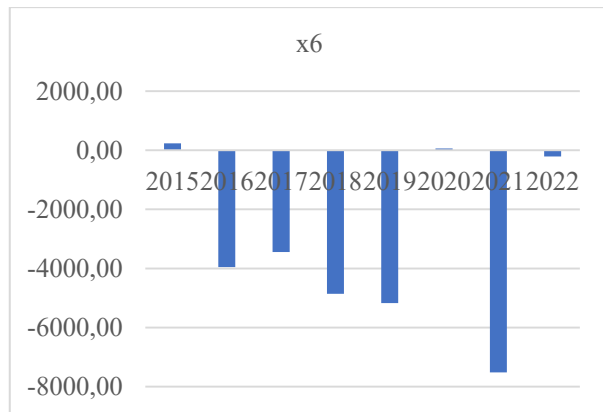


Fig. 7. Dynamics of direct investment (balance)

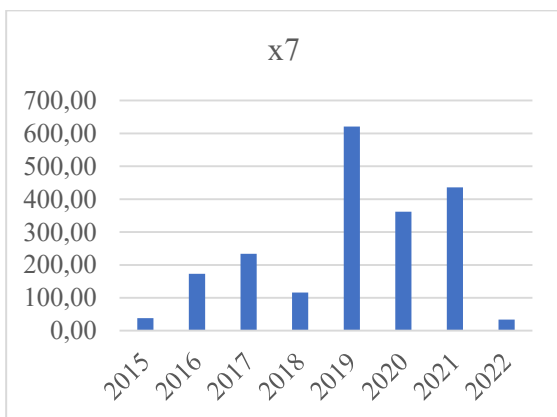


Fig. 8. Dynamics of direct investments (actives)

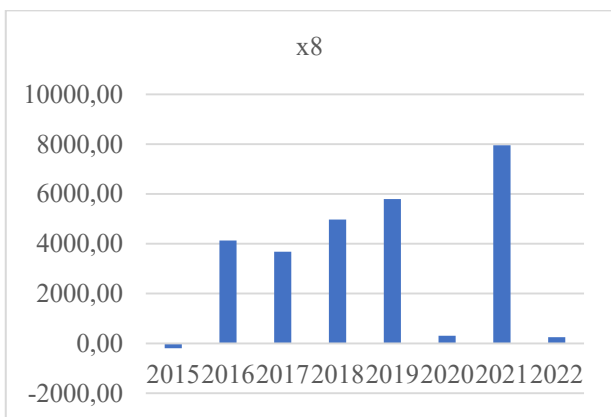


Fig. 9. Dynamics of direct investments (passives)

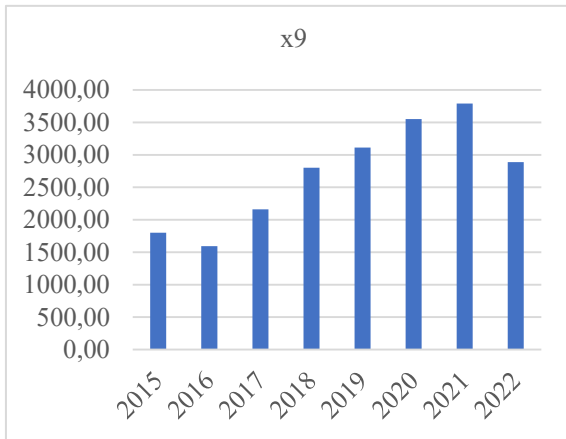


Fig. 10. Dynamics of income from capital transactions



Fig. 11. Dynamics of income

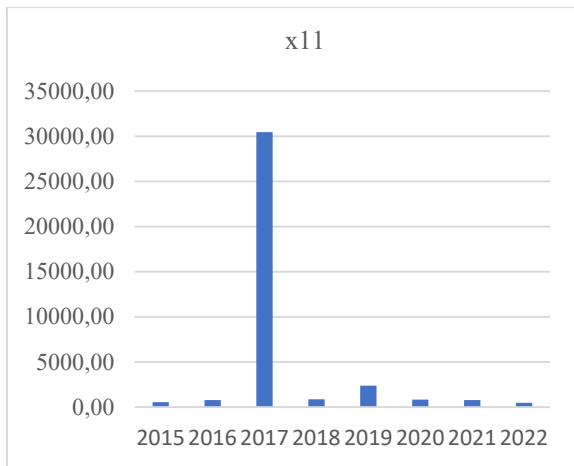


Fig. 12. Dynamics of trust funds

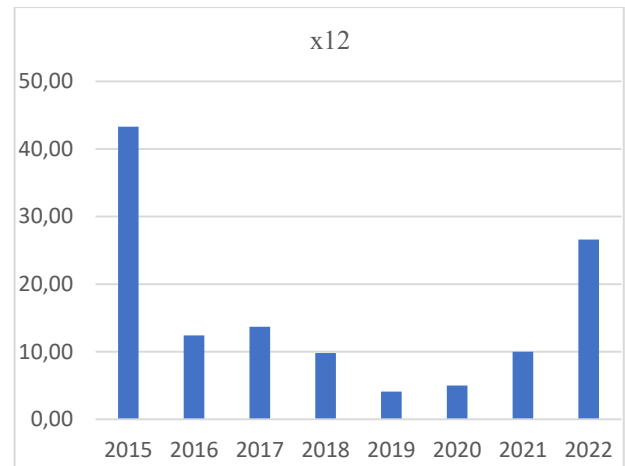


Fig. 13. Dynamics of the consumer price index

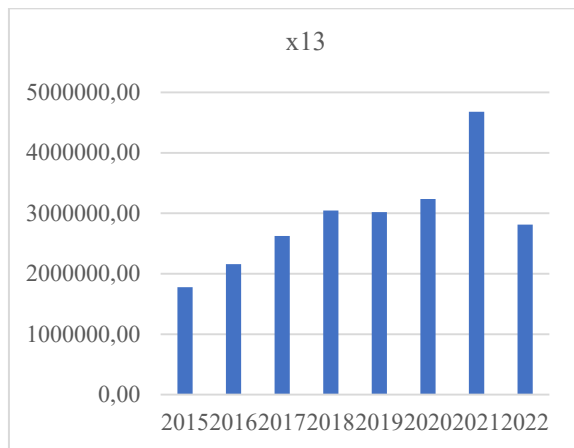


Fig. 14. Dynamics of industrial production sold

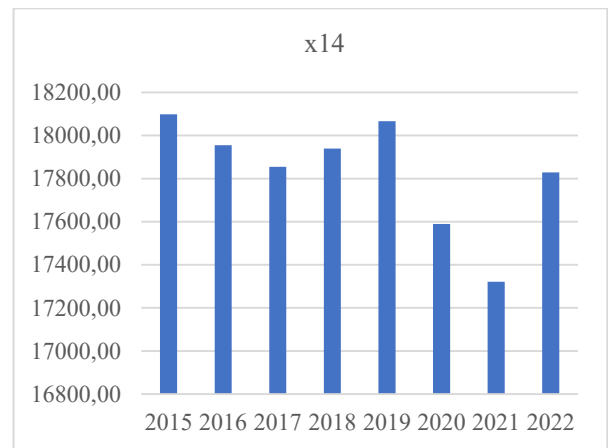


Fig. 15. Dynamics of economically active population

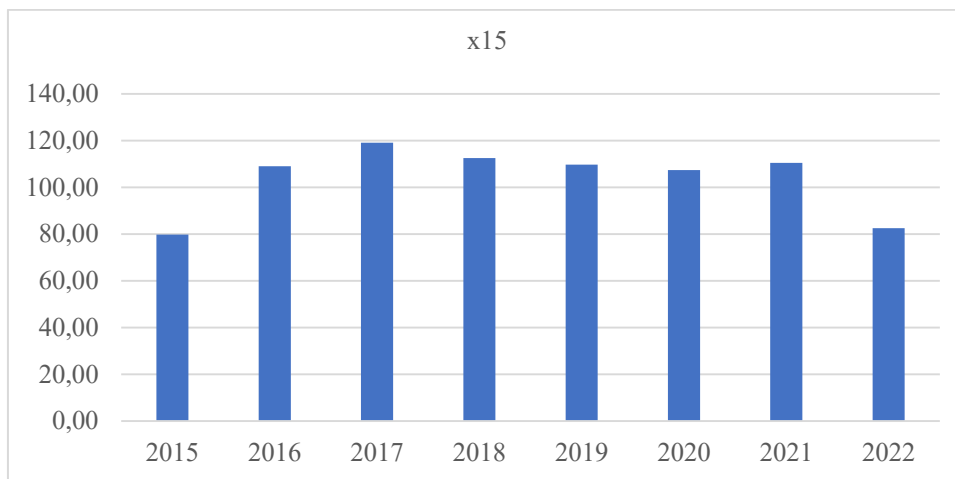


Fig. 16. Dynamics of the real wage index

The analysis of Figs. 2–16 shows that different trends in the factors influencing the global innovation index in Ukraine cause different changes in it. Growth up to 2021 inclusive was characterized by such factors as GDP in actual prices, exports of goods, imports of goods, exports of services, income from capital transactions, and the volume of industrial products (goods and services) sold. In 2022, all of these factors decreased. Other factors of Ukraine's global innovation index had unstable dynamics throughout the study period. To determine the influence of each factor, we should calculate the pairwise correlation coefficients. Table 1 shows the values of the paired correlation coefficients and their importance *P-values*. If the *P-values* are ≤ 0.05 , the coefficient is considered significant and can be used to assess the close relationship between the global innovation index and the factors mentioned above. Thus, Table 1 shows that the global innovation index is most closely correlated directly with the consumer price index (x_{12}) and inversely with the real wage index (x_{15}). Considering the levels of the modulo correlation coefficient in the intervals $[0, 0,4)$ as a weak relationship; $[0,4,0,7)$ as a moderate relationship; and $[0,7,1]$ as a strong relationship, it is advisable to analyze the moderate relationship [24].

Thus, the global innovation index is moderately correlated with such factors as GDP in actual prices, exports of goods, imports of goods, exports of services, direct investment (balance), direct investment (assets), direct investment (liabilities), income from capital transactions, revenues from the European Union, foreign governments, international organizations, donor agencies,

volume of industrial products (goods, services) sold, and economically active population aged 15–70. At the same time, among these factors, there are those that have a positive impact on Ukraine's ranking in the Global Innovation Index (an increase in the value of a factor is interrelated with a decrease in the ranking position, since a ranking of 1 is the highest ranking of a country), and there are those that have a negative impact. The positive factors include the real wage index (x_{15}), direct investment (passives) (x_8), income from capital transactions (x_9), imports of goods (x_3), volume of industrial products (goods, services) sold (x_{13}), exports of services (x_4), direct investment (actives) (x_7), direct investment (passives) (x_8), exports of goods (x_2), and the negative factors are the consumer price index, direct investment (balance) (x_6).

Further, it is advisable to continue the study of the impact of positive factors by calculating a multivariate regression model. First, we should determine the impact of positive factors on the global innovation index, which allowed us to build the following model:

$$Y = 97.549 - 0,329x_{15} - 0.004x_9 .$$

Only two positive factors were significant in the model, while the rest of the factors, according to the Student's criterion, were not significant. The model is statistically significant, as evidenced by both the coefficient of determination ($R^2 = 0.845$) and the Fisher's criterion ($F = 13.64$). Thus, if the real wage index (x_{15}) increases by 10 %, Ukraine's rating will rise

by three positions, holding all other factors constant. And if capital gains (x_9) increase by UAH 1000 million, Ukraine's rating will rise by four positions, all other factors held constant.

$$Y = 43.913 + 0.478x_{12}.$$

Only one negative factor was significant in the model, and the second negative factor (direct investment

(balance)) was not significant according to the Student's criterion. In general, the model is statistically qualitative, as evidenced both by the coefficient of determination ($R^2 = 0.791$), and the Fisher's criterion ($F = 22.73$). So, if the consumer price index (x_{12}) increase by 10 %, Ukraine's rating will fall by almost five positions.

Table 1. Pairwise correlation coefficients of the global innovation index with the factors

	Y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13	x14	x15
Y		-0,420	-0,499	-0,612	-0,584	0,152	0,488	-0,574	-0,509	-0,642	0,324	-0,103	0,890	-0,605	0,366	-0,780
P-values		0,300	0,209	0,107	0,129	0,719	0,220	0,137	0,198	0,086	0,433	0,808	0,003	0,112	0,372	0,023
x1	-0,420		0,779	0,877	0,912	0,590	-0,259	0,365	0,274	0,880	0,475	-0,237	-0,377	0,835	-0,712	0,043
P-values	0,300		0,023	0,004	0,002	0,124	0,536	0,374	0,511	0,004	0,234	0,572	0,358	0,010	0,048	0,919
x2	-0,499	0,779		0,915	0,849	0,051	-0,611	0,600	0,629	0,860	-0,115	-0,155	-0,441	0,974	-0,802	0,316
P-values	0,209	0,023		0,002	0,008	0,905	0,108	0,116	0,095	0,006	0,787	0,714	0,274	0,000	0,017	0,447
x3	-0,612	0,877	0,915		0,978	0,331	-0,620	0,597	0,637	0,877	0,106	-0,118	-0,536	0,926	-0,631	0,334
P-values	0,107	0,004	0,002		0,000	0,424	0,101	0,118	0,089	0,004	0,802	0,781	0,171	0,001	0,094	0,420
x4	-0,584	0,912	0,849	0,978		0,437	-0,482	0,595	0,505	0,904	0,225	-0,187	-0,503	0,862	-0,556	0,219
P-values	0,129	0,002	0,008	0,000		0,279	0,226	0,120	0,202	0,002	0,592	0,658	0,204	0,006	0,153	0,603
x5	0,152	0,590	0,051	0,331	0,437		0,163	-0,245	-0,173	0,209	0,952	-0,133	0,182	0,105	-0,009	-0,446
P-values	0,719	0,124	0,905	0,424	0,279		0,700	0,558	0,681	0,620	0,000	0,753	0,667	0,805	0,983	0,269
x6	0,488	-0,259	-0,611	-0,620	-0,482	0,163		-0,546	-0,998	-0,281	0,409	-0,071	0,581	-0,612	0,304	-0,682
P-values	0,220	0,536	0,108	0,101	0,226	0,700		0,162	0,000	0,500	0,314	0,867	0,131	0,107	0,465	0,062
x7	-0,574	0,365	0,600	0,597	0,595	-0,245	-0,546		0,593	0,589	-0,427	0,011	-0,717	0,569	-0,306	0,560
P-values	0,137	0,374	0,116	0,118	0,120	0,558	0,162		0,121	0,124	0,292	0,979	0,045	0,141	0,462	0,149
x8	-0,509	0,274	0,629	0,637	0,505	-0,173	-0,998	0,593		0,311	-0,423	0,069	-0,608	0,628	-0,313	0,694
P-values	0,198	0,511	0,095	0,089	0,202	0,681	0,000	0,121		0,453	0,297	0,871	0,110	0,096	0,451	0,056
x9	-0,642	0,880	0,860	0,877	0,904	0,209	-0,281	0,589	0,311		0,084	-0,266	-0,528	0,883	-0,720	0,214
P-values	0,086	0,004	0,006	0,004	0,002	0,620	0,500	0,124	0,453		0,843	0,524	0,179	0,004	0,044	0,611
x10	0,324	0,475	-0,115	0,106	0,225	0,952	0,409	-0,427	-0,423	0,084		-0,162	0,338	-0,052	-0,003	-0,597
P-values	0,433	0,234	0,787	0,802	0,592	0,000	0,314	0,292	0,297	0,843		0,702	0,413	0,902	0,994	0,118
x11	-0,103	-0,237	-0,155	-0,118	-0,187	-0,133	-0,071	0,011	0,069	-0,266	-0,162		-0,089	-0,131	0,052	0,451
P-values	0,808	0,572	0,714	0,781	0,658	0,753	0,867	0,979	0,871	0,524	0,702		0,834	0,758	0,903	0,263
x12	0,890	-0,377	-0,441	-0,536	-0,503	0,182	0,581	-0,717	-0,608	-0,528	0,338	-0,089		-0,566	0,376	-0,857
P-values	0,003	0,358	0,274	0,171	0,204	0,667	0,131	0,045	0,110	0,179	0,413	0,834		0,144	0,358	0,007
x13	-0,605	0,835	0,974	0,926	0,862	0,105	-0,612	0,569	0,628	0,883	-0,052	-0,131	-0,566		-0,856	0,410
P-values	0,112	0,010	0,000	0,001	0,006	0,805	0,107	0,141	0,096	0,004	0,902	0,758	0,144		0,007	0,313
x14	0,366	-0,712	-0,802	-0,631	-0,556	-0,009	0,304	-0,306	-0,313	-0,720	-0,003	0,052	0,376	-0,856		-0,277
P-values	0,372	0,048	0,017	0,094	0,153	0,983	0,465	0,462	0,451	0,044	0,994	0,903	0,358	0,007		0,507
x15	-0,780	0,043	0,316	0,334	0,219	-0,446	-0,682	0,560	0,694	0,214	-0,597	0,451	-0,857	0,410	-0,277	
P-values	0,023	0,919	0,447	0,420	0,603	0,269	0,062	0,149	0,056	0,611	0,118	0,263	0,007	0,313	0,507	

According to the constructed multivariate regression models, the indicators that influence the global innovation index of Ukraine's economy are the real wage index, income from capital transactions, and the consumer price index. These indicators should be taken into account as defining features of the innovative development of Ukraine's economy, i.e. in the strategic period, in the

current military and future post-war conditions, and in measures for its implementation, i.e. in the tactical period. Thus, the defining features of the innovative development of Ukraine's economy in the current military conditions in terms of its practical aspect are:

– solvent demand of end users (customers) of innovations for the acquisition of intellectual property,

production and commercialization of innovative technologies and products (in accordance with the indicators of the real wage index, which should be increased, and the consumer price index, whose growth should be restrained)

- investing in innovations that increase income from capital transactions (in accordance with the indicator of income from capital transactions, the growth of which should be promoted).

Thus, it is possible to identify specific regulatory measures to stimulate innovative development, taking into account the following factors: current defining features of innovative development, measures for their implementation and practical implementation in wartime; the experience of leading countries in the global innovation index and the experience of enterprises interested in the final consumption of innovations and in investing in green and ultramodern production, which involves significant additional costs.

To increase the effective demand of consumers (customers) of innovations, the following incentives are available:

- concessional lending to enterprises in key sectors of the economy, such as mechanical engineering, energy, chemical industry, instrumentation, and others, that serve other sectors of the economy;

- government funding for high-tech companies to develop digital skills and adapt employees to meet the needs of Industry 4.0;

- application of incentive programs that provide taxpayers with payroll credits;

- taxation of acquired patents for many years at a reduced tax rate instead of the usual percentage of corporate income tax;

- preferential lending for sales of innovative objects (patents, innovative products and technologies);

- state support for the use of production and infrastructure solutions in the production activities of enterprises involving modern robotic systems, mainly collaborative robots (cobots);

- introduction of free trade zones for innovative products and technologies;

- uniting research institutes, leading universities, and business associations into functional clusters to concentrate creative human capital and create cyclical demand from end users who are themselves developers of innovations;

- organizational support for the functioning of incubators, accelerators, technology parks, centers of expertise and R&D, testbeds, shared resource centers

that support key technologies such as A.I., IoT and others that are targeted for innovation development;

- development of institutions (entities) that ensure innovation development at the macroeconomic level on a national scale and are able to play a key role in the development of certain sectors of the economy and establishing the existing institutional boundaries and institutionalization of the state innovation policy and innovation development programs;

- modernization of production facilities and their accelerated digitalization, which involve the use of mechanisms to encourage end customers and industrial enterprises to use new technologies and products;

- focusing on the development of a policy to support and apply organizational tools for the growth of those enterprises that contribute a larger share of GDP than other business entities.

The following incentives are available to boost investment in innovation:

- reducing taxation on mergers and acquisitions by at least half;

- providing significant benefits (several hundred percent) for patenting and licensing innovations;

- allowing the use of accelerated depreciation methods (e.g., cumulative depreciation rates for innovative products of mechanical engineering as an industry with a slow capital turnover and one that significantly lost consumers during the war) in the case of using environmentally friendly production technologies;

- application of capital super- and hyper-depreciation programs that allow companies to add significant interest (from 40 % to 150 %, following the example of Italy), according to the cost of acquisition of qualified intangible assets;

- use of a preferential tax rate (no more than 10 % on profits or income) from the implementation of Industry 4.0 technologies to invest in domestic production;

- providing preferential or state funding for artificial intelligence projects;

- introduction of special investment regimes for launching innovative production and sales of intellectual property for better penetration into domestic and global value chains.

If the proposed stimulus measures are not adopted and supported at the macroeconomic level, Ukraine's economy will lag behind the developed world on the path of innovation development and will be perceived as a raw material supplement to countries that are more advanced in generating and implementing innovations.

Conclusions and prospects for further development

The relevance of the study of innovative development of Ukraine's economy is determined by the need to clarify its theoretical basis and develop the latest practical measures in accordance with the conditions of wartime. The urgent issues of promoting innovation development are changing approaches to intensifying innovation activity on the part of the State, identifying the influential factors that contribute to export-import activities to find end users interested in innovation, consolidating representatives of innovation systems of different economic levels to focus on achieving joint priority actions to lobby for innovation interests in order to stimulate effective demand for innovative results.

On the basis of generalization of theoretical sources on the definition of the concept of "innovative development", it was improved in accordance with the laws of the dialectic of cognition and the essence of the phenomenon itself. This made it possible to formulate the author's own definition of this concept. The advantages of the proposed definition are its compliance with the dialectical laws of cognition, taking into account the objective essence of innovations in all spheres of life of economic systems at all their levels of functioning and by the main types of innovations.

In accordance with the above generalized concepts of innovation development, it is proposed to study the factors of influence on this phenomenon and its defining features that are significant in modern conditions and

to take them into account in order to improve theory and practice. Using the values of the dynamics of the components of the global innovation index according to the statistical indicators of Ukraine's economy in 2015–2022, the most influential factors according to the relevant indicators are identified and multivariate regression models are built, according to which the main indicators of innovative development are the real wage index, income from capital transactions, and the consumer price index. According to these significant indicators, the author proposes the defining features of the innovative development of Ukraine's economy in the current military conditions, the essence of which is the following measures: formation and support of effective demand of end users (customers) of innovations for the acquisition of intellectual property, production and commercialization of innovative products and innovative technologies (according to the indicators of the real wage index, which should be increased, and the consumer price index, the growth of which should be restrained); investment in the development of the economy of Ukraine. Based on these defining features and taking into account the experience of the world's leading leaders in innovative development, the author proposes measures to ensure and intensify it under martial law in Ukraine.

The directions of further scientific research on innovative development are the development of strategic plans and programs for its provision by economic sectors, taking into account the tactical specification of the proposed and justified measures.

References

1. Adamenko, O. A. (2010), "Conceptual bases of innovative development of enterprises". ["Kontseptualni zasady innovatsiynoho rozvytku pidpriemstv"], *Scientific works of the National University of Food Technologies*. 2010. № 35. P. 5–10. available at: <http://dspace.nuft.edu.ua/jspui/bitstream/123456789/21211/1/1.pdf>
2. Angel, E. "Innovations in time of war – is it timely?" available at: <https://www.clusters.org.ua/blog-single/innovatsiyi-pidchas-viyny/>
3. Belyavtseva, V. V. "Methodology of management of innovative development of the region: monograph". Kharkiv: Printing house Madrid. 2017. 215 p.
4. "Problems of motivation of innovation activity". World practice and Ukrainian realities. available at: <http://dspace.nbu.gov.ua/xmlui/bitstream/handle/123456789/14229/1-Bilobrova.pdf?sequence=1>
5. Nebrat, V. V. "Reconstruction for Development: Foreign Experience and Ukrainian Prospects". An international collective monograph edited by the editorial board, headed by Doctor of Economics V.V. Nebrat; NAS of Ukraine, State Institution "Institute of Economics and Forecasting". National Academy of Sciences of Ukraine. Kyiv, 2023. 571 p. available at: <http://ief.org.ua/wp-content/uploads/2023/08/Reconstruction-for-development.pdf>
6. Geets, V. M. "Instability and economic growth". Monograph. National Academy of Sciences of Ukraine, Institute for Economic Forecasting. Kyiv, Institute for Economic Forecasting. National Academy of Sciences of Ukraine, 2000. 344 p. available at: http://ief.org.ua/?page_id=3625&mid=5

7. Gornyk, V. "Prospects for the development of innovative activity of enterprises in Ukraine". *Scientific notes of Vernadsky TNU. Series: Economics and management*. Vol. 31 (70). № 2, 2020. P. 178–184. DOI: <https://doi.org/10.32838/2523-4803/70-2-29>
8. Dyba, O. M. "Innovative development of enterprises in the context of globalization". Strategy of economic development of Ukraine. 2018. № 42. P. 111–118. available at: http://nbuv.gov.ua/UJRN/seru_2018_42_12
9. Zakharchenko, V. I. *Innovative management: theory and practice in the conditions of economic transformation*. Kyiv: Center for Educational Literature, 2012. 448 p. <https://duikt.edu.ua/ua/lib/1/category/743/view/619?lang=ua&act=view&page=1&category=743&id=619>
10. Illyashenko, N. S. "Theoretical and methodological foundations for determining the leading trajectories of innovative acceleration of industrial enterprises". *Business Inform*. 2019. №2. P. 95–101. DOI: 10.32983/2222-4459-2019-2-95-101
11. Illyashenko, N. S. "Management of advanced innovative development of industrial enterprises". Monograph. Sumy: Tritoria, 2019. 504 p. available at: <https://core.ac.uk/download/pdf/324252798.pdf>
12. Illyashenko, S. M. "Knowledge management in the formation of a strategy for advanced innovative development of the enterprise". Problems of a systematic approach to the economy. 2019. Part 1. P. 215–223. available at: <http://essuir.sumdu.edu.ua/handle/123456789/73966>
13. Illyashenko, S. M. "Analysis of Ukraine's positions in international rankings of innovative development". Sustainable development – XXI century: management, technologies, models. Discussions 2018: a collective monograph NTUU Kyiv Polytechnic Institute. Kyiv, 2018. P. 183–195. available at: https://books.google.com.ua/books?id=Kd6CDwAAQBAJ&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
14. Kvasha, O. S. "Innovative development of the Ukrainian economy: world experience and recommendations for Ukraine. Scientific Bulletin of Uzhhorod National University". Series: International economic relations and world economy. 2016. Issue 6(1). P. 150–154. available at: http://www.visnyk-econom.uzhnu.uz.ua/archive/6_1_2016ua/38.pdf
15. Knyazevich, A. O. "Innovative entrepreneurship in the system of creative management". *Business Inform*. 2022. № 5. P. 33–39. available at: <http://jnas.nbuv.gov.ua/article/UJRN-0001358925>
16. Kovpak, A. "Innovative development of the country as a driving factor for increasing national competitiveness". *Ekonomika. Management. Innovations*. Series: Economic Sciences. Issue 1.(28). available at: <http://eui.zu.edu.ua/article/view/234658/238776>
17. Krylov, D. V. (2022), "Analysis of the rating assessment of the development of innovation activity in Ukraine". *Effective Economy*. № 5. 2022. P. 1–6. DOI: 10.32702/2307-2105-2022.5.8
18. Liudvik, I. (2022), "Innovative development of Ukraine in the context of globalization of the economic space. State and regions". *Series: Economics and entrepreneurship*, № 2 (125). P. 36–40. DOI: <https://doi.org/10.32840/1814-1161/2022-2-6>
19. Mykhailyshyn, L.I. "Transnationalization of the world economy: innovative aspect: monograph". DonNU, 2016. 314 p. available at: <https://abstracts.donnu.edu.ua/article/view/2865/2904>
20. Mochernyi, S. V. (2015), *Methodology of economic research*. System of modern methodologies: a textbook in 4 volumes. Ministry of Education and Science of Ukraine, Ternopil National Economic University. Ternopil: TNEU, Vol. 3. P. 61–76.
21. Ovander, N. L., Orlova, K. E. (2019), "Structural indicators of the innovative model of development of the national economy". *Effective economy*. № 7. P. 1–8. DOI: 10.32702/2307-2105-2019.7.33
22. Ponomarenko, V. S., Malyarets, L. M., Barannik, I. O., Balyuk, Y. S. (2023), "Provisions of the concept of transformation of innovative development of export-import activity of economic entities in wartime and post-war conditions". *Problems of economy*. № 2 (56). P. 87–98. DOI: <https://doi.org/10.32983/2222-0712-2023-2-87-98>
23. Ponomarenko, V. S. (2009), *Data analysis in the study of socio-economic systems: monograph*. Kharkiv: INZHEK Publishing House, 432 p.
24. "Website of the Ministry of Economy of Ukraine". available at: <https://www.me.gov.ua/News/Detail?lang=uk-UA&id=a1554595-5306-44ac-90a7-943745ba06f0&title=UkrainaRozvivatime>
25. "Website of the National Bank of Ukraine". available at: <https://bank.gov.ua/ua/news/all/makroekonomichniy-tamonetarniy-oglyad-cherven-2023-roku>
26. Santo, B. (2004), *The power of innovative self-development*. Innovation. № 2. P. 5–15.
27. Ushkarenko, Y. V., Solovyov, A. I. (2023), "Social innovations in the system of sustainable development". *Problems of modern transformations*. Series: Economics and Management. №7. DOI: <https://doi.org/10.54929/2786-5738-2023-7-03-05>
28. Hovrak, I. V. "Financing of innovative development: realities and prospects". Marketing and management of innovations. 2013. № 1. P. 229–235. available at: http://nbuv.gov.ua/UJRN/Mimi_2013_1_23
29. Schumpeter, J. (2011), *The Theory of Economic Development: A Study of Profits, Capital, Credit, Interest and the Economic Cycle*. Translated from English by V. Stark Kyiv: Kyiv--Mohyla Academy, 2011. 242 p.

30. Yurynets, Z. V. "Formation of innovative strategies: theory, methodology, practice". Monograph. Lviv: SPOLOM. 2016. 412 p. available at: <https://econom.lnu.edu.ua/wp-content/uploads/2018/07/monohrafiia.pdf>
31. Beckman, S. L. "Innovation as a Learning Process". California Management Review. Fall 2007. Vol. 50, № 1. P. 25–56. available at: <https://isfcolombia.uniandes.edu.co/images/documentos/designthinkingdoc.pdf>
32. Cassiman, B., Valentini, G. (2016), "Open innovation: Are inbound and outbound knowledge flows really complementary?" *Strategic Management Journal*. № 37. P. 1034–1046. DOI:10.1002/smj.2375
33. Chesbrough, H., Bogers M. (2014), "Explicating open innovation: Clarifying an emerging paradigm for understanding innovation. New frontiers in open innovation". *Oxford: Oxford University Press*. P. 3–28. available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2427233
34. Dougherty, D. (2017), *Organizing for innovation in complex innovation systems*. Innovation. №19:1. P. 11–15.
35. "Global Innovation Quotient". available at: <http://www.bloomberg.com/slideshow>
36. Iastremska, O. (2019), "Investment and innovative development of industrial enterprises as the basis for technological singularity". *Problems and Perspectives in Management*. Volume 17. Issue 3, P. 477–491. DOI: [http://dx.doi.org/10.21511/ppm.17\(3\).2019.38](http://dx.doi.org/10.21511/ppm.17(3).2019.38)
37. Iastremska, O., Strokovych, H., Iastremska, O., Kalantaridis, C., Nagy, S., & Somosi Veresne, M. (2021), "Formation of Mutual Relations Between Enterprises and Business Partners in the Process of Preparation and Production of New Products". *Marketing and Management of Innovations*. Vol. 1, P. 196–211. DOI:<http://doi.org/10.21272/mmi.2021.1-15>
38. Janáková, H. (2012), "Creative management and innovation". *Creative and Knowledge Society*. Volume 2, Issue 1. P. 95–112. DOI: [10.2478/v10212-011-0019-7](https://doi.org/10.2478/v10212-011-0019-7)
39. Khedhaouria, A., Jamal, A. (2015), "Sourcing knowledge for innovation: knowledge reuse and creation in project teams". *Journal of Knowledge Management*. No 19 (5). P. 932–948. DOI: <https://doi.org/10.1108/JKM-01-2015-0039>
40. Malyarets, L., Iastremska, O., Tutova, A., Dorohov, O. (2023), "Analytical Method of Stimulation for Labour of Top-Managers" *TEM Journal*. Volume 12, Issue 2, P. 1118–1129. DOI: [10.18421/TEM122-58](https://doi.org/10.18421/TEM122-58)
41. Malyarets, L., Iastremska, O., Tutova, A. (2022), "Determination of priority qualitative factors of stimulation of the work of enterprise top managers". *Innovative Technologies and Scientific Solutions for Industries*, No. 2 (20), P. 6–13. DOI: <https://doi.org/10.30837/ITSSI.2022.20.075>
42. Santo, B. (1985), *Innovacio a gazdasagi fejlesztes eszkoze*. Innovation as a Means of Economic Development, Hungarian, by Muszaki konyvkiado, Budapest. 296 p.
43. Shapira, Ph., Youtie, J. "The Innovation System and Innovation Policy in the United States". available at: <http://works.bepress.com/pshapira/19>
44. Sennes, R. "Innovation in Brazil: Public Policies and business Strategies". available at: <http://www.wilsoncenter.org/sites/default/files/Innovation%20Public%20Private%20Strategies%20English.pdf>
45. Stenberg, L. "Government Research and Innovation Policies in Japan". available at: <http://www.innovation.lv/ino2/publications/japan.pdf>
46. Twiss, B. (1974), *Managing technological innovation*. Longman, 237 p.

Received 05.09.2023

Відомості про авторів / About the Authors

Ястремська Олена Миколаївна – доктор економічних наук, професор, Харківський національний економічний університет ім. С. Кузнеця, завідувач кафедри менеджменту, логістики та інновацій, Харків, Україна; e-mail: Iastremska_om@hneu.net; ORCID ID: <http://orcid.org/0000-0002-5653-6301>

Малярець Людмила Михайлівна – доктор економічних наук, професор, Харківський національний економічний університет ім. С. Кузнеця, завідувач кафедри вищої математики та економіко-математичних методів, Харків, Україна; e-mail: malyarets@ukr.net; ORCID ID: <https://orcid.org/0000-0002-1684-9805>

Ястремська Олеся Олександрівна – кандидат економічних наук, доцент, Харківський національний економічний університет ім. Семена Кузнеця, докторант кафедри менеджменту та бізнесу, Харків, Україна; e-mail: Iastremska.o@gmail.com; ORCID ID: <https://orcid.org/0000-0003-1865-0282>

Бараннік Ігор Олексійович – кандидат економічних наук, Харківський національний економічний університет ім. С. Кузнеця, докторант кафедри менеджменту та бізнесу, Харків, Україна; e-mail: barannik@ukr.net; ORCID ID: <https://orcid.org/0000-0001-6364-4768>

Iastremska Olena – Doctor of Sciences (Economics), Professor, Simon Kuznets Kharkiv National University of Economics, Head at the Department of Management, Logistics and Innovation, Kharkiv, Ukraine.

Malyarets Lyudmyla – Doctor of Sciences (Economics), Professor, Simon Kuznets Kharkiv National University of Economics, Head at the Department of Mathematics and Mathematical Methods in Economics, Kharkiv, Ukraine.

Iastremska Olesia – PhD (Economics Sciences), Associated Professor, Simon Kuznets Kharkiv National University of Economics, Candidate on Doctor Degree at the Department of Management and Business, Kharkiv, Ukraine.

Barannik Igor – PhD (Economics Sciences), Simon Kuznets Kharkiv National University of Economics, Candidate on Doctor Degree at the Department of Management and Business, Kharkiv, Ukraine.

СУЧАСНІ ВИЗНАЧАЛЬНІ ХАРАКТЕРИСТИКИ ІННОВАЦІЙНОГО РОЗВИТКУ ЕКОНОМІКИ УКРАЇНИ: ТЕОРЕТИЧНИЙ ТА ПРАКТИЧНИЙ АСПЕКТИ

Предметом статті є теоретичне обґрунтування, методичне забезпечення, практичні пропозиції з дослідження інноваційного розвитку економіки в сучасних умовах. **Мета роботи** – узагальнення та вдосконалення теоретичних засад інноваційного розвитку, виявлення основних впливових факторів і визначальних характеристик на його перебіг та розроблення пропозицій щодо забезпечення інноваційного розвитку економіки України в сучасних умовах. Визначено такі **завдання**: з'ясувати ставлення фахівців до можливості та необхідності підтримки інноваційного розвитку в складних воєнних умовах сьогодення; узагальнити теоретичне підґрунтя інноваційного розвитку, удосконалити його визначення та згрупувати наявні концепції; обґрунтувати найвпливовіші фактори, що зумовлюють активізацію та/або стримання інноваційного розвитку економіки України; виявити його визначальні ознаки, за якими згенеровано заходи забезпечення інноваційного розвитку економіки у воєнних умовах. Для вирішення завдань застосовано такі **методи**: теоретичне узагальнення, аналіз, синтез, індукція, дедукція, структурно-логічний аналіз, системний і ситуаційний підходи, економіко-статистичний аналіз, кореляційний аналіз, багатофакторний регресійний аналіз, графічний метод. Здобуто такі **результати**: удосконалено визначення поняття "інноваційний розвиток"; проаналізовано концепції інноваційного розвитку; виявлено основні фактори, що впливають на перебіг інноваційного розвитку в Україні; побудовано багатофакторні регресійні моделі за впливовими факторами інноваційного розвитку та обґрунтовано найбільш важливі з них; за значущими факторами виявлено сучасні визначальні ознаки інноваційного розвитку економіки України та здійснено їх конкретизацію відповідно до заходів для його забезпечення в умовах воєнних дій. **Висновки**. Доведено, що вдосконалене поняття "інноваційний розвиток" полягає в тому, що воно характеризує розв'язання явних і прихованих суперечностей на основі здійснення якісних змін в економічній системі завдяки накопиченню кількісних перетворень у її складниках і/або факторах інноваційних процесів, інтелектуальної власності, новітніх технологій і продукції. Відповідно до наведених узагальнених концепцій інноваційного розвитку досліджено фактори впливу на це явище за динамікою складників глобального інноваційного індексу й за статистичними показниками економіки України за період 2015–2022 рр. Визначено найбільш впливові фактори за окресленими показниками й побудовано багатофакторні регресійні моделі, за якими виявлено такі сучасні провідні ознаки інноваційного розвитку: формування й підтримування платоспроможного попиту кінцевих споживачів (замовників) інновацій щодо надбання інтелектуальної власності, виробництва й комерціалізації інноваційної продукції та інноваційних технологій; інвестування в інновації. З огляду на це запропоновано заходи щодо забезпечення й активізації інноваційного розвитку в умовах воєнних дій в Україні.

Ключові слова: інноваційний розвиток; глобальний інноваційний індекс; сучасні визначальні ознаки; заходи; воєнні дії.

Бібліографічні описи / Bibliographic descriptions

Ястремська О. М., Мальярець Л. М., Ястремська О. О., Бараннік І. О. Сучасні визначальні характеристики інноваційного розвитку економіки України: теоретичний та практичний аспекти. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 148–162. DOI: <https://doi.org/10.30837/ITSSI.2023.25.148>

Iastremska, O., Malyarets, L., Iastremska, O., Barannik, I. (2023), "Modern defining characteristics of innovative development of the ukrainian economy: theoretical and practical aspects", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 148–162. DOI: <https://doi.org/10.30837/ITSSI.2023.25.148>

УДК 681.5.015.8

DOI: <https://doi.org/10.30837/ITSSI.2023.254.163>

Д. МОСЬПАН

АВТОМАТИЗАЦІЯ ТЕХНОЛОГІЧНОГО ПРОЦЕСУ ФОРМОУТВОРЕННЯ ДЕТАЛЕЙ ІЗ ПРЯМОЛІНІЙНИМИ РИФТАМИ

Предметом дослідження є технологічна лінія виготовлення товстостієвих деталей з рифтами, виготовленими з конструкційних і спеціальних сталей. **Мета роботи** – проаналізувати й створити методи управління основними технологічними операціями на всіх ланках технологічного процесу автоматизованого виготовлення таких деталей і на їх підставі сформулювати передумови розроблення локальних систем керування обладнанням, що забезпечує виконання операцій на кожному визначеному в цій роботі етапі технологічного процесу, а також визначити чинники побудови структури та алгоритму загальної системи керування, яка синхронізуватиме та контролюватиме весь технологічний процес. У статті розв'язуються такі **завдання**: створення моделей основних етапів формоутворення деталей з прямолінійними рифтами та керування динамічними властивостями обраних систем їх реалізації. Використовуються такі **методи**: чисельне інтегрування та розв'язання диференційних рівнянь; аналітичне розв'язання диференційних рівнянь на основі принципу Даламбера або способом використання рівняння Лагранжа; частотний аналіз передатних функцій окремих ланок, поданих у вигляді дробово-раціональних функцій. Досягнуто таких **результатів**: побудовано в першому наближенні склад та послідовність основних операцій узагальненого технологічного процесу автоматизованого виготовлення товстостієвих листових деталей з рифтами; за результатами аналізу процесу пружно-пластичної зміни форми ділянки з багатьма зв'язками визначено вхідні й вихідні параметри, необхідні для побудови моделі керування зазначеним процесом; з'ясовано, що для його реалізації потрібно забезпечити вільне деформування як мінімум 0,25 % площі поверхні рифту поза контактом із формоутворювальним інструментом, і запропоновано необхідні для цього технологічні методи; подано варіанти технічної реалізації обраних технологічних методів; наведено результати моделювання схем і режимів роботи необхідного технічного обладнання. **Висновки**. Застосування запропонованої структури, складу та послідовності основних операцій узагальненого технологічного процесу автоматизованого виготовлення товстостієвих листових деталей з рифтами на основі його реалізації обраними методами на запропонованому обладнанні забезпечить ефективне управління якістю виготовлення зазначених деталей на кожному етапі технологічного процесу.

Ключові слова: автоматизація технологічного процесу; моделювання основних етапів процесу й режимів роботи обладнання; система автоматичного керування; частотні характеристики.

Вступ

Автоматизація процесу виготовлення товстостієвих деталей із конструкційних і спеціальних сталей є актуальним завданням

сучасного машинобудування. У першому наближенні склад та послідовність основних операцій узагальненого технологічного процесу автоматизованого виготовлення таких деталей можна подати так, як зображено на рис. 1.

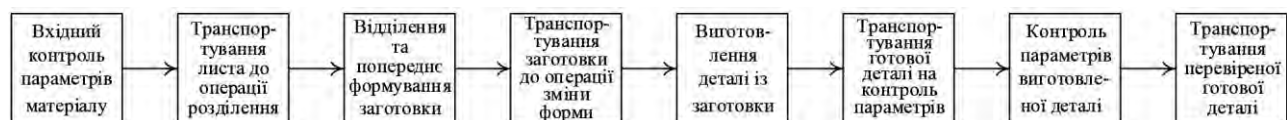


Рис. 1. Склад та послідовність операцій виготовлення деталі

Очевидно, що ключовим моментом є етап виготовлення деталі із заготовки, оскільки можливість його автоматизації та керування ним у реальному часі є вирішальною ланкою системи автоматизації всього процесу.

Аналіз проблеми й наявних методів

Процес виготовлення товстостієвих деталей щодо формоутворення деталей із прямолінійними рифтами детально розглянуто в роботі [1]. Зокрема автори уточнили радіуси кривизни

в меридіональному перерізі та довжину кінця рифту в зоні вільного вигину.

Унаслідок моделювання в праці [2] зроблено висновок про те, що в процесі формоутворення ребер можуть виникати дефекти у вигляді розривів. На вершинах центральних ребер з'являються суттєві стоншування матеріалу. На периферії між ребрами утворюються складки, що погіршують якість готової деталі й завищують вимоги до технологічного обладнання. Для згладжування складок потрібні додаткові зусилля.

У процесі моделювання виявлено, що на формування ребер суттєво впливає величина сил тертя. Так, наприклад, за мінімальних сил тертя матеріал заготовки вільно ковзає вздовж оснастки, запобігаючи надмірному розтягуванню матеріалу в ділянці центральних ребер. Але водночас мінімальні сили тертя зумовлюють значні переміщення матеріалу на периферії.

За умови формування без обважнення плоского елемента заготовки відбувається її руйнування в зоні спряження із кінцем рифту, у разі обважнення понад 8–10 % і формування-витягування по жорсткому пуансону руйнування не відбувається, але утворюються гофри в зоні спряження кінця рифту з плоским елементом. У процесі формування-витягування скороченим пуансоном частина кінця рифту (понад 70 %) формується вільно, розриву заготовки та утворення гофрів не спостерігається.

Було запропоновано такі рекомендації [2]:

- упровадити поетапне формоутворення ребер із проміжним відпадом;

- розглянути інші (неномінальні) геометричні параметри оснастки та готової деталі з метою зниження ступеня деформації;

- мінімізувати тертя між оснасткою та заготовкою.

Після експериментальної перевірки результатів моделювання (деформування здійснювалось на пресі УНМ-100 з відповідною оснасткою до моменту втрати стійкості фланця зразка) зроблено такі висновки:

- граничні відхилення експериментальних значень радіуса кривизни ділянки вільного вигину за умови довжини пуансона, що дорівнює довжині прямолінійного рифту без кінців, обчислені за поданою в роботі [2] залежністю, не перевищують 5 % і лежать в межах допуску на серійно виготовлені деталі. Запобігати утворенню гофрів у процесі штампування товстостієвих деталей можливо способом формування скороченим пуансоном, довжина якого відповідає довжині рифту без кінців,

плюс, як мінімум, 0,7 радіуса циліндричної ділянки рифту. У цьому разі розміри відформованого рифту відповідають заданим у межах дозволу;

- дія додаткових напружень вигину, пов'язаних із зміною радіуса кривизни формоутворювального інструмента є незначними.

У роботі [3] за результатами аналізу процесу пружно-пластичної зміни форми ділянки з багатьма зв'язками зроблений висновок, що для побудови моделі керування таким процесом можна визначити вхідні та вихідні параметри.

Вихідні параметри: швидкість деформації; контактні напруження вздовж осі руху пуансона. Для вимірювання та обчислення значень цих параметрів достатньо встановити на робочий орган преса акселерометр, що дасть змогу в режимі реального часу відстежувати й обчислювати обидва обраних вихідних параметри.

Вхідні параметри: сила, прикладена до заготовки; сили тертя, які перешкоджають витягуванню фланцевої частини заготовки; сили, що діють на заготовку з боку матриці (перетяжного ребра). Вимірювання та керування зазначеними параметрами може бути здійсненим таким чином. Сила, прикладена до заготовки, може змінитися в потрібному діапазоні за допомогою використання керованої зустрічної (відповідно до напрямку дії пуансона) пружної дії з боку матриці, що розміщується на пружній основі, жорсткість якої можна змінювати в певних межах.

Аналогічно може бути вирішене питання керування силами, що діють на заготовку з боку матриці (перетяжного ребра).

Сили тертя, які перешкоджають витягуванню заготовки, також можуть відкоригованим способом керувати величиною сили притискання заготовки.

За інформацією з джерела [4] у процесі витягування циліндричних деталей на пресах подвійної дії з жорстким притисканням важливою є не величина тиску, а встановлення мінімально необхідного для певної товщини листа зазору між матрицею та притиском, який не защемлює заготовку, але запобігає утворенню зморшок.

Сили притискання для першого витягування циліндричних деталей з плоскої заготовки обчислюють за такою формулою [5]:

$$Q = 0,25\pi \left[D^2 - (d_1 + 2r_m)^2 \right], \quad (1)$$

де d_1 – діаметр витягування (матриці) на першій операції, мм;

r_m – радіус закруглення витяжної крайки, мм.

Значення, близькі до обчислених за формулою, дає номограма, зображена на рис. 2. Тиск притискання, знайдений за номограмою, визначається тими самими параметрами.

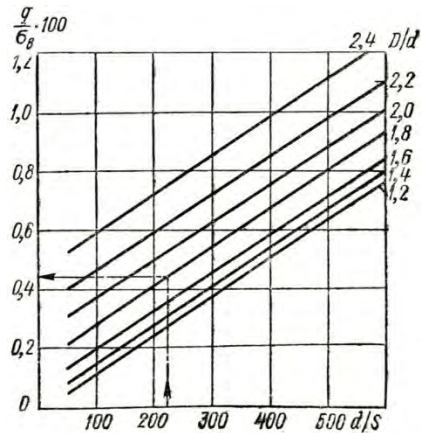


Рис. 2. Номограма для визначення тиску притискання залежно від d/s та D/d (для сталі 10)

Розв'язання завдання

Наведені вище результати аналізу можливості автоматизації етапу виготовлення товстостілової деталі із заготовки свідчать про те, що для його реалізації необхідно забезпечити вільне деформування як мінімум 0,25 % площі поверхні рифту поза контактом з формоутворювальним інструментом. Для цього рекомендовано застосовувати один або декілька з таких технологічних методів:

- 1) використання пристрою для створення перепаду тиску;
- 2) застосування скороченого пуансона;
- 3) керування силою тертя (тобто силою притискання або коефіцієнтом тертя).

Розглянемо можливі варіанти побудови систем автоматичного керування для кожного з наведених вище технологічних методів.

Застосування накладок з еластичного матеріалу

Запропонований у роботі [6] пристрій для створення перепаду тиску з допомогою накладок з еластичного матеріалу, закріплених на заготовці, зображений на рис. 3. Заготовка в цьому разі формується постійним тиском рідини або газу.

Для вирішення поставленого завдання формування товстостілової деталі з рифтами, замість еластичних накладок, можна застосувати спеціальну оснастку, у якій певні ділянки матриці матимуть змогу переміщуватись у напрямку руху пуансона й спиратимуться на основу з керованою пружністю.

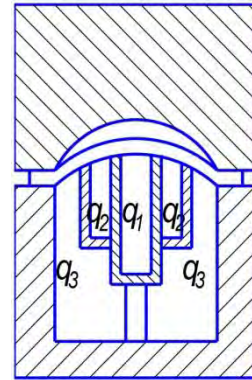


Рис. 3. Пристрій для створення перепаду тиску залежно від d/s та D/d (для сталі 10)

Такою основою можна використати гідроопору із системою активного силового керування (рис. 4).

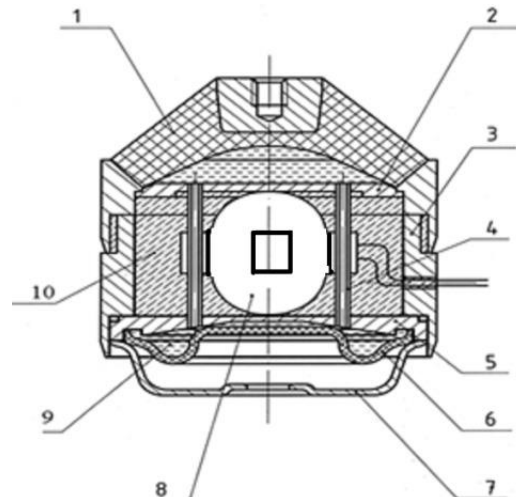


Рис. 4. Розріз керованої гідроопори:

- 1 – гумова обичайка (пружний елемент);
- 2 та 5 – несні перегородки; 3 – корпус;
- 4 – демпферний канал; 6 – мембрана;
- 7 – піддон для кріплення;
- 8 – котушка збудження електромагніту магнітореологічного трансформатора;
- 9 – магнітореологічна рідина;
- 10 – елементи кріплення конструкції

Відповідно до схеми, зображеної на рис. 5, поелементні рівняння динаміки гідроопори мають такий вигляд:

$$m_0 \ddot{x} + c_r x = F(t) - p_1 A \quad (2)$$

$$p_1 - p_2 = r A \dot{x} + L A \ddot{x}. \quad (3)$$

$$p_2 E - A x = \frac{E}{s} u. \quad (4)$$

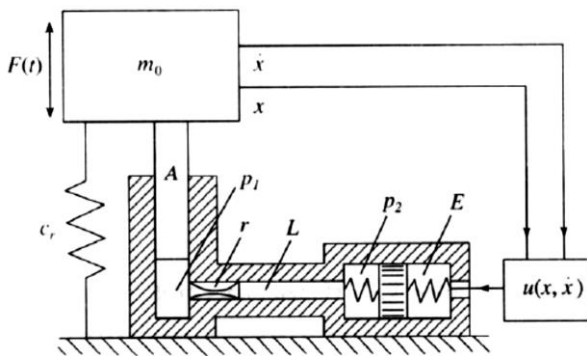


Рис. 5. Гідроопора з активною дією

У формулах (2)–(4):

m_0 – маса об'єкта;

c_r – жорсткість конічної гумової обичайки;

$F(t)$ – сила збурення (сила, прикладена до пуансона);

p_1 – тиск у робочій камері;

A – площа поршня;

p_2 – тиск у додатковій камері;

r – лінійний гідравлічний дисипативний опір каналу;

L – гідравлічна інерційність;

E – ємність додаткової камери;

s – площа перерізу мембрани;

$u(x, \dot{x})$ – активна силова зустрічна дія гідроопори, яка має бути визначеною.

Виразимо з (4) p_2 через x та u і підставимо отриманий вираз у (3), після чого з (3) виразимо p_2 через \dot{x} та \ddot{x} , отримаємо наступне рівняння, яке пов'язуватиме $\ddot{x}, \dot{x}, x, F(t)$ та u :

$$(m_0 + LA^2)\ddot{x} + rA^2\dot{x} + \left(c_r + \frac{A^2}{E}\right)x = F(t) - \frac{A}{s}u(x, \dot{x}). \quad (5)$$

Введемо такі позначки:

$$m = (m_0 + LA^2);$$

$$m_1 = LA;$$

$$2k = rA^2;$$

$$\omega_0^2 = (c_r + A^2/E);$$

$$A/s = N_1.$$

З огляду на наведені позначки рівняння (5) набудатиме вигляду:

$$m\ddot{x} + 2k\dot{x} + \omega_0^2x = F(t) - N_1u(x, \dot{x}). \quad (6)$$

Відповідно до схеми, поданої на рис. 1, величини коефіцієнта передачі сили:

$$n(t) = \frac{m\ddot{x} + 2k\dot{x} + \omega_0^2x}{F(t) - N_1u(x, \dot{x})} = \frac{m\ddot{x} + 2k\dot{x} + \omega_0^2x}{B}.$$

Задача побудови активної зустрічної дії полягає у визначенні такого закону $u^*(x, \dot{x})$, що за умови обмеження на його інтенсивність

$$\int_0^{2\pi/\omega} u^2 dt \leq V \quad (7)$$

забезпечуватиме в усталеному режимі мінімальне значення різниці $\Delta = n(t) - n_{\text{теор.}}$:

$$\min \Delta = \min [n(t) - n_{\text{теор.}}], \quad (8)$$

де $n_{\text{теор.}}$ визначається з теоретичних розрахунків потрібного значення зустрічної дії.

Керування динамічними властивостями системи можна здійснити, використовуючи такі підходи:

- змінювати налаштування системи з допомогою задання певного струму керування демпфера;
- застосовувати канал керування струмом демпфера, водночас замкнути систему за допомогою негативного зворотного зв'язку, наприклад за прискоренням.

Перший підхід відповідатиме керуванню за розімкненим циклом. У другому підході в активній системі створення сили протидії керування способом зміни величини струму в обмотці демпфера здійснюватиметься у функції координат (переміщення, швидкість, прискорення) руху фрагмента матриці. Тобто, на відміну від першого підходу, струм буде змінюватись у процесі керування, реалізуючи адаптивне керування процесом.

У нашій ситуації система керування магніторелогічним трансформатором гідроопори (МРТ гідроопорою) має реалізувати такі функції:

- оцінити ступінь ослаблення гідроопорою рівня силової дії пуансона;
- сформулювати сигнал керування в'язкістю рідини в демпферному каналі магніторелогічного трансформатора.

Для здійснення зазначених функцій система має містити: датчики тиску (сили); систему оброблення сигналів датчиків; потужний напівпровідниковий пристрій з блоками керування силовими ключами. Принцип дії системи автоматичного керування МРТ гідроопори запропонований на рис. 6.

До входу системи надходить інформація про поточні значення величини силової дії (тиску) з датчиків тиску ДТ1, встановленого на гідроопорі (основі 1, на яку діє пуансон), і ДТ2, встановленого на основі 2.

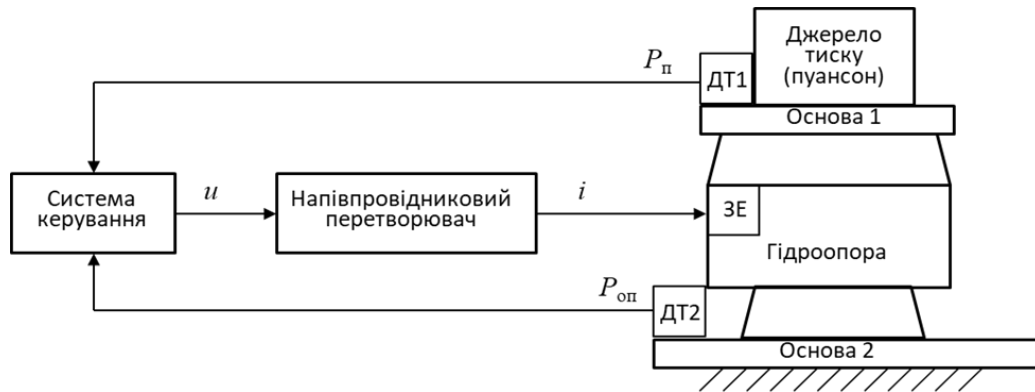


Рис. 6. Функціональна схема системи автоматичного керування гідроопорою формування зустрічної дії

За умови традиційного використання гідроопор у системах віброзахисту, де зовнішня дія є механічними коливаннями в певному спектрі частот, її передатну функцію визначають у такий спосіб. Керовану гідроопору подають у вигляді смугово-загороджувального фільтра. Частота налаштування гідроопори є такою, що придушує смугово-загороджувальний фільтр. Оскільки реальна гідроопора має гумову обичайку, яка гаситиме високочастотні гармоніки, то гідроопору подають як послідовне з'єднання двох ланок: смугово-загороджувального фільтра й передатної функції гумової обичайки. За таких умов передатна функція гідроопори набуває вигляду:

$$W_d = W_f \times W_a = \frac{r_0 s^2 + r_0 \omega_t^2}{s^2 + b \omega_t s + \omega_t^2} \times \frac{0,003s + 1}{0,005s + 1}, \quad (9)$$

де W_f – передатна функція смугово-загороджувального фільтра;

W_a – передатна функція обичайки;

r_0 – коефіцієнт підсилення;

b – ширина загородження;

ω_t – частота налаштування гідроопори.

У нашій ситуації силову дію пуансона під час формоутворення деталі можна подати як одиничну ступінчасту дію, тому частотні властивості гідроопори є менш суттєвими. За таких умов можна скористатися базовою моделлю механічної коливної системи, що містить об'єкт з масою (m) та пружним елементом (k) (рис. 7, а, б).

Структурна схема пояснює місце та роль пружного елемента, до входу якого прикладається зміщення (\bar{z}). Виходом ланки з передатною функцією $W_1 = k$ є сила. Ця сила прикладається

до входу другої ланки з передатною функцією $W_2 = 1/mp^2$. Вихідним сигналом є зміщення (\bar{y}). У такій системі мінімальної складності ланка $W_1 = k$ створює силу пружної взаємодії. Відповідно до теорії автоматичного керування згадана ланка матиме назву підсилювача. Друга ланка є інтегровальною ланкою другого порядку й «трансформує» пружну силу у зміщення.

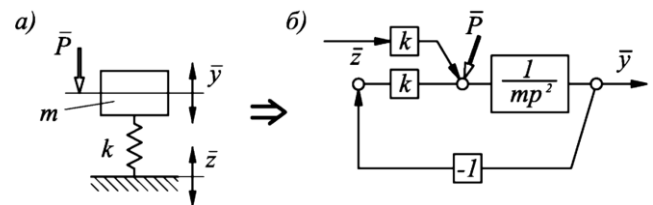


Рис. 7. Коливна механічна система:

а – базова модель; б – її структурна схема

Схеми, зображені на рис. 7, відповідають диференційному рівнянню

$$m\ddot{y} + ky = kz + P(t). \quad (10)$$

Рівняння (10) може бути реалізованим на основі принципу Даламбера або способом використання рівняння Лагранжа. Важливою умовою є те, що в базовій моделі «ключовим» елементом обирається матеріальна точка масою m , щодо якої складається рівняння кінестатики. Пружний елемент з коефіцієнтом жорсткості k таким чином стає елементарною ланкою.

Ускладнення структури може бути здійсненом через паралельне введення до пружного елемента додаткового кола, передатна функція якого може бути подана у вигляді дробово-раціональної функції:

$$W_{\text{од}}(p) = \frac{a_0 + a_1 p + \dots + a_n p^n}{b_0 + b_1 p + \dots + b_m p^m}, \quad (11)$$

де m, n – цілі числа ($n \leq m$);

a_i, b_j – коефіцієнти, що визначаються конструктивними особливостями гідроопори;

$i = 1, \bar{n}; j = 1, \bar{m}$.

Додаткове коло з передатною функцією $W_{\text{од}}(p)$

є додатковим зворотним зв'язком, що вводиться

до структурної схеми (рис. 7, б). У такому разі можна застосовувати відомі в теорії автоматичного керування принципи впровадження керування за абсолютним, відносним відхиленнями, а також за зовнішнім збуренням (рис. 8, а, б).

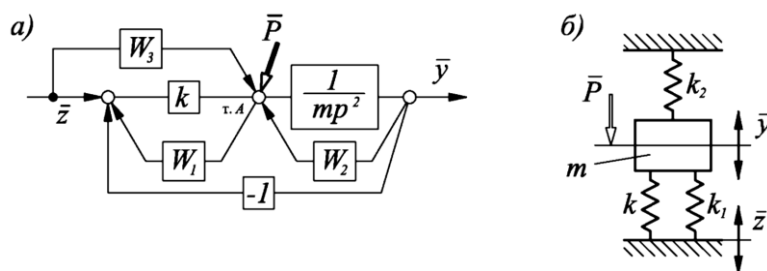


Рис. 8. Види додаткових зв'язків: а – на структурній; б – на розрахунковій схемах

Відповідно до рис. 5 у структурі схеми гідроопори буде присутнім лише один додатковий елемент (гумова обичайка) з передатною функцією $W_{\text{од01}} = W_a = (0,003s + 1)/(0,005s + 1)$ (на рис. 8, а він відповідатиме $W_1 = k_1$).

У такому разі передатна функція структури матиме вигляд:

$$W = \bar{y}/\bar{z} = \frac{k + W_{\text{од01}}}{mp^2 + k + W_{\text{од01}}}. \quad (12)$$

Отже, у розглянутій схемі будуть присутніми два типи елементарних ланок: пружний елемент та інтегровальна ланка другого порядку. Розглянемо детальніше їх частотні характеристики.

Пружний елемент із передатною функцією $W_1 = k$ має такі частотні характеристики:

- дійсна частотна характеристика $U_1 = k$;
- уявна частотна характеристика $V_1 = 0$;
- фазова частотна характеристика $\varphi_1 = 0$,

$\arctg \varphi_1 = 0$;

– амплітудно-частотна характеристика

$$A_1 = \sqrt{U_1^2 + V_1^2} = k;$$

– амплітудно-фазова частотна характеристика (АФЧХ) є вектором, що лежить на осі $0V$ і має довжину k (від частоти не залежить), як показано на рис. 9.

Інтегровальна ланка другого порядку має передатну функцію вигляду $W = A/p^2$ і частотні характеристики:

$$W(j\omega) = -A/\omega^2;$$

$$V(\omega) = 0;$$

$$U(\omega) = -A/\omega^2;$$

$$\operatorname{tg} \varphi = 0;$$

$$\varphi(\omega) = 0;$$

$$A(\omega) = A/\omega^2.$$

АФЧХ інтегровальної ланки другого порядку подана на рис. 10. Вектор зорієнтований у від'ємному напрямку щодо осі $0U$, а його величина є зворотно пропорційною квадрату частоти ω .

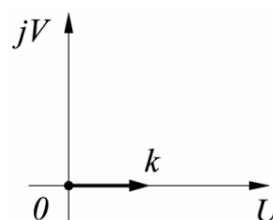


Рис. 9. АФЧХ пружного елемента

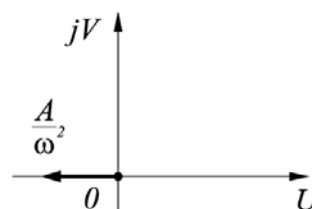


Рис. 10. АФЧХ інтегровальної ланки другого порядку

Елементарні ланки можуть мати різні форми фізичної реалізації. Так, пружний елемент зазвичай подається як гвинтові пружини, листові ресори тощо.

Ланки інтегровального типу мають більш складну природу, що пов'язано з використанням зовнішніх джерел енергії та появою складніших

структур самої ланки. У нашій ситуації це застосування магнітореологічного трансформатора.

**Керування силою тертя
(тобто силою притиску або коефіцієнтом тертя)**

Необхідність забезпечення певного оптимального значення сили тертя в процесі формування товстості деталі з рифтами є актуальною і для способу з використанням скороченого пуансона, і для інших способів. У роботі [7, 8] подані результати спроб керування коефіцієнтом тертя за допомогою нанесення певного шару мастила, але зазначено, що подібні рішення на сьогодні є не достатньо ефективними через неможливість контролювати та утримувати постійну товщину шару мастила протягом усього процесу формоутворення.

Тому був обраний метод, за яким здійснюється кероване притискання полки заготовки. Для його реалізації необхідно побудувати систему автоматичного керування величиною притискання за законом, що забезпечив би рівномірність розподілу механічних напружень на всіх ділянках матеріалу заготовки, які піддаються деформації в процесі виготовлення деталі.

Наразі для керування переміщенням робочих органів за допомогою гідроприводів використовують гідророзподільвачі з дискретними (увімкнено-вимкнено) електромагнітами (ГР) та дроселювальні гідророзподільвачі (ДГР). Приводи з пропорційним керуванням часто називають «інтелектуальна гідравліка».

Гідроприводи з ДГР мають високі точні та динамічні характеристики, однак вони є складнішими, ніж ГР. Можливість високоточного регулювання основних параметрів гідропривода дає змогу забезпечити жорсткі цикли керування. Водночас для більшості мобільних установок необхідна величина тиску в різних переходах робочого циклу не може бути встановленою наперед. У цьому разі доречно застосовувати регулятори *LS* (*Load Sensing* – чутливість до навантаження), у яких перепад тиску на дроселі регулювання швидкості (можливо пропорційному) стабілізується на певному рівні ΔP способом зміни фактичної подачі насоса в режимі online, що забезпечує мінімізацію втрат потужності в гідроприводі.

Під час припинення технологічних операцій центральний контролер подає команду на електронний блок двигуна й насоси. Унаслідок знижуються оберти колінчастого вала до частоти холостого ходу та встановлюються подачі насосів, близькі до нуля. Відновлення технологічної операції супроводжується збільшенням частоти обертання колінчастого вала та встановленням дроселя системи *LS* у робочий стан; у цьому разі тиск на виході з насоса стабілізується на рівні діючого опору навантаження плюс перепад тиску ΔP . На сьогодні гідроприводи з пропорційним електрокеруванням перебувають у стадії інтенсивного розвитку.

На рис. 11–14 зображені схеми гідропривода з *LS*-керуванням.

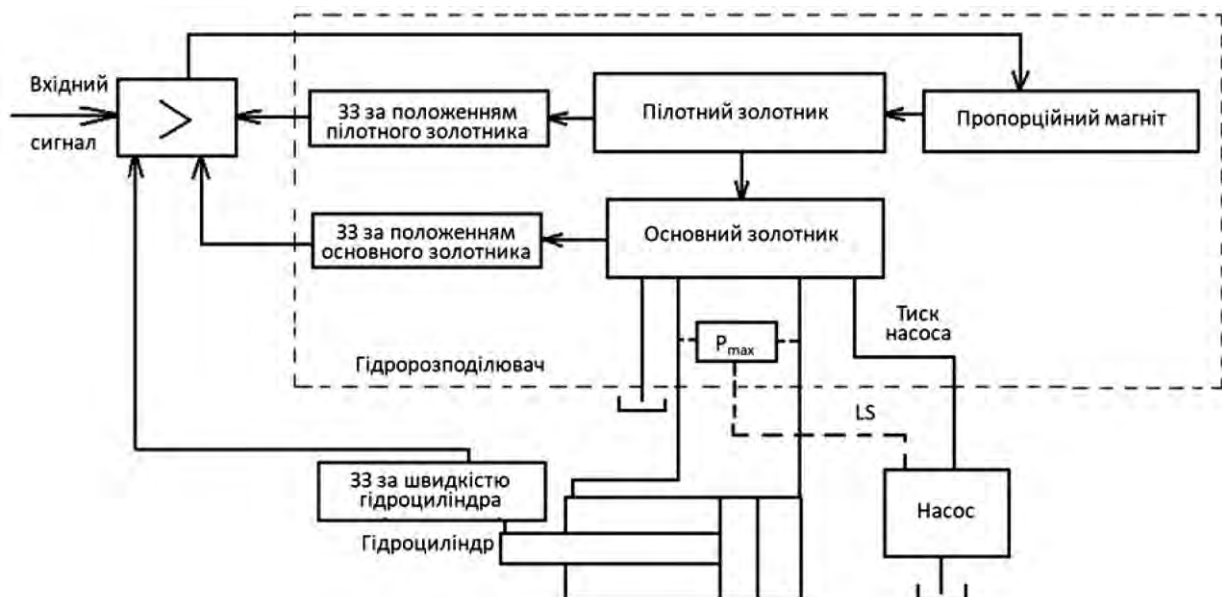


Рис. 11. Структурна схема гідропривода з *LS*-керуванням

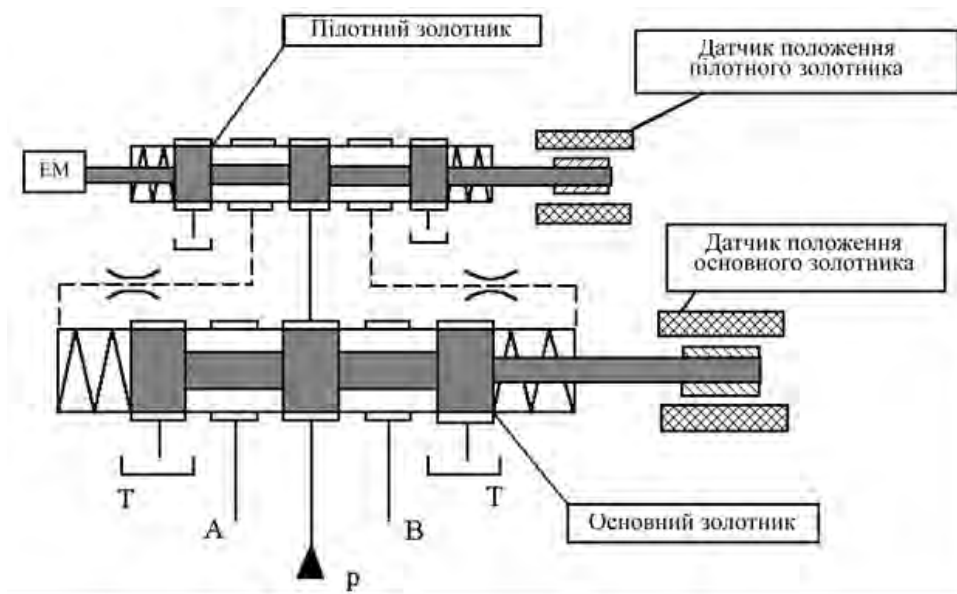


Рис. 12. Принципова схема пропорційного розподільника

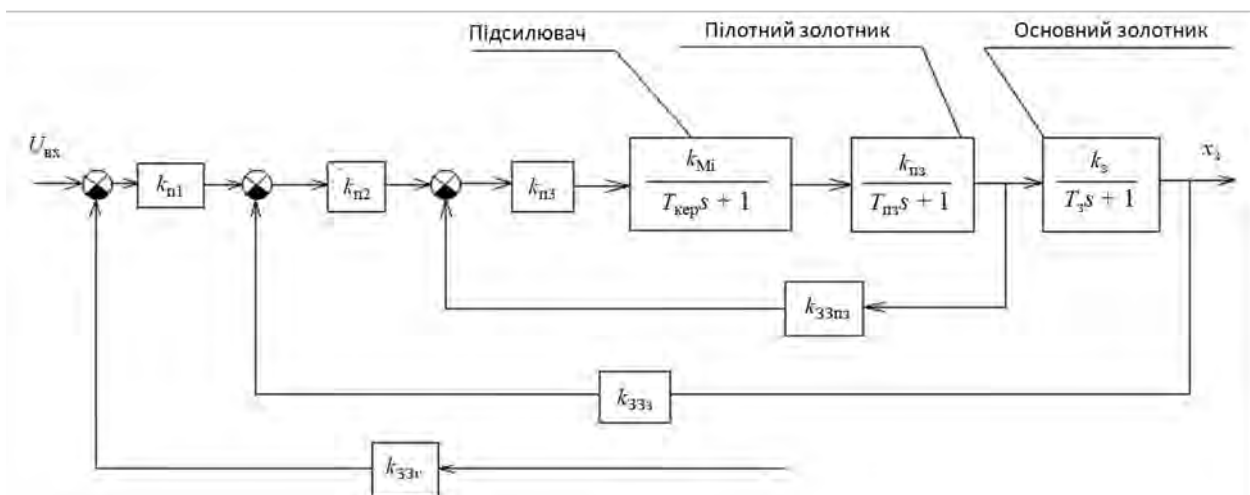


Рис. 13. Структурна схема пропорційного розподільника та електронного підсилювача. У поданій схемі:

$k_{ззпз}$ – коефіцієнт зворотного зв'язку за положенням пілотного золотника;

$k_{ззз}$ – коефіцієнт зворотного зв'язку за положенням основного золотника;

$k_{ззв}$ – коефіцієнт зворотного зв'язку за швидкістю основного золотника

У системах гідропривода з LS -керуванням миттєвий тиск навантаження слугує сигналом зворотного зв'язку для керуючого пристрою, який зі свого боку встановлює необхідний тиск насоса. Тиск підтримується таким, що дорівнює тиску найбільшого навантаження споживача плюс постійний тиск керування. За допомогою компенсаторів тиску підтримується постійний перепад тиску на дроселях A1 та A2 (рис. 13), що й визначає відсутність залежності швидкості споживача від його навантаження. Це і є основний принцип LS -системи.

Тому для реалізації операції керуваного притиску полки заготовки в пропонованій системі автоматичного керування процесом виготовлення товстолистової деталі з рифтами буде застосований саме такий принцип. Структурна схема цієї системи подана на рис. 11. Вхідним сигналом для неї буде сила притискання (тиск) робочого органа гідропривода, яка під час формування деталі підтримуватиметься на рівні, визначеному з попередніх теоретичних розрахунків параметрів формування, або ж змінюватиметься за законом,

визначеним з того самого попереднього аналізу технологічного процесу виготовлення деталі.

Крім цього, кількість точок прикладання сили притискання до полки під час формоутворення та їх просторове розміщення потребують додаткового окремого дослідження з огляду на умови забезпечення необхідних якісних показників виготовленої деталі.

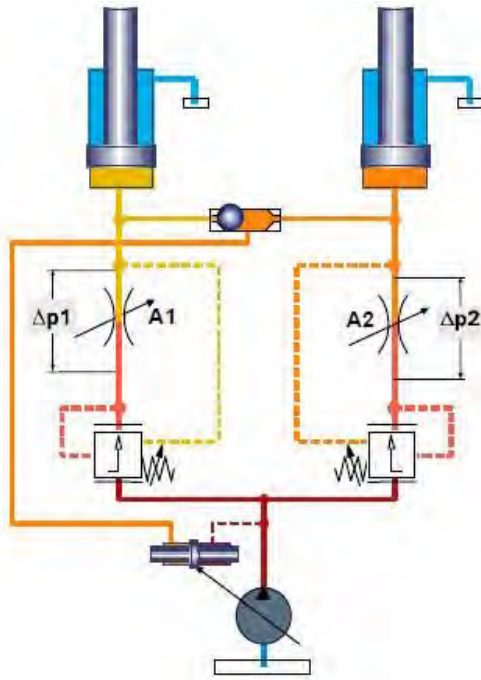


Рис. 14. Принципова схема гідросистеми з *LS*-керуванням

Висновки

Автоматизація процесу виготовлення товстостілових деталей із конструкційних і спеціальних сталей є актуальним завданням сучасного машинобудування.

Для оцінювання можливості побудови системи автоматизованого керування цим процесом у роботі запропоновано схему узагальненої технології автоматизованого виготовлення таких деталей. Розглянуто основні технологічні особливості кожного з етапів та можливі способи їх автоматизації з використанням сучасного обладнання.

Застосування запропонованої структури, складу та послідовності основних операцій узагальненого технологічного процесу автоматизованого виготовлення товстостінних листових деталей з рифтами на основі його реалізації обраними методами на запропонованому обладнанні дасть змогу побудувати систему автоматичного керування, здатну забезпечити ефективне управління якістю виготовлення зазначених деталей на кожному етапі технологічного процесу.

Тому подальший розвиток досліджень буде присвячено розробленню локальних систем керування обладнанням, яке забезпечує виконання операцій на кожному визначеному в цій роботі етапі технологічного процесу, а також побудові структури та алгоритму загальної системи керування, що синхронізуватиме та контролюватиме весь технологічний процес.

Список літератури

1. Черниш А. А. Удосконалення технології формування великогабаритних товстостілових деталей з елементами жорсткості з обмеженням осередку деформації: дис. к-та техн. наук: 05.03.05. Кременчук, 2017. 164 с. URL: <https://donntu.edu.ua/library/novinki-fondu/novinki-fondu-2018-r>
2. Спосіб штампування ребер жорсткості: пат. 89552 У Україна. № u201411388 ; заявл. 20.10.2014; опубл. 25.04.14, Бюл. № 8. URL: <http://specounsils.kdu.edu.ua/vidguk/Aref%20Chernysh.doc>
3. Мосьпан Д.В., Драгобецький В.В., Невлюдов І.Ш. До визначення параметрів системи автоматичного керування технологічним процесом вільного формоутворення. Вісник КрНУ ім. Михайла Остроградського. Кременчук: КрНУ, 2022. Вип. 4 (60). С. 33–38.
4. Тітов В.А., Шамарін Ю.Є., Долматов А.І. Високошвидкісні методи обробки металів тиском. Київ: КВІЦ, 2016. 204 с. URL: <https://docplayer.net/87096991-Tehnichni-nauki-technical-sciences-naukoviy-zhurnal-issn-herald-of-khmelnytskyi-national-university-hmelnickiy-scientific-journal.html>
5. Добрянський С. С., Малафеев Ю. М., Пуховський Є. С. Проектування і виробництво заготовок. Київ: НТУУ «КПІ», 2014. 353 с. URL: <https://ela.kpi.ua/bitstream/123456789/46090/1/Proektuvannia.pdf>
6. Черныш А. А. Формовка прямолинейных рифтов укороченным пуансоном. Вісник Кременчуцького національного університету імені Михайла Остроградського. Кременчук: КрНУ, 2015. Вип. 1/2015 (90). С. 70–74. URL: http://nbuv.gov.ua/UJRN/Vkdpu_2015_1%282%29_12
7. Gorbatyuk S. M., Shapoval A. A., Mos'pan D. V., Dragobetskii V. V. Physical Principles of Manufacturing of Periodic Profile Bars By Vibratory Drawing. *Izvestiya Visshikh Uchebnykh Zavedenii. Chernaya Metallurgiya (Izvestiya. Ferrous Metallurgy)*. 2016. Vol. 59, No. 7. P. 479–484. DOI: 10.17073/0368-0797-2016-7-479-484

8. Dragobetskii V.V., Shapoval A.A., Shchetynin V.T., Agrat R.G., Shlyk S.V., Gorbatyuk S.M., Markov O.E. New Solution For Plastic Deformation Process Intensification. *METALLURGIST*, 2022. Vol. 65. Issue 09-10. P. 1108–1116. DOI: 10.1007/s11015-022-01253-x
9. Huang D., Li R., Huang L., Ji V., Zhang T. Fretting wear behavior of bulk amorphous steel. *Intermetallics*. 2021. 19 (10). P. 1385–1389. DOI: <https://doi.org/10.3390/met13061137>
10. Ramesh R., Gnanamoorthy R. Development of a fretting wear test rig and preliminary studies for understanding the fretting wear properties of steels. *Materials and Design*. 2006. № 27(2). P. 141–146. DOI: 10.1016/j.matdes.2004.09.017
11. Драгобецький В. В., Мороз Н. Н., Троцько О. В. Оптимізація геометричних та технологічних параметрів процесу формозміни листових деталей з раціональним вибором інтенсифікуючих факторів. Вісник Національного технічного університету «ХПІ»: збірник наукових праць. Тематичний випуск: Нові рішення в сучасних технологіях. Харків: «ХПІ», 2009. № 32. С. 38–43. URL: <https://repository.kpi.kharkov.ua/server/api/core/bitstreams/59e4a80b-a09e-4117-a88d-d4318054e43d/content>
12. Lutsenko I., Fomovskaya O. et al. Development of a method for determining the optimal control trajectory for the periodic processes. *EUREKA: Physics and Engineering*. 2019. Vol. 3. P. 59–68. DOI: 10.21303/2461-4262.2019.00900
13. Тришевский О. И. Методика расчета технологических параметров формовки в валках периодически замкнутых гофров на листовых и гнутых профилях. Кузнечно-штамповочное производство. Харьков, 2012. № 1. С. 25–28.
14. Puzyr R., Savelov D., Argat R., Chernish A. Distribution analysis of stresses across the stretching edge of die body and bending radius of deforming roll during profiling and drawing of cylindrical workpiece. *Metallurgical and mining industry*. 2015. No. 1. P. 27–32. URL: https://www.metaljournal.com.ua/assets/Journal/english-edition/MMI_2015_1/6%20Puzyr.pdf
15. Попович В., Кондр А., Плешаков Е. Технологія конструкційних матеріалів і матеріалознавство: навч. посіб. Львів: Папуга, 2004. 422 с.

References

1. Chernysh A. A. "Improvement of the technology of forming large-sized thick-sheet parts with hardness elements with limitation of the center of deformation". Thesis of the candidate of technical sciences: 05.03.05. Kremenchuk, 2017. 164 p. available at: <https://donntu.edu.ua/library/novinki-fondu/novinki-fondu-2018-r>
2. "Method of stamping stiffeners". Patent 89552 U Ukraine. № u201411388 ; declared 20.10.2014; published 25.04.14, Iss. № 8. available at: <http://speccounsils.kdu.edu.ua/vidguk/Aref%20Chernysh.doc>
3. Mospan D., Dragobetskii V., Nevliudov I. (2022), "To determine the parameters of the automatic control system of the technological process of free forming". Scientific journal Transactions of Kremenchuk Mykhailo Ostrohradskyi National University, Issue 4 (60). P. 33–38.
4. V. A. Titov, Yu. Ye. Shamarin, A. I. Dolmatov. "High-speed methods of processing metals by pressure" ["Vysokoshvydkisni metody obrobky metaliv tyskom"]. Kyiv: KVITs, 2016. 204 p. available at: <https://docplayer.net/87096991-Tehnichni-nauki-technical-sciences-naukoviy-zhurnal-issn-herald-of-khmelnyskyi-national-university-hmelnickiy-scientific-journal.html>
5. Dobrianskyi S. S., Malafieiev Yu. M., Pukhovskiy Ye. S. "Design and production of blanks" ["Proektuvannia i vyrobnytstvo zahotovok"]. K.: NTUU «KPI», 2014. 353 p. available at: <https://ela.kpi.ua/bitstream/123456789/46090/1/Proektuvannia.pdf>
6. Chernysh A. A. "Forming rectilinear rifts with a shortened punch". Scientific journal Transactions of Kremenchuk Mykhailo Ostrohradskyi National University. Issue 1/2015 (90). 2015. P. 70–74. available at: http://nbuv.gov.ua/UJRN/Vkdpu_2015_1%282%29_12
7. Gorbatyuk, S. M., Shapoval, A. A., Mospan, D. V., Dragobetskii, V. V. (2016), "Physical Principles of Manufacturing of Periodic Profile Bars By Vibratory Drawing". *Izvestiya Visshikh Uchebnykh Zavedenii. Chernaya Metallurgiya (Izvestiya. Ferrous Metallurgy)*. Vol. 59, No. 7. P. 479–484. DOI: 10.17073/0368-0797-2016-7-479-484
8. Dragobetskii, V.V., Shapoval, A.A., Shchetynin, V.T., Agrat, R.G., Shlyk, S.V., Gorbatyuk, S.M., Markov, O.E. (2022), "New Solution For Plastic Deformation Process Intensification". *METALLURGIST*. Vol. 65. Issue 09-10. P. 1108-1116. DOI 10.1007/s11015-022-01253-x
9. Huang, D., Li, R., Huang, L., Ji, V., Zhang, T. (2021), "Fretting wear behavior of bulk amorphous steel". *Intermetallics*. 19 (10). P. 1385-1389. DOI: <https://doi.org/10.3390/met13061137>
10. Ramesh, R., Gnanamoorthy, R. (2006), "Development of a fretting wear test rig and preliminary studies for understanding the fretting wear properties of steels". *Materials and Design*. № 27(2). P. 141-146. DOI: 10.1016/j.matdes.2004.09.017
11. Dragobetskii V. V., Moroz N. N., Tratsko O. V. (2009). "Optimization of geometric and technological parameters of the process of forming sheet parts with a rational choice of intensifying factors". Bulletin of the National Technical University. KhPI: Collection of scientific papers. Thematic issue: New solutions in modern technologies. No 32. 2009. P. 38–43. available at: <https://repository.kpi.kharkov.ua/server/api/core/bitstreams/59e4a80b-a09e-4117-a88d-d4318054e43d/content>
12. Lutsenko, I., Fomovskaya, O. et al. (2019), "Development of a method for determining the optimal control trajectory for the periodic processes". *EUREKA: Physics and Engineering*. Vol. 3. P. 59–68. DOI: 10.21303/2461-4262.2019.00900

13. Tryshevskiy O. Y. (2012), *Metodyka rascheta tekhnolohycheskykh parametrov formovky v valkakh peryodychesky zamknytykh hofrov na lystovykh y hnutykh profyliakh*. Kuznechno-shtampovochnoe proyzvodstvo. Kharkiv. № 1. P. 25–28.

14. Puzyr R., Savelov D., Argat R., Chernish A. "Distribution analysis of stresses across the stretching edge of die body and bending radius of deforming roll during profiling and drawing of cylindrical workpiece". Metallurgical and mining industry. No. 1. 2015. P. 27–32. available at: https://www.metaljournal.com.ua/assets/Journal/english-edition/MMI_2015_1/6%20Puzyr.pdf

15. V. Popovych, A. Kondyr, E. Pleshakov (2004), *Tekhnolohiia konstruksiiynykh materialiv i materialoznavstvo: navch. posibnyk*. Lviv: Vydavnytstvo «Papuha». 422 p..

Надійшла 15.09.2023

Відомості про авторів / About the Authors

Моспан Денис Владиславович – кандидат технічних наук, доцент, Кременчуцький національний університет імені Михайла Остроградського, доцент кафедри комп'ютерної інженерії та електроніки, Кременчук, Україна; e-mail: denis.mospan@gmail.com; ORCID ID: <https://orcid.org/0000-0001-5473-7874>

Mospan Denis – Phd (Engineering Sciences), Associate Professor, Kremenchuk Mykhailo Ostrohradskyi National University, Associate Professor at the Department of Computer Engineering and Electronics, Kremenchuk, Ukraine.

AUTOMATION OF THE TECHNOLOGICAL PROCESS OF FORMING DETAILS WITH STRAIGHT RIFTS

The **subject matter** of the article is technological line for the production of thick sheet parts with rifts, made of structural and special steels. The **goal** of the work is analysis and creation of methods of managing basic technological operations at all stages of the technological process of the automated production of such parts. The following **tasks** were solved in the article: formation of models of the main stages of forming parts with rectilinear rifts and control of dynamic properties of selected systems of their implementation. The following **methods** used are – methods of numerical integration and numerical methods of solving differential equations; analytical solution of differential equations based on the D'Alembert principle or by using the Lagrange equation; frequency analysis of transfer functions of individual links, presented in the form of fractional-rational functions. The following **results** were obtained – constructed in the first approximation the composition and sequence of the main operations of the generalized technological process of the automated production of thick-walled sheet parts with rifts; according to the results of the analysis of the process of elastic-plastic change of the shape of the region with many connections, the input and output parameters necessary for building a control model of the specified process were determined; it is determined that for its implementation it is necessary to ensure free deformation of at least 0.25 % of the rift surface area outside of contact with the forming tool, and the necessary technological methods for this are proposed; proposed variants of the technical implementation of the selected technological methods; the results of modeling schemes and modes of operation of the necessary technical equipment are presented. **Conclusions:** The application of the proposed structure, composition and sequence of main operations of the generalized technological process of the automated production of thick-walled sheet parts with cracks, based on its implementation by the selected methods on the proposed equipment, will allow to ensure the effective management of the quality of the production of the specified parts at each stage of the technological process.

Keywords: automation of the technological process; simulation of the main stages of the process and equipment operation modes; automatic control system; frequency characteristics.

Бібліографічні описи / Bibliographic descriptions

Моспан Д. В. Автоматизація технологічного процесу формоутворення деталей із прямолінійними рифтами. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 163–173. DOI: <https://doi.org/10.30837/ITSSI.2023.254.163>

Mospan, D. (2023), "Automation of the technological process of forming details with straight rifts", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 163–173. DOI: <https://doi.org/10.30837/ITSSI.2023.254.163>

JIAO HUNKUN, O. AVRUNIN

FEASIBILITY ANALYSIS OF IMPLANT MOVEMENT ALONG ARC TRAJECTORY UNDER NON-CONTACT CONTROL IN MAGNETIC STEREOTAXIC SYSTEM

In this paper, the non-contact control of magnetic implants by changing the external magnetic field in the magnetic stereotaxic system is introduced, and the feasibility of making them move along the arc trajectory is analyzed. Through COMSOL software, the process of moving the miniature magnetic implant along the arc trajectory was simulated, the change of the micro-magnetic implant trajectory after the external magnetic field was changed, the relative position relationship between the large permanent magnets was determined, and the mechanical analysis of the miniature magnetic implant moving along the arc trajectory was carried out. In this experiment, we fix a large permanent magnet, only move the second permanent magnet, first, observe the process of small permanent magnets moving along a straight trajectory, determine the position of the large permanent magnet magnetic field when it contacts the small permanent magnet, and then, analyze the force of the small permanent magnet through the force calculation module, and determine the relative position relationship between the two large permanent magnets by comparing F_x and F_y , and when the small permanent magnet will start to move along the arc trajectory. Then, according to the previous data, we move two adjacent large permanent magnets at the same time at a certain interval, record the movement trajectory of the small magnet, Finally, with the force calculation module of the COMSOL software, force analysis of small permanent magnets moving along arc trajectories. The data from this experiment will be used to determine the relative position relationship between two large permanent magnets adjacent to each other during the actual experiment, and under what conditions the small permanent magnets will move along the arc trajectory. The purpose of this experiment is to provide theoretical and data support for the subsequent practical experiments of the magnetic stereotaxic system, and all parameters in the COMSOL software are derived from the actual measurement data, so as to improve the reliability of the simulation results.

Keywords: Human health; Magnetic field; COMSOL Software; Permanent magnets; Force analysis.

Magnetic stereotaxis system

In traditional neurosurgery for brain tissue, craniotomy is usually performed using wire-guided, mechanically controlled surgical instruments with stereotaxic devices [1–5]. The limitation of this method is that the surgical instruments can only move along a straight trajectory, resulting in limited surgical access, great trauma to the tissues around the movement path, and inability to enter the deep structure of brain tissue, which cannot effectively treat brainstem tumors and some extrapyramidal nerve diseases.

Therefore, the magnetic stereotaxic system [6–11] came into being. This method was proposed in 1990 by Howard M. A., Mayburg M., Grady M. S. At present, the method is still in the experimental stage. The magnetic stereotaxic system consists of an external magnetic field, miniature magnetic implants. It works by probing magnetic surgical instruments into the skull, guiding micro-magnetic implants with catheters along pre-designed and calculated trajectories to the lesion via

an external magnetic field, providing hyperthermia to tumors located in deep structures of brain tissue or delivering targeted drugs through catheters. In contrast to traditional neurosurgery, magnetic stereotaxic systems provide non-contact control of miniature magnetic surgical instruments by changing the external magnetic field, and under computer-aided simulation, multiple surgical pathways can be established, allowing the implant to travel along any route to almost any location in the brain. This minimizes the invasiveness of the implant to surrounding tissues during movement. Since the external magnetic field is controlled by the computer according to a pre-designed program and multiple simulations are required before surgery, the difficulty of operation by medical staff is greatly reduced, and the intraoperative safety is greatly improved. These advantages make the magnetic stereotaxic system one of the least invasive methods for surgical intervention against brain tumors, and one of the most promising methods.

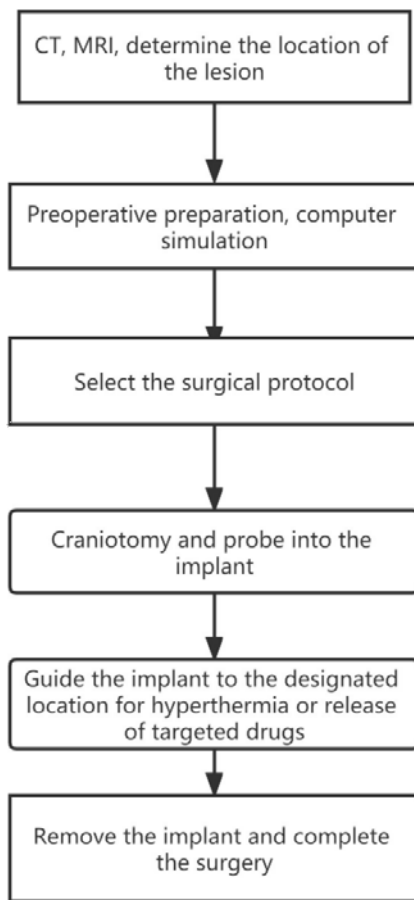


Fig. 1. Introduction to the procedure for magnetic stereotaxic surgery

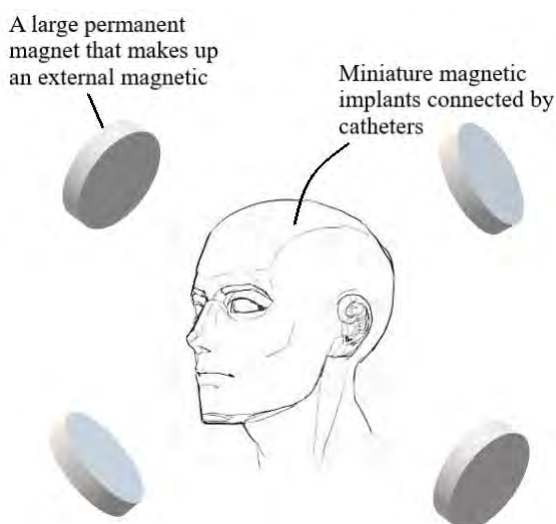


Fig. 2. Conceptual diagram of magnetic stereotaxic system

In this simulation experiment, the external magnetic field of the magnetic stereotaxic system is composed of four large permanent magnets [12, 13], and the small permanent magnets simulate implants. Computer simulation software chooses COMSOL 6.0 [14–16].

Computer simulation experiments

Experimental Objective

To control the miniature magnetic implant non-contact by the magnetic stereotaxic system to make it move along the arc trajectory.

Experimental design

Establish a coordinate system in 3D space, small cylindrical permanent magnets are located in the center of the coordinate system, large permanent magnets are set on the $+X$ and $+Y$ axes, parameters are set in the simulation software according to the experimental material data, and geometric models are constructed.

Simulation experiment parameter settings

- (1) $d_{ion} = 100[\text{mm}] = 0.1[\text{m}]$ (Large permanent magnet diameter);
- (2) $t_{ion} = 10[\text{mm}] = 0.01[\text{m}]$ (Large permanent magnet thickness);
- (3) $d_{NFB} = 1[\text{mm}] = 0.001[\text{m}]$ (Small permanent magnet diameter);
- (4) $l_{NFB} = 2[\text{mm}] = 0.002[\text{m}]$ (Small permanent magnet thickness);
- (5) The small permanent magnet boundary is limited to a cylinder with a radius of 0.1 [m] and a height of 0.5 [m];
- (6) Large permanent magnets have a maximum distance of 0.41 [m] from the center of 3D space and a closest distance of 0.11 [m], the movable distance of large permanent magnets is 0.3 [m].

Simulation experiment material property settings

Permanent magnets:

- (1) Conductivity $\sigma = 1/1.4[\text{uohm}\times\text{m}][\text{S}/\text{m}]$;
- (2) Relative permittivity $\epsilon_r = 1[1]$;
- (3) Recovery permeability $\mu_{rec} = 1.02$;
- (4) Residual flux density norm $normB_r = 1.3[\text{T}]$;

air:

- (1) Conductivity $\sigma = 0[\text{uohm}\times\text{m}][\text{S}/\text{m}]$;
- (2) Relative permittivity $\epsilon_r = 1[1]$;
- (3) Recovery permeability $\mu_{rec} = 1$.

All the above parameters are theoretical parameters and can be set according to the material parameters required for actual experiments.

Depending on the parameter settings, create a geometric model in 3D space:

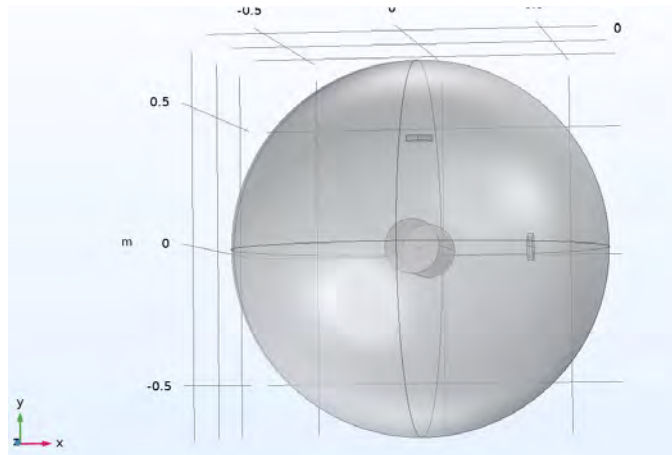


Fig. 3. Geometric models based on parameters in COMSOL software

Equations and formulas for simulation experiments

In this simulation experiment, the external magnetic field of the magnetic stereoscopic positioning system is composed of large permanent magnets, and the magnetic field of the permanent magnets is a static magnetic field. Therefore, "magnetic field, no current (MFNC)" was selected in the COMSOL physics.

In the constitutive relations $B-H$, the magnetization model [17] needs to introduce the magnetization vector field M , and the magnetic field strength H , which are expressed as:

$$H = \frac{B}{\mu_0} - M \tag{1}$$

or:

$$B = \mu_0 (H + M), \tag{2}$$

where μ_0 – is the magnetic permeability;

B is Magnetic flux density.

The remaining flux density is then selected in the magnetization model with the following expression:

$$B = \mu_0 \mu_{rec} H + B_r ; \tag{3}$$

$$B_r = \|B_r\| \frac{e}{\|e\|}, \tag{4}$$

where B is the magnetic flux density; μ_0 is the vacuum permeability; μ_{rec} is recoil permeability; B_r is residual flux density; $\|B_r\|$ is residual flux density norm; e is residual flux direction.

Simulation results

The large permanent magnet located on the $+X$ axis begins to move towards the center position. When the first large permanent magnet is 0.36 [m] from the center position, its magnetic field gradually touches the small permanent magnet.

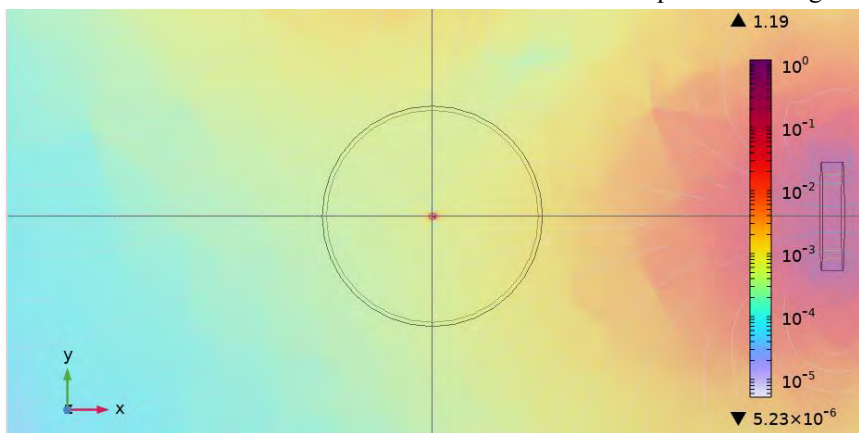


Fig. 4. The magnetic field of a large permanent magnet gradually touches the small permanent magnet

When its magnetic field touches a small permanent magnet, the small permanent magnet begins to approach the first large permanent magnet.

When the small permanent magnet moves to the boundary position, it is blocked by the boundary and stays on the boundary.

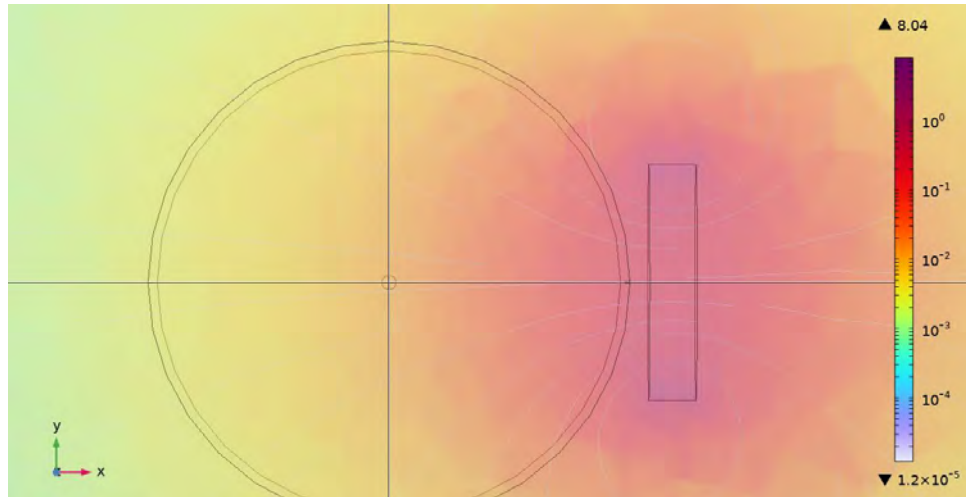


Fig. 5. Small permanent magnets are blocked by the boundary and stay on the boundary

The first large permanent magnet began to move away, and the second large permanent magnet gradually approached, at this time, the magnetic field lines around the small permanent magnet could be clearly seen, but the same simulation software could not intuitively see that the magnetic field of the second large permanent magnet touched the small permanent magnet.

To study the movement of small permanent magnets along arc trajectories under non-contact control, therefore, we use the "Force Calculation" module included in the software to analyze the force of small

permanent magnets located at boundary locations. The distance between the first large permanent magnet and the center is dis_ion , the distance of the second large permanent magnet from the center is dis_ion1 , F_x is the X -axis component of the force experienced by the small permanent magnet, and F_y is the on- Y -axis component of the force experienced by the small permanent magnet.

First, let's assume that the first large permanent magnet stays 0.22 [m] from the center and only the second large permanent magnet is running.

Table 1. Force analysis of a small permanent magnet at the boundary position when the position of the first large permanent magnet remains unchanged and the second large permanent magnet gradually approaches the center position

dis_ion [m]	dis_ion1 [m]	F_x [N]	F_y [N]
0.22	0.24	7.199×10^{-9}	5.408×10^{-9}
0.22	0.23	7.916×10^{-9}	6.877×10^{-9}
0.22	0.22	8.715×10^{-9}	8.561×10^{-9}
0.22	0.21	9.883×10^{-9}	11.541×10^{-9}
0.22	0.20	11.180×10^{-9}	15.123×10^{-9}
0.22	0.19	12.598×10^{-9}	19.332×10^{-9}
0.22	0.18	14.321×10^{-9}	26.112×10^{-9}

From the above table, we find that when the distance from the second large permanent magnet to the center position is less than the distance from the first large permanent magnet to the center position, the Y -axis component of the force of the small

permanent magnet is greater than the X -axis component. That is, when $dis_ion1 < dis_ion$, $F_y > F_x$. To verify this result, we set the first large permanent magnet to be located 0.18 [m] from the center and re-analyzed the force on the small permanent magnet.

Table 2. Force analysis of a small permanent magnet when the first large permanent magnet is 0.18 [m] from the center

dis_ion [m]	dis_ion1 [m]	F_x [N]	F_y [N]
0.18	0.21	22.726×10^{-9}	13.959×10^{-9}
0.18	0.20	24.941×10^{-9}	17.814×10^{-9}
0.18	0.19	27.614×10^{-9}	23.142×10^{-9}
0.18	0.18	31.381×10^{-9}	31.143×10^{-9}
0.18	0.17	35.683×10^{-9}	41.613×10^{-9}
0.18	0.16	40.918×10^{-9}	56.554×10^{-9}
0.18	0.15	48.237×10^{-9}	81.361×10^{-9}

According to Table 1 and Table 2, we can clearly see that when the distance from the first large permanent magnet to the center position is greater than the distance from the second large permanent magnet to the center position, the force on the Y axis of the small permanent magnet is greater than the component force on the X axis, that is, when $dis_ion1 < dis_ion$, $F_x < F_y$.

In order to facilitate practical experiments, we set the distance from the first large permanent magnet to the center position and the distance from the second large permanent magnet to the center position to be 0.05 [m], that is, when the first large permanent magnet begins to move 0.05 [m] to the center position, the second large permanent magnet begins to approach the center

position, and the two large permanent magnets move at the same speed.

During the movement of the two large permanent magnets, we perform a force analysis on the small permanent magnets again. In order to determine the position of the two large permanent magnets when the small permanent magnet initially moves, the position of the small permanent magnet does not change during the analysis, and always stays at the boundary position. It can be determined that when the first large permanent magnet is closest to the center, the small permanent magnet is affected by its magnetic field and stays at the boundary position on the $+X$ axis, so the initial positions of the two large permanent magnets are $dis_ion = 0.11$ [m], $dis_ion1 = 0.16$ [m], respectively.

Table 3. Two large permanent magnets operate at the same time, and the force analysis of small permanent magnets

dis_ion [m]	dis_ion1 [m]	F_x [N]	F_y [N]
0.11	0.16	443.82×10^{-9}	97.217×10^{-9}
0.12	0.15	234.78×10^{-9}	120.96×10^{-9}
0.13	0.14	188.50×10^{-9}	148.66×10^{-9}
0.14	0.13	167.14×10^{-9}	202.43×10^{-9}
0.15	0.12	157.48×10^{-9}	289.53×10^{-9}
0.16	0.11	152.98×10^{-9}	442.38×10^{-9}
0.17	0.12	100.73×10^{-9}	270.51×10^{-9}
0.18	0.13	67.94×10^{-9}	176.42×10^{-9}
0.19	0.14	46.69×10^{-9}	115.81×10^{-9}
0.20	0.15	32.37×10^{-9}	75.34×10^{-9}
0.21	0.16	22.83×10^{-9}	52.09×10^{-9}

It can be clearly seen from Table 3 that when the first large permanent magnet begins to gradually move away, the second large permanent magnet gradually approaches, F_x gradually decreases, F_y gradually increases, when the second large permanent magnet moves 0.11 [m] from the center position, F_y reaches the maximum, and then the second large permanent magnet begins to gradually move away and F_y gradually

decreases. Thus, when the distance from the second large permanent magnet to the center position is less than the distance from the first large permanent magnet to the central position, that is, when $dis_ion1 < dis_ion$, $F_x < F_y$, at which point the small permanent magnet begins to move along the boundary towards the second large permanent magnet.

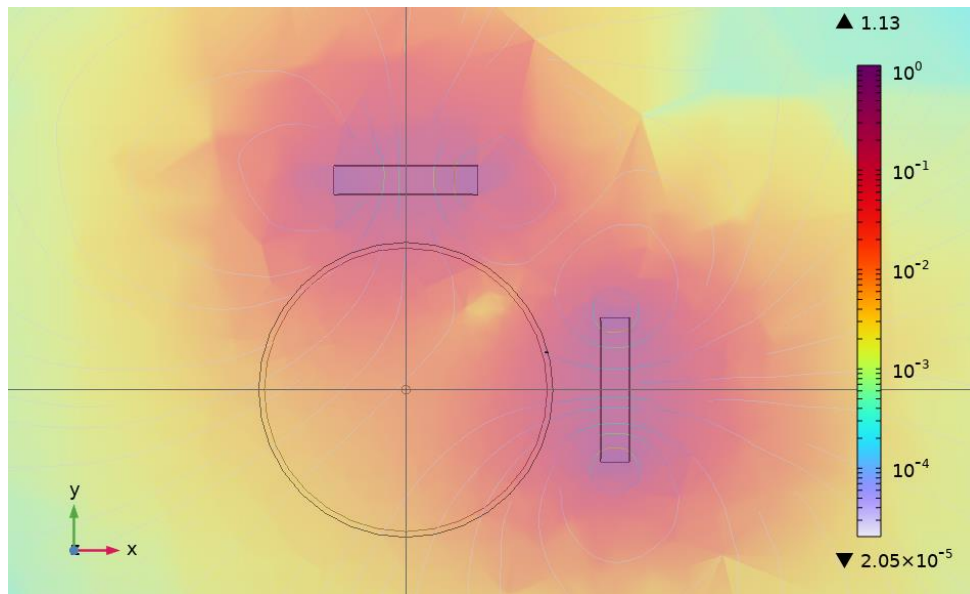


Fig. 6. The small permanent magnet begins to move along the boundary towards the second large permanent magnet

The closer to the second large permanent magnet, the greater the influence of its magnetic field on the small permanent magnet, that is, the greater the F_y , but at this time the small permanent magnet is still affected by the magnetic field of the first large permanent magnet, that is, F_x gradually decreases, which makes the small permanent magnet move along the boundary to the second large permanent magnet.

The small permanent magnet is captured by the magnetic field of the second large permanent magnet and stays on the boundary of the $+Y$ axis. That is, when the second large permanent magnet is closest to the center, i.e., $dis_ion1 = 0.11[m]$, the small permanent magnet has moved to the boundary position along the arc trajectory.

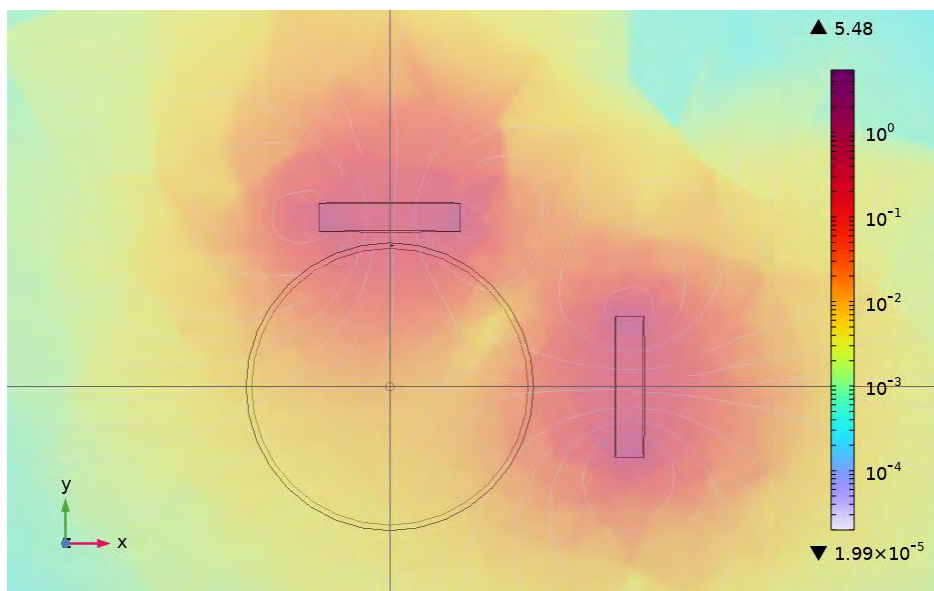


Fig. 7. The small permanent magnet reaches the end of the arc trajectory

As the two large permanent magnets move away, the small permanent magnets remain at the boundary position.

In order to determine the average velocity of a small permanent magnet moving along an arc trajectory towards a second large permanent magnet, we need

to perform a force analysis on the movement of the small permanent magnet. n is the angle of the arc

trajectory of the small permanent magnet moving along the boundary.

Table 4. Force analysis during the movement of small permanent magnets

dis_ion [m]	dis_ionl [m]	n	F_x [N]	F_y [N]
0.130	0.140	0	188.50×10^{-9}	148.66×10^{-9}
0.135	0.135	15	176.37×10^{-9}	172.81×10^{-9}
0.140	0.130	30	167.14×10^{-9}	202.43×10^{-9}
0.145	0.125	45	159.59×10^{-9}	236.59×10^{-9}
0.150	0.120	60	157.48×10^{-9}	289.53×10^{-9}
0.155	0.115	75	155.41×10^{-9}	353.71×10^{-9}
0.160	0.110	90	152.98×10^{-9}	442.38×10^{-9}

In order to intuitively see the force change of the small permanent magnet during the moving process, according to the above table, we draw

a line diagram of the two component forces F_x and F_y of the small permanent magnet during the moving process.

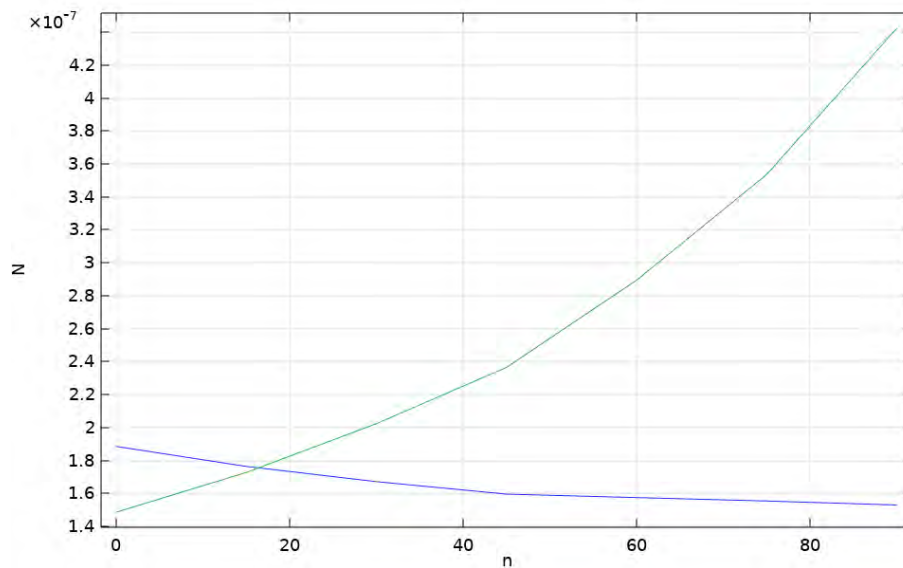


Fig. 8. Force analysis of small permanent magnets as they move along an arc trajectory

Simulation experiment conclusion

Through the above simulation experiments, we find that in the process of moving the small permanent magnet to the second large permanent magnet, as the F_y gradually increases, F_x gradually decreases, and the force direction of the small permanent magnet gradually points to the Y axis direction, therefore, the small permanent magnet moves along the boundary towards the second large permanent magnet, and the trajectory is an arc. According to Tables 1–3, we find that when the distance of the second large permanent

magnet from the center position is less than the distance of the first large permanent magnet from the center, that is, at $dis_ionl < dis_ion$, the direction of the magnetic field around the small permanent magnet located at the boundary position of the X -axis changes, the F_y increases, and the small permanent magnet begins to move towards the second large permanent magnet. Through Table 4, we calculate the changes of F_x and F_y during the movement of small permanent magnets along the boundary when the dis_ion and dis_ionl differ by 0.05 [m], and determine the relative positions of

the two large permanent magnets for subsequent practical experiments.

Through the simulation experiment of COMSOL software, we can clearly observe that by changing the external magnetic field, under non-contact control,

small permanent magnets can move along an arc trajectory. This shows that it is feasible to non-contact control the implant to move along the arc trajectory in the magnetic stereotaxic system, and at the same time provides a theoretical basis for subsequent experiments.

References

1. Kall, B A. (1987), "The impact of computer and imaging technology on stereotactic surgery". *Stereotactic and Functional Neurosurgery*. No. 50(1–6): 9–22. DOI: 10.1159/000100676
2. Avrunin, O., Tymkovich, M., Semenets, V., & Piatyokop, V. (2019), "Computed tomography dataset analysis for stereotaxic neurosurgery navigation". *Paper presented at the Proceedings of the International Conference on Advanced Optoelectronics and Lasers, CAOL*. P. 606–609. DOI:10.1109/CAOL46282.2019.9019459
3. Avrunin, O. G., Alkhorayef, M., Saied, H. F. I., & Tymkovich, M. Y. (2015), "The surgical navigation system with optical position determination technology and sources of errors". *Journal of Medical Imaging and Health Informatics*. No. 5(4), P. 689–696. DOI:10.1166/jmihi.2015.1444
4. Avrunin, O. G., Tymkovich, M. Y., Moskovko, S. P., Romanyuk, S. O., Kotyra, A., & Smailova, S. (2017), "Using a priori data for segmentation anatomical structures of the brain". *Przegląd Elektrotechniczny*. No. 93(5), P. 104–107. DOI:10.15199/48.2017.05.20
5. Chen, Y, Godage, I, Su, H. et al. (2019), "Stereotactic systems for MRI-guided neurosurgeries: a state-of-the-art review". *Annals of biomedical engineering*, No. 47(1–4). P. 335–353.
6. Hunkun, Jiao, Avrunin, O. (2023), "Explore the feasibility study of magnetic stereotaxic system". *Optoelectronic Information-Power Technologies*, No. 45(1). P. 86-96. DOI:10.31649/1681-7893-2023-45-1-86-96
7. Grady, S M, Howard, III M A, Broaddus, W C, et al. "Magnetic stereotaxis: a technique to deliver stereotactic hyperthermia". *Neurosurgery*, 1990, No. 27(6). P. 1010–1016. available at: <https://pubmed.ncbi.nlm.nih.gov/2274121/>
8. Nelson, B. J., Gervasoni, S., Chiu, P. W. Y., et al.(2022), "Magnetically actuated medical robots: An in vivo perspective". *Proceedings of the IEEE*. No. 110(7). P. 1028–1037. DOI:10.1109/JPROC.2022.3165713
9. Grady, M. S., Howard, M. A., Dacey, R. G., et al.(2000), "Experimental study of the magnetic stereotaxis system for catheter manipulation within the brain". *Journal of neurosurgery*. No. 93(2). P. 282–288. DOI: 10.3171/jns.2000.93.2.0282
10. Hunkun, J., Avrunin, O. (2022), "Possibilities of Field Formation by Permanent Magnets in Magnetic Stereotactic Systems", *IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek)*, Kharkiv, Ukraine. P. 1–4. DOI: 10.1109/KhPIWeek57572.2022.9916450
11. Coey, J. M. D. (2002), "Permanent magnet applications". *Journal of Magnetism and Magnetic Materials*. No. 248(3). P. 441–456. DOI: 10.1016/S0304-8853(02)00335-9
12. Calin, M. D., Helerea, E. "Temperature influence on magnetic characteristics of NdFeB permanent magnets". *7th international symposium on advanced topics in electrical engineering (ATEE)*. IEEE, 2011. P. 1–6. available at: https://www.researchgate.net/publication/241186295_Temperature_influence_on_magnetic_characteristics_of_NdFeB_permanent_magnets
13. Pepper, D W, Heinrich, J C. "The finite element method: basic concepts and applications with MATLAB, MAPLE, COMSOL". CRC press. 2017. 628 p. available at: <https://handoutset.com/wp-content/uploads/2022/05/The-finite-element-method-basic-concepts-and-applications-with-MATLAB-MAPLE-and-COMSOL-Heinrich-Juan-C.-Pepper-Darrell-W.pdf>
14. Pryor, R W. "Multiphysics modeling using COMSOL: a first principles approach". Jones & Bartlett Publishers, 2009. 871 p. available at: <http://dl.poweren.ir/downloads/PowerEn/Book/2019/Mar/%D%A9%D8%AA%D8%A7%D8%A8%20%D8%A2%D9%85%D9%88%D8%B2%D8%B4%20%D8%AC%D8%A7%D9%85%D8%B9%20%DA%A9%D8%A7%D9%85%D8%B3%D9%88%D9%84%20%28PowerEn.ir%29.pdf>
15. Chen, W X, Wu, J Y. (2022), "Phase-field cohesive zone modeling of multi-physical fracture in solids and the open-source implementation in Comsol Multiphysics[J]. *Theoretical and Applied Fracture Mechanics*, No. 117. P. 103153. DOI: 10.1016/j.tafmec.2021.103153
16. Zhang, Y, Leng, Y, Zhang, H, et al. "Comparative study on equivalent models calculating magnetic force between permanent magnets". *Journal of Intelligent Manufacturing and Special Equipment*, 2020, No. 1(1). P. 43–65. available at: <https://www.emerald.com/insight/content/doi/10.1108/JIMSE-09-2020-0009/full/html>

Цзяо Ханкунь – Харківський національний університет радіоелектроніки, аспірант кафедри біомедичної інженерії, Харків, Україна; e-mail: 1350829683@qq.com, jiao.hankun@nure.ua. ORCID ID: <https://orcid.org/0000-0003-0992-5344>

Аврунін Олег Григорович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, завідувач кафедри біомедичної інженерії, Харків, Україна; e-mail: oleg.avrunin@nure.ua; ORCID ID: <https://orcid.org/0000-0002-6312-687X>

Jiao Hunkun – Kharkiv National University of Radio Electronics, PhD student at the Department of Biomedical Engineering, Kharkiv, Ukraine.

Avrunin Oleg – Doctor of Sciences (Engineering), Professor, Kharkiv National University of Radio Electronics, Head of the Department of Biomedical Engineering, Kharkiv, Ukraine.

АНАЛІЗ МОЖЛИВОСТІ ПЕРЕМІЩЕННЯ ІМПЛАНТУ ПО ДУГОВІЙ ТРАЄКТОРІЇ ПІД БЕЗКОНТАКТНИМ КОНТРОЛЕМ У МАГНІТНІЙ СТЕРЕОТАКСИЧНІЙ СИСТЕМІ

У статті подано безконтактне керування магнітними імплантатами способом зміни зовнішнього магнітного поля в магнітній стереотаксичній системі та проаналізовано можливість їх переміщення по дуговій траєкторії. За допомогою програмного забезпечення COMSOL змодельовано процес переміщення мініатюрного магнітного імплантата по дуговій траєкторії, досліджено зміну траєкторії мікромагнітного імплантата після зміни зовнішнього магнітного поля, визначено взаємне розташування великих постійних магнітів та проведено механічний аналіз переміщення мініатюрного магнітного імплантата по дуговій траєкторії. У цьому експерименті ми фіксуємо великий постійний магніт, рухаємо лише другий постійний магніт, спочатку спостерігаємо процес руху малих постійних магнітів по прямій траєкторії, визначаємо положення магнітного поля великого постійного магніту, коли він контактує з малим постійним магнітом, а потім аналізуємо силу малого постійного магніту за допомогою модуля розрахунку сили та встановлюємо відносне положення між двома великими постійними магнітами, порівнюючи F_x та F_y , і коли малий постійний магніт почне рухатись по дуговій траєкторії. Далі, згідно з попередніми результатами, ми переміщуємо два сусідні великі постійні магніти одночасно з певним інтервалом, записуємо траєкторію руху малого магніту, і, нарешті, за допомогою модуля розрахунку сил програмного забезпечення COMSOL здійснюємо силовий аналіз руху малих постійних магнітів по дугових траєкторіях. Результати проведеного експерименту будуть використані для визначення взаємного розташування двох великих постійних магнітів, розташованих поруч під час дослідження, для з'ясування того, за яких умов малі постійні магніти будуть рухатися по дуговій траєкторії. Метою цього експерименту є забезпечення теоретичної та інформаційної підтримки для подальших практичних досліджень магнітної стереотаксичної системи, коли всі параметри в програмному забезпеченні COMSOL отримані на основі фактичних показників вимірювань для підвищення вірогідності результатів симуляції.

Ключові слова: здоров'я людини; магнітне поле; програмне забезпечення COMSOL; постійні магніти; силовий аналіз.

Бібліографічні описи / Bibliographic descriptions

Цзяо Ханкунь, Аврунін О. Г. Аналіз можливості переміщення імпланту по дуговій траєкторії під безконтактним контролем у магнітній стереотаксичній системі. *Сучасний стан наукових досліджень та технологій в промисловості*. 2023. № 3 (25). С. 174–182. DOI: <https://doi.org/10.30837/ITSSI.2023.25.174>

Hunkun, J., Avrunin, O. (2023), "Feasibility analysis of implant movement along arc trajectory under non-contact control in magnetic stereotaxic system", *Innovative Technologies and Scientific Solutions for Industries*, No. 3 (25), P. 174–182. DOI: <https://doi.org/10.30837/ITSSI.2023.25.174>

АЛФАВІТНИЙ ПОКАЖЧИК

Аврунін О. Г.	174
Бараннік І. О.	148
Барковська О. Ю.	5
Воропаєва К. А.	5
Гнусов Ю. В.	52
Горбачов В. О.	79
Довгопол Н. В.	129
Єнгалічев С. О.	52
Іваненко В. А.	138
Каратаєв О. А.	15
Козирєв А. Д.	27
Кучук Г. А.	88
Ляшик В. А.	40
Малярець Л. М.	148
Манжай О. В.	52
Мірошник М. А.	98
Можаєв О. О.	52
Мосьпан Д. В.	163
Носов В. В.	52
Пересада О. В.	129
Переяславська С. О.	64
Пономаренко О. Є.	79
Пригара М. П.	114
Прібільнова І. Б.	129
Радченко В. В.	52
Рахліс Д. Ю.	98
Резанов Б. М.	88
Руських О. В.	5
Ситніков Д. Е.	15
Смагіна О. О.	64
Струков В. М.	52
Трифанов О. В.	98
Федорка П. П.	114
Цзяо Ханкунь	174
Шкіль О. С.	98
Шпак О. І.	114
Шубін І. Ю.	27, 40
Ястремська О. М.	148
Ястремська О. О.	148

ALPHABETICAL INDEX

Avrunin Oleg	174
Barannik Igor	148
Barkovska Olessia	5
Voropaieva Kseniia	5
Gnusov Yurii	52
Gorbachov Valeriy	79
Dovgopol Nina	129
Yenhalychev Serhii	52
Ivanenko Viacheslav	138
Karataiev Oleksandr	15
Kozyriev Andrii	27
Kuchuk Heorhii	88
Liashyk Volodymyr	40
Malyarets Lyudmyla	148
Manzhai Oleksandr	52
Miroshnyk Maryna	98
Mozhaiev Oleksandr	52
Mospan Denis	163
Nosov Vitalii	52
Peresada Olena	129
Pereiaslavska Svitlana	64
Ponomarenko Olha	79
Prygara Mykhailo	114
Pribylnova Inna	129
Radchenko Valery	52
Rakhlis Dariia	98
Rezanov Bohdan	88
Ruskikh Oleksandr	5
Sytnikov Dmytro	15
Smahina Olga	64
Strukov Volodymyr	52
Trifanov Oleh	98
Fedorka Pavlo	114
Jiao Hunkun	174
Shkil Alexander	98
Shpak Oleksandr	114
Shubin Ihor	27, 40
Iastremska Olena	148
Iastremska Olesia	148

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

№ 3 (25), 2023

№ 3 (25), 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. 12,3.
Тираж 150 прим.

Format 60×84/8. Conventional printed sheets 12,3.
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