МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛУЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ:
ОСВІТА, НАУКА, ВИРОБНИЦТВО

НАУКОВИЙ
ЖУРНАЛ

Головний редактор – професор, д.т.н., Гордєєв О.О.

№56 2024

м. Луцьк

РЕДАКЦІЙНА КОЛЕГІЯ:

Головний редактор:	
професор, д.т.н. Гордєєв О.О.	(м. Луцьк)
Відповідальний секретар:	
доц., к.т.н. Христинець Н.А.	(м. Луцьк)
Члени редакційної колегії:	
проф., д.т.н. Андрущак І.Є.	(м. Луцьк)
проф., д.т.н. Згуровський М.З	(м. Київ)
Affiliate full professor, Avtandil Gagnidze	(Грузія, м. Тбілісі)
д.т.н., доц. Зеленський К.Х.	(м. Київ)
доц., к.т.н. Суринович О.М.	(м. Луцьк)
Affiliate full professor, Iavich Maksim	(Грузія, м. Тбілісі)
проф., д.т.н. Турбал Ю.В.	(м. Рівне)
доц., к.ф.-м.н. Рибицька О.М.	(м. Львів)
PhD. Milosz Marek	(Польща, м. Люблін)
проф., д.т.н. Мельник А.О.	(м. Львів)
проф., д.т.н. Мороз Б.І.	(м. Дніпро)
проф., д.т.н. Тарасенко В.П.	(м. Київ)
проф, PhD. Alison McMillan	(Великобританія,м. Рексем)
проф., д.т.н. Касянчук М.М.	(м. Тернопіль)
проф., д.т.н. Фауре Е.В.	(м. Черкаси)
проф., д.т.н. Олійников Р.В.	(м. Харків)
проф, д.пед.н. Чернящук Н.Л.	(м. Луцьк)
доц., к.т.н. Назаревич О.Б.	(м. Тернопіль)
PhD. Karim Elish	(США, м. Лейкленд)
PhD. Zbigniew Omiotek	(Польща, м. Люблін)
PhD. Dagmar Cagáňová	(Словаччина, м. Братилава)
PhD. Paweł Komada	(Польща, м. Люблін)
PhD. José Machado	(Португалія, м. Гімарайш)
проф., д.т.н. Сайко В.Г.	(м. Київ)
доц., к.т.н. Приступа С.О	(м. Луцьк)
PhD. Anna Maria Saniuk	(Польща, м. Зелена Гура)
доц., к.т.н. Ткачук А.А. (заступник головного редактора)	(м.Луцьк)

Адреса редколегії:
Луцький національний технічний університет,
кафедра комп'ютерної інженерії та кібербезпеки
вул. Львівська 75, ауд.141
м.Луцьк, 43018
тел. (0332) 74-61-15
E-mail: cit@lntu.edu.ua,
сайт журналу: **cit-journal.com.ua**

КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ: ОСВІТА, НАУКА, ВИРОБНИЦТВО

№56 2024 р.

Зареєстровано Національною радою України з питань телебачення і радіомовлення, як суб'єкт у сфері друкованих медіа (рішення №40 від 11.01.2024 р., ідентифікатор медіа R30-02456)
Рекомендовано до друку Науково-технічною радою Луцького національного технічного університету
(протокол №1 засідання від 11.09.2024р.)
Рішенням МОН України
наказом №515 від 16.05.2016р,
журнал включено в перелік наукових фахових видань

Видання індексується у наукометричних та реферативних базах:
Open Academic Journals Index
Academic Resource Index ResearchBib
Rootindexing
Information Matrix for the Analysis of Journals
Ulrichsweb.

ISSN 2524-0560 (Online)
ISSN 2524-0552 (Print)

ЗМІСТ

АВТОМАТИКА ТА УПРАВЛІННЯ	
Shevtsov I. Actual problems of remote patient monitoring	**5**
Григорович А.Г., Григорович В.Г., Жовнір Ю. І., Грибовський О.М. Формування обернено-адитивної семантичної метрики для аналізу онтологій безпекових систем багатоквартирних будинків	**12**
Коростін О.О. Оптимізація маршрутів морських перевезень за допомогою штучного інтелекту: аналіз можливостей та викликів	**31**
Пасічник М.Ю., Зайцев В.Г. Методи диспетчеризації завдань в системах реального часу	**39**
ІНФОРМАТИКА ТА ОБЧИСЛЮВАЛЬНА ТЕХНІКА	
Sydorchuk V. Deep learning algorithms for predicting routes of snowplows in European cities	**44**
Sidorov D. Best Practices for Developing Secure Front-End Web Applications	**51**
Tsymbal O., Lebedev D. Improving emi performance in automotive data transmission systems using spread spectrum modulation	**59**
Shkitov A. Features of creating a multi-level model of cyber security of critical infrastructure data: an interdisciplinary approach.	**71**
Акімов Д.Д., Гавриленко В.В. Розв'язання транспортної задачі в умовах неоднорідності продукції агропромислових підприємств.	**78**
Багнюк Н. В., Лавренчук С. В., Міскевич О. І., Свистун Б. В. Реалізація системи «Розумний будинок» з використанням Bluetooth, Arduino та Android: від датчиків до мобільного інтерфейсу.	**86**
Баранчук С.А., Бортник К.Я. Система для моніторингу відключень світла на базі платформи Arduino та фреймворку React Native	**93**
Бегліцов С.В. Автоматизація міграції програмного забезпечення у хмарну архітектуру із використанням інструменту інфраструктури як код Terraform у середовищі AWS	**99**
Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г. Аналіз ефективності алгоритмів машинного навчання в обробці великих даних	**107**
Борисов О.В., Артабаєв Ю.З., Сурма А.І. Кібербезпека безпілотних військових апаратів: методи захисту від перехоплення та дистанційного управління	**117**
Веремеєнко А.А. Дослідження використання фрактальних алгоритмів компресії в інформаційних системах, що працюють з великими обсягами даних та вимагають високих коефіцієнтів стиснення при максимальному збереженні якості зображень	**126**
Вознюк А.В., Сачук В.О., Повстяна С.О. Інтерактивні методи навчання веб-розробці: як залучити студентів до активного навчання	**138**
Гулівата І.О. Концепт цифрової освітньої доступності: європейський досвід	**143**
Гусак Л.П., Раздзіховська Л.М. Використання комп'ютерних програм під час викладання курсу «Лінійна алгебра та аналітична геометрія»	**149**
Гусєв О.Ю., Рибальченко Ю.П. Інформаційна технологія підвищення ефективності контролю теплового стану нижньої частини доменної печі	**155**
Загинайло Є.О., Приходченко С.Д. Розвиток засобів локалізації та розгляд базових мБагнетодів акустичної локалізації	**160**
Іваненко А.Р., Марченко О.І. Метод компіляції оголошення класів мов WEB-програмування у мови байт-кодового типу на основі машинного навчання	**165**
Касянчук Д.П., Марченко О.І. Засіб синтаксичного розбору на основі генератора синтаксичних аналізаторів ANTLR та мови Clojure	**174**

Кащенко Д.О. Теорія інформації – суб'єктивні системи	**185**
Коваль І.М., Суринович О.М. Порівняльне дослідження методів попередньої обробки тексту в Orange Data Mining та KNIME	**191**
Лисенко О.О., Бортник К.Я., Багнюк Н.В. Дослідження попиту на медичні товари засобами SQL	**199**
Марценюк В.П., Кіт Н.В. Аналітичний огляд підходів до проектування виробничих мереж	**203**
Марченко О.О. Подальший розвиток пошуку по дереву методом Монте-Карло за допомогою контролю форми дерева пошуку	**213**
Мороз Б.І., Круглик А.С. Проблеми застосування груп безпілотних літальних апаратів. Концептуальні засади їх вирішення	**220**
Панаскін Д.В. Оцінка стійкості нейромережевих моделей до шуму та артефактів у складних умовах експлуатації	**226**
Пахомов С.В., Муляревич О.В., Боярінова Ю.Є. Розробка та оптимізація мобільних додатків для різних платформ із використанням мов програмування C, C++, C#, Java	**236**
Попов О.І. Упровадження штучного інтелекту для прогнозування властивостей металевих матеріалів	**244**
Пундик В. І. Вплив використання безперервної інтеграції, доставки та розгортання на процес розробки програмних систем в середовищі хмарних технологій	**254**
Радченко К. О., Романенко М.В. Прогнозування цін акцій з використанням вейвлет-перетворення та нейронних мереж	**261**
Регеда Ю.О., Регеда В.О. Вплив цифрової еволюції на інформаційні технології	**269**
Семенюк С.Т., Ткачук А.А. Розробка web-браузера та його доповнень для пристроїв ARM x86 (x64)	**276**
Федонюк Ю.А. Мобільний застосунок інформаційно-технологічного супроводу людей із гіпертонічними захворюваннями	**283**
Чепара І.В., Кулаковська І.В. Дослідження інтелектуальної системи для комп'ютерних ігор на основі Q&A-технологій	**292**
Яремко С.А., Новицький Р.М. Актуальні тенденції інформаційних систем для фінансової аналітики.	**304**
Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М. Вимоги та їх реалізація для інтелектуальної системи прийняття рішень ідентифікації та класифікації об'єктів культурних цінностей	**311**
ТЕЛЕКОМУНІКАЦІЇ ТА РАДІОТЕХНІКА	
Bershov V., Yakymchuk N. The method of forming ensembles of complex signals based on multi-scale decomposition of time intervals at different levels of detail	**325**
Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О. Технологічні аспекти впровадження програмно-керованих мереж 5G та 6G	**335**
Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О. Забезпечення якості обслуговування інтернету медичних речей в мережі стандарту 5G	**345**
Круліковський Б.Б., Степанюк Ю.І. Локальна АТС підприємства на базі комп'ютерної мережі	**356**
Мороз С.А., Якимчук Н.М., Селепина Й.Р., Чалий В.Д. Розвиток NFC та RFID: використання індуктивних антен у сучасних пристроях ближньої дії	**363**
Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л. Сучасні можливості управління мобільними телекомунікаційними мережами з використанням безпілотних літальних апаратів	**369**
Радзівілов Г.Д., Масесов М.О., Дєгтяр О.А. Світові тенденції зі створення та розвитку військової техніки.	**381**
Селепина Й.Р., Бондаренко О.О. Інтеграція систем зв'язку на базі Wi-Bas в гібридних мережах	**386**
Soproniuk I., Komar O. Adaptive approach to spectrum monitoring in cognitive radio networks through signal detection optimization	**392**

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-01
UDC 004.09
Shevtsov Ivan, Postgraduate student
https://orcid.org/0000-0003-0597-1589
Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

ACTUAL PROBLEMS OF REMOTE PATIENT MONITORING

Shevtsov I. Actual problems of remote patient monitoring. Remote patient monitoring (RPM) is an innovative technology that provides continuous monitoring of patients' health from a distance. The RPM system is versatile because it can be used in the treatment of acute and chronic conditions, helping doctors keep their patients under control between ward rounds or the absence of personal care. However, the implementation of this system is accompanied by numerous problems: from technical to hospital staff management problems. The object of this study is the technical challenges associated with the use of the RPM system. The problem lies in the imperfection of many elements of RPM, including: data management, sensor accuracy, compatibility and user acceptance of the technology. The research found that these technical challenges arise from insufficient standardization, poor sensor data quality, and interoperability issues between medical platforms. These results are explained by the complexity of the technological infrastructure and the need to improve the qualifications of medical workers. The main goal of the current article is to solve the scientific and technical problem of RPM through the proposal of an effective practical development - an automated system for remote monitoring of patients. Since most processes are moving towards their automation, by using the latest scientific developments, among which we can recall machine learning and neural networks and Internet of Things (IoT) technology, it was decided to consider the factors of their application in remote patient monitoring systems. The results of this study can be used to optimize RPM systems in hospitals, as well as to improve the reliability and efficiency of new automated monitoring systems, which is extremely important in the context of war, which causes a shortage of medical personnel and an increase in the number of potential patients per doctor.
Keywords: remote patient monitoring, data management, interoperability, sensor accuracy, technology adoption, automation, algorithms.

Шевцов І.О. Актуальні проблеми дистанційного моніторингу пацієнтів. Дистанційний моніторинг пацієнтів (ДМП) – це інноваційна технологія, яка забезпечує безперервний нагляд за станом здоров'я пацієнтів на відстані. Система ДМП є універсальною, оскільки її можна використовувати при лікуванні гострих та хронічних станів, що допомагає медикам тримати їх пацієнтів під контролем між обходами палат або відсутністю особистого догляду. Однак впровадження цієї системи супроводжується численними проблемами: від технічних до проблем менеджменту персоналу лікарень. Об'єктом цього дослідження є технічні виклики, які пов'язані з використанням системи ДМП. Проблема полягає у недосконалості багатьох елементів ДПМ, серед яких: управління даними, точність сенсорів, сумісність та сприйняття технології користувачами. У результаті дослідження було виявлено, що ці технічні виклики виникають через недостатню стандартизацію, низьку якість сенсорних даних та проблеми з інтероперабельністю між медичними платформами. Ці результати пояснюються складністю технологічної інфраструктури та потребою в підвищенні кваліфікації медичних працівників. Головною метою поточної статі є вирішення науково-технічної проблеми ДПМ через пропозицію ефективної практичної розробки – автоматизованої системи дистанційного моніторингу пацієнтів. Оскільки більшість процесів рухається до їх автоматизації, шляхом використання новітніх наукових розробок, серед яких можна пригадати машинне навчання та нейромережі і технологію Інтернету речей (IoT), то вирішено було розглянути фактори їх застосування і в системах дистанційного моніторингу пацієнтів. Результати цього дослідження можуть бути використані для оптимізації систем ДМП у лікарнях, а також для підвищення надійності й ефективності нових автоматизованих моніторингових систем, що є надзвичайно важливим в умовах війни, що викликає дефіцит медичних працівників та збільшення потенційних пацієнтів з розрахунку на кожного лікаря.
Ключові слова*:* дистанційний моніторинг пацієнтів, управління даними, інтероперабельність, точність сенсорів, прийняття технологій, автоматизація, алгоритми.

Formulation of the problem.

The rapid evolution of healthcare technology has made remote patient monitoring (RPM) an increasingly crucial element of modern healthcare systems. The ability to monitor patients' health remotely, particularly those with chronic conditions or who require consistent medical supervision, has revolutionized care delivery by reducing hospital visits and enabling continuous health assessment [1]. Provider associations and patient associations claim that RPM systems are effective in improving or maintaining quality of care, further encouraging its use and adoption [2]. However, despite the significant advantages of RPM, it also presents several technical challenges that hinder its widespread adoption and optimal use. Research into the technical challenges of remote patient monitoring has become more critical than ever.

The global rise in chronic diseases, aging populations, and the current burden on healthcare systems necessitate efficient and reliable RPM systems. By addressing and overcoming the technical hurdles, researchers and engineers can improve the quality of patient care and enhance healthcare outcomes [3].

The relevance of studying these challenges is grounded in the increasing reliance on RPM systems, especially in the post-pandemic era [4]. Real-time data transmission, sensor technology, and data security are just a few areas that demand innovation to meet growing healthcare needs. Some groups of technicians

© Shevtsov I.

working at hospitals have already developed their programs for remote monitoring of patients discharged from the hospital after treatment for COVID-19, but they are able to perform a narrow profile of work [5].

The solutions developed from this research will not only reduce the risk of hospitalization but also improve patient management by providing accurate and timely health data for clinicians. Thus, research focused on identifying and resolving technical issues in RPM is highly relevant. Understanding and addressing these issues can facilitate the wider adoption of RPM systems and significantly improve the quality of patient care in a variety of healthcare settings through the automation of these systems.

An analysis of the latest research and publications.

The increasing integration of remote patient monitoring (RPM) technologies into modern healthcare systems highlights both the potential and the challenges associated with real-time data collection and patient management. The following section provides a critical analysis of recent literature on the technical aspects of RPM systems, identifying unresolved problems and offering insights into the reasons for these gaps.

A review of the existing research, particularly in the realm of real-time remote health monitoring systems, underscores significant challenges related to the technical capabilities of sensor technologies and their deployment in priority-based health systems. Albahri et al. emphasize that while real-time RPM can enhance triage efficiency and patient prioritization, significant limitations exist in sensor accuracy, reliability, and interoperability with healthcare infrastructures [6]. The research also highlights the need for more robust algorithms to process the vast amounts of data generated by RPM systems, particularly in critical care settings [7].

Further analysis of the literature reveals that the management and quality of patient-generated health data (PGHD) remain a major issue in RPM. Abdolkhani et al. identify challenges with the integration and standardization of PGHD, which often varies in format and quality, leading to inconsistencies in patient monitoring outcomes [8]. The authors argue that current systems lack the sophistication needed to ensure accurate, timely, and clinically useful data transmission, creating a gap between RPM's potential and its practical application. Moreover, the lack of standardized data protocols contributes to data fragmentation and hampers effective decision-making.

The problem of healthcare practitioner perceptions further complicates the implementation of RPM technologies. Serrano et al. discuss the benefits and challenges as perceived by healthcare professionals, noting that while RPM has been lauded for its potential to reduce hospital readmissions and improve chronic disease management, the perceived technical difficulties often outweigh these benefits [9].

Another important scientific work is the study of Taylor et al. which shows that RPM can reduce acute care use for patients with cardiovascular disease and COPD [10].

Among Ukrainian researchers on the issue of the application of various methods and algorithms for automated monitoring of patients, Boloban and Petrenko should be mentioned and their work studying the features of the application of the EMD algorithm for assessing the level of oxygen in the blood [11].

Uninets I. indicated remote monitoring of patients as one of the components of Smart Healthcare, which has the highest rates of development in the Smart Citizen Services segment [12].

Issues such as inadequate user interfaces, technical malfunctions, and connectivity problems are consistently cited as barriers to broader adoption of RPM technologies. These technical shortcomings are further exacerbated by the lack of training and support for healthcare providers, which limits the effectiveness of RPM in clinical practice.

A systematic review of these studies reveals a common theme: the unresolved issues in RPM primarily stem from technical inadequacies in data collection, integration, and system reliability. Many studies have addressed the need for advancements in sensor technologies and data processing algorithms; however, progress has been slow, primarily due to the complex and multifaceted nature of the problems involved. These challenges are exacerbated by healthcare systems' varying levels of digital readiness and the slow pace of technological standardization.

In conclusion, the literature points to several unresolved problems in the field of RPM. These include sensor accuracy, data standardization, integration challenges, and technical support for healthcare providers. All these problems come down to the main scientific and technical problem of automating patient monitoring, the problem of automating the work of PRM systems. The reasons for these unresolved issues range from technological limitations to systemic barriers within healthcare systems. As such, the next phase of research should focus on addressing these technical shortcomings, particularly in the areas of data standardization, system reliability, and training for healthcare providers, in order to fully harness the

© Shevtsov I.

potential of RPM technologies [13]. It is also necessary to consider the experience of automation of PRM systems for a better understanding of this process and algorithms for its implementation.

Formulation of the purpose and objectives of the research.

The aim of this research is twofold, comprising both scientific and practical components.

The scientific goal of the study focuses on identifying the most significant technical challenges in remote patient monitoring systems, as highlighted in recent literature. This involves analyzing the critical gaps in sensor technology, data management, and system integration that remain unresolved despite recent advancements. By investigating these gaps, the research aims to propose improved methods for enhancing the reliability, efficiency, and accuracy of patient monitoring systems, which will ultimately allow the automation of the remote patient monitoring process.

The practical goal is to provide recommendations for the deployment and use of automated remote patient monitoring systems that optimize healthcare delivery, particularly in resource-limited settings. This includes improving the integration of patient-generated data into clinical workflows while minimizing data loss and inaccuracies. Additionally, the research seeks to suggest best practices for maintaining system functionality over long-term usage, addressing both technical and operational limitations.

To achieve these goals, the following research objectives were set:

- Objective 1: To critically analyze and systematize the existing literature on remote patient monitoring systems, focusing on the technical challenges related to sensor accuracy, data transmission, and real-time monitoring capabilities.
- Objective 2: To identify the key factors that contribute to data management challenges in remote patient monitoring, particularly in handling patient-generated data and ensuring its quality and security.
- Objective 3: To evaluate the current technological solutions for overcoming the identified challenges and propose potential improvements in sensor technology, system architecture, and data integration processes [14].
- Objective 4: To assess the practical applications of remote patient monitoring in healthcare settings, identifying limitations in deployment and offering strategies for optimizing the long-term use of these systems.

Through addressing these objectives, this study aims to contribute to the development of more effective and reliable remote patient monitoring systems that can be utilized in diverse healthcare environments, within the Ukrainian legal practice.

Presentation of the main material and substantiation of the obtained research results.

Remote patient monitoring (RPM) has emerged as a significant innovation in healthcare, offering real-time health data and allowing continuous patient oversight. However, despite its promise, RPM faces several technical challenges that hinder its broad implementation and effectiveness. Among these problems is the problem of data management, sensor technologies, compatibility and user acceptance [15].

Remote patient monitoring is an integral part of any modern patient care system. The entire care system is shown in fig. 1.

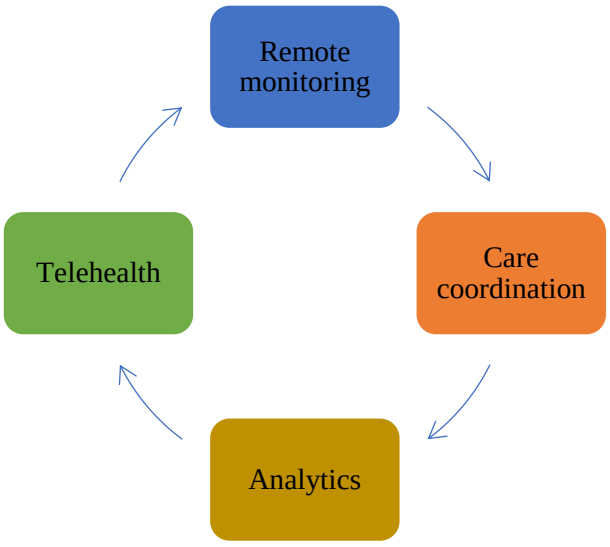


Fig. 1. Modern system of care and diagnosis of patients

© Shevtsov I.

Since domestic legislation in the field of health care is rather confusing and ambiguous, remote monitoring of health status, provision of universal access to electronic medical records of patients, conducting medical research using electronic health records of patients require a flexible approach to legal regulation relations in the field of medical data and updating a number of laws [16].

One of the major technical challenges in RPM is the effective management and integration of patient-generated health data. Handling massive amounts of patient-generated data creates significant challenges related to data quality, interoperability, and privacy. For example, fragmented data from multiple devices and platforms makes it difficult to integrate into clinical systems, resulting in delayed response times and reduced decision-making accuracy. In addition, maintaining data privacy while providing real-time access to healthcare providers is another ongoing challenge in RPM systems.

Focusing on the technological challenges in implementing sensor technology in RPM systems, various open challenges related to the deployment of real-time remote health monitoring systems need to be identified, particularly with respect to sensor accuracy, reliability, and energy efficiency [17]. Many current sensors struggle to maintain stable performance in various environmental conditions, leading to inaccuracies in patient data. In addition, energy consumption in continuous monitoring systems remains a critical issue, as battery limitations often prevent continuous long-term monitoring [8].

Another major problem in RPM is the lack of standardized protocols and system compatibility. The lack of uniform standards for device communication prevents the seamless exchange of data between different healthcare systems. Interoperability issues not only limit the integration of RPM data into electronic health records, but also make it difficult to scale RPM solutions across healthcare providers and institutions. The development of common standards and communication protocols is important for solving this problem [9].

A less technical, but no less important, issue is the user acceptance of RPM systems by both patients and healthcare professionals. The success of RPM systems is largely dependent on patients' ability to properly use the monitoring devices and software. However, insufficient training and support may result in incorrect data collection or device malfunction. Healthcare providers also face challenges in adapting to new technologies, especially in environments with limited IT infrastructure or technical support. Addressing these training needs is critical to ensuring the successful implementation of RPM [14].

Turning to the question of automation of remote monitoring systems, it should be mentioned that among the devices included in such a system there should be only those that show the main indicators of the body, since a large number of devices can overload the system, and this, in turn, will cause delays and obtaining inaccurate results. Devices that are usually included in the system of remote monitoring of the patient are shown in fig. 2.

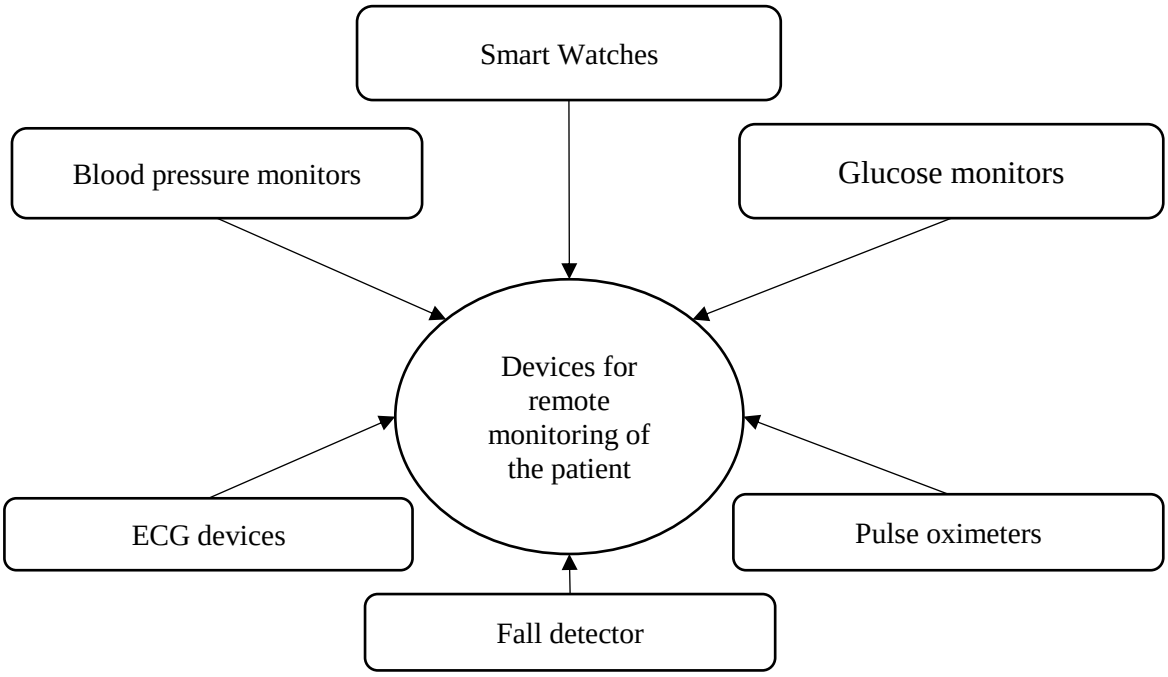


Fig. 2. Main devices used in automated RPM systems

© Shevtsov I.

Below is a description of the value of devices for detecting key indicators in the patient being monitored:

1. Blood pressure monitors: a person can use an IoT blood pressure monitor to record their blood pressure in the app and view any patterns using the generated graphs. If necessary, the doctor can receive blood pressure results by e-mail and provide accurate information about treatment. In terms of connectivity, Bluetooth and Wi-Fi are mainly used by IoT blood pressure devices.
2. Glucose monitors: self-monitoring of blood glucose (SMBG) helps to plan intensive diabetes treatment. White-coat hypertension, which affects those adversely affected by the hospital environment, can be prevented through self-management.
3. ECG devices: The electrical activity of the heart can be measured and recorded using ECG (electrocardiogram) equipment. Electrodes, which are tiny sticky pads, are stuck to the skin of the patient's arms, legs and chest to work with the machine. The electrodes transmit the electrical signals they receive from the heart to the ECG equipment.
4. Pulse oximeters: pulse oximeters are devices that measure the amount of oxygen in a person's blood. The use of pulse oximetry screening as a method for detecting serious congenital heart defects is one of the most remarkable advances in the market for pulse oximetry (CCHDs).
5. Fall detector: can provide a warning signal in the event of a fall. Elderly people, cyclists and hikers who use a fall detector can use it to sound an alarm and seek help.
6. Smart Watches: tracks daily activities such as steps, calorie consumption, heart rate, blood oxygen level, blood pressure and body temperature [18].

These devices are especially useful when working with vulnerable categories of the population (elderly people and persons with disabilities).

Most modern remote patient monitoring systems use artificial intelligence technologies and algorithms. The integration of artificial intelligence (AI) into remote patient monitoring (RPM) has revolutionized healthcare, improving patient care, increasing efficiency and enabling early intervention. The introduction of artificial intelligence algorithms increases the potential of RPM systems by analyzing vast volumes of patient data to identify trends, anomalies and potential problems.

Artificial intelligence algorithms process data on patients' vital signs to establish personalized baselines for each patient, taking into account factors such as age, gender, medical history and current health status. Once these baselines are established, the AI system continuously monitors the input data for abnormalities. There are four key basic components that underpin remote patient monitoring systems using artificial intelligence algorithms (table. 1).

Table. 1. Key components of a remote patient monitoring system using artificial intelligence algorithms

No.	Component name	The essence of the components
1	Near-real-time monitoring	Wearable devices and sensors equipped with artificial intelligence technology continuously collect patient data, ensuring a constant flow of information. This allows artificial intelligence algorithms to detect even the smallest deviations from established baseline patient indicators
2	Pattern recognition	Artificial intelligence models are able to recognize patterns efficiently and quickly. By analyzing patterns in the collected data, such as irregular heart rhythms, sudden spikes or drops in vital signs, and sudden changes in activity levels, AI can detect potential signs of poor health that might otherwise go unnoticed without the use of an algorithm
3	Anomaly detection	Artificial intelligence algorithms are trained to detect anomalies that fall outside of normal variation. When an abnormality is detected, the AI system alerts healthcare providers, allowing them to immediately intervene and take action immediately
4	Predictive analytics	In addition to detecting current abnormalities, AI can also predict potential health problems based on trends and historical data. For example, if a patient's heart rate variability gradually decreases over time, AI can alert healthcare professionals to an increased risk of heart problems

© Shevtsov I.

It should be noted that the use of remote patient monitoring systems has quite serious advantages over classical monitoring tools, among which the following can be recalled:

1. Better management of chronic diseases: patients with diseases such as diabetes, hypertension and heart disease can benefit greatly from RPM. By regularly monitoring their health, patients and their healthcare providers can identify potential problems before they become serious.
2. Improved Convenience: with RPM, patients can manage their health from the comfort of their homes. This eliminates the need for frequent visits to the doctor, saves time and money.
3. Increased peace of mind: for many patients, RPM provides a sense of security and control over their health. They can monitor their own progress and be alerted to potential problems, giving them a greater sense of autonomy.
4. Improved Outcomes: patients using RPM have better outcomes overall. They are more likely to achieve their health goals and experience less complications from chronic diseases [20].

Thus, remote monitoring of patients, as well as systems developed on the basis of this method, have significant prospects for improving patient care, but, unfortunately, its implementation also faces significant technical problems. To fully realize RPM's technical potential, issues such as data management, sensor accuracy, system compatibility, and user adaptability must be addressed. Collaboration between healthcare providers, technologists, and IT teams is essential to overcoming these barriers and advancing the field of medicine.

Conclusions and prospects for further research.

Therefore, the implementation of remote monitoring of patients in Ukraine has several problematic issues, the main ones being the vagueness of the legislation, technical problems, low level of awareness of medical workers and mistrust of patients. In order to solve the problems with the legislation, it is necessary to initiate a change in the current law on the provision of medical services. To solve technical problems, it was proposed to use artificial intelligence algorithms and integrate a limited set of monitoring devices. To solve the problem with the insufficient level of awareness of employees, it is necessary to improve their training in educational institutions by creating a separate course on the basics of remote monitoring of patients. Patient trust can be gained by explaining the benefits to them of using remote monitoring systems, in terms of convenience and time savings. In the following studies, it is necessary to concentrate on the issues of implementation and construction of a system of remote monitoring of patients in an already existing medical institution.

References

1. *Telehealth and remote patient monitoring.* (2024, May 11). telehealth.hhs.gov. Retrieved from https://telehealth.hhs.gov/providers/preparing-patients-for-telehealth/telehealth-and-remote-patient-monitoring [in English].
2. Hayes, C. J., Dawson, L., McCoy, H., Hernandez, M., Andersen, J., Ali, M. M., ... & Eswaran, H. (2023). Utilization of remote patient monitoring within the United States health care system: a scoping review. *Telemedicine and e-Health*, *29*(3), 384-394. Retrieved from https://www.liebertpub.com/doi/abs/10.1089/tmj.2022.0111 [in English].
3. Haleem, A., Javaid, M., Singh, R. P., & Suman, R. (2021). Telemedicine for healthcare: Capabilities, features, barriers, and applications. *Sensors international*, *2*, 1-26. Retrieved from https://www.sciencedirect.com/science/article/pii/S2666351121000383 [in English].
4. DAnza, B., & Pronovost, P. J. (2022). Digital health: unlocking value in a post-pandemic world. *Population Health Management*, *25*(1), 11-22. Retrieved from https://www.liebertpub.com/doi/abs/10.1089/pop.2021.0031 [in English].
5. Aalam, A. A., Hood, C., Donelan, C., Rutenberg, A., Kane, E. M., & Sikka, N. (2021). Remote patient monitoring for ED discharges in the COVID-19 pandemic. *Emergency Medicine Journal*, *38*(3), 229-231. Retrieved from https://emj.bmj.com/content/38/3/229 [in English].
6. Albahri, O. S., Albahri, A. S., Mohammed, K. I., Zaidan, A. A., Zaidan, B. B., Hashim, M., & Salman, O. H. (2018). Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: Taxonomy, open challenges, motivation and recommendations. *Journal of medical systems*, *42*, 1-27. Retrieved from https://link.springer.com/article/10.1007/s10916-018-0943-4 [in English].
7. Aamodt, I. T., Lycholip, E., Celutkiene, J., Strömberg, A., Atar, D., Falk, R. S., ... & Lie, I. (2019). Health care professionals' perceptions of home telemonitoring in heart failure care: cross-sectional survey. *Journal of medical Internet research*, *21*(2). Retrieved from https://www.jmir.org/2019/2/e10362/ [in English].
8. Abdolkhani, R., Gray, K., Borda, A., & DeSouza, R. (2019). Patient-generated health data management and quality challenges in remote patient monitoring. *JAMIA open*, *2*(4), 471-478. Retrieved from https://academic.oup.com/jamiaopen/article/2/4/471/5572201?trk=public_post_comment-text&login=false [in English].
9. Serrano, L. P., Maita, K. C., Avila, F. R., Torres-Guzman, R. A., Garcia, J. P., Eldaly, A. S., ... & Forte, A. J. (2023). Benefits and challenges of remote patient monitoring as perceived by health care practitioners: a systematic review. *The Permanente Journal*, *27*(4), 100-111. Retrieved from https://www.thepermanentejournal.org/doi/pdf/10.7812/TPP/23.022 [in English].

© Shevtsov I.

10. Taylor, M. L., Thomas, E. E., Snoswell, C. L., Smith, A. C., & Caffery, L. J. (2021). Does remote patient monitoring reduce acute care use? A systematic review. *BMJ open*, *11*(3), 1-11. Retrieved from https://bmjopen.bmj.com/content/bmjopen/11/3/e040232.full.pdf [in English].
11. Boloban, O. A., & Petrenko, A. I. (2023). Osoblyvosti zastosuvannya alhorytmu EMD dlya otsinky rivnya kysnyu v krovi [Features of using the EMD algorithm for evaluating the level of oxygen in the blood]. *Vcheni zapysky - Scientific notes*, *34*(2), 125-130. Retrieved from https://www.tech.vernadskyjournals.in.ua/journals/2023/2_2023/part_1/20.pdf [in Ukrainian].
12. Uninets, I. (2022). Rozvytok SMART-CITIES v Ukrayini v konteksti hlobalnykh trendiv [Development of SMART-CITIES in Ukraine in the context of global trends]. *Visnyk Khmelnytskoho natsionalnoho universytetu - Bulletin of the Khmelnytskyi National University*, (4), 223-228. Retrieved from http://journals.khnu.km.ua/vestnik/wp-content/uploads/2022/09/2022-308-35.pdf [in Ukrainian].
13. Yakovyshen, P. O., & Tuzhanskyy, S. Y. (2024). Analiz metodiv peredavannya danykh v telemedychnykh systemakh [Analysis of data transmission methods in telemedicine systems]. *Optyko-elektronni informatsiyno-enerhetychni tekhnolohiyi - Optical-electronic information and energy technologies*, *47*(1), 222-232. Retrieved from https://oeipt.vntu.edu.ua/index.php/oeipt/article/view/697 [in Ukrainian].
14. Maita, K. C., Palmieri-Serrano, L., Avila, F. R., Torres-Guzman, R. A., Garcia, J. P., S. Eldaly, A., ... & Forte, A. J. (2023). Imaging evaluated remotely through telemedicine as a reliable alternative for accurate diagnosis: a systematic review. *Health and Technology*, *13*(3), 347-364. Retrieved from https://link.springer.com/article/10.1007/s12553-023-00745-3 [in English].
15. Alshamrani, M. (2022). IoT and artificial intelligence implementations for remote healthcare monitoring systems: A survey. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 4687-4701. Retrieved from https://www.sciencedirect.com/science/article/pii/S1319157821001385 [in English].
16. Pavlovski, H., Barabash, O., Kovaliv, M., Khmyz, M. YA., Mysyuk, I., & Hryshchuk, A. (2022). Pravovi aspekty informatsiynoyi bezpeky u telemedytsyni v Ukrayini [Legal aspects of information security in telemedicine in Ukraine]. *Path of Science*, *8*(11), 1011-1017. Retrieved from https://doi.org/10.22178/pos.87-2 [in Ukrainian].
17. Abdulkarem, M., Samsudin, K., Rokhani, F. Z., & A Rasid, M. F. (2020). Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction. *Structural health monitoring*, *19*(3), 693-735. Retrieved from https://journals.sagepub.com/doi/abs/10.1177/1475921719854528 [in English].
18. *Prystroyi viddalenoho monitorynhu patsiyentiv (RPM) pokrashchuyut medychne obsluhovuvannya (Povnyy posibnyk u 2024 rotsi) [Remote Patient Monitoring (RPM) Devices Improve Healthcare (The Complete Guide in 2024)]*. (2023, March). Dusun IoT. Retrieved from https://dusuniot.com/uk/blog/remote-patient-monitoring-devices-improves-health-care/ [in Ukrainian].
19. *AI in Remote Patient Monitoring: The Top 4 Use Cases in 2024*. (n. d.). HealthSnap, Inc. Retrieved from https://healthsnap.io/ai-in-remote-patient-monitoring-the-top-4-use-cases-in-2024/ [in English].
20. *The Benefits Of Remote Patient Monitoring: Why It's Time To Invest in Your Health - Health News*. (2023, November 1). CONTEC. Retrieved from https://www.contecmedsystems.com/news/the-benefits-of-remote-patient-monitoring-why-71000503.html [in English].

© Shevtsov I.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-02
УДК 004.8:519.1/.8
Григорович Андрій Геннадійович[1], к.т.н.
https://orcid.org/0000-0002-5361-8854
Григорович Віктор Геннадійович[2], к.ф.-м.н., доцент
https://orcid.org/0000-0002-5828-067X
Жовнір Юрій Іванович[2], аспірант
https://orcid.org/0009-0006-6186-2861
Грибовський Олег Миколайович [2], аспірант
https://orcid.org/0009-0005-6318-3611
[1]Дрогобицький державний педагогічний університет імені Івана Франка, м. Дрогобич, Україна
[2]Національний університет «Львівська політехніка», м. Львів, Україна

ФОРМУВАННЯ ОБЕРНЕНО-АДИТИВНОЇ СЕМАНТИЧНОЇ МЕТРИКИ ДЛЯ АНАЛІЗУ ОНТОЛОГІЙ БЕЗПЕКОВИХ СИСТЕМ БАГАТОКВАРТИРНИХ БУДИНКІВ

Григорович А. Г., Григорович В. Г., Жовнір Ю. І., Грибовський О.М. Формування обернено-адитивної семантичної метрики для аналізу онтологій безпекових систем багатоквартирних будинків. Проблеми побудови метрик є ключовими для розв'язку задачі кількісного оцінювання як в цілому систем об'єктів довільної природи, так і відношень, що описують зв'язки між складовими вказаних систем. Для оцінювання відношень між концептами побудовано метрику для не таксономічної онтології (для довільної семантичної мережі концептів). Практичне значення дослідження полягає у розробленні точної та ефективної семантичної метрики, яка може покращити результати вирішення завдань опрацювання онтологій, зокрема таких як пошук інформації, класифікація та виявлення знань. Запропонована метрика може бути корисною для розробки ефективних інтелектуальних систем, які використовують онтології для подання знань, зокрема при розробленні безпекових систем багатоквартирних будинків чи житлових кварталів. Проведено аналіз та обґрунтування запропонованої метрики.
Ключові слова: онтологія, семантична мережа, метрика, розмір, відстань, концепт.

Hryhorovych A., Hryhorovych V., Zhovnir Yu., Hrybovskyi O.M. Inverse-additive semantic metric and distance calculation between ontology concepts. The problems of metric construction are crucial for solving the task of quantitative evaluation of both systems of objects of any nature and the relationships that describe the connections between the components of these systems. A metric has been developed for evaluating relationships between concepts in a non-taxonomic ontology (for any semantic network of concepts). The practical significance of the research lies in the development of an accurate and efficient semantic metric that can improve the outcomes of ontology processing tasks, such as information retrieval, classification, and knowledge discovery. The proposed metric can be useful for the development of effective intelligent systems that utilize ontologies for knowledge representation, particularly in the development of security systems for apartment buildings or residential neighborhoods. An analysis and justification of the proposed metric have been conducted.
Keywords: Ontology, semantic network, metrics, size, distance, concept.

Постановка наукової проблеми. В сучасних інформаційних системах моделюються предметні області, які містять об'єкти та системи складної будови. При цьому, мережева модель є однією з найбільш адекватних та зручних для опису навколишнього світу. Вважається, що довільну систему можна, зазвичай, описати за допомогою кінцевої множини мережевих структур.

Для кількісного оцінювання моделей систем першочерговим є побудова відповідної метрики, що дозволить визначити поняття «розмір системи» та «відстань між елементами системи», а також більш точно описувати системи та взаємозв'язки між елементами систем, перейти від їх якісних до кількісних характеристик.

Одним з яскравих прикладів таких мережевих структур є онтології – семантичні мережі, які описують поняття (концепти) певних предметних областей та взаємозв'язки між ними.

У випадку дослідження окремих елементів системи та відношень, що існують між цими елементами, важливими є метрики, що описують зв'язки між окремими складовими частинами систем. Для онтологій зазначені метрики дають можливість оцінювати, зокрема, «відстані» між окремими концептами.

Таким чином, вирішення проблеми побудови метрик для кількісного оцінювання мережевих структур, зокрема – онтологій, має важливе науково-методологічне та прикладне значення. Задання формалізмів вказаних метрик дозволить вирішувати ряд задач, пов'язаних із семантичним аналізом текстів – автоматичного реферування, побудови мережевих структур в UML нотаціях, дослідження

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

їх схожості, автоматичного оцінювання відповідей на відкриті тестові завдання, автоматичної побудови семантичної мережі заданого тексту тощо.

В цій роботі розглядається метрика для мережевих структур, що описуються за допомогою орієнтованих графів, придатна як семантична метрика для вимірювання відстані між концептами в довільній (не лише ієрархічній) онтології.

Аналіз досліджень. Метрикам для мережевих та ієрархічних структур, а також метрикам для оцінювання онтологій та концептів в онтологіях присвячено доволі багато публікацій. Описані в джерелах метрики можна поділити на такі види: морфологічні метрики, ймовірнісні (Байесові), метрики для адаптивних онтологій, метрики для оцінювання онтологій, метрики для оцінювання зв'язків між концептами в онтологіях.

Морфологічні метрики для ієрархічних дерев

Fenton N.E. та Pfleeger S.L. [1] описали набір простих морфологічних метрик для ієрархічних дерев, що грунтуються на характеристиках графів:

- розмір *Size* = *n*, де *n* – кількість вершин.
- густина взаємодії *R* = *e* / *n* (відношення кількості ребер до кількості вершин). Для дерева *e* = *n*–1.
- коефіцієнт розгалуження за виходом *Fan_out(i)* – це кількість дочірніх вершин *i*-тої вершини.

Первинними характеристиками графу є кількість вершин *n* та кількість ребер *e*.

Для дерева до них додаються ще дві глобальні характеристики – висота і ширина:

- висота (*depth*) – кількість рівнів (кількість вершин в найдовшому шляху від кореневої вершини до листової);
- ширина (*width*) – максимальна кількість вершин, розміщених на будь якому одному рівні дерева. Ширина рівня – це кількість вершин дерева на даному рівні, тоді ширина дерева – це максимальна ширина на всіх рівнях.

q-метрика для зваженого графа та природня метрика для звичайного графа

В роботі [2] наведено метрику для звичайного графа:

Нехай $L[q] = (X, U; q)$ – звичайний граф з ваговою функцією q, яка ставить у відповідність до кожного ребра $u \in U$ дійсне число $q(u) > 0$ як довжини. Якщо Q – маршрут, то сума $q(Q) \equiv \sum_{u \in Q} q(u)$ по всім його ребрам називається його *q-довжиною*, а просто *довжина* – це кількість ребер маршруту (в обох випадках кожне ребро слід рахувати стільки разів, скільки воно зустрічається в маршруті).

Число

$d(x, y) \equiv d_L^q(x, y) \equiv \min\{q(Q) \mid Q \in \mathbf{Q}(x, y)\}$

де $\mathbf{Q}(x, y)$ – множина всіх простих ланцюгів із x до y, називається *q-відстанню* між вершинами $x, y \in X$ зваженого графа $L[q]$: якщо $x = y$, то Q – ланцюг нульової довжини і його q-довжина $q(Q) \equiv 0$, а якщо вершини x та y – відокремлені, то $\rho(x, y) \equiv +\infty$.

Легко побачити, що *q-відстань* задовольняє трьом аксіомам метрики Фреше (Fréchet):

$\forall\, x, y \in X\ [d(x, y) = 0 \Leftrightarrow x = y]$,

$\forall\, x, y \in X\ [d(x, y) = d(y, x)]$,

$\forall\, x, y \in X\ [d(x, y) + d(y, z) \geq d(x, z)]$,

тобто, є метрикою на множині X. В частковому випадку, коли всі $q(u) = 1$, тобто, коли q-відстань кожного ланцюга збігається з його звичайною довжиною, метрика $d(x, y) \equiv d_L^q(x, y)$ графа $L[q]$ називається *природньою метрикою* звичайного графа $L = (X, U)$.

Ймовірнісні (Байесові) метрики

В роботі К. Нейлора [3] запропоновано ймовірнісний підхід, що грунтується на теоремі Байеса, для оцінювання ієрархічних структур при побудові експертних систем.

Метрика для таких систем грунтується на теоремі Байеса: ймовірність здійснення деякої гіпотези *H* за наявності певних свідчень *E*, які підтверджують цю гіпотезу (тобто, при настанні подій *E*), обчислюється на основі апріорної ймовірності даної гіпотези без свідчень-підтверджень *E* та ймовірності здійснення свідчень за умов, що гіпотеза вірна або хибна:

$$P(H \mid E) = \frac{P(HE)}{P(E)} \quad \Rightarrow \quad P(HE) = P(H \mid E) \cdot P(E) = P(E \mid H) \cdot P(H)$$

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

$$P(H \mid E) = \frac{P(E \mid H) \cdot P(H)}{P(E)}$$

де

$P(H)$ – апріорна ймовірність гіпотези H;
$P(H \mid E)$ – ймовірність гіпотези H при настанні подій E (апостеріорна ймовірність);
$P(E \mid H)$ – ймовірність настання подій E при істинності гіпотези H;
$P(E)$ – ймовірність настання подій E.

Г. С. Теслер в роботі [4] розвиває запропонований К. Нейлором підхід:

Нехай $G = (V, E)$ – зв'язний граф, u та v – дві його різні вершини. Тоді відстанню між вершинами u та v буде довжина найкоротшого маршруту, яка позначається $d(u, v)$. При цьому виконуються всі аксіоми метрики.

Відомо, що всякий граф взаємно однозначно подається бінарним відношенням, яке можна задати матрицею суміжності. Елементи матриці суміжності $A(G)$ мають вигляд

$$a_{ij} = \begin{cases} 1, & \text{якщо вершини з номерами } i \text{ та } j \text{ - сумісні} \\ 0, & \text{в іншому випадку} \end{cases}$$

Рангом графа G називається ранг його матриці суміжності, позначається $rank(G)$.

Якщо u – деяка вершина графу $G = (V, E)$, то величина $e(u) = \max d(u, v)$, $v \in V$ називається ексцентриситетом вершини u. Діаметром графа називають максимальний ексцентриситет серед його вершин і позначають $d(G) = \max e(u)$, $u \in V$.

У цьому випадку як міру можна було б використовувати діаметр графа і побудувати метрику графів на основі їх діаметрів – що еквівалентно одній з морфологічних метрик.

Як приклад ієрархії Г. С. Теслер наводить дерево хвороб людини та їх зв'язки в залежності від причин їх виникнення та механізмів їх розвитку. Як метрику для подібних систем використовують теорему Байеса. Використання такого підходу при побудові експертної системи для медичної бази знань MYCIN наведено в роботі [3]; подібний підхід для аналізу кристалічних структур хімічних сполук описано в роботі [5].

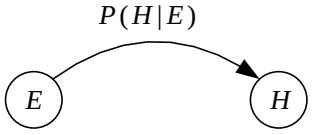


Рис.1. Ілюстрація до ймовірнісної метрики

Слід зауважити, що система MYCIN оперує поняттям «степінь певності»: процедурні правила в ній формулюються у вигляді «ЯКЩО … ТО … З ПЕВНІСТЮ Р», де степінь певності – «приблизно те саме, що ми називаємо умовною ймовірністю $P(H \mid E)$ – ймовірність гіпотези H за умови, що подія E відбулася» [3]. При побудові системи MYCIN експерти-медики пропонували правила і вказували степінь довіри до кожного правила в діапазоні від 1 до 10 – такі експертні оцінки і стали степінню певності для відповідних процедурних правил. Отже, набір процедурних правил такої системи можна описати за допомогою орієнтованого графу, кожне ребро якого має вагу – умовну ймовірність переходу $E \rightarrow H$, тобто ймовірність гіпотези H за умови, що відбулася подія E (рис.1).

Очевидно, що такий підхід дозволяє оцінити лише окремі частини даного орієнтованого графу і не придатний для порівняння різних графів між собою, бо загальна ймовірність повної системи повинна дорівнювати одиниці: $P(G = (V, E)) = 1$.

Метрики для оцінювання онтологій

В роботі [6] запропоновано підхід машинного навчання для поєднання метрик, який використовує різні лінгвістичні та контексті профілі для виявлення відповідності між сутностями різних онтологій. Запропоновано метрики, які стосуються термінологічних та контекстних характеристик сутностей в онтології.

Робота [7] пропонує вирішення проблеми, пов'язаної із збільшенням кількості онтологій та семантичних веб-додатків. Кількість та складність таких онтологій ускладнює розробникам вибір, які саме онтології використовувати. Для спрощення цієї проблеми пропонується алгоритм

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

модуляризації, який може використовуватися для розділення онтологій на набори модулів. Для того, щоб оцінити якість модуляризації, запропоновано нову оціночну метрику, яка кількісно оцінює ефективність модуляризації онтології.

В роботі [8] описано вирішення проблеми, пов'язаної із тим, що різноманітність способів концептуалізації домену призводить до створення різних онтологій із суперечливими частинами або частинами, які перекриваються. З цієї причини онтології потрібно узгодити (вирівняти). Одним з методів вирівнювання онтології є порівняння назв класів та властивостей онтологій за допомогою метрик, що грунтуються на відстанях між літерними рядками.

Робота [9] – це огляд метрик для оцінки якості та правильності онтологій. ONTOMETRIC, OntoQA і Protégé представляють найважливіші інструменти для оцінки онтологій. Також існують різноманітні метрики зчеплення, зв'язку та ранжування, та такі методики, як OntoClean. В статті аналізуються ці інструменти та метрики, та розглядається сучасний стан метрик онтології:

У статті [10] описано деякі метрики для нормалізації онтологій, розглядається сучасний стан та в основному пропонується нормалізація як попередній процес застосування структурних метрик. Цей процес нормалізації складається з п'яти етапів: найменування анонімних класів, найменування анонімних індивідів, ієрархічної класифікації та уніфікації імен, поширення індивідів до найглибших можливих класів та нормалізації властивостей об'єкта.

Роботи [11] та [12] пропонують деякі метрики для ранжування онтологій. В основному ця пропозиція складається з сервлета Java для опрацювання як вхідних даних певних ключових слів, введених користувачем. Потім фреймворк здійснює пошук за допомогою методу Swoogle1 і отримує всі уніфіковані ідентифікатори ресурсу, що представляють онтології, пов'язані з цими ключовими словами. Потім фреймворк здійснює пошук у своїй внутрішній базі даних, чи ці онтології раніше були проаналізовані та отримує їх інформацію. Нарешті, фреймворк ранжує отримані онтології.

Orme та ін. [13] запропонували набір метрик зв'язку для систем на основі онтологій, представлених в мові веб-онтології, такими метриками є: кількість зовнішніх класів, посилань на зовнішні класи та включених посилань. Ця пропозиція визначає новий тип вимірювання зв'язку для розробки системи, який визначає метрику зв'язку на основі даних онтології та її структури. Перша запропонована метрика – представляє кількість різних зовнішніх класів, визначених поза онтологією, які використовуються для визначення нових класів та властивостей онтології. Зовнішні класи можуть включати стандартні класи, визначені як примітиви мови онтології, та визначені користувачем класи з інших онтологій. Це – пряма міра кількості класів онтології. Друга метрика – це кількість посилань на зовнішні класи онтології, – пряма міра кількості розгалужень (у цьому випадку розгалуження – це різні ієрархії класів із зовнішніми коренями). Третя метрика – це пряма міра кількості включених посилань в онтологію. Для поєднання цих метрик використовується стандартний аналізатор на основі XML.

В роботі [14] автори запропонували набір метрик з'єднання онтологій для вимірювання модульної зв'язаності мові веб-онтології. Ці показники – кількість кореневих класів (NoR), кількість листових класів (NoL) та середня глибина ієрархічного дерева успадкування для всіх листових вузлів (ADIT-LN). Автори визначають NoR-метрику як загальну кількість кореневих класів, явно визначених в онтології. Кореневий клас в онтології означає, що клас не має семантичного суперкласу, явно визначеного в онтології. Метрика NoL визначається як кількість класів листів, явно визначених в онтології. Листовий клас в онтології означає, що клас не має семантичного підкласу, явно визначеного в онтології. Нарешті ADIT-LN визначається як сума глибин усіх шляхів, поділена на загальну кількість шляхів.

Yinglong та ін. [15] запропонували інший набір метрик зв'язності онтології для вимірювання модульної зв'язаності онтологій у контексті динамічної та мінливої мережі Web. Ці метрики були визначені з урахуванням принципу зв'язаності з об'єктно-орієнтованого підходу, адаптованого до онтологій. Автори зосереджуються на вимірюванні невідповідностей в онтологіях і повністю розглядають онтологічну семантику, а не структуру. Метрики, які вони пропонують – це кількість онтологічних розділів (NOP), кількість мінімально неузгоджених підмножин (NMIS) та середнє значення аксіомних неузгодженостей (AVAI). У цій роботі також описані алгоритми для обчислення цих метрик та перевірки метрик за допомогою платформ валідації. Автори визначають метрику NOP як кількість семантичних розділів бази знань. NMIS визначається як кількість усіх мінімально несумісних підмножин у базі знань. Цей показник корисний для вимірювання масштабу

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

впливу неузгодженості бази знань. Третя метрика AVAI визначається як: відношення суми значень впливу неузгодженості всіх аксіом і тверджень до кардинальності (кількості елементів) бази знань.

Методологія OntoClean [16] пропонує використовувати деякі визначені мета-властивості: жорсткість, єдиність, ідентичність та залежність. Методологія складається з присвоєння цих мета-властивостей сутностям з метою надання їм логічного та смислового значення. Застосування цих мета-властивостей призводить до накладення декількох обмежень на таксономічну структуру онтології та дозволяє розробити концептуальний аналіз концептів та їх обґрунтованості. Більше того, ця методологія дозволяє аналізувати та виявляти логічно не узгоджені відношення.

Z.Yang та ін. [17] запропонували метрики, які враховують еволюцію онтологій. Автори пропонують набір метрик складності, які в основному розглядають кількість, співвідношення та корелятивність понять та відношень, щоб оцінити онтології з точки зору складності та її еволюції. Ці метрики поділяються на дві групи: примітивні метрики та метрики складності. Примітивні метрики включають загальну кількість понять або класів, загальну кількість відношень, загальну кількість шляхів, де шлях визначається як маршрут від певного конкретного концепту до найбільш загального концепту в онтології. Перша метрика складності визначається як середня кількість відношень на концепт. Друга метрика – це середня кількість шляхів на концепт.

В роботі [18] розглядається питання пошуку ефективного методу оцінки онтологій, наводиться огляд наявних методів оцінки онтології, обговорюються їх переваги та недоліки. Представлені методи оцінки онтології можна згрупувати у чотири категорії: підходи, що грунтуються на золотому стандарті, на основі сукупності елементів, на основі завдань та на основі критеріїв.

Наприклад, Maedche і Staab (2002) розглядають онтології як двошарові системи, що складаються з лексичного та концептуального шару. На основі цієї основної моделі онтології цей підхід вимірює схожість між вивченою онтологією та онтологією сфери туризму, що моделюється експертами. Він вимірює подібність на основі поняття лексикону, опорних функцій та семантичної копії, які детально описані в [19].

Крім того, Ponzetto і Strube [20] оцінюють похідну систематику з Вікіпедії, порівнюючи її з двома базовими таксономіями. Цей підхід відображає вивчену систематику з ResearchCyc за допомогою денотаційного картографу лексеми до концепції.

Treeratpituk та ін. [21] оцінюють якість побудованої систематики з великого текстового корпусу, порівнюючи її з шістьма специфічними таксономіями золотого стандарту. Ці шість еталонних таксономій генеруються з Вікіпедії за допомогою запропонованого ними алгоритму GraBTax.

Zavitsanos та ін. [22] також оцінюють вивчену онтологію на основі золотого відліку. Цей новий підхід перетворює поняття онтології та їх властивості у векторне представлення простору та обчислює схожість та несхожість двох онтологій на лексичному та реляційному рівнях.

Цей тип підходу застосовують також Kashyap та Ramakrishnan [23]. Вони використовують базу даних MEDLINE як корпус документа, а тезаурус MeSH як золотий стандарт для оцінки побудованої систематики. Процес оцінювання порівнює сформовану таксономію з еталонною систематикою з використанням двох класів метрик: (1) Метод якості вмісту: він вимірює перекриття міток між двома таксономіями з метою вимірювання точності та відкликання. (2) Структурний показник якості: він вимірює структурну обґрунтованість етикетки. Тобто, коли в одній таксономії з'являються дві мітки у відносинах батько-дитина, вони повинні з'являтись у послідовному взаємозв'язку (батько-дитина або пращур-нащадок) в іншій таксономії.

Метрики для оцінювання зв'язків між концептами в адаптивних онтологіях

В роботах [24-27] запропоновано метрики на основі адаптивних онтологій для семантичних (наприклад, задач класифікації) та ознакових (наприклад, пошук релевантних прецедентів) задач.

Для семантичних задач відстань між прецедентом і ситуацією визначається як сума відстаней між „найважливішими" поняттями прецедента та поточного випадку. Найважливіше поняття відповідає центру ваг концептуального графа, за допомогою якого подається адаптивна онтологія. Розглядається максимум три «найважливіших» поняття концептуального графа. В цьому випадку отримуємо три центри ваг *i*-го прецедента та три центри ваг поточної ситуації s^1, s^2, s^3. Відстань між прецедентом та поточною ситуацією визначається як

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

$$d(pr,s) = \arg\min \sum_{n=1}^{3} d_n ,\ d_n = d(pr_i^j, s^k),\ j = 1,2,3,\ k = 1,2,3$$

– з дев'яти різних відстаней $d\left(pr_i^j, s^k\right)$, j=1,2,3; k=1,2,3 вибираються такі три, щоб їх сума була мінімальною. Отримана сума і буде відстанню між прецедентом та поточною ситуацією.

Вказана метрика не оцінює всіх шляхів від одного концепту до іншого.

Онтології для опису метрик

В роботі [28] описано розроблення онтології для організації інформації про метрики та її потенційного застосування для визначення та управління метриками в проекті CTSA (Clinical and Translational Science Award). Метою є підтримка інтегрованої бази даних всіх показників, що використовуються компонентами CTSA. Онтологія подається у вигляді концептуальної схеми даних типу «сутність-зв'язок».

Метою даної роботи є розроблення обернено-адитивної семантичної метрики, придатної для обчислення відстаней між концептами онтології з нетаксономічними зв'язками.

Для досягнення мети було необхідно вирішити наступні завдання:

Провести огляд існуючих семантичних метрик, що дозволить проаналізувати семантичні метрики, які використовуються для обчислення відстані між концептами в онтологіях, визначити переваг та недоліків різних підходів.

Розробити обернено-адитивну семантичну метрику, визначити теоретичні основи обернено-адитивної семантичної метрики, формалізувати її за допомогою математичних виразів та алгоритмів.

Провести оцінку обернено-адитивної семантичної метрики шляхом порівняння обернено-адитивної метрики з іншими існуючими метриками за допомогою експериментів та аналізу результатів та оцінювання точності, ефективності та надійності метрики в різних сценаріях.

Провести аналіз впливу обернено-адитивної семантичної метрики, зокрема, дослідження її впливу на виконання різних завдань з опрацювання онтологій, таких як пошук інформації, класифікація та виявлення знань, визначення потенційних сфер застосування метрики.

Аналіз ситуацій та формування прикладів у даній роботі кореспондуються з розробленням реальних інформаційних систем безпекового типу для багатоквартирних будинків, кварталів, мікрорайонів та міст. Водночас, результати дослідження можуть бути використані для розвитку методів семантичного аналізу та опрацювання онтологій, що може бути втілене в системах штучного інтелекту, інженерії даних і знань, в ряді інших галузей.

Онтологія – це «експліцитна специфікація концептуалізації» [29]. Онтологія – формальне та явне визначення концептуалізації, яка містить: концепти (поняття предметної області), їх визначення, ієрархічну організацію понять, відношення між поняттями, аксіоми для формалізації визначень та відношень.

Тут концептуалізація означає введення абстрактних об'єктів для абстрактного, спрощеного опису навколишнього світу.

Для автоматизації семантичного аналізу текстів потрібно ввести метрики на мережевих структурах для оцінювання кількісних характеристик, що описують онтології в цілому, відношення між концептами певної онтології.

Проаналізовані метрики можна поділити на два види:

1. Метрики, що описують сукупні інтегральні характеристики онтології.

Вони дозволяють порівнювати різні онтології.

Інтегральні характеристики – це такі характеристики онтології, які

- оцінюють онтологію в цілому;
- дозволяють порівнювати різні онтології між собою.

Такими характеристиками є, наприклад, наступні характеристики графу онтології:

- кількість вузлів,
- кількість ребер,
- діаметр,
- максимальна, мінімальна, середня степінь вузла, тощо.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

2. Метрики, що описують характеристики відношень між концептами однієї онтології.

Вони дозволяють порівнювати концепти в алгоритмах при розв'язуванні таких задач, як автоматичне реферування, пошук та оцінювання текстів, зокрема таких, як технічні умови, технічні завдання та інші види технічної документації тощо.

Всі відношення між концептами можна поділити на дві категорії:

- Таксономічні відношення.

Вони описують ієрархічні зв'язки «is-a» («…є різновидом…») між концептами, тобто відношення «підмножина – множина», «множина – надмножина».

- Не таксономічні відношення.

Вони описують не ієрархічні зв'язки, наприклад, зв'язок *used-in* («термін_1 зустрічається у визначенні терміну_2») або зв'язок *uses-of* («визначення терміну_1 використовує термін_2», чи «термін_1 посилається на термін_2»).

Наявні метрики цього виду оцінюють відстань між концептами в припущенні існування єдиного шляху від одного концепту до іншого.

У випадку, якщо онтологія є таксономією, тобто всі зв'язки між концептами – ієрархічні, то таке припущення справедливе і зазначені метрики вірні.

Проте, якщо існують не таксономічні відношення між концептами (наприклад, «термін_1 посилається на термін_2»), то така метрика не буде вірною, бо вони передбачають існування багатьох шляхів від одного концепту до іншого.

Метрика, основана на багатьох зв'язках між концептами, має враховувати:

- відстань як кількість переходів в орієнтованому графі онтології на шляху від одного вузла-концепту до іншого,
- кількість таких шляхів.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.

Побудуємо метрику, яка дозволить оцінювати відстані між концептами у випадку нетаксономічних зв'язків в онтології на прикладі онтологічної моделі, що описує та структурує знання про безпекову систему багатоквартирного будинку, а також продемонструємо можливість її адаптації та масштабування, зокрема, для охоплення безпекових систем житлового кварталу, мікрорайону та міста загалом.

Вимогами до такої онтології є повнота, точність, ясність, масштабованість, інтегрованість.

Онтологія повинна охоплювати всі ключові поняття та елементи, пов'язані з безпекою багатоквартирного будинку, також мати потенціал для розширення до рівня житлового кварталу. Концептуальні відношення та властивості в онтології повинні точно відображати реальні відношення та властивості в системах безпеки. Онтологія повинна бути чітко та зрозуміло подана з використанням формальної мови моделювання, та бути чітко документованою. Онтологія повинна бути модульною та розширюваною, щоб її можна було легко адаптувати до потреб безпекових систем житлового кварталу, району та ін. Вона повинна бути сумісною з існуючими стандартами та моделями знань у галузі безпеки житлових будівель. Безперечно, при розробленні онтології важливо врахувати різні типи систем безпеки, які використовуються в багатоквартирних будинках, такі як системи відеоспостереження, системи контролю доступу та системи протипожежної безпеки. Онтологія формується таким чином, щоб вона могла легко інтегруватися з існуючими системами безпеки та інструментами управління в них. Саме такий підхід сприятиме можливості переходу від онтології безпекової системи багатоквартирного будинку до онтології безпекової системи житлового кварталу та ін. Безперечно такий перехід потребує внесення низки додаткових вимог серед яких розширення обсягу охоплення, формування ієрархічної структури, використання багатоагентного підходу, розподіленість архітектури, забезпечення можливості інтеграції, наявність засобів управління динамічними змінами, забезпечення конфіденційності та безпеки.

Процес переходу від однієї онтологічної системи до іншої, ширшої, передбачає включення нових сутностей, до прикладу таких як територія житлового кварталу (будівлі, дороги, зелені зони), в'їзди та виїзди з кварталу, інфраструктура (мережі освітлення, водопостачання, електропостачання), персонал охорони та обслуговування, системи централізованого моніторингу та управління. Удосконалюються ієрархічні відношення між сутностями, наприклад, «будівля є частиною житлового кварталу» та застосовуються принципи композиції та успадкування для презентації спільних характеристик та спеціалізації різних типів сутностей (житлові будинки, адміністративні будівлі, комерційні приміщення). Проводиться моделювання взаємодії між різними

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

агентами системи, такими як системи безпеки окремих будівель, центральний пульт управління, персонал охорони, мешканці житлового кварталу, визначення протоколів та правил взаємодії для забезпечення координації та спільних дій.

Розроблення розширеної онтології відбувається з урахуванням розподіленої архітектури системи безпеки житлового кварталу. Моделювання локальної автономії та децентралізованого прийняття рішень відбувається на рівні окремих будівель з можливістю подальшого централізованого контролю та координації. Важливим є забезпечення можливості інтеграції в онтології описів систем, що використовуються в житловому кварталі (системи диспетчеризації, системи автоматизації будівель, системи відеоспостереження) та використання стандартів та протоколів обміну даними для забезпечення їх сумісності та взаємодії. Розроблення механізмів для оновлення та адаптації онтології відповідно до динамічних змін в житловому кварталі (будівництво нових будівель, зміна інфраструктури, зміна персоналу) та використання методів машинного навчання для автоматичного виявлення та опрацювання змін в даних про систему безпеки є перспективними кроками трансформаційних змін, пов'язаних з набуттям нових знань. При цьому важливим є впровадження механізмів контролю доступу та захисту даних для забезпечення конфіденційності та безпеки інформації про систему безпеки житлового кварталу та використання криптографічних методів для захисту даних під час їх зберігання та передачі. Наявність можливості впровадження зазначених змін дозволить створити гнучку та масштабовану онтологічну систему, яка коректно представлятиме та підтримуватиме роботу складної динамічної системи безпеки житлового кварталу.

Масштабування та формування нових аплікацій онтологічних систем вимагає побудови метрики мережевих структур, що була б придатною для вимірювання відстаней між концептами таких онтологій.

Ця метрика повинна забезпечувати оцінювання відстаней між концептами онтології в припущенні існування не єдиного шляху від одного концепту до іншого. Такі мережні структури описуються довільними орієнтованими графами.

Обернено-адитивна метрика

Позначимо N_i – кількість переходів від концепту A до концепту B по i-му шляху, i=1, …, K, де K – кількість різних шляхів, якими можна перейти по орієнтованому графу певної онтології від концепту A до концепту B.

Визначимо відстань $d(A, B)$ між концептами A та B наступним чином:

$$\frac{1}{d(A,B)} = \sum_{i=1}^{K} \frac{1}{N_i}$$

Приклад:

Припустимо, що між концептами A та B є два шляхи. Перший шлях містить один перехід, а другий – два (рис.2):

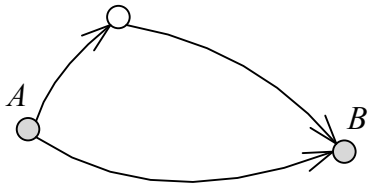


Рис.2. Ілюстрація до прикладу обернено-адитивної метрики

Тоді відстань $d(A, B)$ між ними:

$$\frac{1}{d(A,B)} = \frac{1}{1} + \frac{1}{2} = \frac{3}{2}, \quad d(A,B) = \frac{2}{3}$$

Аналогією до цієї метрики є правило обчислення електричного опору для послідовного і паралельного з'єднання. На основі закону Ома, для послідовного з'єднання опорів R_1 та R_2 загальний опір

$$R = R_1 + R_2,$$

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

для паралельного з'єднання:

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}$$

Обґрунтування

Як відомо, метрика ґрунтується на понятті відстані. Відстань $d(x, y)$ – однозначна, невід'ємна, дійсна функція d: $X \times X \to \mathbb{R}$, визначена для $\forall$ x, y $\in$ X, яка задовольняє трьом аксіомам метрики:

1) $d(x, y) = 0 \Leftrightarrow x \equiv y$ (аксіома тотожності)
2) $d(x, y) = d(y, x)$ (аксіома симетрії)
3) $d(x, z) \le d(x, y) + d(y, z)$ (аксіома трикутника)

Доведемо, що ці аксіоми справджуються для вказаної метрики:

Аксіома тотожності

$d(x, y) = 0 \Leftrightarrow x \equiv y$ (аксіома тотожності)

В нашому випадку $R(A, A) = N$, де N – кількість переходів від вузла A до вузла A, N=0.

Перша аксіома – справджується.

Аксіома симетрії

$d(x, y) = d(y, x)$ (аксіома симетрії)

Слід зауважити, що в загальному випадку для орієнтованого графа не може існувати симетрії в тлумаченні другої аксіоми метрики. Тобто, не може виконуватися правило $R(A, B) = R(B, A)$, оскільки кількість переходів від вузла A до вузла B буде збігатися із кількістю переходів від вузла B до вузла A лише у тому випадку, коли є цикл $A \to B \to A$ та відстань (кількість переходів) на шляху $A \to B$ дорівнює відстані на шляху $B \to A$.

Введення пар симетричних зв'язків, наприклад, для онтології – це пара зв'язків використовує (*uses-of*) – використовується (*used-in*), дозволяє забезпечити виконання аксіоми симетрії для запропонованої метрики в наступному тлумаченні:

$$d_{used-in}(A, B) = d_{uses-of}(B, A)$$

Якщо розглядати пару взаємно симетричних відношень (відношення *uses-of* симетричне до відношення *used-in*), то друга аксіома – справджується.

Аксіома трикутника

$d(x, z) \le d(x, y) + d(y, z)$ (аксіома трикутника)

Розглянемо орієнтований граф, для якого обґрунтуємо правило трикутника (рис.3):

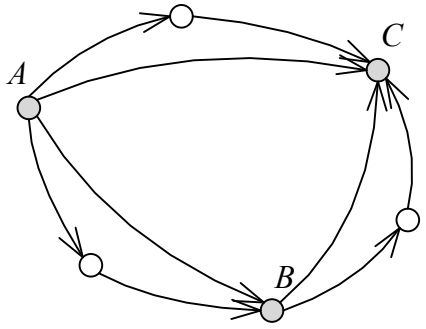


Рис.3.Ілюстрація до правила трикутника для обернено-адитивної метрики

Позначимо:

$d(A, B) = d(\widehat{AB})$ – відстань між вузлами A та B. $\widehat{AB}$ – всі шляхи від вузла A до вузла B.
$d(B, C) = d(\widehat{BC})$ – відстань між вузлами B та C. $\widehat{BC}$ – всі шляхи від вузла B до вузла C.
$d(A, C)$ – відстань між вузлами A та C.
$d(\widehat{AC})$ – відстань між вузлами A та C по шляху $\widehat{AC}$, який не проходить через вузол B.
$d(\widehat{ABC})$ – відстань між вузлами A та C по шляху $\widehat{ABC}$, який проходить через вузол B.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

Згідно з визначенням цієї метрики, після того, як до шляхів $\widehat{AB}$ та $\widehat{BC}$ буде додано шлях $\widehat{AC}$, відстань $d(A,C)$ буде визначатися формулою

$$\frac{1}{d(A,C)} = \frac{1}{d(\widehat{AC})} + \frac{1}{d(\widehat{ABC})}$$

Тоді

$$d(A,C) = \left(\frac{1}{d(\widehat{AC})} + \frac{1}{d(\widehat{ABC})}\right)^{-1} = \left(\frac{1}{d(\widehat{AC})} + \frac{1}{d(\widehat{AB}) + d(\widehat{BC})}\right)^{-1} =$$
$$= \left(\sum_{i=1}^{K_{\widehat{AC}}} \frac{1}{N_i} + \left(\left(\sum_{i=1}^{K_{\widehat{AB}}} \frac{1}{N_i}\right)^{-1} + \left(\sum_{i=1}^{K_{\widehat{BC}}} \frac{1}{N_i}\right)^{-1}\right)^{-1}\right)^{-1}$$

Оскільки

$$d(A,B) = d(\widehat{AB}) = \left(\sum_{i=1}^{K_{AB}} \frac{1}{N_i}\right)^{-1} = \left(\sum_{i=1}^{K_{\widehat{AB}}} \frac{1}{N_i}\right)^{-1}$$

та

$$d(B,C) = d(\widehat{BC}) = \left(\sum_{i=1}^{K_{BC}} \frac{1}{N_i}\right)^{-1} = \left(\sum_{i=1}^{K_{\widehat{BC}}} \frac{1}{N_i}\right)^{-1}$$

то

$$\frac{1}{d(A,C)} \geq \frac{1}{d(A,B) + d(B,C)}$$

Звідси

$$d(A,C) \leq d(A,B) + d(B,C)$$

Що і слід було довести.

Розглянемо випадок, зображеного на рис. 3.
Тут:

$$d(A,B) + d(B,C) = \frac{4}{3}$$
$$\frac{1}{d(A,C)} = \sum_{i=1}^{6} \frac{1}{N_i} = \frac{1}{1} + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} = \frac{24+8+3}{12}$$
$$d(A,C) = \frac{12}{35} \leq R(A,B) + R(B,C) = \frac{4}{3}$$

Пошук відстані між концептами

Як приклад розглянемо концептуальну онтологію, яка відображає зв'язки між ключовими концептами безпекових систем багатоквартирних будинків:

Безпека - основна концепція, що включає всі аспекти захисту багатоквартирних будинків.

Системи безпеки - це узагальнене уявлення про комплекс взаємопов'язаних елементів, заходів, методів та процесів, спрямованих на забезпечення захисту певної системи, організації або середовища від потенційних загроз, ризиків або небезпек. Цей концепт охоплює всі аспекти безпеки, включаючи фізичні, інформаційні, кібернетичні, правові та інші компоненти, які разом формують цілісну структуру для підтримки стійкості та збереження цілісності захищеної системи і включає концепти **відеоспостереження, контроль доступу, сигналізація, протипожежні системи, інтелектуальні системи, аварійні системи.**

Багатоквартирні будинки - об'єкт, для якого розробляються і впроваджуються **системи безпеки**.

Кожна складова концепту **системи безпеки** розкладається на окремі елементи - типи.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

Відеоспостереження - тип **систем безпеки**, який використовується для моніторингу і складається з систем моніторингу, дистанційного контролю.

Контроль доступу -тип систем безпеки, який регулює вхід і вихід в будівлю і містить концепти **домофонні системи**, **інтелектуальні системи**, **інтегровані системи безпеки**.

Охоронні системи-тип систем безпеки, що забезпечує захист від несанкціонованого доступу і містить концепти **охоронні компанії**, **приватна безпека**.

Пожежна безпека- частина безпеки, спрямована на запобігання пожежам і містить концепти **протипожежні системи, аварійні системи**.

Давачі руху - компонент систем безпеки, використовується для виявлення руху і містить концепти сигналізації, інтелектуальні системи моніторингу.

Сигналізація- тип систем безпеки, що попереджає про небезпеку і містить зв'язок з концептом **давачі руху, аварійні системи**.

Інтелектуальні системи - розширені системи безпеки з використанням технологій штучного інтелекту і містить зв'язок з концептами **відеоспостереження, доступ контролю, давачів руху.**

Домофонні системи - тип систем безпеки, що забезпечує аудіо- і відеозв'язок між мешканцями і відвідувачами і містить зв'язок з концептами **контроль доступу**, **інтеграційні систем безпеки.**

Моніторинг - процес спостереження за системами безпеки і передбачає зв'язок з концептами **відеоспостереження, інтелектуальні системи, дистанційний контроль**.

Аварійні системи - системи, що активуються в надзвичайних ситуаціях і передбачає зв'язок з концептами **сигналізація, протипожежна система.**

Вентиляційні системи- інфраструктурні системи, що підтримують якість повітря і передбачає зв'язок з концептом **пожежна безпека**.

Охоронні компанії - організації, що надають послуги безпеки передбачає зв'язок з концептами **охоронні системи, приватна безпека**.

Інтегрована система безпеки – система, що об'єднує різні системи безпеки в єдину структуру передбачає зв'язок з **концептами відеоспостереження, контроль доступу, охоронна система, домофонні системи, інтелектуальні системи** .

Приватна безпека - служби безпеки, що надаються приватними компаніями передбачає зв'язок з концептами **охоронні компанії, охоронні системи**.

Кібербезпека - захист цифрових систем, що використовуються для систем безпеки передбачає зв'язок з концептами **інтелектуальні системи, інтегровані системи безпеки, дистанційний контроль.**

Дистанційний контроль -управління системами безпеки з віддаленого місця передбачає зв'язок з концептами **відеоспостереження, моніторингу, інтелектуальні системи**.

Протипожежні системи - системи, що запобігають або ліквідують пожежі передбачає зв'язок з концептами **пожежна безпека, аварійні системи**.

Наведена онтологія відображає базові взаємозв'язки між ключовими концептами, що дозволяє краще розуміти структуру та взаємодію компонентів у безпекових системах багатоквартирних будинків, кварталів, мікрорайонів та ін.

Фрагмент owl-файлу, що містить онтологію концептів, має вигляд, поданий на рис.4:

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

```
// Основний концепт
file << "  OS [label=\"Охоронні системи\"];\n";

// Компоненти охоронних систем
file << "  D [label=\"Давачі\"];\n";
file << "  S [label=\"Сигналізація\"];\n";
file << "  KD [label=\"Контроль доступу\"];\n";
file << "  VS [label=\"Відеоспостереження\"];\n";
file << "  CSU [label=\"Централізовані системи управління\"];\n";

// Типи давачів
file << "  DR [label=\" Давачі руху\"];\n";
file << "  DO [label=\"Давач відкриття дверей/вікон\"];\n";
file << "  DS [label=\" Давачі розбиття скла\"];\n";
file << "  DD [label=\" Давачі диму\"];\n";

// Типи сигналізації
file << "  ZS [label=\"Звукові сигналізації\"];\n";
file << "  SS [label=\"Світлові сигналізації\"];\n";
file << "  MO [label=\"Мобільні оповіщення\"];\n";

// Типи контролю доступу
file << "  KSD [label=\"Карткові системи доступу\"];\n";
file << "  BZD [label=\"Біометричні системи доступу\"];\n";
file << "  DS1 [label=\"Домофонні системи\"];\n";
file << "  EZ [label=\"Електронні замки\"];\n";
```

Рис.4. Фрагмент owl-файлу, що містить онтологію концепту «Охоронні системи»

Як бачимо, подається основний концепт, а потім його складові. Компоненти охоронних систем», «Типи давачів», «Типи сигналізації», «Типи контролю доступу». Кожний концепт та складова починається тегом //. Концепти, пов'язані з поточним концептом відношенням «складається з», вказані в підсекціях, які відповідають тегам file << . Таким чином, синтаксичний аналіз owl-файлу дозволяє виявляти зв'язки між концептами онтології.

Використовуючи відомі алгоритми обходу орієнтованого графу можемо знайти всі наявні шляхи від одного заданого концепту до іншого, і на їх основі– обчислити відстані між заданими концептами. Фактично, слід розв'язати задачу транспортного потоку в орієнтованому графі онтології від вузла-джерела, що відповідає першому концепту, до вузла-приймача, що відповідає другому концепту.

Обговорення та аналіз отриманих результатів. Є певні проблеми, пов'язані з тим, що в загальному випадку онтологія є орієнтованим графом, вузли якого –концепти, а ребра –зв'язки між концептами.

Онтологія подається у вигляді мультиграфа – оскільки допускається наявність кратних ребер: прості ребра, які мають одні й ті ж самі кінцеві вершини. Іншими словами, дві вершини можуть бути з'єднані більш ніж одним ребром.

Проблема симетрії зв'язків між концептами – проблема симетрії зв'язків для орієнтованого графа

Проблема полягає в тому, що в загальному випадку, для орієнтованого графа онтології правило симетрії не виконується:

$$d(A,B) \neq d(B,A)$$

Тут відстань між концептами визначена на основі морфологічної метрики – як кількість переходів N між вузлами орієнтованого графа онтології на шляху від концепту A до концепту B:

$$d(A,B) = N$$

Цю проблему можна вирішити шляхом побудови пар взаємо-обернених зв'язків, наприклад для кожного зв'язку будуємо обернений до нього зв'язок.

Можна виділити два підходи до побудови не таксономічних відношень між концептами в онтологіях.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

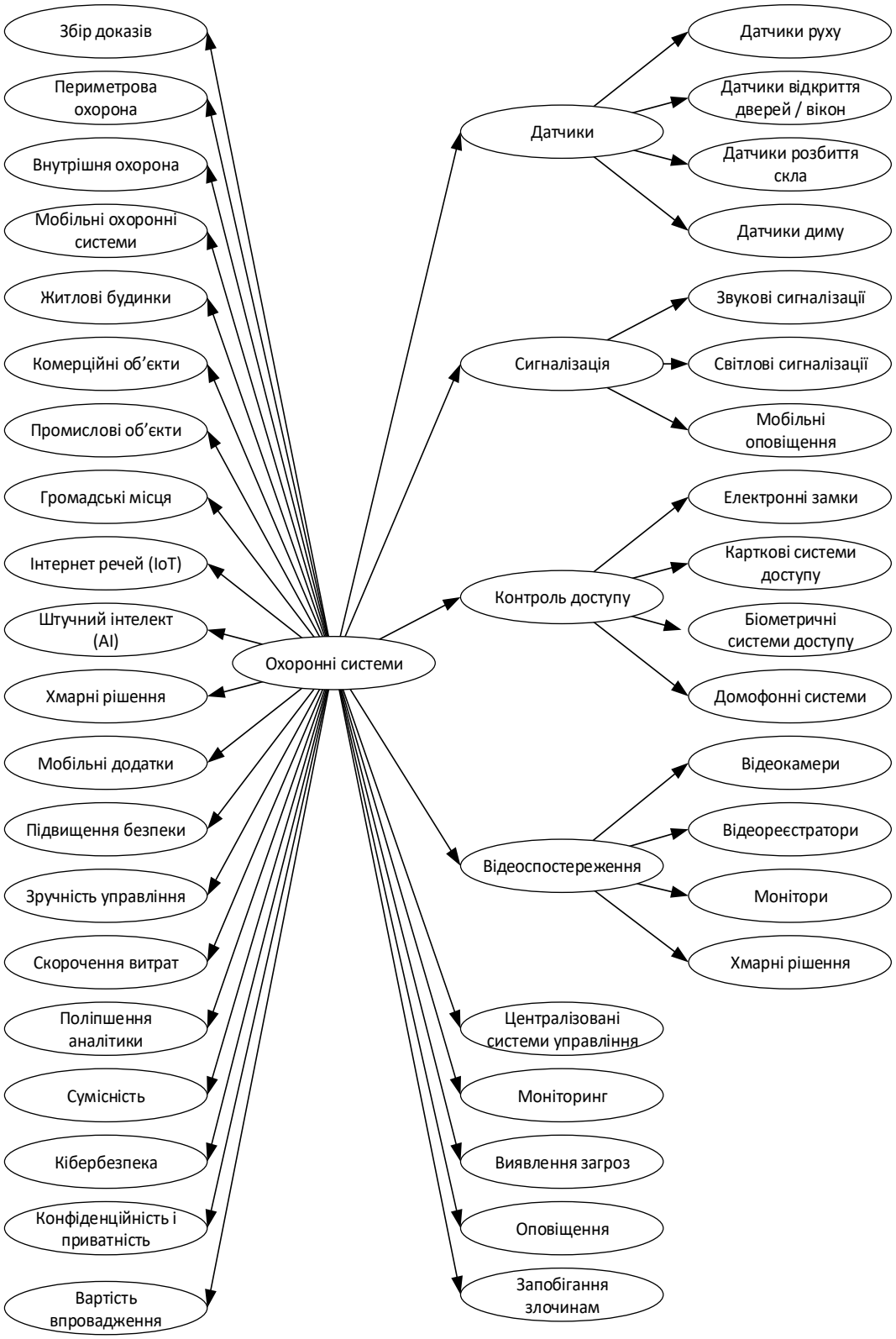


Рис. 5. Онтологія концепту «Охоронні системи»

Перший підхід забезпечує максимальну повноту опису предметної області в онтології. Згідно цього підходу, ми якнайповніше представляємо зв'язки між поняттями предметної області за допомогою відношень між концептами в онтології.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

Якщо розглядати не таксономічні відношення, то кожна пара концептів, які представлені сусідніми вузлами в семантичній мережі онтології, буде представлена парою зв'язків *used-in* та *uses-of*.

При цьому велика кількість зв'язків між концептами робить онтологію занадто «громіздкою», що ускладнює її практичне використання в прикладних системах.

Побудуємо семантичну мережу для понять ПРОГРАМА та АЛГОРИТМ на основі наступного визначення: «програма – це запис алгоритму за допомогою мови програмування».

Визначення терміну ПРОГРАМА використовує (*uses-of*) термін АЛГОРИТМ, а термін АЛГОРИТМ використовується (*used-in*) при визначенні терміну ПРОГРАМА (рис.5):

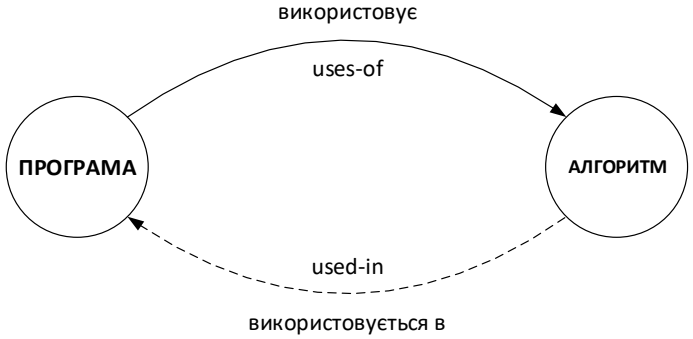


Рис.5 – Ілюстрація до семантичної мережі із зв'язками «використовує» (*uses-of*) та «використовується в» (*used-in*) між концептами

Відстань між концептами *A* та *B* в такій семантичній мережі можна визначити як відстань, що враховує існування двох шляхів, що відповідають симетричним зв'язкам «використовує» (*uses-of*) та «використовується» (*used-in*) між цими концептами:

$$\frac{1}{d(A,B)} = \frac{1}{d_{used-in}(A,B)} + \frac{1}{d_{uses-of}(A,B)}$$

Саме цей підхід дозволяє забезпечити симетрію зв'язків, при цьому

$$d(A,B) = d(B,A)$$

Другий підхід забезпечує простоту подання онтології та, як наслідок, – швидкість обчислень та ефективність при побудові прикладних систем. Оскільки відношення *used-in* виводяться із відношень *uses-of*, можна не визначати їх явно в онтології. При цьому ми навмисно нехтуємо повнотою представлення предметної області задля спрощення онтології.

Для нашого прикладу семантична мережа відповідає фразі «Визначення терміну ПРОГРАМА використовує (*uses-of*) термін АЛГОРИТМ» (рис.6):

Рис.6. Ілюстрація до семантичної мережі, що має лише зв'язки «використовує» (*uses-of*) між концептами

Для такого підходу правило симетрії не виконується:

$$d(A,B) \neq d(B,A)$$

Проблема правила трикутника для орієнтованого графа

Як і з аксіомою симетрії, є проблеми з аксіомою трикутника: в загальному випадку правило трикутника для орієнтованого графа не виконується.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

Знову, розглянемо випадок, коли відстань між концептами визначена на основі морфологічної метрики – як кількість переходів *N* між вузлами орієнтованого графа онтології на шляху від концепту *A* до концепту *B*:

$$d(A, B) = N$$

Розглянемо наступний приклад (рис.7):

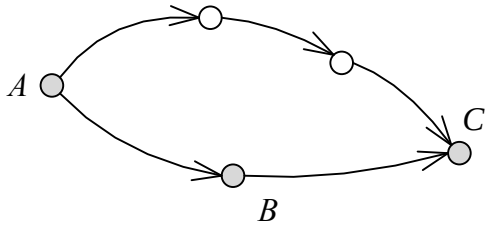


Рис.7. Ілюстрація до проблеми правила трикутника для морфологічної метрики в орієнтованому графі

Для цього прикладу

$$d(A, B) = 1,\ d(B, C) = 1,\ d(A, C) = 3,$$

Очевидно, що у цьому випадку правило трикутника

$$d(A, C) \le d(A, B) + d(B, C)$$

– не виконується.

Природня метрика [1, 2] дозволяє вирішити цю проблему: відстань між вузлами визначена як $d(A, B) = \min\{N_i\}$, де N_i – кількість переходів від концепту *A* до концепту *B* по *i*-му шляху, $i=1 \ldots, K$, де *K* – кількість різних шляхів, якими можна перейти по орієнтованому графу певної онтології від концепту *A* до концепту *B*.

Тоді, враховуючи, що від вузла *A* до вузла *C* існує 2 шляхи: шлях $\widehat{ABC}$, який проходить через концепт *B*, та шлях $\widehat{AC}$, який не проходить через концепт *B*, і що $d(A, B) = 1$, $d(B, C) = 1$, $d(\widehat{ABC}) = 2$, $d(\widehat{AC}) = 3$, $d(A, C) = \min\{ d(\widehat{ABC}), d(\widehat{AC})\} = \min\{2, 3\} = 2$, отримаємо правило трикутника

$$2 = d(A, C) \le d(A, B) + d(B, C) = 2$$

Ще одним варіантом вирішення проблеми може стати введення обернено-адитивної метрики, яка теж враховує ситуацію, при якій є два шляхи від концепту *A* до концепту *C*.

Тоді

$$d(A, B) = 1,\ d(B, C) = 1$$

$$\frac{1}{d(A,C)} = \frac{1}{d(\widehat{ABC})} + \frac{1}{d(\widehat{AC})} = \frac{1}{d(A,B) + d(B,C)} + \frac{1}{d(\widehat{AC})} = \frac{1}{2} + \frac{1}{3} = \frac{5}{6}$$

Звідси

$$d(A, C) = \frac{6}{5} = 1{,}2$$

Тепер правило трикутника виконується:

$$1{,}2 = d(A,\ C) \ \le\ d(A,\ B)\ +\ d(B,\ C) = 2$$

Проблема циклів – циклічних зв'язків між концептами

Це – помилка в організації онтології, – проблема, яка виникає із-за наявності рекурсивних циклів.

Граф зв'язків між такими концептами зображено на рис.8:

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

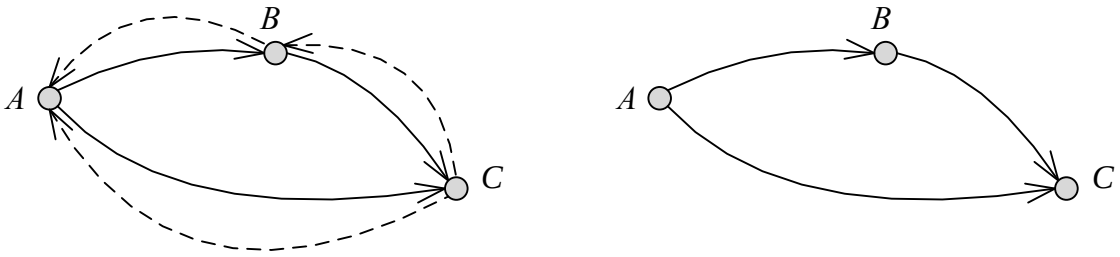


Рис.8. Ілюстрація до семантичних мереж, що мають рекурсивні зв'язки між концептами; ліворуч – «використовує» (*uses-of*) та «використовується в» (*used-in*); праворуч – лише «використовує» (*uses-of*)

Тоді відстань $d(A,B)$ між концептами A та B в цій семантичній мережі можна визначити так:

$$\frac{1}{d(A,B)} = \frac{1}{d_{AB}(A,B)} + \frac{1}{d_{BCA}(B,A)}$$

де $d_{AB}(A,B)$ – відстань від вузла A до вузла B по шляху $A \to B$, $d_{BCA}(B,A)$ – відстань від вузла B до вузла A по шляху $B \to C \to A$.

Визначена таким чином відстань $d(A,B)$ враховує наявність двох шляхів, які з'єднують вузли A та B.

Запропонований в даній роботі підхід за своєю сутністю дозволяє зробити ще один крок до автономної побудови систем алгоритмічних алгебр, які містили б оператори згортання структур проблемних областей для автономної побудови так званих знаннєвих клонів. Засоби алгоритмічної формалізації і синтезу знань предметних областей є тріадою – абстракції, біології та екології програмування. Як абстрактний механізм використовується алгебраїчний апарат теорії клонів. Біологічна компонента відповідає за розповсюдження отриманих алгоритмічних знань на інші задачі. Екологічна компонента призначена для формування інструментальних засобів підтримки методів абстрактної та біологічної компонент, що складають теорію клонів. В рамках екологічної компоненти пропонуються різні інтерпретації алгоритмічних операцій, дослідження паралелізмів та механізмів виводу алгоритмічних знань [30]. В подальшому планується продовжити дослідження методів та засобів «клонування» знань при побудові інтелектуальних інформаційних систем безпекового типу в їх розвитку від систем будинків, розрахованих на проживання однієї сім'ї до багатоквартирних будинків, кварталів, мікрорайонів, регіонів та загалом міст

Висновки. У рамках проведеного дослідження було розроблено методологію формування обернено-адитивної семантичної метрики, що дозволяє здійснювати більш точний та комплексний аналіз онтологій безпекових систем багатоквартирних будинків. Запропонована метрика дозволяє оцінювати відстані між концептами в онтології при наявності багатьох шляхів від одного концепту до іншого, допускає дробові значення відстаней між вузлами-концептами орієнтованого графа онтології та забезпечує вірну оцінку для не таксономічних відношень між концептами.

Запропонована метрика дозволить більш коректно та точно вирішувати задачі, пов'язані із семантичним аналізом текстів – автоматичного реферування, зокрема, текстів технічних умов, технічних завдань, автоматичної побудови семантичних мереж заданих текстів тощо. Основна мета запропонованої метрики полягає в забезпеченні більш високої точності у виявленні та оцінці семантичних зв'язків між різними елементами онтологій, що є критично важливим для процесів системного проектування та управління розлогими безпековими системами.

Продемонстровано можливості адаптації та масштабування онтології для безпекової системи житлового кварталу та проведено оцінювання корисності онтології для представлення, аналізу та управління знаннями про безпеку житлових будівель. Сформовано нові інструменти, які можна використовувати для проектування, розроблення та впровадження більш ефективних та надійних систем безпеки житлових будівель.

Запропонована обернено-адитивна семантична метрика демонструє більш високу ефективність у порівнянні з традиційними підходами до аналізу онтологій. Вона дозволяє знизити

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

рівень семантичної неоднозначності та підвищити точність відповідності елементів онтології, що сприяє більш точному моделюванню безпекових систем. Використання обернено-адитивної семантичної метрики дозволяє адаптувати систему до змін у навколишньому середовищі, забезпечуючи можливість оновлення та вдосконалення онтології відповідно до нових викликів і загроз. Це дозволить робити систему безпеки більш гнучкою та стійкою до можливих змін в умовах експлуатації. Завдяки більш точному аналізу онтологій, запропонована метрика може сприяти прийняттю більш обґрунтованих та ефективних безпекових рішень. Це особливо важливо в контексті багатоквартирних будинків, для яких системи безпеки повинні бути інтегровані та скоординовані з урахуванням інтересів великої кількості мешканців.

Запропонована метрика є одним з інструментів, що дозволяє покращити якість та ефективність аналізу онтологій у контексті систем безпеки багатоквартирних будинків. Водночас, запропонована метрика може використовуватись для побудови систем автоматичного генерування текстів програмного коду з використанням концепції клонування знань на базі формалізмів систем алгоритмічних алгебр, запропонованих свого часу в роботах В.М. Глушкова, Цейтліна, Л. М. Захарії.

Список бібліографічного опису

1. Fenton, N. E. Software Metrics: A Rigorous and Practical Approach. 2nd Edition / N. E. Fenton, S. L. Pfleeger. – Berlin: International Thompson Computer Press, 1997. – 195 с.
2. Wilson, R. J. Introduction to Graph Theory / R. J. Wilson. – Longman, 2010. – 201 с.
3. Naylor C. Build your own PC expert system / C. Naylor. – London: Sigma Press, 1983. – 167 с.
4. Теслер, Г. С. Метрики і норми в ієрархії категоріальних семантик і функцій / Г. С. Теслер // Математичні машини і системи. – 2005. – №2. – С. 65-68.
5. Величко, В. Ю. Розв'язання аналітичних задач в дискретних середовищах методами виведення за аналогією / В. Ю. Величко. – Київ: Інститут кібернетики, 2004. – 170 с.
6. Ngo, D. H. Generic Approach for Combining Linguistic and Context Profile Metrics in Ontology Matching / D. H. Ngo, Z. Bellahsene, R. Coletta // Proceedings of the ODBASE'2011: 10th International Conference on Ontologies, DataBases, and Applications of Semantics, Oct 2011, Crete, Greece. – Crete, 2011. – P. 800-807.
7. Algergawy A. A New Metric To Evaluate Ontology Modularization / A. Algergawy, S. Babalou, B. König-Ries // Proceedings of the 2nd International Workshop on Summarizing and Presenting Entities and Ontologies Co-located with the 13th Extended Semantic Web Conference. – Af Workshop, 15th Int. World Wide Web Conference. – Crete, 2006. – P. 235-240.
8. Stoilos, G. A String Metric for Ontology Alignment / G. Stoilos, G. Stamou, S. Kollias // In Proceedings of the International Semantic Web Conference ISWC 2005: The Semantic Web – ISWC 2005. – 2005. – Pp. 624-637.
9. García J. F. A Survey on Ontology Metrics / J. F. García, F. J. García-Peñalvo, R. Therón // In Proceedings of the World Summit on Knowledge Society WSKS 2010: Knowledge Management, Information Systems, E-Learning, and Sustainability Research. – 2010. – Pp. 22-27.
10. Vrandecic D. How to Design Better Ontology Metrics / D. Vrandecic, Y. Sure // The Semantic Web: Research and Applications. – Springer-Verlag, 2007. – Pp. 311-325.
11. Alani, H., Brewster, C., Shadbolt, N. Ranking Ontologies with AKTiveRank / H. Alani, C. Brewster, N. Shadbolt // In Proceedings of the International Semantic Web Conference, ISWC 2006: 5th International Semantic Web Conference (ISWC). – November 2006, Georgia, USA.
12. Alani, H. Metrics for Ranking Ontologies / H. Alani, C. Brewster // In Proceedings of the 4th International EON Workshop, 15th International World Wide Web Conference. – 2006.
13. Orme, A., Yao, H., Etzkorn, L. Coupling Metrics for Ontology-Based Systems / A. Orme, H. Yao, L. Etzkorn // IEEE Software. – Afiny, 2006. – P. 102-108.
14. Yao, H. Cohesion Metrics for Ontology Design and Application / H. Yao, A. Orme, L. Etzkorn // Journal of Computer Science. – 2005. – Vol. 1(1). – P. 107-113.
15. Ma, Y., Jin, B., Feng, Y. Semantic oriented ontology cohesion metrics for ontology-based systems / Y. Ma, B. Jin, Y. Feng // The Journal of Systems and Software. – Elsevier, 2009.
16. Guarino, N. An Overview of OntoClean / N. Guarino, C. Welty // The Handbook on Ontologies. – Berlin: Springer-Verlag, 2004. – Pp. 151-172; Guarino, N., Welty, C. Evaluating Ontological Decisions with OntoClean / N. Guarino, C. Welty // Communications of the ACM. – ACM Press, 2002. – Pp. 61-65.
17. Yang, Z. Evaluation Metrics for Ontology Complexity and Evolution Analysis / Z. Yang, D. Zhang, C. Ye // Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE'06). – 2006.
18. Raad, J. A Survey on Ontology Evaluation Methods / J. Raad, C. Cruz // Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. – Lisbon, Portugal, November 2015.
19. Maedche, A. Measuring similarity between ontologies / A. Maedche, S. Staab // In Knowledge engineering and knowledge management: Ontologies and the semantic web. – Springer Berlin Heidelberg, 2002. – Pp. 251-263.
20. Ponzetto, S. P. Deriving a large scale taxonomy from Wikipedia / S. P. Ponzetto, M. Strube // In AAAI. – Vol. 7, July 2007. – Pp. 1440-1445.
21. Treeratpituk, P. Graph-based Approach to Automatic Taxonomy Generation (GraBTax) / P. Treeratpituk, M. Khabsa, C. L. Giles // arXiv preprint arXiv:1307.1718. – 2013.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

22. Zavitsanos, E. Gold standard evaluation of ontology learning methods through ontology transformation and alignment / E. Zavitsanos, G. Paliouras, G. A. Vouros // IEEE Transactions on Knowledge and Data Engineering. – 2011. – Vol. 23(11). – Pp. 1635-1648.
23. Kashyap, V. Taxa Miner: an experimentation framework for automated taxonomy bootstrapping / V. Kashyap, C. Ramakrishnan, C. Thomas, A. Sheth // International Journal of Web and Grid Services. – 2005. – Vol. 1(2). – Pp. 240-266.
24. Литвин, В. В. Бази знань інтелектуальних систем підтримки прийняття рішень / В. В. Литвин. – Львів: Видавництво Львівської політехніки, 2011. – 234 с.
25. Інтелектуальні системи, базовані на онтологіях / Д. Г. Досин, В. В. Литвин, Ю. В. Нікольський, В. В. Пасічник. – Львів: Цивілізація, 2009. – 256 с.
26. Литвин, В. В. Спосіб введення метрики для визначення відстані між текстовими документами / В. В. Литвин // Вісник Національного університету «Львівська політехніка». Інформаційні системи та мережі. – 2008. – №621. – С. 162–171.
27. Lytvyn, V. Methods of building intelligent decision support systems based on adaptive ontology / V. Lytvyn, V. Vysotska, D. Dosyn, O. Lozynska, O. Oborska // Proceedings of the IEEE Second International Conference on Data Stream Mining & Processing. – 2018. – Pp. 198-203.
28. Soergel, D. A Metrics Ontology. An intellectual infrastructure for defining, managing, and applying metrics / D. Soergel, O. Helfer // Knowledge Organization for Sustainable World Challenges: Perspectives from Culture, Science, Technology, Sharing, and Connecting Societies. – 2016. – Pp. 333-341.
29. Gruber, T. R. A translation approach to portable ontologies / T. R. Gruber // Knowledge Acquisition. – 1993. – №5(2). – Pp. 199-220.
30. Цейтлін Г.Е. Екологічні аспекти подання знань засобами алгебри алгоритміки / Г.Е. Цейтлін, Л.М. Захарія, О.В. Захарія, Ю.І. Жовнір // Проблеми програмування. - 2010. - №2-3: Спеціальний випуск. - С.369-375.

References

1. Fenton, N. E. Software Metrics: A Rigorous and Practical Approach. 2nd Edition / N. E. Fenton, S. L. Pfleeger. – Berlin: International Thompson Computer Press, 1997. – 195 c.
2. Wilson, R. J. Introduction to Graph Theory / R. J. Wilson. – Longman, 2010. – 201 c.
3. Naylor C. Build your own PC expert system / C. Naylor. – London: Sigma Press, 1983. – 167 c.
4. Tesler, H. S. Metrics and norms in the hierarchy of categorical semantics and functions / H. S. Tesler // Mathematical Machines and Systems. – 2005. – №2. – Pp. 65-68.
5. Velichko, V. Yu. Solving analytical problems in discrete environments by methods of derivation by analogy: Ph.D. thesis / V. Yu. Velichko. – Kyiv: Institute of Cybernetics, 2004..
6. Ngo, D. H. Generic Approach for Combining Linguistic and Context Profile Metrics in Ontology Matching / D. H. Ngo, Z. Bellahsene, R. Coletta // Proceedings of the ODBASE'2011: 10th International Conference on Ontologies, DataBases, and Applications of Semantics, Oct 2011, Crete, Greece. – Crete, 2011. – P. 800-807.
7. Algergawy A. A New Metric To Evaluate Ontology Modularization / A. Algergawy, S. Babalou, B. König-Ries // Proceedings of the 2nd International Workshop on Summarizing and Presenting Entities and Ontologies Co-located with the 13th Extended Semantic Web Conference. – Af Workshop, 15th Int. World Wide Web Conference. – Crete, 2006. – P. 235-240.
8. Stoilos, G. A String Metric for Ontology Alignment / G. Stoilos, G. Stamou, S. Kollias // In Proceedings of the International Semantic Web Conference ISWC 2005: The Semantic Web – ISWC 2005. – 2005. – Pp. 624-637.
9. García J. F. A Survey on Ontology Metrics / J. F. García, F. J. García-Peñalvo, R. Therón // In Proceedings of the World Summit on Knowledge Society WSKS 2010: Knowledge Management, Information Systems, E-Learning, and Sustainability Research. – 2010. – Pp. 22-27.
10. Vrandecic D. How to Design Better Ontology Metrics / D. Vrandecic, Y. Sure // The Semantic Web: Research and Applications. – Springer-Verlag, 2007. – Pp. 311-325.
11. Alani, H., Brewster, C., Shadbolt, N. Ranking Ontologies with AKTiveRank / H. Alani, C. Brewster, N. Shadbolt // In Proceedings of the International Semantic Web Conference, ISWC 2006: 5th International Semantic Web Conference (ISWC). – November 2006, Georgia, USA.
12. Alani, H. Metrics for Ranking Ontologies / H. Alani, C. Brewster // In Proceedings of the 4th International EON Workshop, 15th International World Wide Web Conference. – 2006.
13. Orme, A., Yao, H., Etzkorn, L. Coupling Metrics for Ontology-Based Systems / A. Orme, H. Yao, L. Etzkorn // IEEE Software. – Afiny, 2006. – P. 102-108.
14. Yao, H. Cohesion Metrics for Ontology Design and Application / H. Yao, A. Orme, L. Etzkorn // Journal of Computer Science. – 2005. – Vol. 1(1). – P. 107-113.
15. Ma, Y., Jin, B., Feng, Y. Semantic oriented ontology cohesion metrics for ontology-based systems / Y. Ma, B. Jin, Y. Feng // The Journal of Systems and Software. – Elsevier, 2009.
16. Guarino, N. An Overview of OntoClean / N. Guarino, C. Welty // The Handbook on Ontologies. – Berlin: Springer-Verlag, 2004. – Pp. 151-172; Guarino, N., Welty, C. Evaluating Ontological Decisions with OntoClean / N. Guarino, C. Welty // Communications of the ACM. – ACM Press, 2002. – Pp. 61-65.
17. Yang, Z. Evaluation Metrics for Ontology Complexity and Evolution Analysis / Z. Yang, D. Zhang, C. Ye // Proceedings of the IEEE International Conference on e-Business Engineering (ICEBE'06). – 2006.
18. Raad, J. A Survey on Ontology Evaluation Methods / J. Raad, C. Cruz // Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management. – Lisbon, Portugal, November 2015.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

19. Maedche, A. Measuring similarity between ontologies / A. Maedche, S. Staab // In Knowledge engineering and knowledge management: Ontologies and the semantic web. – Springer Berlin Heidelberg, 2002. – Pp. 251-263.
20. Ponzetto, S. P. Deriving a large scale taxonomy from Wikipedia / S. P. Ponzetto, M. Strube // In AAAI. – Vol. 7, July 2007. – Pp. 1440-1445.
21. Treeratpituk, P. Graph-based Approach to Automatic Taxonomy Generation (GraBTax) / P. Treeratpituk, M. Khabsa, C. L. Giles // arXiv preprint arXiv:1307.1718. – 2013.
22. Zavitsanos, E. Gold standard evaluation of ontology learning methods through ontology transformation and alignment / E. Zavitsanos, G. Paliouras, G. A. Vouros // IEEE Transactions on Knowledge and Data Engineering. – 2011. – Vol. 23(11). – Pp. 1635-1648.
23. Kashyap, V. Taxa Miner: an experimentation framework for automated taxonomy bootstrapping / V. Kashyap, C. Ramakrishnan, C. Thomas, A. Sheth // International Journal of Web and Grid Services. – 2005. – Vol. 1(2). – Pp. 240-266.
24. Lytvyn, V. V. Knowledge bases of intelligent decision-making support systems / V. V. Lytvyn. – Lviv: Lviv Polytechnic Publishing House, 2011. – ISBN 978-617-607-059-7.
25. Dosyn, D. G., Lytvyn, V. V., Nikolskyi, Yu. V., Pasichnik, V. V. Intelligent systems based on ontologies / D. G. Dosyn, V. V. Lytvyn, Yu. V. Nikolskyi, V. V. Pasichnik. – Lviv: Civilization, 2009.
26. Lytvyn, V. V. A method of entering a metric to determine the distance between text documents / V. V. Lytvyn // Visnyk of the National Lviv Polytechnic University. Information Systems and Networks. – 2008. – №621. – Pp. 162–171.
27. Lytvyn, V. Methods of building intelligent decision support systems based on adaptive ontology / V. Lytvyn, V. Vysotska, D. Dosyn, O. Lozynska, O. Oborska // Proceedings of the IEEE Second International Conference on Data Stream Mining & Processing. – 2018. – Pp. 198-203.
28. Soergel, D. A Metrics Ontology. An intellectual infrastructure for defining, managing, and applying metrics / D. Soergel, O. Helfer // Knowledge Organization for Sustainable World Challenges: Perspectives from Culture, Science, Technology, Sharing, and Connecting Societies. – 2016. – Pp. 333-341.
29. Gruber, T. R. A translation approach to portable ontologies / T. R. Gruber // Knowledge Acquisition. – 1993. – №5(2). – Pp. 199-220.
30. Tseitlin, H. E., Zakharija, L. M., Zakharija, O. V., Zhovnir, Yu. I. Ecological aspects of knowledge representation using algorithmic algebra / H. E. Tseitlin, L. M. Zakharija, O. V. Zakharija, Yu. I. Zhovnir // Problems of Programming. – 2010. – №2-3: Special Issue. – Pp. 369-375.

© Григорович А.Г., Григорович В.Г., Жовнір Ю.І., Грибовський О.М.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-03
УДК 004.8:656.61
Коростін Олександр Олексійович, магістр, провідний інженер-програміст
https://orcid.org/0009-0007-7510-6757
Shipnext BV, м. Антверпен, Бельгія

ОПТИМІЗАЦІЯ МАРШРУТІВ МОРСЬКИХ ПЕРЕВЕЗЕНЬ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ: АНАЛІЗ МОЖЛИВОСТЕЙ ТА ВИКЛИКІВ

Коростін О.О. Оптимізація маршрутів морських перевезень за допомогою штучного інтелекту: аналіз можливостей та викликів. У сучасних умовах глобалізації та щораз вищих обсягів міжнародної торгівлі ефективність морських перевезень набуває надзвичайної важливості для світової економіки. З огляду на складність і багатофакторність процесів у морській логістиці використання новітніх технологій стає невід'ємною частиною розвитку галузі. Однією з таких технологій є штучний інтелект (ШІ), який відкриває нові можливості для оптимізації маршрутів морських перевезень, зменшення витрат та підвищення ефективності. У статті проаналізовано основні принципи ШІ та його роль у сучасних логістичних рішеннях. Розглянуто такі інноваційні рішення, як алгоритми машинного навчання для прогнозування погодних умов, системи управління рухом суден на основі ШІ, цифрові двійники для моделювання та оптимізації маршрутів, а також використання інтернету речей та великих даних для аналізу та управління в реальному часі. Зокрема, проаналізовано випадки застосування ШІ для покращення маршрутів, підвищення безпеки судноплавства та управління ланцюгами постачання. Наголошено, що оптимізація маршрутів морських перевезень за допомогою ШІ є перспективним напрямом, який дозволяє підвищити ефективність, знизити витрати та покращити безпеку в логістичній галузі. Використання сучасних технологій ШІ відкриває нові можливості для компаній, що займаються морськими перевезеннями, та сприяє розвитку інноваційних рішень у галузі.
Ключові слова: логістика морських перевезень, штучний інтелект, оптимальний маршрут, моделі глибокого навчання для прогнозування ризиків, нейронні мережі

Korostin O. Optimization of maritime transportation routes using artificial intelligence: analysis of opportunities and challenges. In today's globalized world and ever-increasing volumes of international trade, the efficiency of maritime transportation is of paramount importance to the global economy. Given the complexity and multifactorial nature of maritime logistics processes, the use of the latest technologies is becoming an integral part of the industry's development. One of these technologies is artificial intelligence (AI), which opens up new opportunities for optimizing maritime transportation routes, reducing costs and increasing efficiency. The article analyzes the basic principles of AI and its role in modern logistics solutions. Innovative solutions such as machine learning algorithms for forecasting weather conditions, AI-based ship traffic management systems, digital twins for modeling and optimizing routes, and the use of the Internet of Things and big data for real-time analysis and management are considered. In particular, the author analyzes cases of AI application for route improvement, enhanced safety of navigation, and supply chain management. It is emphasized that optimization of maritime transportation routes using AI is a promising area that can increase efficiency, reduce costs and improve safety in the logistics industry. The use of modern AI technologies opens up new opportunities for maritime transportation companies and promotes the development of innovative solutions in the industry.
Keywords: maritime transportation logistics, artificial intelligence, optimal route, deep learning models for risk prediction, neural networks

Постановка проблеми. Міжнародні морські перевезення є невід'ємною частиною глобальної торгівлі, забезпечуючи транспортування понад 80% світового вантажу. Зважаючи на складність та динамічність цього процесу, ефективне управління маршрутами морських перевезень стає критично важливим для зменшення витрат, зниження викидів парникових газів та підвищення загальної ефективності логістичних ланцюгів. Традиційні методи планування маршрутів часто не можуть урахувати всі можливі змінні та непередбачувані фактори, що призводить до затримок, неефективного використання ресурсів та збільшення операційних витрат.

Штучний інтелект (далі – ШІ) пропонує нові можливості для оптимізації маршрутів морських перевезень, дозволяючи аналізувати великі обсяги даних, прогнозувати ризики та приймати обґрунтовані рішення в режимі реального часу. Впровадження ШІ в цей процес може значно покращити точність планування, адаптивність до змінних умов та знизити загальні витрати, а це робить дослідження в цій галузі актуальними та перспективними, відкриваючи нові можливості для подальшого розвитку інноваційних технологій у сфері логістики та транспорту.

Аналіз останніх досліджень і публікацій. Розвиток штучного інтелекту та його впровадження в логістичні процеси, зокрема у сфері морських перевезень, є предметом активних досліджень. Важливість оптимізації маршрутів за допомогою ШІ зумовлена необхідністю підвищення ефективності, зниження витрат та забезпечення стійкості логістичних ланцюгів.

Одним з основних напрямів є використання алгоритмів машинного навчання для оптимізації маршрутів морських перевезень. У дослідженні М. Тимотеуш, І. Дурлік та П. Козловської [1]

© Коростін О.О.

розглядаються різні алгоритми машинного навчання, як-от генетичні алгоритми, алгоритми мурашиного пошуку та нейронні мережі, для розроблення найбільш оптимальних маршрутів, що дозволяє значно знизити витрати на паливо та скоротити час доставки.

У роботі А. Хог та Х. Андерсен [2] розглянуто вплив впровадження ШІ на управління логістичними ланцюгами. Автори акцентують увагу на використанні ШІ для аналізу великих даних, прогнозування попиту та оптимізації вантажних потоків. Дослідження підтверджує, що впровадження ШІ дозволяє знизити витрати на управління запасами та покращити ефективність логістичних операцій. Проблема прогнозування та управління ризиками в морських перевезеннях досліджується в статті С. Бєльгінова, І. Увалієва, С. Рустамова [3]. Науковці використовують моделі глибокого навчання для прогнозування ризиків, пов'язаних із погодними умовами, затримками в портах та іншими непередбачуваними подіями. Дослідження показує, що застосування ШІ для управління ризиками дозволяє значно знизити вплив негативних факторів на логістичні процеси.

У дослідженні В. Чжан, В. Ян та Ю. Ян [4] розглянуто можливості зниження викидів парникових газів шляхом оптимізації маршрутів з використанням ШІ. Автори показують, що впровадження ШІ дозволяє ефективно планувати маршрути, зменшуючи витрати палива та знижуючи викиди CO2, що підкреслює важливість екологічного аспекту оптимізації маршрутів.

Отже, наукові праці підтверджують актуальність і перспективність використання ШІ для оптимізації маршрутів морських перевезень, надаючи наукове підґрунтя та практичні рекомендації для подальших досліджень і впроваджень.

Виділення не вирішених раніше частин загальної проблеми, котрим присвячено означену статтю. Незважаючи на значні досягнення у впровадженні штучного інтелекту у сферу морських перевезень, залишаються певні аспекти, які потребують подальшого дослідження. У статті зосереджено увагу на кількох ключових питаннях, що не отримали належного висвітлення в попередніх дослідженнях, але мають важливе значення для оптимізації маршрутів морських перевезень, насамперед це такі питання, як [4–7]:

1. Інтеграція динамічних даних у реальному часі. Попередні дослідження переважно фокусувалися на використанні історичних даних для оптимізації маршрутів морських перевезень. Проте морські перевезення характеризуються високою динамічністю та непередбачуваністю, зокрема через зміну погодних умов, затримки в портах, зміну попиту тощо. Інтеграція динамічних даних у реальному часі залишається викликом, який потребує додаткових досліджень для розроблення адаптивних систем, що зможуть швидко реагувати на зміни.

2. Забезпечення безпеки та конфіденційності даних. Застосування ШІ вимагає оброблення великої кількості даних, включаючи конфіденційну інформацію про вантажі, маршрути, клієнтів та постачальників, а це питання поки що не отримало достатньої уваги в контексті морських перевезень. Розроблення ефективних методів захисту даних та запобігання кібератакам є однією з невирішених проблем.

3. Екологічні аспекти оптимізації маршрутів. Хоча багато досліджень зазначають про потенціал ШІ в зниженні витрат палива та викидів парникових газів, відсутні систематичні підходи до інтеграції екологічних критеріїв в алгоритми оптимізації маршрутів. Необхідно розробити моделі, які би враховували не лише економічні, але й екологічні аспекти, забезпечуючи стійкий розвиток морських перевезень.

4. Взаємодія з іншими видами транспорту. Морські перевезення є частиною глобальних логістичних ланцюгів, які включають різні види транспорту: залізничний, автомобільний, авіаційний. Оптимізація маршрутів морських перевезень повинна враховувати взаємодію з іншими видами транспорту для забезпечення ефективності всього логістичного процесу, що вимагає розроблення комплексних моделей з огляду на мультимодальні перевезення.

Перелічені проблеми демонструють складність та багатогранність завдання оптимізації маршрутів морських перевезень за допомогою ШІ. Інтеграція динамічних даних у реальному часі, забезпечення безпеки та конфіденційності даних, урахування екологічних аспектів та взаємодія з іншими видами транспорту – все це є ключовими питаннями, вирішення яких сприятиме створенню більш ефективних, безпечних та стійких систем морських перевезень.

Отже, мета дослідження полягає в обґрунтуванні нових підходів до оптимізації маршрутів міжнародних морських перевезень за допомогою штучного інтелекту.

Дослідження спрямоване на вивчення сучасних методів та алгоритмів ШІ для управління логістичними процесами, розроблення моделей, що враховують економічні та екологічні критерії,

© Коростін О.О.

а також на визначення ефективних стратегій інтеграції ШІ з іншими інформаційними системами для забезпечення комплексного підходу до оптимізації маршрутів морських перевезень.

Виклад основного матеріалу дослідження. Морські перевезення відіграють ключову роль у світовій торгівлі та економіці. Більше 90% світових вантажів транспортується морем, що робить цей вид транспорту найефективнішим та найекономічнішим способом доставлення товарів на великі відстані. Важливість морських перевезень важко переоцінити, оскільки вони забезпечують безперервність постачання таких ресурсів, як нафта, природний газ, зерно та інші сировинні товари. Без перевезень морем багато країн не змогли би підтримувати свої економічні потреби та забезпечувати населення необхідними продуктами.

Морські перевезення поділяються на кілька основних категорій: контейнерні перевезення, перевезення нафтопродуктів, газу, сухих насипних вантажів (зерно, вугілля, руда) та спеціалізовані перевезення (автомобілі, ліс, живі тварини тощо). Кожен із цих видів перевезень має свої особливості й вимоги, які диктують певні стандарти безпеки, зберігання та транспортування.

Порти та супутні їм індустрії створюють мільйони робочих місць по всьому світу. Крім того, морські перевезення сприяють розвитку суміжних галузей, як-от суднобудування, ремонт суден, логістика, страхування та фінансові послуги. Ефективність морських перевезень безпосередньо впливає на вартість товарів та рівень інфляції в різних країнах.

Інновації у сфері морських перевезень постійно розвиваються, будучи спрямовані на підвищення ефективності, безпеки та екологічності цього виду транспортування. Використання великих даних (Big Data), автоматизації та штучного інтелекту відкриває нові можливості для оптимізації маршрутів, прогнозування погодних умов та управління флотом.

Оптимізація маршрутів морських перевезень є ключовим фактором, що впливає на ефективність та витрати в морській логістиці. В умовах глобалізації та зростання обсягів міжнародної торгівлі важливість ефективного управління маршрутами набуває ще більшого значення. Оптимізація маршрутів дозволяє досягти кількох важливих цілей [2–6]:

– вибір оптимального маршруту дозволяє мінімізувати відстань та час у дорозі, що безпосередньо зменшує витрати на паливо. Наприклад, урахування погодних умов, течій та швидкості вітру може суттєво вплинути на споживання палива та підвищити економічність рейсу;

– оптимізація маршрутів сприяє більш ефективному використанню суден. Планування маршруту з урахуванням завантаження та розвантаження в портах, часу перебування в черзі та інших логістичних факторів дозволяє зменшити простої суден, що не тільки підвищує ефективність роботи флоту, але й дозволяє краще управляти ресурсами та забезпечує більш стабільний графік перевезень;

– оптимально сплановані маршрути дозволяють зменшити час доставлення вантажів, що є критично важливим для багатьох галузей, як-от виробництво, де своєчасне постачання сировини або готової продукції може вплинути на безперервність виробничих процесів. Крім того, зменшення часу в дорозі дозволяє швидше обертати судна, підвищуючи їх продуктивність;

– оптимізація маршрутів сприяє підвищенню безпеки морських перевезень. Урахування погодних умов, зон з підвищеним ризиком (наприклад, районів піратства) та інших факторів допомагає уникнути небезпечних ситуацій, знижує ризики аварій, втрат вантажів та пошкоджень суден, що у свою чергу зменшує страхові витрати та підвищує надійність перевезень;

– оптимізація маршрутів має значний позитивний вплив на навколишнє середовище. Зменшення витрат на паливо призводить до зниження викидів парникових газів та інших шкідливих речовин, що сприяє зменшенню екологічного сліду морських перевезень та допомагає відповідати міжнародним екологічним стандартам;

– сучасні технології, як-от штучний інтелект, великі дані (Big Data) та інтернет речей (IoT), відіграють важливу роль в оптимізації маршрутів. Вони дозволяють аналізувати великі обсяги інформації в реальному часі, прогнозувати погодні умови, оцінювати ефективність різних маршрутів та приймати оптимальні рішення, а це підвищує точність планування та дозволяє більш гнучко реагувати на зміни умов у морі;

– зменшення витрат на паливо, підвищення ефективності флоту та зниження часу доставки ведуть до зменшення загальних витрат на морські перевезення, що дозволяє знизити вартість товарів для кінцевих споживачів і підвищити конкурентоспроможність компаній на міжнародному ринку.

© Коростін О.О.

Отже, в сучасних умовах глобалізації та швидкого розвитку технологій оптимізація маршрутів морських перевезень стає критично важливою для забезпечення ефективності логістичних процесів. Це складне завдання, що вимагає врахування великої кількості змінних факторів та обмежень, як-от погодні умови, завантаженість портів, економічні показники тощо. Штучний інтелект надає нові можливості для вдосконалення планування маршрутів, підвищення безпеки та зниження витрат.

У загальному розумінні штучний інтелект – це галузь комп'ютерних наук, яка займається розробленням алгоритмів та систем, здатних виконувати завдання, що зазвичай вимагають людського інтелекту. До таких завдань належать розпізнавання мови, візуальне сприйняття, прийняття рішень, вирішення проблем та навчання на основі досвіду. ШІ включає різні підходи, як-от машинне навчання, нейронні мережі, оброблення природної мови та робототехніка.

Розглянемо основні підходи до використання ШІ для оптимізації маршрутів морських перевезень більш детально, зокрема алгоритми машинного навчання, моделі глибокого навчання та гібридні моделі (табл. 1).

Таблиця 1. Основні підходи до використання ШІ для оптимізації маршрутів морських перевезень

Характеристика	Напрями використання
Алгоритми машинного навчання	
Генетичні алгоритми, засновані на принципах природного відбору та генетичних операцій, можуть використовуватися для оптимізації складних логістичних задач	Ефективно справляються із задачами оптимізації, що включають велику кількість можливих рішень, як-от маршрути з багатьма портами заходу
Алгоритми мурашиного пошуку імітують поведінку мурах, що знаходять найкоротші шляхи до джерела їжі, вони ефективні для задач оптимізації маршрутів	Можуть використовуватися для динамічної оптимізації маршрутів, ураховуючи зміни в завантаженості портів та погодних умовах
Алгоритми рою частинок базуються на моделюванні поведінки рою комах або птахів, які координують свої дії для досягнення спільної мети. Вони використовуються для знаходження оптимальних рішень у задачах із багатьма змінними	Здатні швидко адаптуватися до змін та знаходити ефективні рішення з огляду на різні фактори, що впливають на маршрут
Моделі глибокого навчання	
Нейронні мережі можуть використовуватися для прогнозування попиту на вантажні перевезення та оптимізації маршрутів. Вони здатні обробляти великі обсяги даних та виявляти складні залежності між різними параметрами	Можуть навчатися на історичних даних, що дозволяє їм точно прогнозувати попит та оптимізувати маршрути з урахуванням зміни умов
Рекурентні нейронні мережі ефективні для оброблення часових рядів даних, таких як зміни погодних умов або завантаженість портів. Вони можуть використовуватися для прогнозування майбутніх змін та адаптації маршрутів у реальному часі	Здатні враховувати залежності між минулими та майбутніми станами системи, що робить їх корисними для динамічної оптимізації маршрутів
Конволюційні нейронні мережі, що використовуються переважно для оброблення зображень, можуть застосовуватися для аналізу супутникових знімків та карт морських шляхів	Дозволяють виявляти об'єкти та структури, що можуть впливати на безпеку та ефективність маршрутів. Використання CNN дозволяє покращити точність прогнозів та планування маршрутів
Гібридні моделі	

© Коростін О.О.

Гібридні моделі комбінують різні підходи та алгоритми для досягнення найкращих результатів в оптимізації маршрутів морських перевезень	Наприклад, можна використовувати генетичні алгоритми для початкової оптимізації, а потім застосовувати нейронні мережі для подальшого покращення маршрутів на основі реальних даних. Такий підхід дозволяє ефективно аналізувати різні фактори та адаптуватися до змінних умов.

Оптимізація маршрутів морських перевезень вимагає оперативного аналізу безлічі таких змінних, як погодні умови, трафік у портах, навігаційні обмеження та економічні показники. Інтеграція динамічних даних у реальному часі завдяки можливостям штучного інтелекту та алгоритмам машинного навчання забезпечує більш точне та адаптивне управління перевезеннями. На рисунку 1 наведено основні джерела динамічних даних та їх характеристика.

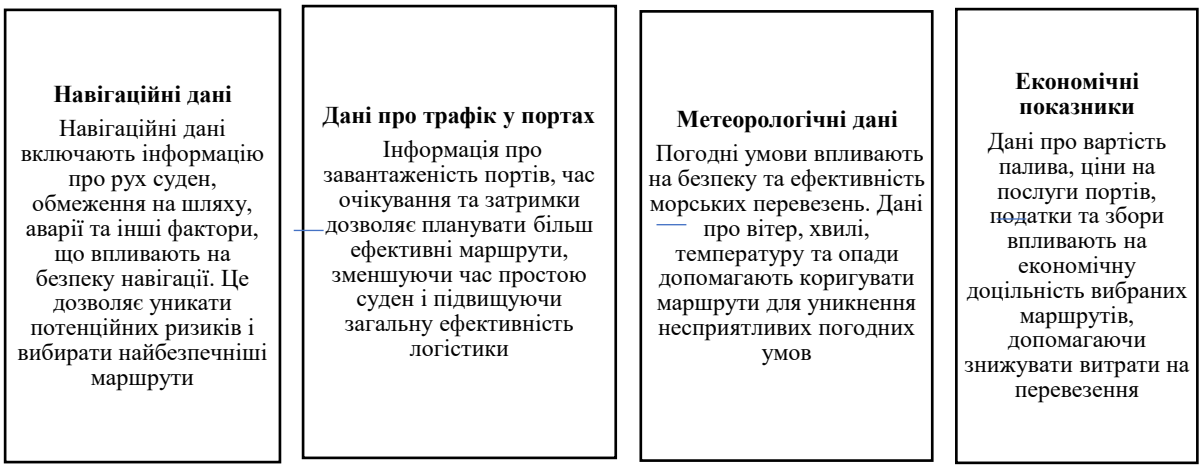


Рис. 1. Основні джерела динамічних даних

Необхідність інтеграції динамічних даних у реальному часі для оптимізації маршрутів морських перевезень зумовлена певною сукупністю факторів.

По-перше, ці дані дозволяють оперативно приймати рішення на основі актуальної інформації, що особливо важливо в умовах морських перевезень, де ситуація може змінюватися швидко та непередбачувано. Інформація про погодні умови, стан моря, завантаженість портів та інші фактори може змінюватися щогодини, і здатність реагувати на ці зміни в реальному часі може значно підвищити ефективність та безпеку перевезень.

По-друге, ШІ може використовувати динамічні дані для прогнозування та попередження можливих загроз. Наприклад, аналізуючи дані про погодні умови та стан моря, система може попередити про наближення шторму або інших небезпечних явищ, що дозволить капітану судна вчасно змінити маршрут або вжити необхідних заходів для забезпечення безпеки екіпажу та вантажу.

По-третє, оптимізація маршрутів морських перевезень з використанням динамічних даних у реальному часі може сприяти зменшенню викидів вуглекислого газу та інших шкідливих речовин, що важливо в умовах щораз більших вимог до екологічної відповідальності бізнесу. ШІ може визначити найбільш економічні маршрути, що дозволить зменшити споживання палива та, відповідно, знизити негативний вплив на довкілля.

Отже, інтеграція динамічних даних у реальному часі для оптимізації маршрутів морських перевезень за допомогою ШІ є важливим кроком у підвищенні ефективності та безпеки цього процесу, проте він супроводжується й низкою викликів, які потрібно враховувати для досягнення оптимальних результатів. Основні з них наведено в таблиці 2.

Таблиця 2. Можливості та виклики впровадження технологій ШІ для оптимізації маршрутів морських перевезень

© Коростін О.О.

Можливості	Виклики
Підвищення точності прогнозів – використання актуальних даних дозволяє робити більш точні прогнози щодо погодних умов, затримок у портах та інших факторів, що впливають на маршрути, що сприяє більш ефективному плануванню	*Захист інформації* – збирання й оброблення великих обсягів даних у реальному часі потребують забезпечення високого рівня безпеки та конфіденційності, що включає використання передових методів захисту даних, як-от шифрування, аутентифікація та контроль доступу
Гнучкість та адаптивність – можливість швидко реагувати на зміни умов та адаптувати маршрути в реальному часі підвищує ефективність перевезень, знижує час простою та забезпечує безпеку	*Складність оброблення великих обсягів даних* у реальному часі вимагає значних обчислювальних ресурсів. Організації повинні інвестувати в потужні сервери, хмарні платформи та інші технології для забезпечення швидкого та ефективного оброблення інформації
Зниження витрат – оптимізація маршрутів з урахуванням динамічних даних дозволяє знижувати витрати на паливо, час очікування в портах та інші операційні витрати, що сприяє економічній ефективності морських перевезень	*Відповідність регуляторним вимогам,* що вимагає розроблення політики та процедур для управління даними, а також навчання персоналу з питань безпеки та конфіденційності
Підвищення безпеки – моніторинг навігаційних умов та вчасне виявлення потенційних ризиків сприяє підвищенню безпеки морських перевезень, зменшуючи ризик аварій та інших інцидентів	*Супротив з боку персоналу* – впровадження технологій ШІ вимагає значних інвестицій, що може бути складним для невеликих компаній. Високі витрати на розроблення, впровадження та підтримку систем ШІ можуть стримувати їх поширення

Використання штучного інтелекту (ШІ) у морській логістиці вже приносить значні результати, покращуючи ефективність, знижуючи витрати та підвищуючи безпеку. Нижче наведено кілька реальних прикладів, які ілюструють, як ШІ змінює зміст морських перевезень.

Одна з найбільших судноплавних компаній світу, Maersk, активно використовує ШІ для оптимізації своїх маршрутів. Використовуючи алгоритми машинного навчання, Maersk аналізує дані про погоду, течії, вітри та інші морські умови для визначення найбільш економічно вигідних маршрутів. Це дозволяє зменшити споживання палива та знизити викиди CO_2. За допомогою ШІ компанія також може прогнозувати можливі затримки та уникати зон підвищеного ризику [2].

Порт Роттердам, один із найбільших портів Європи, впровадив систему прогнозування, яка базується на штучному інтелекті. Ця система використовує великі дані та алгоритми ШІ для прогнозування завантаженості порту, часу очікування суден та інших ключових показників. Завдяки цьому порт може краще планувати свою роботу, зменшувати затримки та підвищувати ефективність оброблення вантажів. Такий підхід також допомагає зменшити витрати на логістику та покращити обслуговування клієнтів [5]. Mediterranean Shipping Company (MSC) використовує штучний інтелект для автоматизації процесів управління контейнерними перевезеннями. За допомогою алгоритмів ШІ компанія аналізує дані про завантаження та розвантаження контейнерів, що дозволяє оптимізувати використання контейнерних суден та зменшити час простоїв. Крім того, ШІ допомагає MSC прогнозувати попит на контейнерні перевезення, що дозволяє краще планувати свої ресурси та підвищувати ефективність роботи [7].

Rolls-Royce спільно з іншими компаніями розробляє автономні судна, які використовують штучний інтелект для навігації та управління. Такі судна обладнано численними сенсорами та системами ШІ, які аналізують навколишнє середовище, прогнозують погодні умови та вибирають оптимальні маршрути. Автономні судна можуть значно знизити витрати на екіпаж, підвищити безпеку та зменшити ризики людського фактору [9].

Французька судноплавна компанія CMA CGM використовує штучний інтелект для моніторингу та прогнозування стану контейнерів. Система ШІ аналізує дані із сенсорів,

© Коростін О.О.

установлених на контейнерах, та прогнозує можливі проблеми, як-от пошкодження або перевищення температурних режимів. Це дозволяє компанії своєчасно реагувати на проблеми, забезпечуючи збереження вантажу та підвищуючи якість обслуговування клієнтів.

IBM Watson Supply Chain – це платформа, яка використовує штучний інтелект для оптимізації логістичних процесів, зокрема морських перевезень. Watson аналізує великі обсяги даних, включаючи інформацію про погоду, стан портів, транспортні коридори та інші фактори, що впливають на морські перевезення. Це дозволяє компаніям прогнозувати затримки, оптимізувати маршрути та приймати більш обґрунтовані рішення щодо управління логістикою [3].

Hapag-Lloyd, одна з провідних судноплавних компаній світу, впровадила систему ШІ для прогнозування затримок у морських перевезеннях. Система аналізує дані про фактори, що можуть вплинути на доставку вантажів. Завдяки цьому компанія може реагувати на потенційні проблеми та інформувати клієнтів про можливі затримки, що підвищує рівень обслуговування [10].

Ці приклади демонструють, як штучний інтелект може радикально змінити морську логістику, зробивши її більш ефективною, економічною та безпечною. Впровадження ШІ в морські перевезення відкриває нові можливості для оптимізації процесів, зниження витрат та підвищення якості обслуговування клієнтів, що є ключовим фактором для конкурентоспроможності на глобальному ринку.

Технології штучного інтелекту розвиваються стрімкими темпами, відкриваючи нові можливості для різних галузей, включаючи морські перевезення. Цифрові двійники (digital twins) – це віртуальні моделі фізичних об'єктів, які дозволяють моделювати їх поведінку в реальному часі. Використання цифрових двійників суден та маршрутів дозволяє проводити симуляції та оптимізувати маршрути, враховуючи різні сценарії, що допомагає зменшити ризики та підвищити ефективність морських перевезень. Розроблення автономних суден та їх інтеграція в глобальну логістичну систему є важливим напрямом досліджень. Це включає не тільки технічні аспекти автономії, але й створення нових стандартів, нормативних актів та методів управління. Інтеграція автономних суден може значно знизити витрати на екіпаж, підвищити безпеку і поліпшити морські перевезення. Колаборативні мережі та розподілені системи на основі ШІ дозволяють суднам обмінюватися інформацією та координувати свої дії для оптимізації маршрутів та уникнення конфліктних ситуацій. Це може включати використання блокчейн-технологій для забезпечення безпеки та прозорості даних.

Висновки та перспективи подальшого дослідження. Оптимізація маршрутів морських перевезень за допомогою штучного інтелекту (ШІ) відкриває нові горизонти для підвищення ефективності та зниження витрат у логістичній галузі. Використання інноваційних технологій ШІ, як-от машинне навчання, нейронні мережі, оброблення природної мови та комп'ютерний зір, дозволяє вирішувати складні завдання управління маршрутами, прогнозування попиту та забезпечення безпеки.

Використання алгоритмів машинного навчання для аналізу великих обсягів даних дозволяє вибирати найкращі маршрути з урахуванням погодних умов, морських течій та трафіку суден, що дозволяє зменшити час у дорозі, знизити витрати на паливо і підвищити ефективність перевезень.

Розвиток технологій автономних суден обіцяє революцію в морській логістиці, адже такі судна можуть самостійно навігувати, уникати перешкод та вибирати оптимальні маршрути, що значно знижує витрати на екіпаж та підвищує безпеку.

Алгоритми ШІ можуть аналізувати історичні дані та поточні тенденції для точного прогнозування попиту на морські перевезення. Це дозволяє компаніям краще планувати свої ресурси, уникати перевантажень та забезпечувати своєчасну доставку вантажів.

Також системи ШІ можуть виявляти потенційні загрози та попереджати аварії, аналізуючи дані із сенсорів та інших джерел, що підвищує безпеку морських перевезень та зменшує ризики.

Попри значні можливості, впровадження ШІ в морські перевезення стикається з низкою викликів. Передусім це складність інтеграції нових технологій у наявну інфраструктуру та необхідність значних інвестицій. Крім того, виникають питання безпеки даних та захисту від кіберзагроз. Регуляторні бар'єри та стандартизація також є важливими аспектами, які потребують вирішення для успішного впровадження автономних суден та інших інноваційних рішень.

З розвитком технологій ШІ та їх вдосконаленням очікується подальше зростання їх ролі в морській логістиці. Такі інноваційні рішення, як цифрові двійники, інтернет речей (IoT) та адаптивні системи управління, можуть суттєво покращити ефективність та безпеку морських

© Коростін О.О.

перевезень. Перспективні дослідження в галузі глибокого навчання, автономних суден та колаборативних мереж відкривають нові можливості для оптимізації логістичних процесів та підвищення конкурентоспроможності транспортних компаній.

Отже, оптимізація маршрутів морських перевезень за допомогою ШІ є перспективним напрямом, який дозволяє підвищити ефективність, знизити витрати та покращити безпеку в логістичній галузі. Використання сучасних технологій ШІ відкриває нові можливості для компаній, що займаються морськими перевезеннями, та сприяє розвитку інноваційних рішень у галузі. Попри виклики та ризики, перспективи розвитку технологій ШІ в морській логістиці виглядають надзвичайно обнадійливими, що створює підґрунтя для подальших досліджень та впровадження інновацій.

Список бібліографічного опису:

1. Tymoteusz M., Durlik I., Kozlovska P., Krzemińska A., Jawor M., Cembrowska-Lech D., Kisiel A., Mosiundz S., Kołodziejczak M. Optimizing ship logistics through advanced AI algorithms: revolutionizing the future of maritime research. *Collection of Scientific Papers «SCIENTIA»*, March 24, 2023; Zagreb, Croatia. P. 126–131. URL: https://previous.scientia.report/index.php/archive/article/view/826. (date of access: 16.07.2024).
2. Hogh A., Andersen H. B. Seafarer fatigue: A review of risk factors, consequences for seafarers' health and safety, and options for mitigation. *International Maritime Health*. 2018. Vol. 69, No. 2. P. 104–114.
3. Belginova S., Uvaliyeva I., Rustamov S. The application of data mining methods for the process of diagnosing diseases. *Journal of Theoretical and Applied Information Technology*. 2019. Vol. 97, No. 7. P. 1980-1998.
4. Zhang W., Yan W., Yang Y. A review of the recent research on the maritime integrated transportation system. *Transport Reviews*. 2020. Vol. 40, No. 4. P. 524–548.
5. Rahikainen M., Luoma E. Cybersecurity in the maritime industry: A systematic literature review and future research agenda. *Journal of Marine Science and Engineering*. 2020. Vol. 8, No. 7. P. 511.
6. Nguyen H. P., Nguyen C. Th. U., Tran Th. M., Dang Q. H., Pham N. D. Artificial Intelligence and Machine Learning for Green Shipping: Navigating towards Sustainable Maritime Practices. *International Journal on Informatics Visualization*. 2024. Vol. 8, No. 1. P. 1-17.
7. Imran M., Ayob A., Jamaludin S. Applications of artificial intelligence in ship berthing: A review. *Indian Journal Geo Mar. Sci.* 2021. Vol. 50. P. 855–863.
8. Bendul J. C., Kreutzfeldt J., Michalski F. Digital transformation in maritime logistics: The case of smart container terminals. *Computers & Industrial Engineering*. 2020. Vol. 139. P. 105-117.
9. Ellegaard O., Wallin J.A. The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*. 2015. Vol. 105. P. 1809–1831.
10. Kilibarda M., Andrejić M., Popović V. Research in logistics service quality: A systematic literature review. *Transport*. 2020. Vol. 35. P. 224–235.

References:

1. Tymoteusz, M., Durlik, I., Kozlovska, P., Krzemińska, A., Jawor, M., Cembrowska-Lech, D., Kisiel, A., Mosiundz, S., & Kołodziejczak, M. (2023). Optimizing ship logistics through advanced AI algorithms: Revolutionizing the future of maritime research. *Collection of Scientific Papers «SCIENTIA»*, 126–131. https://previous.scientia.report/index.php/archive/article/view/826
2. Hogh, A., & Andersen, H. B. (2018). Seafarer fatigue: A review of risk factors, consequences for seafarers' health and safety, and options for mitigation. *International Maritime Health*, 69(2), 104–114.
3. Belginova, S., Uvaliyeva, I., & Rustamov, S. (2019). The application of data mining methods for the process of diagnosing diseases. *Journal of Theoretical and Applied Information Technology*, 97(7), 1980-1998.
4. Zhang, W., Yan, W., & Yang, Y. (2020). A review of the recent research on the maritime integrated transportation system. *Transport Reviews*, 40(4), 524–548.
5. Rahikainen, M., & Luoma, E. (2020). Cybersecurity in the maritime industry: A systematic literature review and future research agenda. *Journal of Marine Science and Engineering*, 8(7), 511.
6. Nguyen, H. P., Nguyen, C. Th. U., Tran, Th. M., Dang, Q. H., & Pham, N. D. (2024). Artificial Intelligence and Machine Learning for Green Shipping: Navigating towards Sustainable Maritime Practices. *International Journal on Informatics Visualization*, 8(1), 1-17.
7. Imran, M., Ayob, A., & Jamaludin, S. (2021). Applications of artificial intelligence in ship berthing: A review. *Indian Journal Geo Mar. Sci.*, 50, 855–863.
8. Bendul, J. C., Kreutzfeldt, J., & Michalski, F. (2020). Digital transformation in maritime logistics: The case of smart container terminals. *Computers & Industrial Engineering*, 139, 105-117.
9. Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, 105, 1809–1831.
10. Kilibarda, M., Andrejić, M., & Popović, V. (2020). Research in logistics service quality: A systematic literature review. *Transport*, 35, 224–235.

© Коростін О.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-04
УДК 004.2
Пасічник Максим Юрійович, аспірант
https://orcid.org/0009-0006-7885-3507
Зайцев Володимир Григорович, д.т.н., професор
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

МЕТОДИ ДИСПЕТЧЕРИЗАЦІЇ ЗАВДАНЬ У СИСТЕМАХ РЕАЛЬНОГО ЧАСУ

Пасічник М.Ю., Зайцев В.Г. Методи диспетчеризації завдань в системах реального часу. Системи реального часу є невід'ємною частиною сучасних технологій, охоплюючи сфери автоматизації виробництва, медичні пристрої, транспортні системи та інші критично важливі області. Ці системи вимагають забезпечення своєчасного та надійного виконання завдань, що робить питання ефективної диспетчеризації особливо актуальним. У даній статті представлено огляд і аналіз сучасних методів диспетчеризації завдань у системах реального часу, зокрема статичних, динамічних і комбінованих підходів. У статті підкреслено важливість забезпечення балансу між передбачуваністю та гнучкістю методів диспетчеризації для підвищення ефективності та надійності систем. Особливу увагу приділено ключовим викликам, таким як збалансування навантаження між процесорами, забезпечення надійності та відмовостійкості, ефективне використання ресурсів, динамічна адаптація до змінних умов і врахування енергоспоживання. Розробка нових методів і вдосконалення існуючих підходів має враховувати ці виклики для досягнення оптимальної продуктивності. Особливу увагу приділено перспективам застосування машинного навчання для прогнозування навантаження та динамічної оптимізації алгоритмів диспетчеризації. Використання таких технологій дозволяє значно підвищити ефективність і надійність систем реального часу, забезпечуючи адаптивне управління завданнями в реальному часі. Математичне моделювання, проведене у статті, дозволяє оцінити ефективність різних методів диспетчеризації за ключовими критеріями, такими як час відгуку, пропускна здатність, здатність до масштабування, використання ресурсів, надійність і гнучкість. Результати показують, що жоден метод не є універсально найкращим, і вибір оптимального підходу залежить від специфічних вимог конкретної системи. У статті окреслено напрями для подальших досліджень, включаючи інтеграцію методів машинного навчання для підвищення адаптивності та продуктивності систем реального часу. Це є важливим кроком у подоланні викликів, пов'язаних з оптимізацією використання ресурсів і дотриманням дедлайнів у змінних умовах, що сприятиме подальшому прогресу в цій галузі.
Ключові слова: системи реального часу, диспетчеризація завдань, балансування навантаження, динамічна адаптація, пропускна здатність, час відгуку, здатність до масштабування, машинне навчання

Pasichnyk M., Zaitsev V. Task scheduling methods in real-time systems. Real-time systems are an integral part of modern technologies, covering the areas of production automation, medical devices, transport systems and other critical areas. These systems require ensuring timely and reliable execution of tasks, which makes the issue of effective dispatching particularly relevant. This article presents an overview and analysis of modern methods of dispatching tasks in real-time systems, including static, dynamic and combined approaches. The paper highlights the importance of balancing predictability and flexibility in dispatching methods to improve system efficiency and reliability. Special attention is paid to key challenges such as load balancing between processors, ensuring reliability and fault tolerance, efficient use of resources, dynamic adaptation to changing conditions and consideration of power consumption. The development of new methods and the improvement of existing approaches must take these challenges into account to achieve optimal performance. Special attention is paid to the prospects of using machine learning for load forecasting and dynamic optimization of dispatching algorithms. The use of such technologies allows to significantly increase the efficiency and reliability of real-time systems, providing adaptive management of tasks in real time. The mathematical modelling carried out in the article allows to evaluate the effectiveness of different dispatching methods according to key criteria such as response time, throughput, scalability, resource utilization, reliability and flexibility. The results show that no method is universally best, and the choice of the optimal approach depends on the specific requirements of a particular system. The paper outlines directions for further research, including the integration of machine learning techniques to improve the adaptability and performance of real-time systems. This is an important step in overcoming the challenges of optimizing the use of resources and meeting deadlines in a changing environment, which will contribute to further progress in this field.
Keywords: real-time systems, task scheduling, load balancing, dynamic adaptation, throughput, response time, scalability, machine learning

Постановка проблеми. Системи реального часу відіграють ключову роль у сучасних технологіях, знаходячи застосування у багатьох критично важливих галузях, таких як автоматизація виробництва, медичні пристрої, транспортні системи та інші. Ці системи потребують забезпечення своєчасного та надійного виконання завдань, оскільки навіть незначні затримки можуть призвести до серйозних наслідків. Наприклад, у медичних системах пропущений дедлайн може вплинути на життя пацієнта, а в системах управління транспортом – спричинити аварії.

Однією з основних проблем у системах реального часу є диспетчеризація завдань, яка визначає порядок їх виконання, розподіл ресурсів та забезпечення дотримання часових обмежень. Ефективні методи диспетчеризації повинні враховувати різні фактори, такі як пріоритети завдань, змінні навантаження та обмежені ресурси. Існуючі методи диспетчеризації, хоча й досягли значного

@ Пасічник М.Ю., Зайцев В.Г.

прогресу, все ще стикаються з численними викликами, зокрема зі збалансуванням навантаження, забезпеченням надійності та адаптивності систем до змінних умов.

Розробка нових і вдосконалення існуючих методів диспетчеризації є важливим науковим і практичним завданням. Це дозволить підвищити продуктивність і надійність систем реального часу, забезпечивши їхню здатність до роботи в різних умовах. Успішне вирішення цієї проблеми сприятиме покращенню якості та безпеки таких систем, що є критично важливим для багатьох сфер людської діяльності.

Аналіз останніх досліджень і публікацій. У роботі «Scheduling in Real-Time System» [1] автори досліджують різні алгоритми планування, що застосовуються в системах реального часу. Вони підкреслюють важливість своєчасного виконання завдань, критичного для систем, що вимагають високої надійності. У статті обговорюються планування з фіксованим пріоритетом (RMS) і планування з найпершим дедлайном (EDF), ілюструючи їх застосування та ефективність у різних сценаріях. Автори наголошують на важливості оптимальних стратегій планування для забезпечення стабільності та продуктивності системи в умовах жорстких часових обмежень.

У статті «Task Scheduling in Real-Time Systems with Energy Harvesting and Energy Minimization» [2] представлений інноваційний підхід, що інтегрує енергозбирання з плануванням завдань. Автори пропонують метод динамічного масштабування напруги (DVS) у поєднанні з плануванням EDF для підвищення енергоефективності. Їх алгоритм пріоритетизує завдання на основі доступної енергії та дедлайнів, значно знижуючи споживання енергії при збереженні продуктивності системи. Експериментальні результати демонструють ефективність алгоритму в балансуванні використання енергії та планування завдань, що робить його придатним для систем реального часу з обмеженою енергією.

У статті «A New FPGA-Based Task Scheduler for Real-Time Systems» [3] представлений планувальник завдань на базі FPGA, розроблений для покращення продуктивності систем реального часу. Дослідження детально описує архітектуру та впровадження планувальника FPGA, підкреслюючи його швидкість, гнучкість та надійність. Автори порівнюють продуктивність планувальника FPGA з традиційними програмними планувальниками, показуючи значні покращення в управлінні завданнями та часі виконання. У висновку зазначається, що планування на базі FPGA може суттєво покращити ефективність систем реального часу.

У статті «Aperiodic Task Scheduling for Hard-Real-Time Systems» [4] автори вирішують проблему інтеграції неперіодичних завдань у системи з жорсткими реальними часовими обмеженнями. Вони пропонують метод планування, що забезпечує своєчасне виконання як періодичних, так і неперіодичних завдань без шкоди для продуктивності системи. Підхід включає гібридний алгоритм планування, який динамічно коригує пріоритети на основі дедлайнів завдань та навантаження системи. Результати моделювання підтверджують здатність алгоритму задовольняти суворі часові вимоги, доводячи його придатність у різних реальних часових середовищах.

Постановка завдання. Реалізація ефективної диспетчеризації завдань в системах реального часу стикається з низкою викликів, які впливають на загальну продуктивність та надійність систем. Основні виклики включають:

- Збалансування навантаження: Забезпечення рівномірного розподілу навантаження між процесорами є критично важливим для уникнення перевантаження окремих компонентів системи. Нерівномірний розподіл може призвести до затримок у виконанні завдань та зниження загальної ефективності.
- Дотримання дедлайнів: У системах реального часу завдання повинні бути виконані у визначені часові рамки. Пропущення дедлайнів може призвести до серйозних наслідків, особливо у критичних додатках, таких як медичні системи або системи управління транспортом.
- Ефективне використання ресурсів: Максимальне використання доступних ресурсів, таких як процесорний час, пам'ять та енергетичні ресурси, є важливим для забезпечення ефективної роботи системи. Недостатнє використання ресурсів може призвести до марнування потужностей, а надмірне використання – до перегріву та зниження продуктивності.
- Динамічна адаптація: Системи реального часу повинні бути здатні адаптуватися до змінних умов, таких як зміна навантаження, виникнення нових завдань або зміна пріоритетів. Це вимагає використання адаптивних алгоритмів диспетчеризації, які можуть оперативно реагувати на такі зміни.

@ Пасічник М.Ю., Зайцев В.Г.

- Надійність та відмовостійкість: Забезпечення надійності та здатності системи до відновлення після збоїв є важливим аспектом диспетчеризації завдань. Системи повинні мати механізми для виявлення та корекції помилок, щоб запобігти втратам даних та забезпечити безперервну роботу.
- Врахування енергоспоживання: Зниження енергоспоживання стає все більш важливим, особливо для мобільних та вбудованих систем. Ефективні алгоритми диспетчеризації повинні враховувати енергетичні обмеження та оптимізувати використання енергії без шкоди для продуктивності.
- Складність реалізації: Впровадження складних алгоритмів диспетчеризації може вимагати значних витрат ресурсів на їх розробку та тестування. Необхідно знайти баланс між складністю алгоритму та його ефективністю, щоб забезпечити практичну реалізацію в реальних системах.

Подолання цих викликів є важливим для забезпечення ефективної та надійної роботи систем реального часу. Розробка нових методів диспетчеризації та вдосконалення існуючих підходів потребує врахування цих аспектів для досягнення оптимальної продуктивності.

Виклад основного матеріалу. Ефективність методів диспетчеризації завдань у системах реального часу можна оцінити за допомогою кількох ключових критеріїв:

- Час відгуку: важливий показник, що визначає час між надходженням запиту і початком його обробки. Методи, які забезпечують мінімальний час відгуку, є більш ефективними для систем реального часу.
- Пропускна здатність: Визначає кількість завдань, які система може обробити за одиницю часу. Висока пропускна здатність вказує на ефективне використання ресурсів.
- Здатність до масштабування: Важливо, щоб метод диспетчеризації міг ефективно працювати як при низькому, так і при високому навантаженні. Це показує гнучкість і адаптивність методу до змінних умов системи.
- Використання ресурсів: Оптимізація використання процесорного часу, пам'яті та енергетичних ресурсів є критичною для підтримки стабільної роботи системи. Методи, які максимально використовують доступні ресурси без перевантаження, є більш ефективними.
- Надійність: Здатність методу забезпечити стабільну роботу системи без збоїв або втрат даних. Надійні методи мають механізми для виявлення та корекції помилок.
- Гнучкість: Важливо, щоб метод диспетчеризації міг адаптуватися до змінних умов і нових вимог. Це включає можливість динамічної зміни пріоритетів завдань та адаптацію до нових типів завдань.

Для оцінки ефективності методів диспетчеризації завдань у системах реального часу розглянемо модель. Нехай є система з n завдань, де кожне завдання i характеризується:

- T_i — час виконання завдання i
- D_i — дедлайн завдання i
- P_i — період завдання i
- U_i — використання процесора завданням i
- R_i— час відгуку завдання i

Згідно теореми про верхню межу використання процесора:

$$U(n) = \sum_{i=1}^{n} \frac{T_i}{P_i} \leq n\left(2^{\frac{1}{n}} - 1\right)$$

Верхня межа $U(n)$ наближається до 69% (ln 2), коли число завдань наближається до нескінченності. Це оцінка для найгіршого випадку, але, як показано в роботі [5], для випадково обраної групи завдань ймовірна верхня межа дорівнює 88%.

Для кожного завдання і час відгуку можна оцінити з урахуванням інших завдань у системі. Припустимо, що завдання виконуються з пріоритетами. Нехай $HP(i)$ — множина завдань з вищим пріоритетом, ніж у завдання i.

Час відгуку R_i для завдання і можна знайти розв'язавши нерівність:

@ Пасічник М.Ю., Зайцев В.Г.

$$R_i \leq T_i + \sum_{j \in HP(i)}^{n} \left\lceil \frac{R_i}{P_j} \right\rceil T_j$$

Ця нерівність відображає час, який завдання і витратить на своє виконання та час очікування через завдання з вищими пріоритетами.

Затримка визначається як максимальний час від моменту надходження завдання до його завершення. Для системи з пріоритетами затримка для завдання і визначається як:

$$L_i = R_i + T_i$$

Розглянемо адаптивний метод, де пріоритети завдань можуть змінюватися залежно від стану системи. Нехай $\pi_i(t)$ — пріоритет завдання i в момент часу t. Тоді час відгуку для адаптивного методу можна визначити як:

$$R_i(t) = T_{start_i}(t) + T_i + \sum_{j \in HP_i(t)}^{n} \left\lceil \frac{R_i}{P_j} \right\rceil T_j$$

де $HP_i(t)$ — множина завдань з вищим пріоритетом у момент часу t.

Для оцінки ефективності введемо функцію ефективності E, яка враховує час відгуку і використання процесора:

$$E = \alpha \left(1 - \frac{1}{n} \sum_{i=1}^{n} \frac{R_i}{D_i}\right) + \beta(1 - U)$$

де α та β — вагові коефіцієнти, що визначають важливість відповідних критеріїв.

Ця модель дозволяє врахувати різні аспекти ефективності методів диспетчеризації та оптимізувати їх для забезпечення максимальної продуктивності та надійності систем реального часу.

Припустимо, що в системі є три завдання із наступними характеристиками (таблиця 1), які незалежні від стану системи і з'являються зі своїм періодом.

Завдання	T_i	D_i	P_i	R_i
1	2	5	10	3
2	3	3	6	2
3	3	8	12	5

Таблиця 1. Характеристики завдань

Загальне використання процесора U можна обчислити як:

$$U = \sum_{i=1}^{n} \frac{T_i}{P_i} = \frac{2}{10} + \frac{3}{6} + \frac{3}{12} = 0.95$$

Припустимо, що $\alpha = 0.5$ та $\beta = 0.5$. Тоді функція ефективності буде:

$$E = 0.5 \left(1 - \frac{1}{3}\left(\frac{3}{5} + \frac{2}{3} + \frac{5}{8}\right)\right) + 0.5\,(1 - 0.95) \cong 0.2097$$

@ Пасічник М.Ю., Зайцев В.Г.

Отримане значення ефективності E вказує, що система працює на 20.97% від можливої максимальної ефективності за заданими критеріями. Таким чином, значення E надає загальну оцінку ефективності алгоритму диспетчеризації, що допомагає виявити можливості для оптимізації і покращення роботи системи реального часу. Подальші напрямки досліджень у сфері диспетчеризації завдань у системах реального часу можуть зосередитися на інтеграції методів машинного навчання для динамічного прогнозування навантаження та оптимізації алгоритмів у режимі реального часу. Використання моделей глибокого навчання для аналізу даних та прогнозування майбутніх станів системи дозволить більш точно і своєчасно адаптувати розподіл завдань, підвищуючи загальну ефективність і надійність. Крім того, розробка енергоефективних алгоритмів, які враховують змінні енергетичні потреби і можливості, стане ключовим аспектом для мобільних та вбудованих систем. Іншим перспективним напрямом є впровадження адаптивних методів, здатних автоматично налаштовувати параметри диспетчеризації у відповідь на зміни в робочому середовищі, що дозволить забезпечити оптимальну продуктивність навіть за умов непередбачуваних змін. Дослідження в цих напрямках сприятимуть створенню більш гнучких, стійких та енергоефективних систем реального часу, здатних ефективно виконувати критично важливі завдання у сучасних технологічних умовах.

Висновок. Методи диспетчеризації у системах реального часу є критично важливими для забезпечення своєчасного та ефективного виконання завдань. Статичні методи забезпечують передбачуваність і простоту впровадження, тоді як динамічні методи пропонують вищу гнучкість та адаптивність, але потребують більше ресурсів. Комбіновані підходи поєднують переваги обох методів, забезпечуючи баланс між передбачуваністю та гнучкістю. Основні виклики включають збалансування навантаження, дотримання дедлайнів, ефективне використання ресурсів, динамічну адаптацію, надійність, відмовостійкість та врахування енергоспоживання. Оцінка ефективності методів диспетчеризації може базуватися на таких критеріях, як час відгуку, пропускна здатність, здатність до масштабування, використання ресурсів, надійність та гнучкість. Подальші дослідження у цій галузі дозволять створити більш адаптивні, енергоефективні та стійкі до збоїв системи, що є критично важливим для сучасних технологій у різних сферах застосування.

Список бібліографічного опису:

1. Sirshar, M.; Rizwan, H.; Shabbir, M.; Nawaz, H. Scheduling in Real Time System. Preprints 2019, 2019120072. URL: https://doi.org/10.20944/preprints201912.0072.v1
2. Kumar, A. & Alam, B. (2018). Task Scheduling in Real Time Systems with Energy Harvesting and Energy Minimization. Journal of Computer Science, 14(8), 1126-1133. URL: https://doi.org/10.3844/jcssp.2018.1126.1133
3. Kohútka L. A New FPGA-Based Task Scheduler for Real-Time Systems. Electronics. 2023, 12(8), 1870. URL: https://doi.org/10.3390/electronics12081870
4. Sprunt, B., Sha, L. & Lehoczky, J. Aperiodic task scheduling for Hard-Real-Time systems. Real-Time Syst 1, 27–60 (1989). https://doi.org/10.1007/BF02341920
5. J. Lehoczky, L. Sha and Y. Ding, «The rate monotonic scheduling algorithm: exact characterization and average case behavior,» [1989] Proceedings. Real-Time Systems Symposium, Santa Monica, CA, USA, 1989, pp. 166-171. URL: https://doi.org/10.1109/REAL.1989.63567

References:

1. Sirshar, M.; Rizwan, H.; Shabbir, M.; Nawaz, H. Scheduling in Real Time System. Preprints 2019, 2019120072. URL: https://doi.org/10.20944/preprints201912.0072.v1
2. Kumar, A. & Alam, B. (2018). Task Scheduling in Real Time Systems with Energy Harvesting and Energy Minimization. Journal of Computer Science, 14(8), 1126-1133. URL: https://doi.org/10.3844/jcssp.2018.1126.1133
3. Kohútka L. A New FPGA-Based Task Scheduler for Real-Time Systems. Electronics. 2023, 12(8), 1870. URL: https://doi.org/10.3390/electronics12081870
4. Sprunt, B., Sha, L. & Lehoczky, J. Aperiodic task scheduling for Hard-Real-Time systems. Real-Time Syst 1, 27–60 (1989). https://doi.org/10.1007/BF02341920
5. J. Lehoczky, L. Sha and Y. Ding, «The rate monotonic scheduling algorithm: exact characterization and average case behavior,» [1989] Proceedings. Real-Time Systems Symposium, Santa Monica, CA, USA, 1989, pp. 166-171. URL: https://doi.org/10.1109/REAL.1989.63567

@ Пасічник М.Ю., Зайцев В.Г.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-05
UDC 621.391
Sydorchuk Vasyl, Postgraduate
https://orcid.org/0009-0000-6587-3046
Zhytomyr Polytechnic State University, Zhytomyr, Ukraine

DEEP LEARNING ALGORITHMS FOR PREDICTING ROUTES OF SNOWPLOWS IN EU CITIES

Sydorchuk V. Deep learning algorithms for predicting routes of snowplows in European cities. Most cities are faced with annual snowfall but are finding it difficult to deal with their snow plowing activities. Even though winter road maintenance has been studied for many decades, most of the papers do not present models that can be scaled up to allow for incorporation of side constraints that are often met in practical applications. Subject of the paper: Comparison of deep learning algorithms for forecasting snowplow routes in European cities in terms of spatial analysis, routing and traffic flow. The study focuses on extending neural networks including Recurrent Neural Network (RNN), Convolutional Neural Network (CNN) and Graph Neural Network (GNN) to overcome the problematics of snowplow in urban setting. Temporal and spatial data are incorporated using sophisticated models such as Spatio-Temporal Graph Convolutional Networks (STGCNs) using examples that capture the capacity to control the changes in the snowplow routes in reaction to the real-time traffic flow and weather information. The report also looks at how transformer models among other emerging technologies could be harnessed to improve predictive accuracy as well as efficiency. Comparisons made to these other methods evaluating deep learning methods' benefits and drawbacks and their application to urban infrastructure and traffic flow. The results, therefore, point out that integrating these complicated formulas can go a long way in enhancing the efficiency and safety of snowplowing in the European cities to create better lit and safer transport networks.
Keywords. spatial modeling, route optimization, neural networks, urban infrastructure, traffic management

Сидорчук В.О. Алгоритми глибокого навчання для прогнозування маршрутів снігоприбиральної техніки в містах Євросоюзу. Алгоритми глибокого навчання для прогнозування маршрутів снігоприбиральних машин у європейських містах. Більшість міст стикаються зі щорічними снігопадами, а також із труднощами в управлінні снігоприбиральною технікою. Незважаючи на те, що зимове утримання доріг вивчається вже багато десятиліть, більшість робіт не представляють моделей, які можна масштабувати для врахування побічних обмежень, що часто зустрічаються в практичних застосуваннях. Предмет дослідження: Порівняння алгоритмів глибокого навчання для прогнозування маршрутів снігоприбиральної техніки в європейських містах з точки зору просторового аналізу, маршрутизації та транспортних потоків. Дослідження фокусується на розширенні нейронних мереж, включаючи рекурентні нейронні мережі (RNN), згорткові нейронні мережі (CNN) та графові нейронні мережі (GNN), для подолання проблематики снігоприбиральної техніки в міських умовах. Часові та просторові дані інтегруються за допомогою складних моделей, таких як просторово-часові графові згорткові мережі (STGCN), на прикладах, які демонструють здатність контролювати зміни в маршрутах снігоприбиральної техніки у відповідь на транспортні потоки в реальному часі та інформацію про погоду. У звіті також розглядається, як трансформаторні моделі серед інших нових технологій можуть бути використані для підвищення точності та ефективності прогнозування. Проведено порівняння з іншими методами, оцінено переваги та недоліки методів глибокого навчання, а також їх застосування до міської інфраструктури та транспортних потоків. Таким чином, результати показують, що інтеграція цих складних формул може значно підвищити ефективність і безпеку снігоприбиральної техніки в європейських містах для створення краще освітлених і безпечніших транспортних мереж.
Ключові слова: просторове моделювання, оптимізація маршрутів, нейронні мережі, міська інфраструктура, управління дорожнім рухом

The problem statement. Among the largest costs that regional or local authorities bear during winter to maintain the roads is removing snow and ice from the roads. That is why much emphasis is placed on the search for procedures to improve the distribution of resources since the possible cost-reduction benefits may be quite modest, while the overall benefit of their application can be very high. The priorities of concern in snow removal involve the people (the drivers, the mechanics, the supervisors,) the consumables (the road salt, the road sand, other chemicals used in combating ice and snow,) the equipment (the snowplows, the dump trucks, the supervisor trucks, and so on,) the financial aspect in terms of the wages for the people, the consumables, the equipment's maintenance and repair, and last, but not the One of them is the set of weather coefficients, and the other one is geographical data Forecasting implies numerous factors; they include the weather coefficients and the geographical indicators. And by resources – this is again something that you have to comprehend – where you are going to need it most, particularly if it comes to the points of snow drifts.

At the moment, AI is recasting enumerable practices in any discipline in a way that machines can learn from exposure, create, and solve problems regarded hitherto as requiring intelligence. AI has been conceptualized over 60 years ago and the development of technologies and product based on AI and interest in these technologies and applications have witnessed a number of fluctuations. Several factors have contributed to a recent revival of interest in AI over the last decade: more computational capabilities, more and/or larger mass storage devices, and advances in the computational formalization of AI techniques

© Sydorchuk V.

including a subfield of this area, ML [1]. AI has been received warmly by the public and policy makers with signs of huge possible benefits in terms of productivity, resource management and people's quality of life. In the banking and financial sector, artificial intelligence is crucial for high-frequency stock trading and fraud detection. AI is utilized in national security for object/threat recognition and cybersecurity. Artificial intelligence (AI) is utilized in the medical field to evaluate patient data, assist in diagnosis, and forecast which treatments will work best for each patient [2].

With potentially revolutionary effects, the current generation of AI subfields and methodologies is ready to expand into the transportation environment. By utilizing cutting-edge technologies like shared mobility services, automated vehicles (AV), connected vehicles (CV), and accessible transportation capabilities, the Intelligent Transportation Systems (ITS) Joint Program Office (JPO) and its modal partners have taken the lead in addressing fundamental issues in mobility, safety, and equity. The United States Department of Transportation (USDOT) has significantly increased its research and development of artificial intelligence (AI) in the past few years [3]. The Federal Highways Administration (FHWA), Federal Aviation Administration (FAA), and Federal Railroad Administration (FRA) are a few of the USDOT's modal administrations that have led the way in using AI technologies for mission execution. Data fusion, anomaly detection, safety analysis, and video analytics have all been applied using AI. For instance, the development of AI technologies for the collection of copious amounts of traffic data, including safety data, to recognize patterns and link seemingly unrelated data streams, and for video analytics to assist in predicting driver behavior in various driving scenarios, was funded by the Federal Highway Administration's Exploratory Advanced Research Program [4]. The use of AI to the development of prediction methods and assessment instruments is being studied by the FHWA's Traffic Analysis Tools (TAT) Program [5]. More than $16 million in awards were recently given out by the FHWA's Advanced traffic and Congestion Management Technologies Deployment (ATCMTD) Program to develop AI-powered solutions for multimodal traffic management [6]. Using artificial intelligence (AI) and unmanned aerial systems (UAS), FRA is creating a suite of technologies for predictive analytics and intrusion detection [7]. Other government organizations are investigating the use of artificial intelligence (AI) in citizen-facing services. These organizations include the Federal Transit Administration (FTA), the Federal Motor Carrier Safety Administration (FMCSA), and the Pipeline and Hazardous Materials Safety Administration (PHMSA) [8].

Having effective and timely snow management in the European countries can be an important concern specifically during cold season because heavy accumulation of snow can hinder the life within urban environment, mobility and in extreme cases even endanger lives. The conventional ways employed in planning snowplows routes are usually slow and ineffective and can in many a times fail to meet the dynamic conditions of the urban environment such as, traffic jam, closed roads, and differences in the intensity of snowfalls. This requires the enhancement of better ways to improve on the functionality of snowplow, in order to free the priority sectors and areas much faster [9].

The problem under consideration is the development of deep learning models that will allow determining the optimal routes for snowplows in the context of spatial configurations of big city environments. The problem is to combine spatial modeling with the techniques of route optimization with the help of neural networks to analyze the data on historical snow, traffic and available spatial layouts. These algorithms should therefore have the ability of coping with the volatility of the urban environment; provide turn by turn route guidance based on the current and future state of the roads [10].

However, one should mention that traffic management data should be incorporated to avoid interruptions during the procedure of snow clearing. This was attributed to the need for the growth of neural networks that would be capable of handling spatial data so as to deliver proper and dependable route estimations. In overcoming these challenges, the proposed deep learning models assets organizing the operation of snowplows, reducing response time and thereby improving the fluidity of cities during snow conditions for a safer and more sustainable environment in the European cities.

The research purpose formulation. The purpose of this work is to consider and design the deep learning algorithms that can be used for the forecast of the snowplow paths in the European cities, thus improving the effectiveness of the snow management in the urban environment. Through the adoption of important ideas like spatial modeling, route optimization, neural networks, and traffic management, the report is meant to offset the dynamism of urban environment during snow events as a challenge. The objective is to give an understanding of how the use of such high-level mathematical algorithms can be utilized to enhance the functionality of snowplows to allow snow clearance within the shortest possible

© Sydorchuk V.

time on essential sections. This report also wishes to stress on the necessity of including present road traffic information and city infrastructure plan in to the derationing process in order more effectively guide the snowplows within the cities with least or no disturbance of the general traffic flow. By this way, the report wants to bring new inputs related to the applicability of the deep learning and spatial modeling in the field of the management of the urban infrastructures so as to create the basis for building smarter and safer cities that are capable to address extreme climate conditions with more effectiveness and accuracy.

The latest research and publications analysis. There is a good deal of research available on the use of geographic information systems (GIS) in transportation asset management, the incorporation of databases and expert systems, and the fundamentals of applying optimization approaches. Recent papers in the field of expert system development include [11] work, which integrated GIS information to optimize visiting nurse routing [12], optimization of Sears service van routing using operations research techniques and GIS, which resulted in $42 million in annual savings. [11] have combined a GIS database with a knowledge base for toxin cleanup in the context of municipal public works projects. and the integration of GIS with other datasets for managing public transportation has been demonstrated [13].

Expert systems are an AI technique that various transportation researchers have employed. For instance, an intelligent system for choosing deicing material for winter road maintenance has been created [14]. Geographical and demographic information may be utilized to project future transportation needs in Switzerland. Making decisions about how to improve the transportation infrastructure depends on this demand estimate. Individuals are categorized based on the daily commutes they take. The data is first sorted using a decision tree approach, and the classification is then further refined using a support vector machine (SVM). Both models are based on supervised neural networks. To create a forest, many de-correlated decision trees are integrated using a decision tree. The accuracy of the results increases with the number of trees. This technique assigns a classification based on the distinct attributes of each x-tree [15, 16].

The main research material presentation. Due of the abundance of operational data on snowplow operations, these issues may be data-driven resolved using predictive machine-learning (ML) algorithms. In this work, we used the widely used machine learning approach known as random forest (RF) to forecast snowplow vehicle performance, which was measured by the rank of major repair times. Support vector machines (SVM) are another machine learning approach. Both models were applied to the fleet of snow plough trucks operated by the Utah Department of Transportation, and it was shown that RF performs better in terms of prediction accuracy than linear SVM. Furthermore, by identifying critical performance characteristics, a feature significance analysis using the RF model can assist transportation organizations in improving their truck replacement strategies. Finally, an example of applying the created prediction model with RF indicates the threshold of labor intensity to stop trucks' performance from rapidly declining in different types of working conditions. When compared to life-cycle cost evaluations from earlier research, the prediction model presented here can assist transportation authorities in setting more appropriate fleet replacement priorities [17].

Presentation of main material. Introduction to Route Optimization and Neural Networks

In European cities, precise organization and logistics of snowplows during winter is something very important for traffic safety and road usability. It involves the forecast, of routes that would take least time and use less fuel while at the same time covering all or most of the important areas. Conventional approaches of route planning do not consider traffic conditions, weather and other conditions in cities and infrastructure. One of the most successful subtypes of deep learning, namely recurrent neural networks, can also accommodate these components into the models [18].

The Methodology of Neural Networks applied to Spatial Analysis and Traffic Control

Artificial neural networks operate on the principles of the biological neural networks and have become popular in spatial modeling because of their capability of developing complex patterns from large data sets. For example, in the case of selecting the best route for snowplow, the neural networks can take data regarding the weather forecast, historical traffic data or configuration of the road networks and come up with an optimized route. With respect to the time-series data such weather patterns and the traffic flow which are vital in the anticipation of the route to be followed by the snowplow, the long short term memory networks which are a class of the Recurrent neural networks can be implemented. Since bidirectional LSTM networks are used, the direction of data can be seen, and using this data, much more precise route recommendations can be obtained [19, 20].

Some of the developments in CNN for Modeling Spatial Dependence

CNNs have in the past been utilized in image processing but current being employed in traffic and

© Sydorchuk V.

urban planning. In snowplow route optimization, one of the most important factors is the spatial dependency feature; CNNs can therefore be employed to capture this feature. For instance, the road network can be limited to grids [21], and CNNs can then identify spatial hieroglyphs of each grid, for example, road density, interconnection, and traffic. Nevertheless, standard CNNs fail to handle non-Euclidean geometries such as a road network, which has given rise to GCNs [22]. GCNs generalize CNNs for working with graphs and can be used for the modeling of transportation networks for they are based on graphs using intersections and road segments [23].

Spatio-temporal models for dynamic route prediction has been proposed

The features such as spatial and temporal models should be combined in order to design a good dynamic route prediction in the snowplow infrastructure. Similarly, Spatio-Temporal Graph Convolutional Networks (STGCNs) capitalize on the GCNs and LSTM networks to model traffic and road conditions spanned spatially and temporally. This mixed model may facilitate the estimation of traffic capability and road access in real time, by which means the snowplows may alter their pathways responding to the current circumstances.

Case Study: Implementation in EU Cities

Some of the large European cities have already ventured into deploying these contemporary forms of neural network models of traffic and routing. For example, when employing LSTM and CNN models, cities can predict not only the find the routes for snowplows but to estimate the amount of time it will take to clear the snow. They are built from data sets which consist of previous traffic information, weather data and actual traffic pattern information and has enhanced the efficiency and safety of the roads.

New Happenings and Future Direction

New work demonstrates the ability of Transformer models that employ attention mechanisms in capturing dependency structures in traffic data. These models have been found to perform better than the conventional RNNs with regards to traffic flow predication, and hence could be implemented in snowplow route optimization in the future. Further, real time information streams, social media feeds along with the IoT sensors is becoming popular in use. For example, tweeting about road conditions can be used in improving models used to predict states thereof.

Table 1. Comparison of Neural Network Models for Route Prediction [9, 11, 19, 20, 23]

Model Type	Strengths	Weaknesses	Application in Snowplow Route Optimization
LSTM Networks	Handles time-series data effectively	Requires large datasets	Predicting traffic flow and weather impact
CNNs	Extracts spatial features from grids	Limited in non-Euclidean structures	Initial spatial modeling
GCNs	Models graph structures of road networks	Computationally intensive	Advanced spatial dependency modeling
STGCNs	Combines spatial and temporal data	Complex implementation	Real-time dynamic route optimization
Transformer Models	Captures complex data dependencies	High resource requirements	Future applications in route prediction

Comparative Analysis of Deep Learning Techniques for Traffic Prediction

The method of machine learning nonlinear models is still being used in traffic prediction mostly in the city center where factors like congestion, closing of some roads, and events are likely to occur. Below is a deeper explanation of some of the most prominent techniques: Below is an elaboration of some of the more common techniques that have been described above:

Recurrent Neural Networks (RNNs)

Use Case

The other reason why use of RNNs is suitable when modeling traffic flow is because they are suitable for sequence prediction, which is especially useful when predicting the next traffic in the subsequent time steps. They are meant for data that is in some way a function of time, for example traffic density as a function of the time of day or vice versa and weather changes.

Strengths

© Sydorchuk V.

RNNs can operate over time, that is, RNNs can 'learn' order or the temporal dimension of phenomena. This is important in the projection of current traffic pattern trends brought about by the elements of traffic flow.

Limitations

RNNs have issues like vanishing and exploding gradient which might affect its training for the particular long-range dependencies. This results to them being less efficient when used for estimating traffic over longer times intervals majoring components such as LSTM (Long Short-Term Memory).

Temporal Convolutional Networks (TCNs)

Use Case

To the RNNs, there are TCNs which are well known to be used in sequence modeling. They use convolutional layer that helps to capture temporal dependencies and it awards them the ability to use many sequences at once.

Strengths

It is far better than the RNN in that it poses no problem of vanishing gradient even for long sequences of information input. These are less sequential in computing and as such are faster to train and more scalable as compared to most of the deep learning models.

Limitations

However, it is need to mention some problems about TCNs: they require much computational resources and can have some difficulties with the some immensely complex temporal relations that can be met in traffic data.

Therefore, although they provide some advantages, TCNs require much computational power and cab encounter difficulties in some, very complex temporal relations that one can encounter in traffic data.

Multilayer Perceptron (MLP)

Use Case

MLPs are simple neural networks confined usually for and used primarily for classification and regression tasks. In traffic prediction, MLPs are able to map simpler dependent variables and the sort of independent variables as the impact of weather on the rate of flow of traffic.

Strengths

They are also easy to implement and unlike some of the more complex HLLs are not computationally very intensive. They are still applicable when the relations in the data are not of a relatively high degree of complexity.

Limitations

An MLP would be insufficient in estimating traffic, especially if the data acquisition is in the form of a sequence because MLPs are incompetent of processing such forms of data. What they cannot do, however, is modelling temporal dependencies, which other examples of models such as the Recurrent Neural Networks (RNNs) or Temporal Convolutional Networks (TCNs).

Graph Neural Networks (GNNs)

Use Case

Intersection and road segments can be viewed as nodes and links respectively and data such as the road networks of cities are an input to GNNs. This makes them suitable for problems where there are dependencies in space i.e. where at least some of the operations require argument values that are dependent upon the value of the previous operation's argument.

Strengths

They can include the topology of the network in a road map and thus enable prediction of the flow of traffic and congestion in any given area of a city. They do well in problems where traffic movement is most endangered by the current layout of the network.

Limitations

GNNs are very flexible with very high non-linearity and, therefore, are very resource demanding networks to train. They also work best when they are well implemented by people with a good knowledge of graph theory; thus, they have rather steep learning curves, especially when it comes to general traffic prediction tasks

Transformers

Use Case

© Sydorchuk V.

At present, Transformers which has become famous in natural languages processing are considered for traffic prediction due to the capability of attention-based mechanisms to take care of long-range temporal dependencies.

Strengths

Compared to Recurrent Neural Networks or Temporal Convolutional Networks, the Transformers have a higher potential of capturing nonlinear features of data; therefore, it might be more suitable for forecasting of some data. They also have certain very useful features – they can take complete sequences in one go, which as you know is faster.

Limitations

Issues of transformer interference arise in the case of an absence of the data and computational power in traffic managing systems which is necessary for their support. They also serve newer in this area, so their use is still in the process if being discovered and being experiment.

Table 2. Comparative Analysis of Deep Learning Techniques for Traffic Prediction

Technique	Use Case	Strengths	Limitations	Example in Snowplow Routing
Recurrent Neural Networks (RNNs)	Predicting sequential data (e.g., traffic over time)	Captures temporal dependencies	Prone to vanishing gradient issues	Traffic flow prediction
Temporal Convolutional Networks (TCNs)	Traffic flow prediction using sequence data	Handles long sequences with parallelism	High computational resources	Medium to long-term predictions
Multilayer Perceptron (MLP)	Basic traffic prediction and classification	Simple and efficient for small datasets	Limited in handling sequential data	Initial route planning
Graph Neural Networks (GNNs)	Analyzing road networks as graphs	Handles non-Euclidean data effectively	Complex model structure and training process	Road network analysis
Transformers	Advanced traffic prediction with attention mechanisms	Captures complex dependencies effectively	High resource consumption	Future application potential

The shown above tables additional provide an elevated and accurate representation of the fundamental parameters, compare methods and data fundamentals necessary to derive and evaluate deep learning models specifically tailored to the optimization of snowplow route. They demonstrate that within the same problem, different model can be used according to the state of the context of the city 's infrastructure and the type of data with which one has to operate in. All these are very important in rendering and finally implementing right course of action of snow removal policy that would have least inconvenience and maximum safety to the public in winter in the cities of Europe.

Conclusion and prospects for further research. Each of the earlier explained deep learning methods has its own strengths and weaknesses, and, therefore, the methods can be used in traffic prediction and route planning in the following manner. Even today, there is a very high demand to work with and improve RNNs, and their derivatives LSTM because of their ability to work with sequences. Despite TCNs improving so much the ability to handle sequences of significant length, GNNs enable to capture intrinsic spatial relations present in road networks. Auto encoder is the basic model and Transformer is the state of the art with relatively high resource requirement. The choice of this model is very much influenced by the selected traffic prediction mission or task, the nature or the complexity of data and computation power available. Deep learning algorithms especially neural networks have been applied especially in snowplow routes, it is a pioneering example of improved urban structure information management.

Indeed, through the use of spatial modeling and, traffic data, these models present a dynamic and effective way to deal with winter impacts on the European city. This becomes even more apparent as more innovative models such as transformers are integrated more into the system where diverse data feed is embraced to improve the accuracy of the predictions.

References

1. Hallmark, B., & Dong, J. (2020). Examining the effects of winter road maintenance operations on traffic safety through visual analytics. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*.

© Sydorchuk V.

IEEE.
2. Hatamzad, M., Pinerez, G. P., & Casselgren, J. (2021). Using slightly imbalanced binary classification to predict the efficiency of winter road maintenance. *IEEE Access, 9*, 160048–160063.
3. Vasudevan, M., et al. (2020). Identifying real-world transportation applications using artificial intelligence (AI)-real-world AI scenarios in transportation for possible deployment. United States Department of Transportation. Intelligent Transportation.
4. U.S. Department of Transportation. (2020). Federal Highway Administration Office of Operations Recent TSMO Resources. Retrieved August 21, 2024, from https://ops.fhwa.dot.gov/plan4ops/docs/recent_resources_winter_2024.pdf
5. Federal Highway Administration Office of Operations. (2020). Tools for tactical decision-making/advancing methods for predicting performance. Retrieved August 21, 2024, from https://ops.fhwa.dot.gov/fastact/atcmtd/2018/awards/index.htm
6. U.S. Department of Transportation. (2020). U.S. DOT Artificial Intelligence Research Highlights. Retrieved August 21, 2024, from https://ops.fhwa.dot.gov/fastact/atcmtd/2018/awards/index.htm
7. Baillargeon, J. (2019). Artificial intelligence research at Federal Railroad Administration (FRA). In *ITS America Annual Meeting*.
8. Borener, S. (2019). The safety data transformation. In *ITS America Annual Meeting*.
9. Carrillo, J., et al. (2020). Design of efficient deep learning models for determining road surface condition from roadside camera images and weather data. *arXiv preprint arXiv:2009.10282.*
10. El-Rayes, K., & Ignacio, E.-J. (2022). Evaluating the benefits of implementing mobile road weather information sensors. *FHWA-ICT-22-004*.
11. Asamani, G. P., & Sun, Z. (2023). Actionable science for snow monitoring and response. In *Actionable Science of Global Environment Change: From Big Data to Practical Research* (pp. 229-259). Springer.
12. Kim, N., et al. (2021). Hyperconnected urban fulfillment and delivery. *Transportation Research Part E: Logistics and Transportation Review, 145*, 102104.
13. Jiao, W., et al. (2022). Optimization design and prediction of the snow-melting pavement based on electrical-thermal system. *Cold Regions Science and Technology, 193,* 103406.
14. Ng Corrales, L. d. C., et al. (2020). Overall equipment effectiveness: Systematic literature review and overview of different approaches. *Applied Sciences, 10*(18), 6469.
15. Ernest, A. (2021). An artificial intelligence-based approach to estimating time of arrival and bus occupancy for public transport systems in Africa. *arXiv preprint arXiv:2101.07674.*
16. Vermesan, O. (2022). *Artificial intelligence for digitising industry–applications.* CRC Press.
17. Gong, T., et al. (2023). Edge intelligence in intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems, 24*(9), 8919-8944.
18. Yu, Y., et al. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation, 31*(7), 1235-1270.
19. Almasan, P., et al. (2022). Deep reinforcement learning meets graph neural networks: Exploring a routing optimization use case. *Computer Communications, 196*, 184-194.
20. Bi, H., et al. (2021). GIS aided sustainable urban road management with a unifying queueing and neural network model. *Applied Energy, 291*, 116818.
21. Deneu, B., et al. (2021). Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLoS Computational Biology, 17*(4), e1008856.
22. Du, P., et al. (2020). Advances of four machine learning methods for spatial data handling: A review. *Journal of Geovisualization and Spatial Analysis, 4*, 1-25.
23. Bhatt, D., et al. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics, 10*(20), 2470.

© Sydorchuk V.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-06
UDC 004.42:004.92
Sidorov Denys, Bachelor's Degree
https://orcid.org/0009-0004-8363-6553
Odesa I.I. Mechnikov National University, Odesa, Ukraine

BEST PRACTICES FOR DEVELOPING SECURE FRONT-END WEB APPLICATIONS

Sidorov D. Best Practices for Developing Secure Front-End Web Applications. In the course of the work, a review of trends and best practices in creating user interfaces was conducted. To achieve this goal, attention was focused on the use of advanced technologies and design methods by interface developers, which have been permanently evolving in recent years. The methodology used to create intuitive and convenient interfaces is revealed. The main trends and innovations in the design of user interfaces were mentioned. The impact of user experience (UX) on the development process is indicated. Practical examples and cases of implementation of the latest approaches in real projects are given. It is also proposed to apply the selected best practices for creating a user interface in accordance with the characteristics of each of them, which will make the development of user interfaces easier and more efficient, based on which practical recommendations were provided for optimizing the development process. These guidelines can be used both at the departmental level of large corporations and by individual developers and freelancers, making them universal in terms of accessibility and comprehensibility. The received recommendations are the basis for the newest paradigm of development and optimization of interactive user interfaces, which can change the approach to the development of user interfaces in general. The very question of changing the paradigm of creating interactive user interfaces based on information about the interaction of applications with the user, not only in the framework of web development, but also in desktop development, is determined to be relevant for further research, which will be considered in future scientific works.
Keywords: web development, frontend, user interaction, web design, browser, mobile optimization, applications, cross-browser compatibility.

Сідоров Д. В. Кращі практики розробки безпечних інтерфейсних веб-додатків. В ході роботи було проведено огляд трендів та найкращих практик у створенні користувацьких інтерфейсів. Для досягнення цієї мети було акцентовано увагу на використанні розробниками інтерфейсів передових технологій і методів дизайну, що перманентно еволюціонують в останні роки. Виявлено методологію, що використовується для створення інтуїтивно зрозумілих та зручних інтерфейсів. Було згадано основні тенденції та інновації у дизайні користувацьких інтерфейсів. Зазначено вплив користувацького досвіду (UX) на процес розробки. Наведено практичні приклади та кейси впровадження новітніх підходів у реальних проектах. Також запропоновано застосування обраних найкращих практик створення користувацького інтерфейсу в відповідності до особливостей кожної з них, що зробить розробку користувацьких інтерфейсів простішою та ефективнішою, на основі чого були надані практичні рекомендації для оптимізації процесу розробки. Ці рекомендації можуть бути задіяні як на рівні відділів великих корпорацій, так і індивідуальними розробниками та фрілансерами, що робить їх універсальними в контексті доступу та зрозумілості. Отримані рекомендації є основою для новітньої парадигми розробки та оптимізації інтерактивних користувацьких інтерфейсів, що здатне змінити підхід до розробки користувацьких інтерфейсів в цілому. Саме питання зміни парадигми створення інтерактивних користувацьких інтерфейсів на основі інформації про взаємодію додатків з користувачем, не тільки в рамках веб-розробки, але і десктопної розробки, визначено актуальним для наступних досліджень, що будуть розглянуті в майбутніх наукових працях.
Ключові слова: веб-розробка, фронтенд, взаємодія з користувачем, веб-дизайн, браузер, мобільна оптимізація, застосунки, кросбраузерна сумісність.

The problem statement. Modern programs contain not only a visual representation of elements on the screen. Thanks to the interactive design, the programs also take into account the interaction of the user with the graphic elements of the program. Interaction design is a specialized field of design that focuses on how users interact with certain products and services. Digital products have greatly benefited from the implementation of this area of design by integrating the user interface into the overall user interaction [1]. It is this approach that allows interactive interface developers to combine software and hardware into digital solutions that users love. Interactive interface designers can encourage users to engage in certain types of interactions by using various elements in the interface. In this way, they help them avoid unwanted interaction, reducing the risk of a negative user experience.

As of 2022, 67% of people worldwide used mobile phones. 80% of them had smartphones, and 96% of Internet users [2]. This shows the importance these gadgets play in society, from making calls and playing games to getting credit information from the bank or the weather forecast, and the direct correlation between having a mobile phone and using web applications. An increasing number of different mobile applications are being created to help their users perform a wide range of tasks, and most of these applications have an interactive interface.

To ensure the success of using the application, developers strive to maximize the quality of the user experience, and therefore, user satisfaction [3]. In 2018, the total number of mobile applications

© Sidorov D.

downloaded was 194 billion, with each smartphone user having many applications installed on their phone [4].

Interactive design has become one of the most important elements in the process of web development and mobile application development. Typically, users interact with interface elements to perform certain actions they need. If they can't interact with the user interface the way they want, the results are usually suboptimal. Disappointment appears, which negatively affects interaction with the user and calls into question the reputation of the developer in front of the customer of the application or site.

During user interface (UI) development, developers create the visual elements of a web application, such as colors, typography, icons, and graphics. The design of the interface should correspond to the style of the brand and reflect the needs of the target audience. The interface design should also be adaptive, i.e. adapt to different screen sizes and resolutions, meet the requirements of mobile optimization [5].

The main problem in web development and web design has now become the issue of insufficient competence of front-end developers and designers. Among all the reasons that lead sites to failure, it is necessary to highlight errors in its design, lack of cross-browser compatibility (does not work in all browsers) and lack of adaptive layout [6]. One way to solve this triple challenge is to create interactive user interfaces, based on existing best practices in accordance with the characteristics of each one, which will make web development easier and more efficient. Websites and applications developed based on these practices will be optimized for use on mobile devices and in a variety of browsers, making them more accessible to potential users. All this requires a paradigm shift in the creation of user interfaces and their unification, which is possible only when the main trends and innovations in the design of user interfaces are identified.

Formulation of the purpose and objectives of the research. The purpose of the article is to research and reveal modern innovative approaches in the creation of user interfaces, with an emphasis on the use of advanced technologies and design methods. To achieve the goal, the following tasks of the current research were solved:

1. The peculiarities of creating user interfaces and interactive interfaces as areas of web design are explained.
2. The methodology used to create intuitive and convenient interfaces is revealed.
3. The main trends and innovations in the design of user interfaces are considered.
4. The impact of user experience (UX) on the development process is determined.
5. Practical examples and cases of implementation of the latest development approaches in real projects are mentioned.
6. Based on the previous steps, practical recommendations were developed to optimize the web development process.

Among the general scientific methods that were used to solve research tasks and achieve its goal, the following should be noted:

- monitoring method: used to collect, systematize and analyze information about user interfaces and their development features;

- comparison method: came in handy when researching different methodologies for creating interfaces and comparing them;

- the method of abstraction: was used during the research in order to highlight the main concepts and categories;

- methods of analysis and synthesis: used in the process of identifying the stages and factors of development, as well as the most influential elements of the researched object.

To solve the general tasks of the current research, the following groups of special methods were used: information collection methods; information processing methods; methods of analytical work; justification method.

The relevance of the researched topic is argued by the importance of quality control of websites and applications that allow users to make their lives easier and more convenient, and developers to attract funds and, based on previous experience, to create increasingly high-quality and modern websites and applications that will have intuitive and convenient user interfaces.

Analysis of recent research and publications. Today, the topic of web development and optimization of mobile and web applications for various platforms and browsers in the context of the user interface, as well as its derivatives, is of interest to many domestic and foreign researchers and practitioners

© Sidorov D.

in the field of information technology, and especially front-end development specialists. Among the Ukrainian scientists who researched this topic, it is necessary to mention H. Chemerys, O. Barkalov, R. Babakov, O. Blahodelskyy, O. Pushkar, Y. Hrabovskyy, I. Bardus, and O. Pryvezentsev.

H. Chemerys in her scientific publication considered the principle of the «golden» ratio in the process of designing the user interface of websites. She proved that the application of classical principles of composition in the process of prototyping the user interface of websites has a positive effect on the harmonious perception of design as a whole [7].

In the joint article of O. Barkalov and R. Babakov, the issue of development and use of cross-platform software was considered, which was based on a study of the speed of various libraries designed for building a graphical user interface of applications written in the Python language [8].

In his scientific work, O. Blahodelskyy conducted a literature review and analysis of the features of the use of modern computer technologies to create the external interface of web applications based on the application of machine and deep learning models [9].

O. Pushkar and Y. Hrabovskyy, in their article, analyzed the characteristic features and principles of building a user interface and substantiated their own author's approach to the development of an intellectual user interface of educational publications in the e-learning system [10].

I. Bardus and O. Pryvezentsev reviewed the job duties of specialists who develop user interfaces (UI designer, UX designer, front-end developer, graphic designer and user experience researcher) and described the full process of their development, from concept to cross-platform support [11].

From among foreign researchers, it is worth noting M. Lazuardy, D. Anggraini, E. Planas, G. Daniel, M. Brambilla, J. Cabot, A .Narayanan, A. Mathur, M. Chetty, M. Kshirsagar, W. Nasution, P. Nusa.

The article by M. Lazuardy and D. Anggraini presents a comprehensive overview of the development of SIASN web applications. In addition, the advantages and disadvantages of React.js and Next.js as the main technologies in building the user interface of the SIASN program are considered [12].

E. Planas, G. Daniel, M. Brambilla, and J. Cabot investigated the application of model-driven techniques to the development of software applications with embedded multi-functional user interfaces. They also proposed a new domain-specific language (DSL) to define several types of dialog interface (CUI) [13].

A. Narayanan, A. Mathur, M. Chetty, and M. Kshirsagar conducted research on the evolution of «dark patterns» (user interfaces that encourage users to make decisions they would not otherwise make). During the research, it was determined that a number of world-famous online services are trying to mislead users through the use of «dark templates» [14].

In the joint scientific work of W. Nasution and P. Nusa, a design prototype (user interface)UI/(user experience)UX of an Indonesian educational web program called "IdeIn" was considered, which was developed using the method of design thinking. All stages of design development and its features are also described [15].

Having analyzed the most up-to-date materials that are currently publicly available, such as scientific works, monographs, articles on specialized sites, we can conclude that the topic of web development and optimization of mobile and web applications for various platforms and browsers in the context of the user interface is relevant and interesting for many researchers around the world.

Also, mentioning and evaluating the significant contribution of the researchers listed above to the review of the current topic, it is necessary to note the lack of a single universal approach to the study of issues and features of web development through the prism of user interfaces and interactive user interfaces as elements of web design.

Presentation of the main material and substantiation of the obtained research results.

First of all, it is necessary to explain the meaning of the phrase «user interface». A user interface (UI) is generally anything that facilitates interaction between a user and a machine. In the world of computers, it can be anything from a keyboard, a joystick, a screen or a program. In the case of computer software, this can be a command line, a web page, a user input form, or any program interface [16].

The design of the interface, created according to all the canons of UI design, helps the user to quickly understand what function, which icons, buttons perform, what follows. Whether the user continues to use the application or closes it after a couple of minutes depends on the UI design. Today, user interface is not only about interactivity and animation. This is what allows a brand to stand out, convey its mission and emotions, and create additional value for its customers. The stages of the development and approval processes of the user interface can be seen in the table. 1.

© Sidorov D.

Table. 1. Step-by-step development and approval of the user interface design

№	The name of the development stage	The essence of the step
1	Transfer of materials and briefing on the project	After completing the work on UX design, all the obtained data is transferred to the UI team: research, conclusions, insights, artifacts and prototypes with an explanation of all user paths and business features. This helps the UI team quickly dive into project materials, specifications, niche trends, and other relevant information to understand the needs of the target audience, project goals, and client requirements
2	Interviews and insights	Developers ask questions that can help choose the optimal style of the future product. They ask about the mood, emotions, brand book, logo, current site (if any) and the tasks they need to solve at this stage. If there is a site, ask the owners what they like and don't like about the current design. In addition, developers are asked to describe how they see the future site using metaphors. In this way, the owners can convey the desired mood of the product – it should be restrained and professional, or light and relaxed. All of this helps to understand the overall vision and design goals. It can be a simple desire to have a fresh, stylish interface, or something more specific. For example, stand out among competitors with the help of non-standard design solutions
3	Moodboards	After receiving all the necessary information, developers start creating moodboards to finally define the style of the future product. That is, customers get several options to choose from and can choose what they like. They also add some sketches of market trends and proven solutions to Figma as references. It is there that developers decide which elements to include in the design. After approving the details of the interface, they proceed to direct work on the design, taking into account all the requirements and preferences of customers
4	Concept development and page design	The creation of the interface itself begins with the main page and some key pages. In the case of e-commerce, this can be the main page, product card, cart and checkout. Developers create their design and style, including the smallest details such as layout, fonts, color palette, icons, etc. After that, they show the design of key pages to the customer, justifying the chosen design solutions and their importance for achieving the project's goals. If the client is satisfied with everything, the developers move on to developing the design of the rest of the site's pages. When key pages are finalized, work begins on designing other pages/screens in all possible states and with all details in mind. Secondary pages include: blog, service pages, personal account, content and service pages, etc
5	Animation	Depending on the project and the chosen solutions, developers can lay down and develop CSS animations to make the site more interactive and interesting. Animated prototypes can be created to visually demonstrate specific user interactions with the interface that are difficult to explain in words. This includes smooth transitions between pages, attractive animation during page scrolling, pop-up elements, drop-down menu
6	UI guide	At this stage, a detailed UI guide is created, covering all the main aspects of the design: from colors, fonts and line spacing, indents to interacting with information based on user behavior scenarios, interactive/pop-up elements, button clicks, transitions, and more. This step-by-step guide ensures a smooth transition from design to development, leaving no room for guesswork around design features and user interaction
7	Author's supervision of execution	After all stages of design creation and approval are completed, the project is handed over to the development team. Since the entire development team is responsible for the final design of the project, it is important for them that all UI solutions and ideas are correctly implemented. That is why they carry out the author's supervision of the work at all stages of development.

© Sidorov D.

Next, it is necessary to determine the specifics of creating user interfaces and interactive interfaces as areas of web design. These features are determined by the following development goals of these interfaces:

- convenience and practicality: the product is easy to use, which means that more people will be inclined to use it;

- beauty and uniqueness: the product is more pleasant to use and looks aesthetic, which attracts more and more users.

It is now important for any developer to understand the characteristics of a good user interface and why it is so important to the success of their product. Features of creating a user interface are shown in fig. 1.

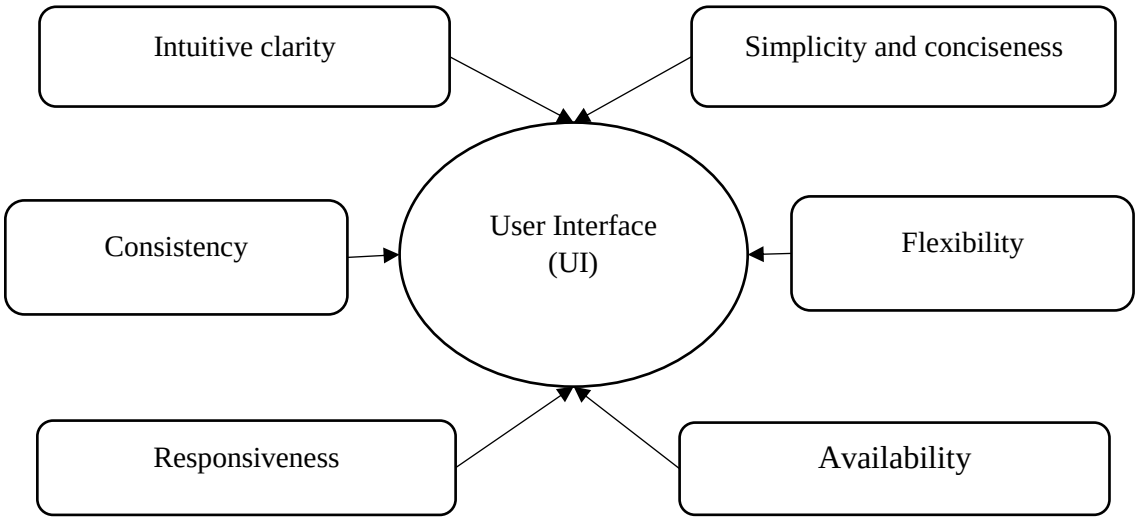


Fig. 1. Features of a qualitatively created user interface

The essence of the features of qualitatively developed user interfaces is explained in more detail below:

1. Intuitiveness: users should be able to understand how to use the product without reading long instructions. The best way to achieve this is to use familiar conventions and make functions self-evident (use easily recognizable symbols).

2. Simplicity and conciseness: a good user interface is always simple and clear. It should include only the elements that are absolutely necessary and nothing else, which will make it easier for users to find what they are looking for and allow them to use the product without getting tired.

3. Consistency: all elements of the user interface should be consistent with each other. This includes schemas, typography, iconography, and the overall design language; all of these elements create a better experience that is pleasing to the eye and easy to use.

4. Responsiveness: a good user interface should be attractive and work well on all devices and platforms, being cross-browser and cross-platform (optimized for mobile applications as well). This is important because, as previously identified, more and more people are using mobile devices to access the Internet.

5. Flexibility: a good user interface should be flexible and be able to adapt to different screen sizes, resolutions, and orientations (landscape or portrait), as developers do not know what device users will use to access the site or application.

6. Accessibility: a good user interface should be suitable for use by people with disabilities or special needs. This means making text alternatives available for images and allowing users to adjust the font size of these text alternatives. [18].

Summarizing the peculiarities of user interface development, it should be noted that it is conceptually defined as a combination of understanding the needs of the target audience and the skills of the frontend designer and developer (UI team).

The next step in the research is to review the user interface elements. UI components are an integral part of the design of a website or application. They are the building blocks for all products. Any web

© Sidorov D.

designer or front-end developer needs to have a deep understanding of user interface elements and how users interact with them. Designers typically don't draw UI elements themselves when they create web pages or mobile apps. They usually start with a ready-made repository of user interface elements. As a rule, there are three main categories of user interface elements, all these groups are shown in fig. 2.

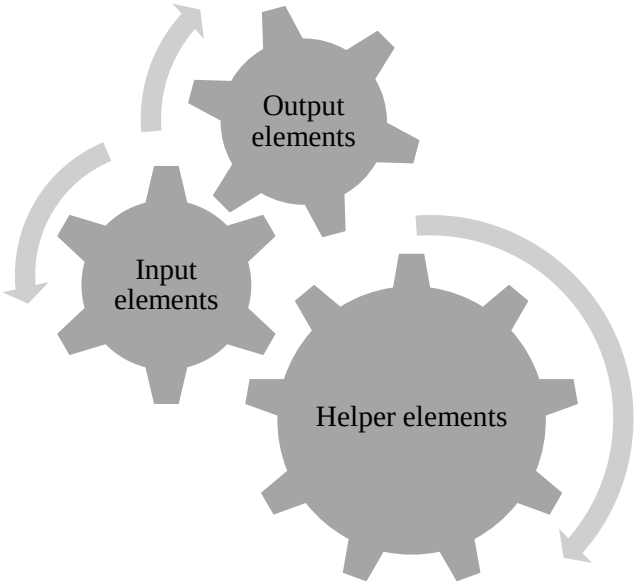


Fig. 2. Groups of user interface elements

The functions of the user interface elements can be briefly explained as follows:

1. Users use them to add the information they need and move on to the next step.
2. Elements show the result of a previous user action.
3. They help to move around the site or application, get information and draw the user's attention to a certain element [19].

Next, the main trends and innovations in the development and design of user interfaces will be described and listed:

- in 2024, design integrated with artificial intelligence has become more common, allowing designers to solve complex problems and expand the boundaries of their creativity;

- in 2024, animated icons have become a staple of user interface design, breathing new life into interfaces. Tools like Midjourney have helped designers create stunning 3D visuals that captivate users;

- typography plays a vital role in user interface design, and in 2024, big and bold fonts will become the dominant trend;

- dark mode has gained immense popularity in the last decade. Its development continued in 2024. This user interface trend offers a sleek and modern aesthetic, reducing eye strain in low-light conditions;

- the trend towards minimalism in design continued in 2024. Conciseness of interfaces helps to ensure the desired level of interaction with the user, allowing him to focus on working with the main elements [20].

In conclusion, it should be noted that the design of user interfaces is constantly evolving, and in 2024 there will be significant achievements and trends. For the designers of the future, the main task has become to combine their creativity and technical expertise to create interfaces that bring pleasure to the users of websites and applications.

Also, it should be noted that the user experience (UX) plays a decisive role in the success of creating user interfaces. Prioritizing UX can lead to increased user engagement, better retention and loyalty, better brand reputation, and lower development costs in the long run.

UX methods can be used to define design, such as:

– user research;
– persona creation;
– prototyping and testing.

© Sidorov D.

User-centered design is a common process today, in which developers and designers take the needs of users as the starting point for all design and further development of a product.

Practical examples and cases of implementation of the latest development approaches to user interfaces will be shown based on the overview of the website and the application.

Popular social media management service Hootsuite uses a fixed menu for its landing page and offers three eye-catching social sign-up buttons. But if the user wants to register using e-mail, then after clicking «create a new account», he will not go to a new page, because the system will display a window with illustrated fields. The Hootsuite app encourages users to fill out their details and start using the service right away, showing how much time they can save. After registering and adding social networks, the user will see an interactive guide to start working with the application.

The My Energy website, created to track the consumption and costs of energy, gas and water, clearly explains to users why certain data is needed and shows that registration is performed simultaneously in three steps: classifying the house, entering the address, verifying the location. Site users can use the tips section to mark tips they are already following, add them to to-do lists, and quickly view tips using filters and Saving Impact and Effort Required indicators presented in the form of gradient circles.

To speed up and simplify the process of developing user interfaces, it was decided to provide a number of recommendations, among which are the following:

– it is necessary to achieve maximum interaction between all members of the UI team, which allows to reduce time spent and the number of possible errors during development;

– it is necessary to conduct research on the market of technical innovations, looking for more powerful means of development, which will allow to win the competition from other developers;

– it is necessary to develop the creativity of designers by analyzing the works of other creators or artists during exhibitions;

– for frontend developers, as well as designers, their soft skills are important, which allow them to effectively interact with a manager or client, listening to his user experience.

Most of the recommendations given do not apply to hard skills, because due to the constant development of the website and application market, all developers must constantly improve their professional skills, regardless of frontend or backend.

Conclusions and prospects for further research.

In today's conditions of intensive development of information technologies, it is impossible to imagine a person's life without websites or mobile applications. During the work, an overview of the features and stages of user interface development, from the idea to direct implementation by the designer and front-end developer, was conducted. The main trends and innovations in the design of user interfaces in recent years and decades were mentioned. The impact of user experience (UX) on the development process is indicated as the main factor of successful development. Practical recommendations for speeding up and simplifying the development of user interfaces were also provided. These recommendations can be used both at the level of corporate departments and by individual developers or freelancers, which makes them universal. The received recommendations should become the basis for changing the paradigm of front-end development, which can change the paradigm of creating websites and applications as a whole. The very question of changing the paradigm of creating interactive user interfaces based on information about the interaction of applications with the user, not only in the framework of web development, but also in desktop development, is determined to be relevant for further research, and will be considered within the framework of these studies.

References

1. Gomez, J. (2021, November 9). *Interactive UI Design*. Koombea. Retrieved from https://www.koombea.com/blog/interactive-ui-design/ [in English].
2. Nahorna, A. (2022). Doslidzhennya: 63% naselennya planety mayut dostup do internetu, a v 67% ye mobilnyy telefon [Research: 63% of the world's population has access to the Internet, and 67% have a mobile phone]. Retrieved from https://dev.ua/news/63-naselennia-planety-maiut-internet [in Ukrainian].
3. Al-Subaihin, A. A., Sarro, F., Black, S., Capra, L., & Harman, M. (2019). App store effects on software engineering practices. *IEEE Transactions on Software Engineering*, *47*(2), 300-319. DOI: 10.1109/TSE.2019.2891715 [in English].
4. Simon, J. P. (2020). The global internet market (s): a reconstruction of the views of the industry. *Digital Policy, Regulation and Governance*, *22*(2), 109-133. Retrieved from https://www.emerald.com/insight/content/doi/10.1108/DPRG-10-2019-0092/full/html [in English].
5. *Web App Development: 8 Critical Stages to Build Amazing Apps*. (2023, March 9). While Web Production. https://whileweb.com/blog/web-app-development-8-critical-stages-to-build-amazing-app/ [in English].

© Sidorov D.

6. Kobylyanskyy, I. (n. d.). *20+ pomylok pry stvorenni saytu, yaki slid unykaty [20+ website design mistakes to avoid]*. Impulse Design. Retrieved from https://impulse-design.com.ua/ua/20-oshibok-pri-sozdanii-sajta.html [in Ukrainian].
7. Chemerys, H. Y. (2019). Zastosuvannya pryntsypiv kompozytsiyi u proektuvanni korystuvatskoho interfeysu veb-saytiv [Application of the principles of composition in the design of the user interface of websites]. Informatsiyni tekhnolohiyi v osviti ta nautsi - Information technologies in education and science, (11), 323-326. Retrieved from http://surl.li/atqtbo [in Ukrainian].
8. Barkalov, O. O., & Babakov, R. M. (2022). Doslidzhennya produktyvnosti krosplatformnykh bibliotek dlya pobudovy hrafichnoho korystuvatskoho interfeysu [Study of the performance of cross-platform libraries for building a graphical user interface. Applied aspects of modern interdisciplinary research]. *Prykladni aspekty suchasnykh mizhdystsyplinarnykh doslidzhen. Kiberbezpeka, prykladni informatsiyni tekhnolohiyi, tekhnolohiyi internetu rechey - Applied aspects of modern interdisciplinary research. Cyber security, applied information technologies, Internet of Things technologies*, 78-81. Retrieved from https://jpasmd.donnu.edu.ua/article/view/12921 [in Ukrainian].
9. Blahodelskyy, O. (2023). Zastosuvannya suchasnykh kompyuternykh tekhnolohiy z metoyu stvorennya veb-rozrobky y zovnishnoho interfeysu (front-end) [Application of modern computer technologies with the purpose of creating web development and external interface (front-end)]. *Nauka i tekhnika sohodni - Science and technology today*, *9*(23), 526-540. Retrieved from http://perspectives.pp.ua/index.php/nts/article/view/5626/5658 [in Ukrainian].
10. Pushkar, O., & Hrabovskyy, Y. (2019). Metodyka rozrobky intelektualnoho korystuvatskoho interfeysu navchalnykh vydan v systemi e-learning [Methodology for developing an intelligent user interface for educational publications in the e-learning system]. *Development Management, 17*(3), 23-34. doi:10.21511/dm.17(3).2019.03 [in Ukrainian].
11. Bardus, I., & Pryvezentsev, O. (2023). Analiz profesiynoyi diyalnosti fakhivtsya z rozrobky korystuvalnytskykh interfeysiv [Analysis of the professional activity of a specialist in the development of user interfaces]. *Naukovi zapysky Berdyanskoho derzhavnoho pedahohichnoho universytetu. Pedahohichni nauky - Scientific notes of Berdyan State Pedagogical University. Pedagogical Sciences*, (1), 199-209. Retrieved from https://dspace.bdpu.org.ua/server/api/core/bitstreams/487c8783-b42e-48cc-8653-4fbf61a6f2a3/content [in Ukrainian].
12. Lazuardy, M. F. S., & Anggraini, D. (2022). Modern front end web architectures with react. js and next. js. *Research Journal of Advanced Engineering and Science*, *7*(1), 132-141. Retrieved from https://irjaes.com/wp-content/uploads/2022/02/IRJAES-V7N1P162Y22.pdf [in English].
13. Planas, E., Daniel, G., Brambilla, M., & Cabot, J. (2021). Towards a model-driven approach for multiexperience AI-based user interfaces. *Software and Systems Modeling*, *20*(4), 997-1009. Retrieved from https://link.springer.com/article/10.1007/s10270-021-00904-y [in English].
14. Narayanan, A., Mathur, A., Chetty, M., & Kshirsagar, M. (2020). Dark Patterns: Past, Present, and Future: The evolution of tricky user interfaces. *Queue*, *18*(2), 67-92. Retrieved from https://dl.acm.org/doi/pdf/10.1145/3400899.3400901 [in English].
15. Nasution, W. S. L., & Nusa, P. (2021). UI/UX design web-based learning application using design thinking method. *ARRUS Journal of Engineering and Technology*, *1*(1), 18-27. Retrieved from https://pdfs.semanticscholar.org/d35d/cb384e80fb3abad1146815740a045f0048d3.pdf [in English].
16. *UI - MDN Web Docs Glossary: Definitions of Web-related terms*. (n. d.). MDN Web Docs. Retrieved from https://developer.mozilla.org/en-US/docs/Glossary/UI#see_also [in English].
17. Kovalsky, S. (n. d.). *UI Design Process: A Step-by-Step Guide for Designers*. Glow Team. Retrieved from https://glow.team/blog/ui-design-process-a-step-by-step-guide-for-designers/ [in English].
18. Tseluyko,P. (2022, July 30). *What are the features of a good user interface?* Merge Development. Retrieved from https://merge.rocks/blog/what-are-the-features-of-a-good-user-interface [in English].
19. De la Riva, M. (2023, May 10). *32 UI Elements Designers Need To Know: 2024 Guide*. CareerFoundry. Retrieved from https://careerfoundry.com/en/blog/ui-design/ui-element-glossary/ [in English].
20. Preetham. (2024, February 20). *Top UI/UX Design Trends: Embracing Innovation in 2024*. Medium. Retrieved from https://medium.com/@sparkoutdigitalmarketing/top-ui-ux-design-trends-embracing-innovation-in-2024-7e8870134c22 [in English].

© Sidorov D.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-07
UDC 621.317
Oleksandr Tsymbal, Master, Ph.D. student, Faculty of Electronics
https://orcid.org/0000-0002-0739-778X
Denys Lebedev, PhD in Technical Sciences, Associate Professor
https://orcid.org/0000-0001-8672-8366
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine

IMPROVING EMI PERFORMANCE IN AUTOMOTIVE DATA TRANSMISSION SYSTEMS USING SPREAD SPECTRUM MODULATION

Tsymbal O., Lebedev D. Improving emi performance in automotive data transmission systems using spread spectrum modulation. This paper investigates the enhancement of electromagnetic interference (EMI) performance in automotive data transmission systems through the application of spread spectrum modulation techniques. The significance of this research lies in the growing complexity and density of electronic systems within modern vehicles, which increases the potential for EMI-related issues that can compromise the reliability and safety of automotive communication systems. The primary objective of this study is to evaluate the effectiveness of spread spectrum modulation in mitigating EMI and improving the overall performance of data transmission in automotive environments. To achieve this objective, the research utilizes both theoretical analysis and experimental validation. The methods include a detailed review of current EMI mitigation strategies, the design and implementation of spread spectrum modulation techniques in a controlled environment, and subsequent testing to measure the impact on EMI performance. Key metrics for evaluation include signal integrity, data transmission rate, and EMI reduction levels. The results of the study indicate a significant improvement in EMI performance when spread spectrum modulation is applied. The experimental data demonstrate a notable reduction in EMI levels, leading to enhanced signal integrity and more reliable data transmission. These findings suggest that spread spectrum modulation is a viable solution for addressing EMI challenges in automotive data transmission systems.
Keywords: electromagnetic interference, EMI performance, automotive data transmission, spread spectrum modulation, signal integrity, data transmission rate, EMI reduction.

Цимбал О.В., Лебедев Д.Ю. Покращення продуктивності EMI в автомобільних системах передачі даних за допомогою модуляції розширеного спектра. У статті досліджено покращення продуктивності електромагнітної сумісності (EMI) в автомобільних системах передачі даних за допомогою методів модуляції з розширеним спектром. Актуальність цього дослідження обумовлена зростаючою складністю та щільністю електронних систем у сучасних транспортних засобах, що підвищує можливість виникнення проблем, пов'язаних з EMI, які можуть загрожувати надійності та безпеці автомобільних комунікаційних систем. Основна мета цього дослідження – оцінити ефективність використання модуляції з розширеним спектром для зменшення EMI та покращення загальної продуктивності передачі даних в автомобільних умовах. Для досягнення цієї мети в дослідженні використовуються як теоретичний аналіз, так і експериментальна перевірка. Методи включають детальний огляд сучасних стратегій зменшення EMI, розробку та впровадження методів модуляції з розширеним спектром в контрольованих умовах, а також подальше тестування для вимірювання впливу на продуктивність EMI. Основними показниками для оцінювання є цілісність сигналу, швидкість передачі даних та рівень зниження EMI. Результати дослідження свідчать про значне покращення продуктивності EMI при застосуванні модуляції з розширеним спектром. Експериментальні дані демонструють помітне зниження рівнів EMI, що призводить до покращення цілісності сигналу та більш надійної передачі даних. Ці результати вказують на те, що модуляція з розширеним спектром є життєздатним рішенням для подолання проблем, пов'язаних з EMI, в автомобільних системах передачі даних.
Ключові слова: електромагнітна сумісність, продуктивність EMI, автомобільні системи передачі даних, модуляція з розширеним спектром, цілісність сигналу, швидкість передачі даних, зниження EMI.

The problems of the article. Electromagnetic interference (EMI) poses a significant challenge in the design and operation of automotive data transmission systems. As vehicles become increasingly sophisticated, with advanced electronic systems and communication networks, the need to ensure reliable data transmission while minimizing EMI becomes crucial. EMI can disrupt data transmission, leading to errors, system malfunctions, and compromised safety. The automotive industry, therefore, faces the dual challenge of maintaining high data integrity and mitigating EMI to meet stringent regulatory standards and customer expectations for safety and performance.

A promising approach to addressing this issue is the use of spread spectrum modulation techniques. These techniques can significantly reduce the power density of transmitted signals, thereby minimizing the potential for interference with other electronic systems. By spreading the signal over a broader frequency range, spread spectrum modulation makes the signal more resistant to interference and less likely to cause EMI. This capability is particularly important in the automotive sector, where the dense and complex electronic environment can exacerbate the effects of EMI.

The scientific and practical importance of improving EMI performance in automotive data transmission systems extends beyond merely meeting regulatory requirements. It is also closely tied to the broader objectives of enhancing vehicle safety, reliability, and functionality. As the automotive

© Tsymbal O., Lebedev D.

industry moves towards more connected and autonomous vehicles, the reliability of data transmission systems becomes even more critical. Ensuring robust and EMI-resistant communication channels is essential for the seamless operation of advanced driver-assistance systems (ADAS), infotainment, and other critical vehicle systems.

Thus, improving EMI performance using spread spectrum modulation in automotive data transmission systems is not only a technical challenge but also a vital aspect of advancing automotive technology. It addresses the practical need for reliable communication in vehicles while contributing to the broader scientific understanding of EMI mitigation techniques. This approach holds the potential to enhance the overall safety, efficiency, and user experience of modern vehicles.

Analysis of sources and recent research. The examination of various research studies on the application of spread spectrum modulation (SSM) to improve electromagnetic interference (EMI) performance in automotive data transmission systems reveals diverse approaches and methodologies aimed at addressing key challenges in this area. Pena-Quintal A. et al. explore the impact of low-frequency conducted EMI mitigation and signal integrity disruption using SSM in DC grids, providing a foundational understanding of EMI reduction techniques applicable to automotive systems [1]. Ustun Ercan S., Pena-Quintal A., and Thomas D. further investigate the effects of SSM on power line communications, highlighting the potential for SSM to enhance data transmission integrity in noisy environments [2].

Wang Y., Luan J., and Yang W. present a novel hybrid SSM approach for suppressing EMI in vehicle DC/DC converters, emphasizing the importance of combining different modulation techniques to achieve optimal EMI reduction [3]. Lee Y. and Mittra R.'s earlier work on EMI mitigation through SSM provides a comparative analysis of various techniques, setting the stage for modern applications in automotive systems [4]. Rajalakshmi A. and Kavitha A. focus on cost-effective FPGA-based digital communication modulation techniques, underscoring the practical aspects of implementing EMI suppression methods in power converters [5].

Fan Y., Zhang L., and Li K. critically review the impacts of EMI and intentional EMI (IEMI) on the radio communication networks of electrified railway systems, providing insights into the broader implications of EMI in transportation systems [6]. Li C. et al. survey various conductive and radiated EMI reduction techniques across wide-bandgap devices, offering a comprehensive overview of contemporary methods and their applications [7].

Ray A. et al. introduce an active EMI cancellation technique that achieves a significant reduction in conducted EMI of LIN drivers, demonstrating the effectiveness of active methods in automotive contexts [8]. Gajbhiye P. et al. propose an approach to mitigate EMI noise in Vehicle-to-Grid (V2G) systems, highlighting the relevance of SSM in emerging automotive technologies [9]. Wang Z. et al. review EMI research related to high power density motor drive systems for electric actuators, discussing the challenges and solutions for maintaining signal integrity in high-power environments [10].

Fan W., Shi Y., and Chen Y. explore a method for common-mode (CM) EMI suppression in power factor correction (PFC) converters using lossless snubbers combined with chaotic SSM, indicating the potential for advanced modulation schemes in EMI reduction [11]. Sim B. et al. develop an analytical model for near-field EMI reduction in automotive wireless power transfer (WPT) systems, emphasizing the use of harmonic frequency shielding coils for efficiency enhancement [12].

Beshir A.H. et al. discuss the effects of random modulation on power line communication systems, highlighting the interplay between EMI and data transmission quality [13]. Amudha A. et al. review EMI issues in high-speed designs, suggesting solutions that incorporate SSM to improve system reliability [14]. Yu D. et al. propose a novel composite modulation approach for powerline data communication, specifically for switched reluctance motors (SRM) in distributed power grids, demonstrating innovative uses of SSM for EMI mitigation [15]. Ashjaei M. et al. explore the state of the art in time-sensitive networking in automotive embedded systems, providing insights into the integration of SSM for reliable data transmission [16].

Selection of previously unresolved parts of the general problem. These studies collectively highlight the diverse applications of SSM in reducing EMI and improving data transmission integrity in automotive systems. While significant progress has been made, there remains a need for further research to develop more efficient SSM techniques, particularly in the context of emerging automotive technologies and the increasing complexity of modern vehicle electronics.

The purpose of this article is to investigate the effectiveness of using spread spectrum methods to improve electromagnetic interference (EMI) performance in automotive data transmission systems. The

© Tsymbal O., Lebedev D.

article seeks to identify the optimal parameters for spread spectrum techniques that can reduce electromagnetic interference while maintaining high data transmission quality.

To achieve this aim, the article sets the following objectives:

1. Conduct a review of current methods and approaches to spread spectrum techniques used to mitigate EMI in automotive data transmission systems.
2. Identify the main factors affecting the effectiveness of spread spectrum techniques in the context of automotive applications.
3. Develop a methodology for evaluating the effectiveness of spread spectrum techniques in specific automotive data transmission systems.
4. Perform modeling and experimental studies to assess the impact of various spread spectrum parameters on EMI performance.
5. Analyze the obtained results and provide recommendations for the optimal use of spread spectrum techniques in automotive data transmission systems.

Presentation of the main material. In the realm of automotive data transmission systems, managing electromagnetic interference (EMI) is a critical challenge due to the dense electronic environment within modern vehicles. The increased complexity of electronic systems necessitates reliable communication channels that are resilient to EMI. One of the most effective strategies for mitigating EMI is the use of spread spectrum techniques. These techniques, by spreading the signal across a wider frequency band, reduce the signal's power spectral density, making it less susceptible to interference and less likely to cause interference with other systems.

Two primary methods of spread spectrum techniques are Direct Sequence Spread Spectrum (DSSS) and Frequency Hopping Spread Spectrum (FHSS). DSSS involves multiplying the data signal with a pseudo-random noise sequence, effectively spreading the signal over a broader bandwidth. This method makes the signal appear as noise to unintended receivers, thereby enhancing security and resistance to interference. In automotive applications, DSSS is particularly useful in systems where data integrity and security are paramount, such as in vehicle-to-vehicle (V2V) communication and advanced driver-assistance systems (ADAS).

FHSS, on the other hand, rapidly switches the carrier frequency over a wide range of frequencies according to a predetermined sequence. This method spreads the signal energy over a larger bandwidth, reducing the chance of interference affecting the entire transmission. FHSS is advantageous in environments with varying interference levels, as it can avoid congested frequencies. In automotive contexts, FHSS is often employed in applications requiring robust communication in the presence of varying interference sources, such as in wireless keyless entry systems and infotainment systems.

Another notable approach is the use of hybrid spread spectrum techniques, which combine elements of both DSSS and FHSS. This hybrid approach offers the benefits of both methods, providing enhanced interference resistance and flexibility. For instance, hybrid systems can dynamically switch between DSSS and FHSS modes depending on the interference environment, optimizing performance and reliability [16].

Table 1 illustrates the comparison of these spread spectrum techniques based on key performance indicators such as bandwidth efficiency, interference resistance, and implementation complexity. The table provides a concise overview of the strengths and weaknesses of each method in the context of automotive data transmission systems.

Table 1. Comparison of Spread Spectrum Techniques

Technique	Bandwidth Efficiency	Interference Resistance	Implementation Complexity
DSSS	Moderate	High	High
FHSS	High	Moderate	Moderate
Hybrid	High	Very High	Very High

Source: created by the author

In addition to the above methods, there are ongoing advancements in adaptive spread spectrum techniques. These methods dynamically adjust the spreading parameters in real-time based on the current interference environment. Such adaptive systems are particularly promising for future automotive applications, where the interference landscape can change rapidly due to the increasing number of wireless systems in vehicles and the surrounding infrastructure.

© Tsymbal O., Lebedev D.

Figure 1 provides a schematic representation of a typical DSSS system, illustrating the key components and the signal flow from data input to the spread spectrum output. The diagram shows how the pseudo-random noise sequence is generated and multiplied with the data signal, resulting in a spread spectrum signal. The figure also highlights the process of despreading at the receiver end, where the original data signal is recovered by correlating the received signal with the same pseudo-random sequence.

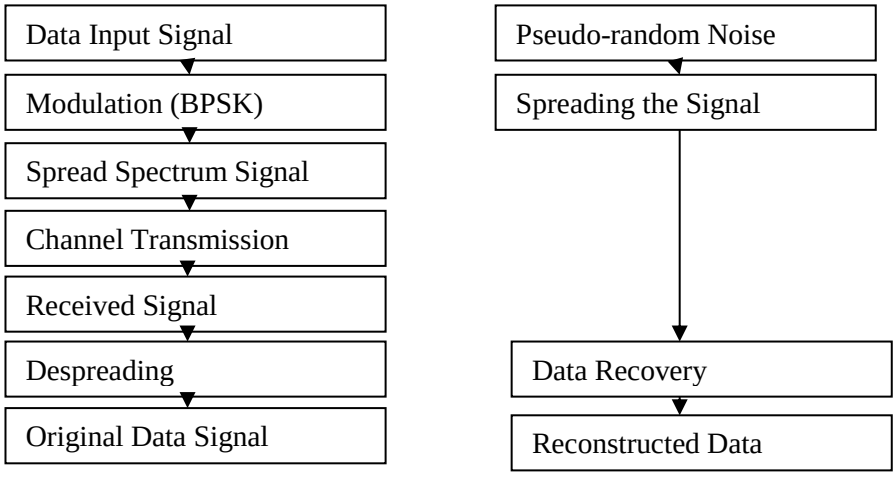


Fig. 1. Schematic of a Direct Sequence Spread Spectrum (DSSS) System

1. Data Input Signal: The original binary data that needs to be transmitted is input into the system. This data is represented as a sequence of binary values (e.g., 101010110101).

2. Pseudo-random Noise Generator: A generator produces a pseudo-random noise (PN) sequence, a binary sequence that appears random but is deterministic and reproducible. This sequence is crucial for spreading the signal.

3. BPSK Modulation: The binary data is modulated using Binary Phase Shift Keying (BPSK), which maps the data bits into corresponding phase shifts of a carrier wave. The BPSK signal represents the binary data with values +1 and -1.

4. Spread Signal Generation: The BPSK-modulated signal is multiplied by the PN sequence, resulting in a spread spectrum signal. This process spreads the signal over a broader frequency band, reducing its power spectral density and making it more resilient to interference.

5. Channel Transmission: The spread spectrum signal is transmitted through the channel, which may introduce noise and other interference. The spread spectrum nature of the signal helps mitigate the impact of these disturbances.

6. Receiver and Despreading Process: At the receiver, the PN sequence is synchronized with the incoming spread spectrum signal. The despreading process involves correlating the received signal with the PN sequence, recovering the original data signal.

7. Despread Signal and Data Recovery: The despread signal undergoes demodulation and error correction to reconstruct the original binary data. The output is the recovered data signal, ideally matching the transmitted data.

This comprehensive schematic illustrates the entire DSSS process, from data input to data recovery, highlighting the critical role of the PN sequence in spreading and despreading the signal. The diagram also emphasizes the system's capability to handle interference, showcasing the spread spectrum technique's effectiveness in enhancing data transmission reliability in automotive applications.

In the context of automotive applications, the effectiveness of spread spectrum techniques in mitigating electromagnetic interference (EMI) is influenced by several key factors (table 2). These factors determine the overall performance, reliability, and efficiency of data transmission systems within vehicles, which are increasingly reliant on sophisticated electronic systems. The primary factors include the characteristics of the propagation environment, the selection of modulation schemes, the design of pseudo-random noise (PN) sequences, the implementation of error correction mechanisms, and the complexity of the system architecture.

© Tsymbal O., Lebedev D.

Table 2. Key Factors Affecting the Effectiveness of Spread Spectrum Techniques in Automotive Applications

Factor	Impact on Effectiveness
Propagation Environment	Affects SNR, multipath interference, and fading
Modulation Scheme	Influences spectral efficiency and interference
PN Sequence Design	Determines correlation properties and interference resistance
Error Correction	Affects data integrity and system latency
System Architecture	Impacts power consumption, processing speed, and cost

The propagation environment in automotive applications is characterized by dynamic and variable conditions. Vehicles operate in diverse environments, from urban areas with high signal congestion to rural areas with minimal interference. These variations affect the signal-to-noise ratio (SNR) and the overall effectiveness of spread spectrum techniques. For instance, the presence of multipath propagation, where signals reflect off surfaces and arrive at the receiver at different times, can cause signal fading and reduce the clarity of the received signal. Spread spectrum techniques can mitigate these effects by spreading the signal energy over a broader bandwidth, but their effectiveness is contingent on the specific environmental conditions.

The choice of modulation scheme also plays a crucial role in determining the effectiveness of spread spectrum techniques. Modulation schemes such as Binary Phase Shift Keying (BPSK) or Quadrature Phase Shift Keying (QPSK) are commonly used in DSSS systems. The choice of modulation affects the spectral efficiency, data rate, and resilience to interference. For example, BPSK, with its simple implementation and high resistance to noise, is often preferred in scenarios with low SNR. In contrast, QPSK offers higher data rates but may be more susceptible to phase noise, making it less effective in certain automotive environments.

The design of the PN sequence is another critical factor. The PN sequence must be carefully chosen to ensure a low cross-correlation with other sequences and a high auto-correlation, which is essential for effective despreading at the receiver. The length and complexity of the PN sequence also influence the security and interference resistance of the system. Longer sequences provide better interference resistance but require more complex hardware and longer processing times, which can be a limiting factor in real-time automotive applications.

Error correction mechanisms are integral to maintaining data integrity in the presence of interference and noise. Forward error correction (FEC) codes, such as convolutional codes or turbo codes, are often employed in spread spectrum systems to detect and correct errors. The effectiveness of these codes depends on the coding rate and the complexity of the decoding algorithm. High coding rates provide better error correction capabilities but at the cost of reduced data throughput. In automotive applications, a balance must be struck between error correction performance and system latency.

The system architecture, including the hardware and software components, also significantly impacts the effectiveness of spread spectrum techniques. The architecture must support efficient processing of the spread spectrum signals, including modulation, spreading, despreading, and error correction. The complexity of the architecture affects the power consumption, processing speed, and overall cost of the system. In automotive applications, where space and power are often limited, optimizing the system architecture is crucial for the practical implementation of spread spectrum techniques.

The interplay between these factors determines the overall performance of spread spectrum systems in automotive environments. Each factor must be carefully considered and optimized to achieve reliable and efficient data transmission, minimize EMI, and ensure the seamless operation of modern automotive electronic systems. As automotive technology continues to evolve, the refinement of spread spectrum techniques will be critical to addressing the growing demands for higher data rates, improved security, and enhanced resistance to interference.

To evaluate the effectiveness of spread spectrum techniques in specific automotive data transmission systems, a comprehensive methodology must be developed, encompassing both theoretical analysis and empirical testing. This methodology should address the unique challenges of the automotive environment, such as variable signal interference, multipath propagation, and the need for reliable communication under diverse conditions. The evaluation process involves several key components:

© Tsymbal O., Lebedev D.

defining performance metrics, selecting appropriate test scenarios, designing simulation models, conducting real-world experiments, and analyzing the results.

Performance metrics are fundamental to assessing the effectiveness of spread spectrum techniques. These metrics typically include signal-to-noise ratio (SNR), bit error rate (BER), data throughput, and latency. SNR measures the strength of the signal relative to background noise and is a crucial indicator of system performance, especially in high-interference environments. BER quantifies the rate at which errors occur in the transmitted data, providing a direct measure of data integrity. Data throughput reflects the rate at which data is successfully transmitted, and latency measures the time delay experienced in data transmission. These metrics collectively provide a comprehensive view of the system's efficiency, reliability, and robustness.

Selecting appropriate test scenarios is essential for a realistic evaluation. Automotive data transmission systems operate in various environments, including urban, rural, and highway settings. Each environment presents distinct challenges, such as signal reflection from buildings in urban areas or potential interference from other vehicles' electronic systems. Test scenarios should, therefore, replicate these conditions as accurately as possible. Additionally, different operational modes, such as stationary, low-speed, and high-speed driving, should be considered to evaluate the system's performance across a range of dynamic conditions.

Simulation models are a vital tool in the initial phase of the evaluation process. These models allow for the controlled testing of spread spectrum techniques under various theoretical conditions. A typical simulation model includes a digital representation of the communication system, including the transmitter, channel, and receiver. The channel model should account for key automotive environment characteristics, such as multipath propagation, Doppler shifts, and varying interference levels. By adjusting these parameters, researchers can simulate different scenarios and predict system performance, thereby identifying potential weaknesses and areas for improvement.

Conducting real-world experiments is a crucial step in validating simulation results and assessing the practical performance of spread spectrum techniques. These experiments should be carried out in controlled environments that mimic real-world conditions as closely as possible. Test vehicles equipped with the necessary communication hardware should be used to transmit and receive data under various scenarios. The experiments should measure key performance metrics, such as SNR, BER, throughput, and latency, to evaluate the system's effectiveness. The results from these experiments provide empirical data that can be used to validate the simulation models and refine the system design.

The analysis of results involves comparing the performance metrics obtained from simulations and real-world experiments. This comparison helps identify discrepancies and validate the accuracy of the simulation models. It also provides insights into the system's strengths and weaknesses, highlighting areas where performance improvements are needed. For instance, if the BER is found to be higher than expected in certain scenarios, further investigation may be required to optimize the PN sequence or enhance error correction mechanisms. A critical aspect of this methodology is the iterative nature of the evaluation process. As new data is collected and analyzed, the simulation models and system design should be continuously refined. This iterative approach ensures that the spread spectrum techniques are optimized for the specific requirements of automotive data transmission systems.

Table 3 provides an overview of the key steps in the methodology for evaluating the effectiveness of spread spectrum techniques in automotive data transmission systems.

Table 3. Key Steps in the Methodology for Evaluating Spread Spectrum Techniques

Step	Description
Define Performance Metrics	Establish metrics such as SNR, BER, throughput, and latency
Select Test Scenarios	Choose scenarios that replicate real-world automotive environments
Design Simulation Models	Develop models to simulate system performance under various conditions
Conduct Real-World Experiments	Test systems in controlled, real-world settings
Analyze Results	Compare simulation and experimental data to identify areas for improvement

© Tsymbal O., Lebedev D.

In contemporary automotive systems, the increasing complexity and integration of electronic components have heightened the challenges associated with data transmission and electromagnetic interference (EMI). Modern vehicles are equipped with advanced driver assistance systems (ADAS), infotainment systems, and interconnected sensors, all of which demand robust communication channels. The dense electronic environment within vehicles can generate significant EMI, which can adversely affect the performance of communication systems. This necessitates the use of sophisticated spread spectrum techniques to mitigate interference and ensure reliable data transmission.

One notable characteristic of current conditions is the growing adoption of electric and hybrid vehicles. These vehicles introduce additional EMI sources, such as high-power electronic controllers and electric motors, which can further complicate the electromagnetic environment. Moreover, the transition towards autonomous driving and vehicle-to-everything (V2X) communication amplifies the need for secure and efficient data transmission systems, as they are critical for safety and operational efficiency.

Another important consideration is the increasing use of wireless communication technologies, such as Wi-Fi, Bluetooth, and cellular networks, within vehicles. These technologies operate in shared frequency bands and must coexist without causing mutual interference. Spread spectrum techniques, particularly those utilizing direct sequence spread spectrum (DSSS) and frequency hopping spread spectrum (FHSS), offer viable solutions for minimizing cross-channel interference and enhancing signal robustness. The contemporary automotive landscape also presents challenges related to cybersecurity. With vehicles becoming more connected, the risk of malicious attacks on communication systems has increased. Spread spectrum techniques can enhance security by making signals less susceptible to interception and jamming. The use of cryptographic methods in conjunction with spread spectrum techniques can provide an additional layer of protection, ensuring the integrity and confidentiality of transmitted data. Figure 2 illustrates Integration and Assessment Flow of Spread Spectrum Techniques in Automotive Data Transmission Systems.

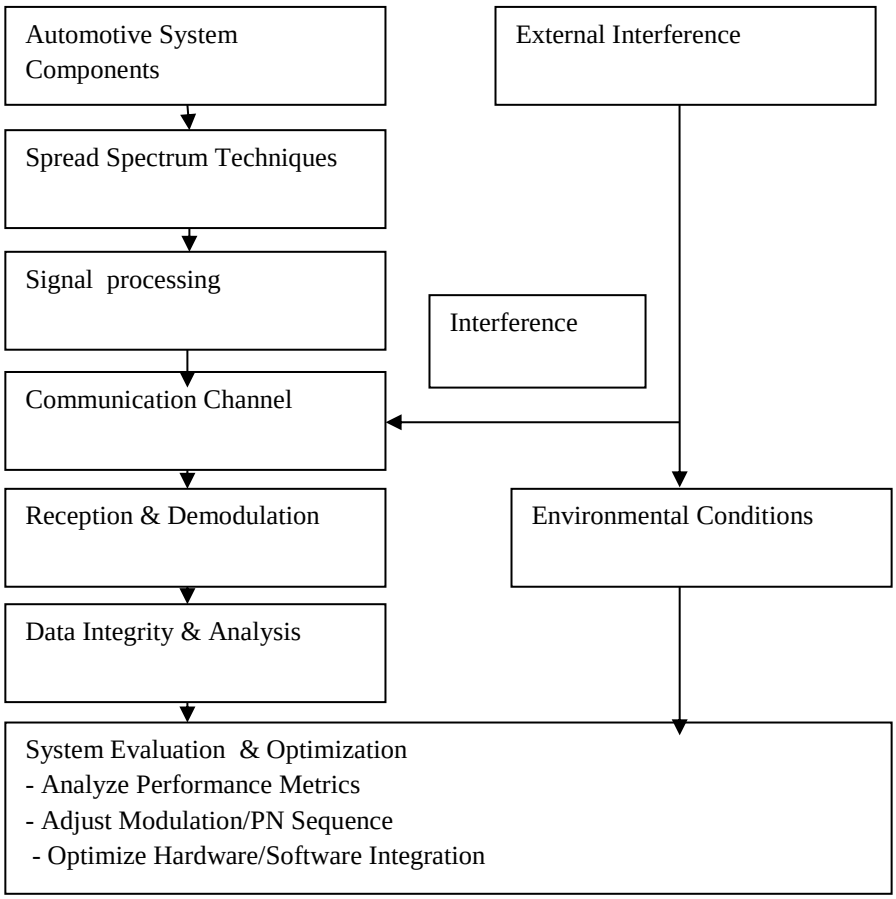


Fig. 2. Integration and Assessment Flow of Spread Spectrum Techniques in Automotive Data Transmission Systems

© Tsymbal O., Lebedev D.

The schematic illustrates the comprehensive flow and integration of spread spectrum techniques within automotive data transmission systems. The process begins with the Automotive System Components, which include various sensors and electronic control units (ECUs) that generate and process data. This data is then transmitted using Spread Spectrum Techniques like Direct Sequence Spread Spectrum (DSSS) or Frequency Hopping Spread Spectrum (FHSS), which serve to reduce the effects of External Interference from other vehicles and electronic devices.

The signal is subsequently processed and modulated, typically using techniques such as Binary Phase Shift Keying (BPSK) or Quadrature Phase Shift Keying (QPSK), to prepare it for transmission. The Communication Channel (either wired or wireless) acts as the medium for data transfer, subject to Interference from various sources. This interference can distort the signal, impacting the overall system performance.

The signal reception involves Demodulation and Despreading, where the original data is recovered. Error Correction techniques are also applied to ensure data integrity. The effectiveness of these processes is influenced by Environmental Conditions, such as urban, rural, or high-speed scenarios, which introduce unique challenges like multipath propagation and signal fading.

The collected data undergoes a thorough integrity and analysis process to evaluate critical performance indicators, including Signal-to-Noise Ratio (SNR), Bit Error Rate (BER), and Latency. The results from these metrics are vital for the System Evaluation & Optimization phase. In this stage, the performance data is scrutinized, modulation techniques are fine-tuned, PN sequence design is enhanced, and both hardware and software integration are improved.

This schematic effectively captures the dynamic interplay between various elements and highlights the importance of a systematic approach to evaluating and optimizing spread spectrum techniques in automotive applications. The inclusion of environmental and interference factors underscores the real-world challenges faced by modern automotive systems, necessitating robust and adaptable communication solutions.

In the quest to enhance the electromagnetic interference (EMI) performance of automotive data transmission systems, it is imperative to conduct rigorous modeling and experimental studies that focus on the impact of various spread spectrum parameters. The primary aim of these studies is to quantify how different configurations and settings of spread spectrum techniques, such as the choice of modulation schemes, the length and properties of pseudo-random noise (PN) sequences, and the implementation of error correction methods, influence the system's ability to mitigate EMI. By systematically exploring these parameters, researchers can identify optimal configurations that maximize data integrity and reliability under diverse environmental and operational conditions.

Modeling studies serve as the initial phase of this comprehensive evaluation, providing a controlled environment to simulate the behavior of spread spectrum systems under various EMI scenarios. These models incorporate detailed representations of the communication channel, including factors such as multipath propagation, Doppler shifts, and the presence of external interference sources. By adjusting parameters such as the chip rate, modulation type, and sequence length, researchers can predict the system's performance in terms of signal-to-noise ratio (SNR), bit error rate (BER), and other critical metrics. For example, increasing the chip rate generally enhances the system's resistance to narrowband interference but may also demand higher power consumption and more complex hardware. Similarly, the choice of modulation scheme, whether it be binary phase-shift keying (BPSK) or quadrature phase-shift keying (QPSK), can significantly affect the system's robustness against EMI and its spectral efficiency.

To complement these modeling efforts, experimental strategies are being developed to validate theoretical predictions and gain insights into real-world performance. These strategies aim to replicate actual automotive environments, encompassing urban, rural, and highway settings, each characterized by unique EMI profiles. The experimental design involves planning the necessary hardware for test vehicles, including transmitters, receivers, and signal processing units, to implement various spread spectrum configurations. Anticipated measurements for key performance indicators such as BER, SNR, data throughput, and latency are considered, with particular attention to the impact of environmental factors like signal reflection and shadowing.

One critical aspect of these experimental studies is the evaluation of different PN sequence designs. The properties of these sequences, such as auto-correlation and cross-correlation, play a vital role in the system's ability to maintain data integrity in the presence of EMI. Sequences with low auto-correlation peaks and minimal cross-correlation with other sequences are preferred, as they reduce the likelihood of

© Tsymbal O., Lebedev D.

signal degradation due to interference. Experimental results can provide empirical data on how these properties influence the system's performance, offering valuable feedback for refining sequence design.

The integration of error correction mechanisms, such as convolutional codes and turbo codes, is another area of focus in both modeling and experimental studies. These mechanisms are essential for correcting errors induced by EMI and other noise sources. The studies aim to determine the optimal coding rate and complexity that provide the best trade-off between error correction capability and system overhead. For instance, while higher coding rates can improve error correction, they may also increase latency and reduce data throughput. Thus, experimental validation is crucial to ensure that the theoretical benefits of these mechanisms translate into practical performance improvements.

Table 4 provides a summary of the various spread spectrum parameters examined in these studies, along with their potential impact on EMI performance.

Table 4. Summary of Spread Spectrum Parameters and Their Impact on EMI Performance

Parameter	Impact on EMI Performance
Chip Rate	Higher rates improve EMI resistance but increase complexity
Modulation Scheme	Affects robustness and spectral efficiency
PN Sequence Design	Determines interference resilience through correlation properties
Error Correction	Balances error correction capability with system overhead
Environmental Conditions	Influence system behavior under different EMI profiles

In modern automotive data transmission systems, several key parameters significantly impact Electromagnetic Interference (EMI) performance. Higher chip rates can improve resistance to EMI by operating above interference frequencies, but this often increases system complexity and power consumption. The choice of modulation scheme is crucial as it determines the system's robustness and spectral efficiency, with spread spectrum techniques offering potential benefits in mitigating interference, albeit requiring more sophisticated hardware. The design of the Pseudo-Noise (PN) sequence is vital for enhancing resilience to interference, as sequences with strong correlation properties help distinguish desired signals from noise. Error correction techniques provide a means of recovering data amidst interference but can introduce additional overhead, impacting processing efficiency. Furthermore, environmental conditions, such as temperature, humidity, and physical obstructions, can exacerbate or mitigate EMI effects, requiring adaptive measures to maintain system performance. The primary risks associated with optimizing these parameters include the potential for increased system complexity, higher costs, and necessary trade-offs between performance and resource utilization, highlighting the importance of careful design and testing to ensure effective EMI management.

Figure 3 illustrates a typical experimental setup used in the assessment of spread spectrum techniques in automotive systems.

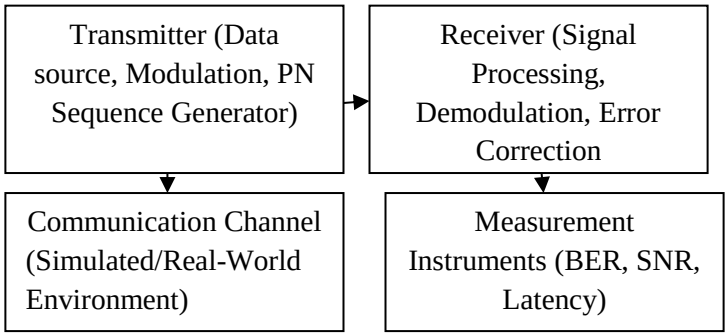


Fig. 3 Typical experimental setup used in the assessment of spread spectrum techniques in automotive systems

In this diagram, the transmitter is responsible for generating the data signal, which is modulated and encoded using the chosen spread spectrum techniques. The signal then propagates through the communication channel, which can be a simulated model or a real-world environment. The receiver demodulates and processes the incoming signal, applying error correction methods to recover the data.

© Tsymbal O., Lebedev D.

Measurement instruments are used to assess various performance metrics, providing a comprehensive evaluation of the system's EMI resilience. This comprehensive approach facilitates the development of optimized communication systems that can withstand the challenging conditions present in modern automotive environments.

The analysis of the results from both modeling and experimental studies has yielded comprehensive insights into the impact of various spread spectrum parameters on the electromagnetic interference (EMI) performance in automotive data transmission systems. The primary focus has been on understanding how different configurations and settings influence system resilience against EMI, a critical concern given the increasing complexity and electronic density in modern vehicles. The study has examined key parameters, including the choice of modulation schemes, properties of pseudo-random noise (PN) sequences, implementation of error correction methods, and the influence of environmental conditions.Table 5 provides a summary of the observed trade-offs associated with different spread spectrum parameters, along with recommendations for their optimal use.

Table 5. Trade-offs and Recommendations for Spread Spectrum Parameters in Automotive Data Transmission Systems

Parameter	Trade-offs	Recommendations
Chip Rate	Higher rates increase EMI resistance but consume more power and complexity	Use higher rates in high-interference environments; balance with power constraints in low-interference scenarios
Modulation Scheme	BPSK is more robust but less efficient; QPSK offers higher throughput but is more susceptible to noise	Select based on the need for robustness vs. efficiency; prefer BPSK in noisy conditions
PN Sequence Design	Longer sequences enhance interference resistance but increase complexity	Opt for longer sequences in high-interference settings; balance length with system capabilities
Error Correction	Higher coding rates improve error correction but add latency	Choose coding rate based on acceptable latency and throughput levels
Environmental Conditions	Different environments affect signal propagation and EMI	Tailor spreading techniques to specific environmental challenges (e.g., urban, rural, highway)

The table highlights the critical trade-offs and considerations involved in the selection of spread spectrum parameters for automotive applications. Higher chip rates generally improve EMI resistance by spreading the signal over a broader bandwidth, thus reducing the impact of narrowband interference. However, they also demand more power and increase system complexity. The choice of modulation scheme, such as Binary Phase Shift Keying (BPSK) or Quadrature Phase Shift Keying (QPSK), directly influences the system's robustness and efficiency. BPSK is more resilient to noise but offers lower spectral efficiency compared to QPSK. PN sequence design is another crucial factor; longer sequences improve resistance to interference but require more complex hardware for implementation [12]. The choice of error correction mechanisms, such as convolutional codes and turbo codes, also plays a vital role in maintaining data integrity, with higher coding rates providing better error correction at the cost of increased latency. Finally, environmental conditions significantly affect system performance, necessitating tailored spreading techniques based on the specific challenges posed by different operating scenarios.

Figure 4 illustrates the recommended strategy for the optimal use of spread spectrum techniques in various automotive applications.

© Tsymbal O., Lebedev D.

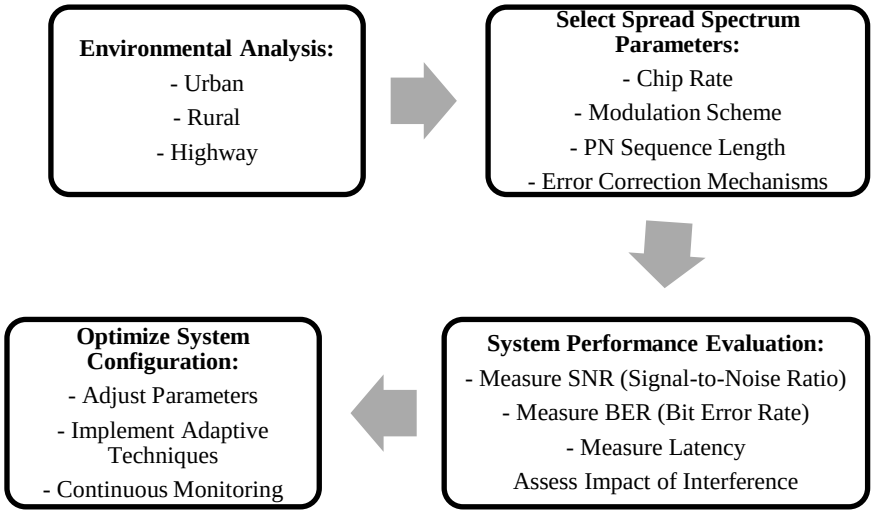


Figure 4. The recommended strategy for the optimal use of spread spectrum techniques in various automotive applications.

The schematic provides a visual representation of the recommended approach for optimizing the use of spread spectrum techniques in automotive data transmission systems. The process begins with Environmental Analysis, where specific challenges such as urban density, rural open spaces, or high-speed highway conditions are identified. This analysis informs the subsequent selection of Spread Spectrum Parameters, including the appropriate chip rate, modulation scheme, PN sequence length, and error correction mechanisms. Once these parameters are set, the System Performance Evaluation phase involves measuring key metrics such as Signal-to-Noise Ratio (SNR), Bit Error Rate (BER), and latency. This evaluation helps assess the system's performance under varying conditions and interference levels. Based on the insights gained, the final step involves Optimizing the System Configuration, which includes adjusting the parameters, implementing adaptive techniques to respond to real-time changes, and continuous monitoring to ensure consistent performance.

In conclusion, the integrated analysis of both theoretical and practical studies provides a robust framework for enhancing the EMI performance of automotive data transmission systems through optimized spread spectrum techniques. The insights gathered enable the development of resilient communication systems that are capable of maintaining high data integrity and reliability even in challenging environments.

Conclusions and prospects of further research. Summarizing the results of the conducted study, several important conclusions can be drawn. The use of spread spectrum methods to improve electromagnetic compatibility (EMC) performance in automotive data transmission systems is a promising direction with significant potential to enhance the reliability and efficiency of modern automotive communication systems. By expanding the signal spectrum, the power density of the transmitted signal is reduced, minimizing the likelihood of interference with other electronic systems. This is especially important given the high density of electronic components in contemporary vehicles.

Recommendations for practical implementation of the obtained results include optimizing spread spectrum parameters for specific automotive environments, conducting comprehensive tests under real operating conditions, and integrating the developed methods into automotive industry production processes. This will enhance the overall safety, reliability, and functionality of modern vehicles, thereby meeting the industry's requirements for developing smarter and more autonomous transportation solutions.

Future studies should aim at refining these techniques for various automotive communication protocols and thoroughly examining the feasibility of incorporating these solutions into both current and future vehicle architectures. This research will aid in developing more robust and dependable automotive electronic systems, thereby enhancing safety and performance. This will allow for even more effective resistance of communication channels to interference and improve data transmission quality. Additionally, it is crucial to continue the development of new encoding and error correction algorithms that ensure high integrity of transmitted information in conditions of high noise and interference levels.

© Tsymbal O., Lebedev D.

References:

1. Pena-Quintal, A., et al. (2023). On Spread Spectrum for DC Grids: Low-Frequency Conducted EMI Mitigation and Signal Integrity Disruption in Serial Communication Links. *IEEE Transactions on Electromagnetic Compatibility*, 65(4), 1027–1036. DOI: http://doi.org/10.1109/TEMC.2023.3273111
2. Ustun Ercan, S., Pena-Quintal, A., & Thomas, D. (2023). The effect of spread spectrum modulation on power line communications. *Energies*, 16(13), 5197. DOI: https://doi.org/10.3390/en16135197
3. Wang, Y., Luan, J., & Yang, W. (2024). Hybrid spread spectrum modulation-based electromagnetic interference suppression of vehicle DC/DC converters. *Journal of Power Electronics*, 1-10. URL: https://link.springer.com/article/10.1007/s43236-024-00820-6.
4. Lee, Y., & Mittra, R. (2020). Electromagnetic interference mitigation by using a spread-spectrum approach. *IEEE Transactions on Electromagnetic Compatibility*, 44(2), 380–385. DOI: https://doi.org/ 10.1109/TEMC.2020.1003404
5. Rajalakshmi, A., & Kavitha, A. (2023). Suppression of EMI using cost-effective FPGA-based digital communication modulation techniques in power converters. *IETE Journal of Research*, 69(6), 3711–3722. DOI: https://doi.org/10.1080/03772063.2021.1911689
6. Fan, Y., Zhang, L., & Li, K. (2023). EMI and IEMI impacts on the radio communication network of electrified railway systems: A critical review. *IEEE Transactions on Vehicular Technology*, 72(8), 10409–10424. DOI: https://doi.org/10.1109/TVT.2023.3260582
7. Li, C., et al. (2023). A survey of conductive and radiated EMI reduction techniques in power electronics converters across wide-bandgap devices. *IET Power Electronics*, 16(13), 2121–2137. DOI: https://doi.org/10.1049/pel2.12532
8. Ray, A., et al. (2022). An Active EMI Cancellation Technique Achieving a 25-dB Reduction in Conducted EMI of LIN Drivers. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 69(9), 3829-3840.DOI: https://doi.org/10.1109/TCSI.2022.3187068
9. Gajbhiye, P., et al. (2021). An Approach to Mitigate the EMI Noise Effect in Vehicle-to-Grid (V2G) System. *2021 2nd Global Conference for Advancement in Technology* (GCAT). IEEE, p. 1-6. DOI: https://doi.org/10.1109/GCAT52182.2021.9587736
10. Wang, Z., et al. (2023). A Review of EMI Research of High Power Density Motor Drive Systems for Electric Actuator. *Actuators*. MDPI, p. 411. DOI: https://doi.org/10.3390/act12110411.
11. Fan, W., Shi, Y., & Chen, Y. (2023). A Method for CM EMI Suppression on PFC Converter Using Lossless Snubber with Chaotic Spread Spectrum. *Energies*, 16(8), 3583. DOI: https://doi.org/10.3390/en16083583
12. Sim, B., et al. (2020). A near field analytical model for EMI reduction and efficiency enhancement using an nth harmonic frequency shielding coil in a loosely coupled automotive WPT system. *IEEE Transactions on Electromagnetic Compatibility*, 63(3), 935-946. DOI: https://doi.org/10.1109/TEMC.2020.3039412
13. Beshir, A.H., et al. (2021). Electromagnetic interference of power converter with random modulation on the power line communication system. *Electronics*, 10(23), 2979. DOI: https://doi.org/10.3390/electronics10232979
14. Amudha, A., et al. (2022). A Review on EMI Issues in High speed Designs and Solutions. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 4(4), 191-203. DOI: https://doi.org/10.35882/jeeemi.v4i4.253
15. Yu, D., et al. (2021). A novel power and signal composite modulation approach to powerline data communication for SRM in distributed power grids. *IEEE Transactions on Power Electronics*, 36(9), 10436-10446. https://doi.org/10.1109/TPEL.2021.3061469
16. Ashjaei, M., et al. (2021). Time-Sensitive Networking in automotive embedded systems: State of the art and research opportunities. *Journal of systems architecture*, 117, 102137. https://doi.org/10.1016/j.sysarc.2021.102137

© Tsymbal O., Lebedev D.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-08
UDC 004.49
Shkitov Andrii, Postgraduate student
https://orcid.org/0009-0005-4600-8467
Open International University of Human Development «Ukraine», Kyiv, Ukraine

FEATURES OF CREATING A MULTI-LEVEL MODEL OF CYBER SECURITY OF CRITICAL INFRASTRUCTURE DATA: AN INTERDISCIPLINARY APPROACH

Shkitov A. Features of creating a multi-level model of cyber security of critical infrastructure data: an interdisciplinary approach. The article examines the model of ensuring cyber security of Ukraine and documents the organizational and legal support for the protection of critical infrastructure objects from cyber attacks. The main concepts are described: multi-level data model, identification of cyber security, and critical infrastructure facilities. The levels of protection of critical infrastructure objects and measures at each level that ensure cyber security are highlighted and described. The stability of the critical infrastructure object is substantiated, and the formula for determining risks in critical infrastructure objects is described. The developed basic tuple model for the classification of critical information infrastructure objects by multi-level data models is highlighted. The prospects and risks of implementing multi-level data models in critical infrastructure facilities have been identified. It was concluded that multi-level data models are an important tool for ensuring cyber security in today's digital world and critical infrastructure enterprises in private. They allow for the effective protection of confidential information by isolating and controlling different levels of data access.
Keywords: cyber security, critical infrastructure, identification model, regulatory and legal support, multi-level data model

Шкітов А.А. Особливості створення багаторівневої моделі кібербезпеки даних критичної інфраструктури: міждисциплінарний підхід. У статті досліджено модель забезпечення кібербезпеки України та документи організаційно-правового забезпечення захисту об'єктів критичної інфраструктури від кібератак. Описано основні поняття: багаторівнева модель даних, ідентифікація кібербезпеки та об'єкти критичної інфраструктури. Висвітлено та описано рівні захисту об'єктів критичної інфраструктури та заходи на кожному рівні, що забезпечують кібербезпеку. Обґрунтовано стійкість об'єктів критичної інфраструктури та описано формулу для визначення ризиків в об'єктах критичної інфраструктури. Висвітлено розроблену базову модель кортежу для класифікації об'єктів критичної інформаційної інфраструктури за багаторівневими моделями даних. Визначено перспективи та ризики впровадження багаторівневих моделей даних на об'єктах критичної інфраструктури. Було зроблено висновок, що багаторівневі моделі даних є важливим інструментом для забезпечення кібербезпеки в сучасному цифровому світі та на підприємствах критичної інфраструктури, а заходи кібербезпеки забезпечують ефективний захист конфіденційної інформації шляхом ізоляції та контролю різних рівнів доступу до даних.
Ключові слова: кібербезпека, критична інфраструктура, модель ідентифікації, нормативно-правове забезпечення, багаторівнева модель даних

Introduction. In today's digital world, where cyber threats are becoming more complex and dangerous, ensuring the cyber security of critical infrastructure becomes an extremely important task. This requires the classification and improvement of approaches to protection against cyber threats and the implementation of effective innovative strategies. One of the classification leitmotifs of this problem is the development and implementation of a multilevel data model.

Cyber wars of the 21st century are relevant for the whole world and increase the importance of this issue. Ukrainian organizations should be maximally protected from cyber attacks. As noted in the State Service of Special Communications and Information Protection of Ukraine, «cyber security of critical infrastructure is one of the priorities of the national security of Ukraine». It is also worth noting that in today's world, where the regulatory environment is constantly changing and becoming increasingly complex, the identification of requirements becomes extremely important for businesses, government bodies, and critical infrastructure in general (Fig. 1).

Under the conditions of cyber modernity, there are a number of normative documents that determine the classification and categorization leitmotifs of cyber protection of critical infrastructure objects, which are mandatory for enterprises, institutions, and organizations classified as critical infrastructure objects according to the legislation. These include: Resolution of the Cabinet of Ministers of Ukraine dated June 19, 2019, No. 519 «On approval of General requirements for cyber protection of critical infrastructure facilities»; Resolution of the Cabinet of Ministers of Ukraine dated November 11, 2020, No. 1176 «On approval of the Procedure for conducting a review of the state of cyber protection of critical information infrastructure state information resources and information the requirement for the protection of which is established by law»; Resolution of the Cabinet of Ministers of Ukraine dated December 23,

© Shkitov A.

2020, No. 1295 «Some issues of ensuring the functioning of the system for detecting vulnerabilities and responding to cyber incidents and cyber attacks»; Resolution of the Cabinet of Ministers of Ukraine dated December 29, 2021, No. 1426 «On approval of the Regulation on the organizational and technical model of cyber protection»; Methodological recommendations for increasing the level of cyber protection of critical information infrastructure approved by the order of the State Special Communications Administration dated 06.10.2021 No. 601 (with changes introduced in accordance with the orders of the State Special Communications Administration dated 10.07.2022 No. 343); Order of the State Special Communications Administration dated December 1, 2023, No. 1011 «On the approval of Recommendations for the development of a plan for the protection of a critical infrastructure facility based on the project threat of a national level cyber attack/cyber incident» and others.

Fig. 1. General sectors of critical infrastructure

Implementing cyber protection measures involves: identification - detection of real and potential cyber threats for prevention and their neutralization; protection - development and implementation of methods, means, and procedures of cyber protection directed at software sustainability and reliability, functioning of information telecommunications, information, and telecommunications and technological systems; detection - conducting monitoring, definition, collection, and processing of atypical events in cyberspace; response - application of measures aimed at the prevention of cyber incidents, cyber attacks, minimization of their possible consequences (prevention of threats to life or health of people and infliction of damage to property), improvement of systems of cyber protection taking into account the necessity of proportionality and opportunities of such systems against real and potential risks; recovery - renewal of regular functioning of information and telecommunications technological systems after cyber attacks, recovery of information and information in case of their damage or deletion, creation of prerequisites for carrying out an investigation by consequences of cyber attacks. During the time of software functioning, the basic infrastructure of cyber protection has to provide protection in cyberspace, national electronic informative resources, communication, and technological systems; protection of objects of critical infrastructure; implementation of activities on forming a culture of cyber security on objects of critical infrastructure and enterprises regardless of their forms of property; informing citizens about the consequences of cyber incidents [1-3].

© Shkitov A.

So, in the modern cyber age, protecting critical infrastructure is one of the primary tasks of the state, taking into account the fact that the latest information and communication technologies are being implemented in Ukraine in all areas. At the same time, the issue of classification and regulation of the multi-level data model, specifically regarding the cyber protection system of critical infrastructure objects and the diagnosis of the state of cyber security, is one of the important and relevant in the conditions of Ukraine.

Literature review. The following scientists investigated the issue of cyber security of critical infrastructure in their works: Hnatiuk S.O. [4], Ivanyuta S.P. [5], Kondratov S.I. [6], Lukyanchuk R.V. [7], Leonov B.D. [8], Ryzhov I.M. [9], Seryogin V.S. [6], Sukhodolya O.M. [10], Tkachuk N.A. [11], Tsiapa S.M. [12], etc. The directions of the development of cyber security were considered in their works by V. Lebedev, D. Ogorodnikov, M. Oleynik, D. Prozorov, A. Svishchev, E.V. Kovalenko, O.O. The question of the multi-level model in cyber security Kharchenko V.P., Korchenko O.G., Hnatyuk S.O. [13, 14] and others.

In the research work [15], a logical-probabilistic model of cyber security for critical infrastructure objects in the power industry was developed for protection against cyber threats and the study of event probabilities. A research article [16] presents a simulator based on agents for evaluating cyber threats in interconnected critical infrastructures that support Italian legislative requirements of cyber protection measures. The article [17] suggests a taxonomy for modeling relationships in critical infrastructure to improve analysis of cyber security by detecting vulnerabilities and critical components in complex systems. Another paper [18] proposes a model for protecting critical informative infrastructure by analyzing functions, connections, threat objects, and implementing methods for minimizing damage from security violations and material costs.

From a literature review, it is evident that scientists all over the world are dealing with cyber security issues and forming information protection models for higher authorities and powerful commercial structures. Since the main damage to information is usually caused by criminal actions (viruses, hacking of secret keys, data theft, etc.), various security mechanisms are created to combat them, including organizational, technical, and software measures and means of information protection [6]. However, a systematic analysis of the multi-level data model in cyber security identification needs further research. The purpose of this work is to consider the main features of creating a multi-level model of cyber security of critical infrastructure data and to determine the main directions of development of this model

Methodology. The model of the regulatory environment allows one to view data at multiple levels of abstraction, from specific requirements to high-level policies and strategies. Fundamentally, the multi-level security model is built on models of authority and truth verification, providing authorized access to users to closed information in databases based on predetermined powers. However, this is not enough for protection, as such models do not have secrecy classes, and databases usually store information from open to completely confidential. For such purposes, the Bell-LaPadula multi-level security model, which is also called the «model of the highest level of secrecy» [19, 20], is intended, which provides users with access to secret data in databases according to different classes of secrecy and is a classic model of an authorized demarcation of access to data. The Bell-LaPadula model uses concepts such as the level of secrecy, the set and current levels of user access, the level of the hierarchy of objects, and others [21].

Thus, the benefits of using a multi-level data model include enabling a more precise definition of requirements, improving the effectiveness of risk analysis, and ensuring compliance with regulatory requirements.

The use of a multi-level data model also helps to avoid duplication of information and improves knowledge management in the organization, which is the prerogative of special services.

In the classification-categorical nature of priority algorithms in information protection, it is worth noting that Identification (lat. Identifico) — is the recognition of something or someone. In this sense, the term is used in general engineering and legal psychology, where it is understood as the process of comparing one object with another on the basis of any feature or property, as a result of which their similarity or difference is established. Identification is the act of establishing an identity [22]. In this context, Identification in particular (information security) is a procedure for recognizing a user in the system, as a rule, with the help of a preventively determined name (identifier) or other a priori information about them, which is perceived by the system.

© Shkitov A.

Identification is used to obtain information about the subject of the system based on the identifier provided by them. It is the initial procedure for granting access to the system. After it, authentication and authorization are carried out [23].

In our opinion, the multi-level data model of cyber security provides for the integration of classification-categorical leitmotifs of protection levels and measures at each level of implementation. This model is based on the idea that no single defense can be completely effective in addressing all cyber threats, so a combination of measures at different levels of the infrastructure must be used.

The first level of logical conditioning with respect to the multi-level model is network and systems level safeguards. This includes the use of network firewalls, interface filters, user authentication and authorization, and other technologies to prevent unauthorized access to the system.

The second level is protection at the level of applications and services. This includes the use of anti-virus programs, data integrity mechanisms, protection against software vulnerabilities, and other technologies to prevent cyber attacks on applications and services.

The third level is protection at the data level. This includes data encryption, access control to confidential information, auditing, and monitoring of user activity to identify suspicious activities.

In the classification-categorical nature, one of the main advantages of the multi-level data model of cyber security is its comprehensive approach to protection against cyber threats. The integration of different levels of protection allows you to create a reliable barrier against various types of attacks and ensure full protection of critical infrastructure.

In addition, multi-level data models allow you to effectively manage access rights by identifying them in accordance with regulatory and legal provisions and providing the ability to precisely control who has access to which information. This provides an additional layer of protection, preventing possible security breaches due to unauthorized access.

In view of this, modern domestic scientists O.G. Korchenko, Yu.O. Dreys, and Romanenko O.O. proposed a basic tuple model for the classification of critical information infrastructure (CII) objects of the state, which contains the main identifiers of the object and is suitable for the topic of our research [24]:

$$ID =\leq ID_1, ID_2, \ldots, ID_i, \ldots, ID_n >, \quad (1)$$

where $ID_i \subseteq ID \ (i = 1, n)$ is the component of the tuple displaying the i-th object identifier, and n is their number.

For example, to form the list of the state [19, 20, 5, 11], with n =≤ 8, tuple (1) is defined as:

$$ID = < ID, ID, ID, ID, ID, ID, ID, ID >=<S,U,O,N,I,R,H,M >,1\ 2\ 3\ 4\ 5\ 6\ 7\ 8 \quad (2)$$

The first component of the tuple S - the KII sector can be represented as a set of sectors:

Taking (4) into account, expression (3) can be represented in the following form: 3 {,,..., }},

$$S = \left\{\bigcup_{i=1}^{n_1} S_i\right\} = \{S_1, S_2, \ldots, S_{n_1}\}, \quad (3)$$

Where $S_i \subseteq S \ (i = \overline{1, n_1})$ is the i-th subset of groups of KII sector identifiers, and n_1 is the total number of groups. For the i-th subset S_i we define as:

$$S_i = \left\{\bigcup_{j=1}^{n_{1j}} S_{ij}\right\} = \{S_{i1}, S_{i2}, \ldots, S_{in_{1i}}\}, \quad (1)$$

Where $S_{ij} \subseteq S_i \ (j = \overline{1, n_{1i}})$ are identifiers of sectors of the i-th group, and n_{1i} - their number. Taking into account **Помилка! Джерело посилання не знайдено.**, the expression **Помилка! Джерело посилання не знайдено.** can be presented in the following form:

$$S = \{\bigcup_{i=1}^{n_1} S_i\} = \left\{\bigcup_{i=1}^{n_1} \left\{\bigcup_{j=1}^{n_1 j} S_{ij}\right\}\right\} = \left\{\{S_{11}, S_{12}, \ldots, S_{1n_{11}}\}, \{S_{21}, S_{22}, \ldots, S_{2n_{12}}\}, \ldots, \left\{S_{n_1 1}, S_{n_1 2}, \ldots, S_{n_1 n_{1mu1}}\right\}\right\}, \left(i = \overline{l, n_1}, \ j = \overline{l, n_{1i}},\right). \quad (5)$$

The next component of the tuple U is the set of identifiers of the administrative-territorial units of Ukraine, within which the OKI is located and displayed as:

$$S = \left\{\bigcup_{i=1}^{n_2} U_i\right\} = \{U_1, U_2, \ldots, U_{n_2}\} = \{»01», »02», »03», \ldots, »n_2»\} \quad (6)$$

Where $Ui \subseteq U \ (i = \overline{1, n_2})$ is the identifier of the administrative-territorial unit, and n_2 is their total number.

© Shkitov A.

It is also necessary to consider it expedient to disclose the concept of stability of a critical infrastructure object, which was clearly disclosed by O.P. Ermenchuk.

In other words, the stability of a critical infrastructure object is its ability to counter threats, minimize the consequences of their impact and negative factors, and quickly recover. For operators, the ideal model for ensuring the stability of CI objects is one where even the active direct action of various threats does not interfere with guaranteeing the provision of basic functions and services, and the restoration of basic functions and services is carried out in the shortest possible time. It is important to identify risks at the earliest possible stage using the proposed formula:

$$P = f(\Pi, B, \text{C}, T) \quad (7)$$

Each of these risk components can be evaluated by experts on the appropriate point scale using the methods evaluations mentioned above. In the simplest version, taking into account the coefficients of significance, b_i the function P has the following form:

$$P = b_1\Pi + b_2B + b_3C + b_4T, \text{де } b_1 + b_2 + b_3 + b_4 = 1 \quad (2)$$

Where risk is *P*, the state of protection of the object (C) against a certain threat, the potential of the threat (P) and the duration of its action, the predicted period of restoration of the functioning of the CI object (T), the importance of the object (B) for a certain type subjects (state, society, business) [11].

However, it must be taken into account that the implementation of multi-level data models can require significant efforts and resources. It requires careful design and implementation, as well as constant monitoring and adaptation to changes in cyber security threats.

In the future, the development of artificial intelligence and machine learning technologies may contribute to further improvements of multi-level data models, specifically for critical infrastructure enterprises. They can become more adaptive and effective in responding to new cyber threats by automating the processes of detection and response to potential attacks.

Results and Discussion. Multi-level data models are an important tool for ensuring cyber security in today's digital world and critical infrastructure enterprises in particular. They allow one to effectively protect confidential information by isolating and controlling different levels of data access. With the right approach and the use of modern technologies, multi-level data models can provide a high level of protection against cyber threats.

Therefore, the normative-legal regulation of the classification-categorical nature of the multi-level data model should be understood as a subject of jurisdiction enshrined in the Constitution of Ukraine regarding the foundations of national security, the foundations of internal and external policy, electronic communications, and the protection of state information resources and information. In accordance with this, the legal and organizational bases for ensuring the protection of the vital interests of man and citizen, society, and the state, national interests of Ukraine in cyberspace, the main goals, directions, and principles of state policy in the field of cyber security, the powers of state bodies, enterprises, institutions, and organizations must be determined [22]. At the same time, cyber security protects the vital interests of a person and citizen, society, and the state in cyberspace, which ensures the sustainable development of the information society and digital communication environment, timely detection, prevention, and neutralization of real and potential threats to the national security of Ukraine in cyberspace [22].

In our opinion, the ISO/IEC 27032 standard defines «cyber security» through the category of cyberspace security as preserving the confidentiality, integrity, and availability of information in cyberspace.

According to the Law of Ukraine «On the Basic Principles of the Development of the Information Society in Ukraine for 2007-2015», the problem of information security should be solved by:

- creating a fully functional information infrastructure of the state and ensuring the protection of its critical elements;
- increasing the level of coordination of the activities of state bodies regarding the identification, assessment and forecasting of threats to information security, prevention of such threats and ensuring the elimination of their consequences, implementation of international cooperation on these issues;

© Shkitov A.

- improvement of the legal framework for ensuring information security, in particular protection of information resources, countering computer crime, protection of personal data, as well as law enforcement activities in the information sphere;
- deployment and development of the National Confidential Communication System as a modern protected transport base capable of integrating territorially distributed information systems in which confidential information is processed [25].

Critical infrastructure facilities are strategically important enterprises and institutions necessary for the functioning of the country's society and its economy.

Enterprises related to critical infrastructure objects:

1. In order and provision of the most important public (administrative) services;
2. Energy supply (including the supply of thermal energy);
3. Water supply and drainage;
4. Genealogical support;
5. Health care;
6. Pharmaceutical industry.

The term «critical infrastructure» usually includes those objects, systems, networks, or their parts, the malfunctioning or destruction of which will lead to the most serious consequences for the social and economic sphere of the state, will negatively affect the level of its defense capability and national security. In addition, the functioning of critical infrastructure in peacetime is associated with the maintenance of vital functions in society, the protection of the basic needs of its members, and the formation of a sense of safety and security in them. The Green Paper on Critical Infrastructure Protection interprets Ukraine's critical infrastructure as a system and resources, physical or virtual, so vitally important to the country that their incapacity or destruction undermines national security, the national economy, the health or safety of the population, or has as a result any combination of the above [6].

Today, the state actively addresses of the questions of implementation and increasing the efficiency of the national systems of cyber security for critical infrastructure objects with the purpose of ensuring stable and safe functioning of national critical infrastructure in cyberspace, creating prerequisites for the association of efforts of subjects responsible for cyber security to solve the task of increasing the cyber resilience of the critical infrastructure of the state. This includes communication and information systems, whose sustainability and reliability are critically important for the functioning of state bodies, enterprises, institutions, and organizations.

The main task of technological infrastructure for cyber defense is the operational and effective protection of cyberspace in opposition to cyberattacks, cybercrimes, cyberterrorism, cyberespionage, including: collection, analysis, evaluation, generalization and spread of information about cyber incidents; granting methodical help to other subjects of cyber protection; mutual informing of subjects of cyber protection about new real and potential threats; creation of conditions for responsible and trusted exchange of information between subjects of cyber protection in every aspect [26, 27]. Thus, reviewing the state of cyber security in relevant sectors of the Ukrainian economy is crucial for strengthening the protection of critical infrastructure. It will help to improve state cyber security, increase the security of information resources and communication systems of critical infrastructure objects, as well as raise the level of the fuel and energy sector to provide a functioning safe integrated informative environment of communication.

Conclusion. The multi-level model data is a powerful tool for identifying requirements for regulatory and legal provision. Its implementation will become a key step in improving the process of identifying requirements for regulatory and legal provision, contributing to more effective organizations and analysis of information, and may improve business efficiency, provide compliance with regulatory requirements, and reduce risks for organizations in all industries. In particular, such a model allows for integrating different data sources, increasing the accuracy and completeness of analysis, and contributing to more operational decision-making. The multi-level data model also provides flexibility in adaptation to changes in the legislative and regulatory base, which is important for maintaining actuality and compliance with normative requirements.

As a result, the multi-level data model for the software cyber security of critical infrastructure is an effective approach to protection from cyber threats. The integration of different levels of protection allows creating reliable and comprehensive protection that ensures the safety and stability of critical infrastructure systems in the digital world. It is especially relevant in the conditions of growing complexities and

© Shkitov A.

interdependence of informative systems, where even insignificant violations can have serious consequences.

References

1. Limba, T., et al. (2017) Cyber security management model for critical infrastructure. Entrepreneurship and Sustainability Issues, 4(4),P. 559-573. https://doi.org/10.9770/jesi.2017.4.4(12)
2. Ten, C.-W., Govindarasu, M., & Liu, C.-C. (2010) Cybersecurity for critical infrastructures: Attack and defense modeling. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 40(4), P.853-865.
3. Sarker, I. H., Janicke, H., Ferrag, M. A., & Abuadbba, A. (2024) Multi-aspect rule-based AI: Methods, taxonomy, challenges, and directions towards automation, intelligence, and transparent cybersecurity modeling for critical infrastructures. Internet of Things, 25, 101110.
4. Hnatyuk, S.O. (2013) Cyberterrorism: History of development, current trends, and countermeasures. Information Security, 19(2), P. 118-129.
5. Ivanyuta, S.P. (2018) Priority areas of legislative and organizational support for certification of critical infrastructure objects.
6. Biryukov, D.S., Kondratov, S.I., Nasvit, O.I., & Sukhodolya, O.M. (2015) Green book on critical infrastructure protection in Ukraine: Analytical report. Kyiv: NISD.
7. Lukyanchuk, R.V. (2016) State strategic planning in the field of cyber security: Realities today. Herald National Academy of State Management at the President of Ukraine. Series: State Management, 3, P. 131-137.
8. Leonov, B.D., Shostak, R.M., & Seryogin, V.S. (2020) Development of methodical provision of anti-terrorist protection of critical infrastructure objects (using the example of the USA). Information and Law, 3(34),P. 88-95.
9. Ryzhov, I.M. (2016) Basic concepts of anti-terrorist security: Monograph. Kyiv: National Academy of SBU.
10. Kondratov S. I., Sukhodolia O. M. (2020) The state system of critical infrastructure protection in the national security system: an analyst. add. / edited by OHM. Dried fruits Kyiv: NISD, P. 28.
11. Yermenchuk, O.P. (2018) Basic approaches to the organization of critical infrastructure protection in European countries: Experience for Ukraine: Monograph. Dnipro: Dniprop. State University of Internal Affairs.
12. Alekseieva O. (2021) Legal and organizational support for the protection of critical information infrastructure objects from cyber attacks. Information and law. No. 4(39). https://doi.org/ 10.37750/2616-6798.2023.4(47).291633
13. Kharchenko, V.P., Korchenko, O.G., & Hnatyuk, S.O. (2016) The basic model of the formation of requirements for ensuring cybersecurity of civil aviation. Security of Information, 22(2).
14. Kharchenko, V.P., Korchenko, O.G., & Hnatyuk, S.O. (2016) Multi-level data model for identification of the security of requirements in accordance with the regulatory and legal provision of cybersecurity of civil aviation information protection.
15. Alekseichuk, L., Novikov, O., Yakobchuk, D., & Rodionov, A. (2023).Cyber security logical and probabilistic model of a critical infrastructure facility in the electric energy industry. Theoretical and Applied Cybersecurity.
16. Bonagura, V., Foglietta, C., Panzieri, S., Rossi, M., Santini, R., & Scannapieco, M., et al. (2023) Modeling and assessing the impacts of cyber threats on interdependent critical infrastructures. In Proceedings of the 2023 International Conference on Cybersecurity.
17. Jiang, Y., Jeusfeld, M. A., & Ding, J. (2023) Model-based cybersecurity analysis. Business & Information Systems Engineering.
18. Gasimov, V., & Mammadov, J. I. (2023) Model and method for determining the optimal structure of the security system for critical information infrastructure. Reports of the Belorussian State University of Computer Science and Radio Electronics.
19. Decree of the President of Ukraine No. 155. (2002, February 19) On the procedure for organizing and monitoring the execution of decrees, orders, and instructions of the President of Ukraine.
20. Decree of the President of Ukraine No. 447. (2021, August 26) On the decision of the National Security and Defense Council of Ukraine dated May 14, 2021 «On the Cybersecurity Strategy of Ukraine».
21. Netesin, I. (2001). Approach to security of distributed databases. Legal Regulatory and Metrological Support of the Information Protection System in Ukraine: Scientific and Technical Collection, 2, P.118-124.
22. Electronic resource, access mode URL: https://zakon.rada.gov.ua/laws/show/2163-19#Text
23. Electronic resource, access mode URL: https://uk.wikipedia.org/wiki/ Identification
24. Electronic resource, access mode URL: https://jrnl.nau.edu.ua/index.php/ZI/article/view/12448
25. Electronic resource, access mode URL: https://coordynata.com.ua/pravove-zabezpecenna-kiberzahistu-v-ukraini
26. Tsantikidou, K., & Sklavos, N. (2024) Threats, attacks, and cryptography frameworks of cybersecurity in critical infrastructures. Cryptography, 8(1).
27. Varma, V. V. (2024) Cybersecurity of critical infrastructure. International Machine Learning Journal and Computer Engineering, 7(7), P.1-17.

© Shkitov A.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-09
УДК 811.161.2'373.72:070
Акімов Дмитро Дмитрович, аспірант
http://orcid.org/0000-0003-2440-1864
Гавриленко Валерій Володимирович, д. фіз.-мат. н., професор
http://orcid.org/0000-0001-9682-4204
Національний транспортний університет, м. Київ, Україна

РОЗВ'ЯЗАННЯ ТРАНСПОРТНОЇ ЗАДАЧІ В УМОВАХ НЕОДНОРІДНОСТІ ПРОДУКЦІЇ АГРОПРОМИСЛОВИХ ПІДПРИЄМСТВ

Акімов Д.Д. Гавриленко В.В. Розв'язання транспортної задачі в умовах неоднорідності продукції агропромислових підприємств. У статті досліджено сутність транспортної задачі, її мету та завдання, а також методи розв'язання в умовах неоднорідності продукції. Особлива увага приділена агропромисловим підприємствам, де чинники, пов'язані з якістю продукції, строками її придатності та дотриманням режиму вологості під час перевезення, суттєво впливають на логістичні процеси. Специфіка агропромислової продукції полягає в її різноманітності за типами, якістю, термінами зберігання та іншими характеристиками, що створює додаткові виклики для оптимізації логістичних операцій. Важливим аспектом є також інтеграція екологічно безпечних практик у логістичні процеси агропромислових підприємств, що сприяє не лише збереженню якості продукції, але й зменшенню впливу на навколишнє середовище [1]. Використання спеціалізованих програмних комплексів для моделювання логістичних сценаріїв, що враховують всі можливі фактори ризику дозволяють більш ефективно планувати маршрути перевезень [2]. У статті проведено детальний аналіз існуючих методів розв'язання транспортних задач та досліджено проблему оптимізації транспортних витрат у агропромисловому секторі з урахуванням різноманітності продукції. Представлено декілька реальних прикладів, які ілюструють застосування запропонованих методів. Зокрема, розглянуто задачі для різних типів продукції, де враховано не лише транспортні витрати, але й втрати від псування продукції. Окреслено практичні рекомендації щодо вдосконалення логістичних процесів в агропромислових підприємствах з урахуванням неоднорідності продукції, що дозволяє знизити витрати на транспортування та підвищити ефективність постачання. Отримані результати демонструють ефективність запропонованих підходів у зниженні загальних витрат і мінімізації втрат продукції. Ця стаття буде корисною для науковців та практиків, які займаються питаннями логістики і управління в агропромисловому комплексі, а також для фахівців з оптимізації виробничо-транспортних систем.
Ключові слова: транспортна задача, агропромислові підприємства, неоднорідність продукції, логістика, оптимальний план розподілу

Akimov D., Gavrilenko V. Solving the transportation problem in the conditions of heterogeneity of products of agro-industrial enterprises. The article investigates the essence of the transportation problem, its purpose and objectives, as well as methods of solving it in conditions of product heterogeneity. Particular attention is paid to agro-industrial enterprises, where factors related to product quality, shelf life and compliance with the humidity regime during transportation have a significant impact on logistics processes. The specificity of agricultural products lies in their diversity in terms of types, quality, shelf life, and other characteristics, which creates additional challenges for optimizing logistics operations. An important aspect is also the integration of environmentally friendly practices into the logistics processes of agro-industrial enterprises, which contributes not only to maintaining product quality but also to reducing environmental impact [1]. The use of specialized software systems for modeling logistics scenarios that take into account all possible risk factors allows for more efficient planning of transportation routes [2]. The article provides a detailed analysis of existing methods for solving transportation problems and investigates the problem of optimizing transportation costs in the agricultural sector, taking into account the diversity of products. Several real-life examples are presented to illustrate the application of the proposed methods. In particular, the article considers problems for different types of products, which take into account not only transportation costs but also losses from product spoilage. Practical recommendations for improving logistics processes in agro-industrial enterprises, taking into account product heterogeneity, are outlined, which allows reducing transportation costs and increasing supply efficiency. The results demonstrate the effectiveness of the proposed approaches in reducing overall costs and minimizing product losses. This article will be useful for scientists and practitioners involved in logistics and management in the agro-industrial complex, as well as for specialists in optimizing production and transportation systems.
Keywords: transportation problem, agricultural enterprises, product heterogeneity, logistics, optimal distribution plan

Постановка проблеми. Процес транспортування є однією з найважливіших складових всіх виробничих та торгівельних процесів будь-якого підприємства. Тому перед багатьма агропромисловими підприємствами стоїть питання вибору оптимального варіанту логістики (перевезення) товарів від пункту виробництва (там, де безпосередньо виготовляються товари) до пунктів споживання (наприклад, у продуктові магазини), з урахуванням усіх реальних можливостей та з мінімальними витратами на це транспортування [3]. Формування та розв'язання транспортної задачі допомагає у вирішенні даного завдання. Особливу складність становить необхідність врахування неоднорідності продукції, що може відрізнятися за типом, якістю, термінами зберігання та іншими характеристиками. Враховуючи ці аспекти, вибір оптимального варіанту логістики для

© Акімов Д.Д. Гавриленко В.В.

неоднорідних товарів є надзвичайно актуальним і важливим для забезпечення ефективної роботи агропромислових підприємств.

Аналіз останніх публікацій. Багато вчених, здебільшого зарубіжних, таких як Монтлевич В.М., Бородінова І.О., Копфер Х.В., Танг К.С., Марстер Р.Є., Фішер М.Л. та ін., розглядали сутність транспортних задач та їх розв'язання саме для неоднорідної продукції. Особливості логістики саме агропромислових підприємств розглянуто у публікаціях українських вчених Радченко О. П., Нечипоренко К. В.

Постановка завдання. Метою статті є розгляд транспортних задач в умовах неоднорідності продукції агропромислових підприємств.

Виклад основного матеріалу. Перш ніж розглянути транспортні задачі в умовах неоднорідності продукції агропромислових підприємств, дамо визначення, в чому полягає сутність транспортної задачі взагалі.

Транспортна задача – задача, яка ставиться логістами, для знаходження оптимального варіанту доставки товарів з пункту А в пункт Б [4]. Оптимальність в задачі визначається мінімізацією витрат на цю доставку. Це є ключовим завданням, але транспортні задачі також використовують для таких цілей:

- досягнення регулярності вантажних перевезень;
- оптимізація складських запасів;
- зниження собівартості перевезень;
- оптимізація роботи транспорту.

Найчастіше при розгляді дисципліни, пов'язаної з логістикою, студенти розглядають транспортні задачі з однорідними товарами, тоді як на практиці логісти мають справу саме з транспортними задачами в умовах неоднорідності продукції [5].

Одним з варіантів розв'язку транспортної задачі в умовах неоднорідності продукції є її вирішення як єдиної задачі в тому випадку, коли виникає необхідність пошуку базового плану перевезень взаємозамінних видів продукції. В цій ситуації треба це питання вирішувати як єдину задачу, так як в ній різні товари можуть умовно прирівнюватись один до одного через коригуючі коефіцієнти (які переводять умовно один товар в інший). Розв'язання цієї задачі не має принципових відмінностей від вирішення закритої однопродуктової задачі. Існують лише методичні прийоми обробки вихідної інформації [6].

В агропромисловому секторі існує низка специфічних факторів, які впливають на логістичні процеси. Продукція агробізнесу характеризується значною різноманітністю видів, якості та термінів зберігання. Це включає продукти, які мають різні вимоги до температурного режиму, вологості та умов зберігання. Додатково, агропродукція підлягає сезонним коливанням обсягів виробництва та споживання, що ускладнює планування логістичних операцій. Важливим аспектом є також швидке псування багатьох видів продукції, що вимагає швидких та ефективних рішень щодо перевезень. Крім того, агропромислові підприємства часто розташовані в сільській місцевості з обмеженою інфраструктурою, що впливає на доступність і ефективність транспортних мереж. Усі ці фактори вимагають спеціалізованого підходу до розв'язання транспортних задач, який би враховував унікальні умови агропромислового сектору [7].

Такий вчений як Танг К.С. [8] привів наступний спосіб вирішення багатопродуктової транспортної задачі, який розглядається в прикладі 1.

Приклад 1. Споживачу потрібно доставити взаємозамінний корм, а саме соя та ячмінь, для худоби,. Необхідна умова: загальна потреба в цих двох продуктах в умовних одиницях калорійності повинна бути повністю задоволена. Відомо, що 1 кг умовного корму, що задовольняє в добовій нормі калорійності однієї корови дорівнює 7 000 ккал, 1 кг ячменю – 2 800 ккал, 1 кг сої – 4 200 ккал. Звідси коригуючий коефіцієнт за теплопровідною здатністю (калорійний еквівалент) для ячменю дорівнює:

$$К1 = \frac{2\,800}{7000} = 0{,}4,$$

а для сої:

$$К2 = \frac{4\,200}{7000} = 0{,}6.$$

© Акімов Д.Д. Гавриленко В.В.

В таблиці 1.1 представлені потужності та попит ячменю в кілограмах і показано оптимальний план перевезення з функціоналом, який дорівнює $F1 = 13\,980$, а в таблиці 1.2 – аналогічні дані для сої з функціоналом $F2 = 10\,620$.

За двома планами обсяг перевезення дорівнює:

$$F1 + F2 = 13\,980 + 10\,620 = 24\,600 \text{ (км/кг)}$$

У постачальників А та $Л_2$ є ячмінь та соя, у постачальника $А_3$ – тільки ячмінь, а у постачальника $А_4$ – тільки соя. В обох таблицях відстань між постачальниками А і $Л_2$ і споживачами однакова, так як ці два види продукції будуть перевозитись один і тим же транспортним шляхом.

Таблиця 1.1. Потужності та попит на ячмінь

b_j / a_i	B_1 100	B_2 180	B_3 120
А, 150	12 100	72	6 050
$Л_2$, 75	48	245	4 870
$А_3$, 175	72	36 175	60

Таблиця 1.2. Потужності та попит на сою

b_j / a_i	B_1 60	B_2 210	B_3 125
А, 130	1 260	72	6 070
$Л_2$, 100	48	2 445	4 855
$А_4$,, 165	36	12 165	72

Використовуючи коригуючі коефіцієнти, розрахованої вище (для ячменю – $К_1 = 0{,}4$, для сої – $К_2 = 0{,}6$), та дані таблиці 1.1 та 1.2 робимо перерахунок і формулюємо таблицю 1.3, в якій указані дані в умовних (перерахованих) одиницях.

Нижче приведемо пояснення, пов'язані з цим перерахунком:

1. Розрахунок попиту споживачів в умовних одиницях (у.о.) проведемо на прикладі B_1 (див. табл. 1.1 та 1.2). Попит споживача B_1 на ячмінь дорівнює 100 кг, на сою – 60 кг. Використовуючи коригуючі коефіцієнти, розраховуємо його потребу в умовному паливі:

$$B_1 = 100 \cdot 0{,}4 + 60 \cdot 0{,}6 = 76$$

2. Потужність постачальника $А_1$ в умовних одиницях розраховується окремо для кожного виду палива, тому що вона виступає як обмеження на можливий розмір поставок k-ого виду продукції, який знаходиться у j-ого постачальника. Потужність розраховується шляхом множення величини потужності постачальника на коригуючий коефіцієнт по ячменю та по сої окремо:

- по ячменю:

$$А_1 = 150 \cdot 0{,}4 = 60$$

- по сої:

$$А_1 = 160 \cdot 0{,}6 = 96.$$

3. Показники відстані ($с_{ij}$ – правий верхній кут в клітинах матриці) в умовних одиницях отримаємо шляхом ділення їх на коригуючі коефіцієнти:

-по ячменю:

$$c11 = \frac{12}{0{,}4} = 30$$

© Акімов Д.Д. Гавриленко В.В.

-по сої:

$$c11 = \frac{12}{0{,}6} = 20$$

Після переведення всіх показників матриці в умовні одиниці, як було показано в пунктах 1-3, отримаємо нову матрицю, в якій проводимо перерозподіл умовних одиниць до отримання оптимального варіанту (див. табл. 1.3).

Таблиця 1.3. Оптимальний варіант розподілу поставок в умовних одиницях

a_i, у.о.		b_j, у.о.	B_1 76	B_2 198	B_3 123
А	Ячмінь	60	3 060	150	120
	Соя	78	2 016	100	8 062
$Л_2$	Ячмінь	30	150	6 029	901
	Соя	60	100	40	6 060
$А_3$	Ячмінь	70	180	9 070	150
$А_4$	Соя	99	60	2 099	100

Функціонал оптимального плану, виражений в умовних одиницях в таблиці 1.3, складає 20 790. Порівняно з загальним потенціалом минулих таблиць по ячменю та сої, обсяг транспортної роботи в останньому варіанті з умовними одиницями зменшився на 3 810 одиниць, що дає економію по обсягу вантажообігу більш, ніж на 15%.

Як бачимо, що автор, приклад якого описано вище, розглядали продукцію, яку можна привести до умовно одного виду. Тому розглянемо і іншого автора, який розглядав два і більше принципово різних видів продукції (для яких коригуючий коефіцієнт неможливо вирахувати).

Багато авторів, розглядаючи саме багатопродуктові транспортні задачі, беруть за приклад транспортування палива. І такий автор як Копфер Х.В., є не виключенням.

Розглянемо запропонований і цим автором розв'язок наступного прикладу транспортної задачі в умовах неоднорідності продукції.

Приклад 2 [9]**.** Морський корабель місткістю 6 560 т з 11 відсіками перевозить і доставляє п'ять різних зернових культур, які не можна змішувати, від пункту їх виробництва до пункту призначення. Добові норми попиту подано наступним чином (в тонах/день):

$$(d_j) = (66, 71, 72, 76, 81)$$

із загальною потребою в 366 т/добу. А місткість відсіків такі:

$$(q_i) = (844, 826, 764, 675, 661, 626, 626, 565, 373{,}313, 287)$$ (в тонах).

Загальний час доставки цих культур і кількість поставок знайдемо далі.

Набір з 11 відсіків буде розділений на п'ять культур і кількість усіх можливих варіантів цього розподілу дорівнює 29 607 600 ($5!\,S_{11}^{(5)}$).

Динамічне програмування дає наступне рішення:

$(1217 = q_1 + q_8, 1252 = q_6 + q_7, 1275 = q_4 + q_{10} + q_{11}, 1301 = q_2 + q_8, 1425 = q_3 + q_5)$ з $t = \frac{1425}{81} = 17{,}593.$

Цей розрахунок здійснений за 1776,51 процесорних секунд на ПК. Існує $2^{11} - 1 = 2047$ змінних стану і це становить 465135 статичних оцінок.

Автор запропоновує наступний алгоритм для пошуку оптимального вирішення цієї проблеми. Так, початковий розв'язок наведено нижче (місткість виділених відсіків і кількість кожної культури).

© Акімов Д.Д. Гавриленко В.В.

Це не є структурованим поділом, оскільки поділ набору відсіків не виконується відповідно до їх порядку чи рейтингу. Початковий розв'язок отримується як:

$$(1199 = q_2 + q_9,, 1261 = q_5 + q_{10} + q_{11}, 1301 = q_4 + q_7, 1329 = q_3 + q_8, 1470 = q_1 + 6)$$

з часом поповнення $\frac{1329}{76}$ =17,486 днів.

Нижня межа вартості часу поповнення становить $t_0 = 17{,}486$, і отримано верхню межу $t_{UB} = \frac{6560}{366} = 17{,}923$ днів. Домінуюча культура $j^* = 4$. Початкові значення $\mu_j = \frac{1}{366}$ для $j = 1,2,3,4,5$ та $\lambda_i = \frac{q_i}{366}$ для $i = 1,2,\ldots,11$. Розв'язування $DSP(t_0, X_0)$ і накладення обмеження $v_4 \leq 0$ дають $u_0 = 17{,}761$. PLP(17,761) розв'яжімо спочатку, вирішивши кожну $SP_j(17{,}757)$, перерахувавши всі комбінації відсіків знаходження мінімального рівня розподілу $A_i \geq td_j$ для кожного $j = 1,2,3,4,5$. Обсяги поставки при $t = 17{,}761$ задані таким вектором:

$$(D_j) = (td_j) = (1171{,}96;\ 1260{,}75;\ 1278{,}50;\ 1349{,}53;\ 1438{,}32).$$

Розв'язування кожного $SP_j(17{,}761)$ дає мінімальні вимоги до розподілу потужностей для культур; $(A_j) = (1191, 1261, 1279, 1350, 1439)$. Домінуюча культура має індекс $j^* = 2$. Час поповнення становить 17,761 днів із цими розподілами (якщо це можливо), шляхом повного використання відсіків, відведених для культури $j = 2$. Оскільки $6538 < 6560$ та $S = \frac{6560\ 6538}{366} > 0$ і умова, за якої відсіки відведені таким чином, що будь-які культури не перекривають відсіки накладається на домінуючий продукт. Розв'язування $SP_j(17{,}761)$, (крім $j = 2$) з цією умовою знову дає мінімальні вимоги до розподілу $(R_j) = (1191, 1261, 1301, 1390, 1439)$. Далі $6582 > 6560$ і $S = \frac{6560 - 6582}{366} = -0{,}060$, тоді $t_{UB} = 17{,}761$ і процес буде повторюється з $t = 17{,}761 - 0{,}060 = 17{,}701$. Обсяги поставки визначено $(D_j) = (1168{,}27;\ 1256{,}77;\ 1274{,}47;\ 1345{,}27;\ 1433{,}78)$.

Таблиця 1.4

(A_j)	$(t_j = A_j/d_j)$	$\min(t_j)$	j^*
(1157,1251,1261,1335,1424)	(17,560;17,620;17,514;17,566;17,580)	17,514	3
(1157,1251,1275,1335,1424)	(17,560;17,620;17,708;17,566;17,580)	17,530	1
(1165,1251,1275,1335,1424)	(17,652;17,620;17,708;17,566;17,580)	17,566	4
(1165,1251,1275,1336,1424)	(17,652;17,620;17,708;17,579;17,580)	17,579	4
(1165,1251,1275,1347,1424)	(17,652;17,620;17,708;17,524;17,580)	17,580	5
(1165,1251,1275,1347,1425)	(17,652;17,620;17,708;17,524;17,593)	17,593	5
(1165,1251,1275,1347,1426)	(17,652;17,620;17,708;17,524;17,605)	17,605	5
(1165,1251,1275,1347,1439)	(17,652;17,620;17,708;17,524;17,765)	17,620	5

Тоді мінімальні вимоги до розподілу, отриманого розв'язуванням кожного $SP_j(17{,}701 + \epsilon)$ за $(A_j) = (1191, 1261, 1275, 1347, 1439)$. Час поповнення перераховується як

$$t = min(18{,}045, 17{,}761, 17{,}708, 17{,}724, 17{,}765) = 17{,}708,$$

і домінуючою культурою є $j^* = 3$. $\sum_{j \epsilon J} Aj = 6513 < 6560$ означає накладення зазначеної умови неперекриття, зазначеної вище, а потім наведено мінімальні вимоги $(A_j) = (1191, 1287, 1275, 1390, 1450)$ з $\sum_{j \epsilon J} Aj = 6593$. $S = \frac{6560 - 6593}{366} = -0{,}090$, і t_{UB}= 17,708 і $t = 17{,}708 - 0{,}090 = 17{,}618$. Повторення процесу відбувається наступним чином. Кількість поставок задається вектором

$$(D_j) = (1162{,}79;\ 1250{,}88;\ 1268{,}50;\ 1338{,}97;\ 1427{,}06),$$

© Акімов Д.Д. Гавриленко В.В.

а отже мінімальні вимоги до розподілу $(A_j) = (1165, 1251, 1275, 1347, 1439)$, отримано за допомогою розв'язування кожного SP_j (17.618). Час поповнення перераховується і $t = min(17{,}652;\ 17{,}620;\ 17{,}708;\ 17{,}724;\ 17{,}765) = 17{,}620$. Оскільки $\sum_{j\epsilon J} Aj = 6477$, умова неперекриття виконана, і мінімальні вимоги тепер визначаються як $(A_j) = (1252, 1251, 1287, 1390, 1439)$. $\sum_{j\epsilon J} Aj =$ 6619 означає, що $S = \frac{6560-6619}{366} = -0{,}161$ і $t_{UB} = 17{,}620$ але $t = 17{,}620 - 0{,}161 = 17{,}459 < 17{,}486$. Отже, інтервал від 17,486 до 17,620 є інтервалом невизначеності. Вичерпний пошук буде здійснюватися в цей інтервал, як описано нижче.

При $t = 17486$, $(A_j) = (1157, 1251, 1261, 1329, 1424)$ обчислення часу t дає $t = min(17{,}530, 17{,}620, 17{,}514, 17{,}486, 17{,}580)$ =17,486 з $j^* = 4$. Таким чином, наступний найменший розподіл до 1329 для культури $j = 4$ буде отримано розв'язуванням $SP_4(t + \in)$ і буде дорівнювати 1335. Схоже міркування повторюється, доки t не стане рівним або перевищить 17,620, і ці результати представлені в таблиці 1.4.

Вже відомо, що немає рішень при $t = 17{,}620$, отже останній інтервал ігнорується. В інтервалі при = $t = 17{,}605$ будуть визначені можливий розподіл. Можна отримати кілька комбінацій рівнів розподілу, наприклад 1286 = $q_6 + q_8$ або 1286 = $q_7 + q_8$. Верхній індекс відповідає кількості різних комбінацій відсіків: $E_1 = \{1191^{(2)}\}$, $E_2 = \{1252\}$, $E_3 = \{1287^{(2)}, 1301^{(2)}\}$, $E_4 = \{1390^{(2)}\}$, і, нарешті, $E_5 = \{1426\}$ з $j^* = 5$. Помітивши, що $1191 = q_6 + q_8 = q_7 + q_8$ і $1252 = q_6 + q_7$, що означає, що не існує допустимих рішень і інтервал при $t = 17{,}593$ тестується далі. Набори відсіків такі:

$$E_1 = \{1165, 1191^{(2)}, 1199, 1217, 1225, 1226^{(2)}\},$$
$$E_2 = \{1251, 1252, 1275, 1286^{(2)}, 1301^{(2)}, 1312^{(2)}\},$$

розподіли для $j = 3$ знаходяться в наборі $E_3 = \{1275, 1286^{(2)}, 1301^{(2)}, 1312^{(2)}, 1335\}$ та для $j = 4$ знаходяться в наборі $E_4 = \{1361, 1391, 1409, 1426, 1444\}$ і $E_5 = \{1425\}$. Цей результат полягає в знаходженні можливих рішень в 108 випробувань і $t = 17{,}593$. Інше оптимальне рішення отримано та наведено нижче. Час виконання цієї задачі становить 32,10 процесорних секунд.

$$(x_{i1}) = (0, 1{,}0{,}0{,}0{,}0{,}0{,}0, 1{,}0{,}0)$$
$$A_1 = 1199\ D_1 = 17{,}593 \cdot 66 = 1161{,}14,$$

$$(x_{i2}) = (0{,}0{,}0{,}0{,}0, 1, 1{,}0{,}0{,}0, 0)$$
$$A_2 = 1252\ D_2 = 17{,}593 \cdot 71 = 1249{,}10,$$

$$(x_{i3}) = (0{,}0{,}0, 1{,}0{,}0, 0{,}0{,}0, 1, 1)$$
$$A_3 = 1275\ D_3 = 17{,}593 \cdot 72 = 1266{,}70,$$

$$(x_{i4}) = (1{,}0{,}0{,}0{,}0{,}0, 0, 1{,}0{,}0{,}0)$$
$$A_4 = 1409\ D_4 = 17{,}593 \cdot 76 = 1337{,}07,$$

$$(x_{i5}) = (0{,}0, 1{,}0, 1{,}0{,}0{,}0{,}0{,}0{,}0)$$
$$A_5 = 1425\ D_5 = 17{,}593 \cdot 81 = 1425{,}00,$$

Цей графік доставки буде повторений кожні 17.593 днів і культура 5 є домінуючою культурою. Якщо всі можливості в інтервалі (17.486,17.620) тестуються без цього ефективного методу, кількість перевірених можливостей буде 99792 (= 123 $12 \cdot 11 \cdot 21$). Запропонована процедура пошуку перевірила лише 108 можливостей. У гіршому випадку це буде перевірка $2 \cdot 1 \cdot 4 \cdot 2 \cdot 1 + 8 \cdot 9 + 8 \cdot 5 \cdot 1 = 2880 + 16 = 2896$ випадків, що становить 2,90% від 99792 і це приблизно 0,0098% від усіх 29607600 можливостей.

Обчислювальна робота алгоритму перевіряється шляхом запуску випадково згенерованих проблем, результати якого узагальнено в таблиці 1.5.

© Акімов Д.Д. Гавриленко В.В.

Таблиця 1.5

n	m	9	10	11	12	13
4	mean	4,11	13,68	27,55	101,52	
	std	1,42	4,01	6,38	27,44	
5	mean	8,39	17,58	53,94	180,14	
	std	4,34	5,67	12,03	117,59	
6	mean		32,81	67,29	350,68	
	std		16,29	52,91	164,48	
7	mean				215,74	399,22
	std				31,54	194,79

Практичні задачі різного розміру генеруються випадковим чином та розв'язуються на ПК. У поколінні випадкових проблем передбачається норма попиту рівномірно розподілених в інтервалі від 20 до 100, і місткість відсіків також уніфікована в інтервалі від 150 до 950. Тоді випадкові числа округляються до найближчого цілого числа. Середнє виконання і його стандартне відхилення для кожної групи з 5 проблем наведено в таблиці 1.5. Через бінарний характер проблем, існує висока мінливість часу виконання. Подальший аналіз регресії означає, що час виконання може бути виражений
як експоненціальна функція від n (кількість культур) і m (кількість відсіків) наступним чином: час $= 0{,}0002(1{,}3198^n)\,(2{,}7229^n)$. Метод знаходить оптимальні рішення (оптимальний час поповнення та оптимальне призначення відсіків для виробу продуктів) задачі. Ронен (1995) вказує на це у використовуваних транспортних засобах з 13 відсіками або менше в нафторозподільних фірмах. Рішення більшого розміру вимагатиме набагато більше пам'яті та швидкість виконання.

Висновок. У статті розглянуто сутність та особливості транспортної задачі в умовах неоднорідності продукції, зокрема в агропромисловому секторі. Автори дослідили чинники, що впливають на логістичні процеси, такі як якість продукції, строки її придатності та умови перевезення, що створюють додаткові виклики для оптимізації. Значна увага приділена інтеграції екологічно безпечних практик, які сприяють не лише збереженню якості продукції, але й зменшенню негативного впливу на довкілля. Завдання транспортної задачі полягає в знаходженні оптимального плану доставки продукції, яке забезпечує досягнення мінімальних витрат на транспортування, мінімальних витрат часу тощо. Так яка часто зустрічається на практиці випадки, коли треба транспортувати не один вид продукції, а декілька, то в таких ситуаціях допоможе розгляд багатопродуктової транспортної задачі, два принципово різні підходи розв'язання якої було представлено.

Список бібліографічного опису

1. Маргіта Н. О., Білоніжка У. З. Сучасні тенденції впровадження «зеленої» логістики. *Маркетинг і менеджмент інновацій.* 2014. № 1. С. 279-284.
2. Величко О. П. Логістика в системі менеджменту підприємств аграрного сектору економіки: Монографія. Дніпропетровськ: Акцент ПП, 2015. 525 с.
3. Нечипоренко К. В. Проблеми розвитку транспортної логістики сільськогосподарських товаровиробників. *Економічні науки. Серія «Облік і фінанси»*. Збірник наукових праць. Луцький національний технічний університет. Вип. 10(37). Ч. 4. Редкол.: відп. ред. д.е.н., професор Герасимчук З. В. Луцьк, 2013. С. 40-46.
4. Молодід О. К. Транспортна задача: навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення», спеціалізації «Програмне забезпечення розподілених систем», «Програмне забезпечення Web-технологій та мобільних пристроїв». КПІ ім. Ігоря Сікорського. Київ: КПІ ім. Ігоря Сікорського, 2018. 37 с.
5. Наконечний С. І., Савіна С. С. Математичне програмування: навч. посіб. Київ: КНЕУ, 2003. 253 с.
6. Radchenko O. P. Problems and perspectives of the functioning of national logistics systems on agricultural product markets. Market economy: modern management theory and practice. 2023. Т. 21, № 2(51). С. 126–136.
7. Акімов Д. Д. Застосування технологій інтернету речей для оптимізації логістики в сільському господарстві. *Computer integrated technologies: education, science, production*. 2023. No 53. С. 9–15. URL: https://doi.org/10.36910/6775-2524-0560-2023-53-02
8. Tang C. S. A Max-Min Allocation Problem: Its Solutions and Applications. *Operations Research.* 1988. Vol. 36, No. 2. P. 359-367.

© Акімов Д.Д. Гавриленко В.В.

9. Kopfer H. W., Kopfer H. Emissions minimization vehicle routing problem in dependence of different vehicle classes. *Dynamics in logistics.* 2013. P. 49-58.

References:

1. Margita, N. O., & Bilonizhka, U. Z. (2014). Suchasni tendentsii vprovadzhennia "zelenoi" lohistyky [Modern trends in the implementation of "green" logistics]. Marketing i menedzhment innovatsii – Marketing and Management of Innovations, (1), 279-284. [in Ukrainian].
2. Velychko, O. P. (2015). Lohistyka v systemi menedzhmentu pidpryiemstv ahrarnoho sektoru ekonomiky: Monohrafiia [Logistics in the management system of enterprises in the agricultural sector of the economy: Monograph]. Dnipro: Aktsent PP. [in Ukrainian].
3. Nechyporenko, K. V. (2013). Problemy rozvytku transportnoi lohistyky silskohospodarskykh tovarovyrobnykiv [Problems of the development of transport logistics of agricultural producers]. Ekonomichni nauky. Seriia "Oblik i finansy". Zbirnyk naukovykh prats. Lutskyi natsionalnyi tekhnichnyi universytet – Economic Sciences. Series "Accounting and Finance". Collection of Scientific Papers. Lutsk National Technical University, 10(37), Ch. 4, 40-46. [in Ukrainian].
4. Molodid, O. K. (2018). Transportna zadacha: navch. posib. dlia stud. spetsialnosti 121 "Inzheneriia prohramnoho zabezpechennia", spetsializatsii "Prohramne zabezpechennia rozpodilenykh system", "Prohramne zabezpechennia Web-tekhnolohii ta mobilnykh prystroiv" [Transportation problem: textbook for students majoring in 121 "Software Engineering", specializing in "Software for distributed systems", "Software for Web technologies and mobile devices"]. KPI im. Ihoria Sikorskoho. Kyiv: KPI im. Ihoria Sikorskoho. [in Ukrainian].
5. Nakonechnyi, S. I., & Savina, S. S. (2003). Matematychne prohramuvannia: navch. posib. [Mathematical programming: textbook]. Kyiv: KNEU. [in Ukrainian].
6. Radchenko, O. P. (2023). Problems and perspectives of the functioning of national logistics systems on agricultural product markets. Market economy: modern management theory and practice, 21(2)(51), 126–136.
7. Akimov, D. D. (2023). Zastosuvannia tekhnolohii internetu rechei dlia optymizatsii lohistyky v silskomu hospodarstvi [Application of Internet of Things technologies for logistics optimization in agriculture]. Computer Integrated Technologies: Education, Science, Production, (53), 9–15. URL: https://doi.org/10.36910/6775-2524-0560-2023-53-02 [in Ukrainian].
8. Tang, C. S. (1988). A Max-Min Allocation Problem: Its Solutions and Applications. Operations Research, 36(2), 359-367.
9. Kopfer, H. W., & Kopfer, H. (2013). Emissions minimization vehicle routing problem in dependence of different vehicle classes. Dynamics in Logistics, 49-58.

© Акімов Д.Д. Гавриленко В.В.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-10
УДК 004.72
Багнюк Наталія Володимирівна, к.т.н., доцент
https://orcid.org/0000-0002-7120-5455
Лавренчук Світлана Василівна, к.т.н., доцент
https://orcid.org/0000-0002-5453-3924
Міскевич Оксана Іванівна, ст. викладач
https://orcid.org/0000-0002-5009-2391
Свистун Богдан Васильович, магістрант
Луцький національний технічний університет, м.Луцьк, Україна

РЕАЛІЗАЦІЯ СИСТЕМИ «РОЗУМНИЙ БУДИНОК» З ВИКОРИСТАННЯМ BLUETOOTH, ARDUINO ТА ANDROID: ВІД ДАТЧИКІВ ДО МОБІЛЬНОГО ІНТЕРФЕЙСУ

Багнюк Н. В., Лавренчук С. В., Міскевич О. І., Свистун Б. В. Реалізація системи «Розумний будинок» з використанням Bluetooth, Arduino та Android: від датчиків до мобільного інтерфейсу. Описана розробка програмного забезпечення для системи «Розумний будинок» на базі платформи Arduino Uno. Основна увага приділяється написанню скетчу для взаємодії з датчиками руху, температури та вологості, які забезпечують моніторинг і контроль умов у приміщенні. Створено мобільний додаток для операційної системи Android, що функціонує як Bluetooth термінал для збору, відображення та аналізу даних, отриманих від датчиків. Реалізовано інтерфейс для зручного налаштування та підключення пристроїв, а також забезпечено можливість візуалізації зібраної інформації в реальному часі. Проведено тестування системи в умовах реальної експлуатації, що підтвердило її стабільну роботу та точність вимірювань. У статті також наведено аналіз можливостей подальшого розширення функціоналу, таких як інтеграція з іншими системами автоматизації та підтримка додаткових типів датчиків. Представлена розробка має практичне значення для побудови інтелектуальних систем управління середовищем у житлових та комерційних приміщеннях.
Ключові слова: Arduino, розумний будинок, Bluetooth, мобільний додаток, датчики, автоматизація, контроль середовища

Bahniuk N., Lavrenchuk S., Miskevych O., Svystun B. Implementation of the «Smart House» system using Bluetooth, Arduino, and Android: from sensors to mobile interface. The article describes the development of software for the Smart Home system based on the Arduino Uno platform. The main focus is on writing a sketch for interaction with motion, temperature and humidity sensors that provide monitoring and control of indoor conditions. A mobile application for the Android operating system has been created that functions as a Bluetooth terminal for collecting, displaying and analysing data received from the sensors. An interface was implemented for easy configuration and connection of devices, as well as the ability to visualise the collected information in real time. The system has been tested in real-world operation, which has confirmed its stable operation and measurement accuracy. The article also analyses the possibilities of further expanding the functionality, such as integration with other automation systems and support for additional types of sensors. The presented development is of practical importance for the construction of intelligent environmental management systems in residential and commercial premises.
Keywords: Arduino, smart home, Bluetooth, mobile application, sensors, automation, environment control

Постановка наукової проблеми. Система «Розумний будинок» набуває все більшої популярності завдяки можливості автоматизації рутинних завдань і підвищенню рівня комфорту та безпеки житла. Одним з важливих аспектів таких систем є моніторинг навколишнього середовища в приміщенні та забезпечення своєчасного реагування на зміни умов, таких як температура, вологість та рух. Існуючі рішення, часто складні та дорогі, вимагають застосування спеціалізованого обладнання та програмного забезпечення. Таким чином, постає наукова задача розробити прототип доступної та простої у використанні системи «Розумний будинок» на базі мікроконтролера Arduino Uno з використанням модуля Bluetooth для передачі даних. Основною метою є створення системи, яка зможе ефективно моніторити рух, температуру та вологість у приміщенні і передавати ці дані на мобільний додаток, забезпечуючи зручний інтерфейс для користувача.

Аналіз останніх досліджень і публікацій. Існуючі дослідження у сфері систем «Розумний будинок» демонструють широкий спектр технологій та підходів до автоматизації домашнього середовища [1]. Більшість комерційних рішень включають складні системи з багатьма датчиками та контролерами, що взаємодіють через різні протоколи зв'язку, такі як Wi-Fi, ZigBee, або Z-Wave [2-3]. Проте ці рішення часто є дорогими та вимагають спеціалізованих знань для їх встановлення та налаштування. Arduino Uno, у поєднанні з доступними датчиками та модулями зв'язку, пропонує економічно ефективну та гнучку платформу для побудови системи «Розумний будинок». Використання модуля Bluetooth HC-05 для передачі даних забезпечує простоту підключення та управління пристроями за допомогою мобільних додатків [4]. Попередні дослідження показали, що

© Багнюк Н.В., Лавренчук С.В., Міскевич О.І., Свистун Б.В.

такий підхід дозволяє створювати функціональні прототипи систем моніторингу та управління домашнім середовищем з мінімальними витратами та складністю.

Виділення невирішених раніше частин загальної проблеми. Попри численні дослідження в області автоматизації та контролю, існує низка невирішених питань, що стосуються інтеграції платформи Arduino з різноманітними датчиками та передачею даних у реальному часі через Bluetooth-з'єднання. Однією з основних проблем є забезпечення стабільної і надійної передачі даних між датчиками, встановленими в системі «Розумний будинок», і мобільним додатком, який виконує роль терміналу для їхнього моніторингу. Традиційні рішення для систем домашньої автоматизації часто вимагають складних і дорогих апаратних компонентів або мережевих протоколів, таких як Wi-Fi або Ethernet. Однак вони можуть бути надмірними для невеликих систем контролю, де головна мета полягає в моніторингу декількох параметрів навколишнього середовища. Водночас використання Bluetooth для передачі даних у таких системах залишалося недостатньо дослідженим через обмежену дальність дії і складнощі забезпечення безперервного зв'язку в умовах перешкод і обмежень енергоспоживання.

Мета дослідження полягає у розробці та впровадженні прототипу системи «Розумний будинок» на базі плати Arduino Uno з використанням стандарту зв'язку Bluetooth для моніторингу та передачі даних про рух за допомогою інфрачервоного датчика руху SR312, температуру та вологість за допомогою датчика DHT11 [5-6]. Система повинна забезпечувати збір даних, надсилання зібраних даних через модуль Bluetooth на мобільний додаток, розроблений на платформі Android, відображення та інтерпретація даних в читабельному вигляді в додатку для користувача [7].

Основна частина дослідження. Для написання прошивки для плати Arduino Uno було використано інтегроване середовище розробки Arduino IDE. Це середовище спеціально створене для написання скетчів для плат сімейства Arduino і містить вбудований редактор коду, консоль, бібліотеки тощо. Процес написання програмного коду для мікроконтролера Arduino, який виконує роль головного модуля системи, починається з підключення бібліотек та створення констант. Бібліотека <DHT.h> необхідна для роботи з датчиком вологості і температури DHT11. Бібліотека <SoftwareSerial.h> потрібна для реалізації послідовного зв'язку на цифрових контактах плати Arduino Uno R3. Процес підключення бібліотек показано на рисунку 1.

```
#include <SoftwareSerial.h>
#include <DHT.h>
#include <DHT_U.h>
```

Рис. 1. Підключення бібліотек у середовищі розробки Arduino

Після підключення бібліотек небхідно ініціалізувати змінні з певними значеннями, а саме, присвоїти їм номера пінів на платі Arduino Uno R3, які будуть використовувати в роботі для підключення елементів системи до мікроконтролера, об'єкти класів для використання елементів, таких як датчики та адаптер стандарту зв'язку Bluetooth. Ініціалізуються піни RX та TX для Bluetooth-адаптера для реалізації обміну даними з іншими пристроями. RX та TX піни відповідають за прийом та відправку даних відповідно. Після цього створюється клас для Bluetooth-модуля. Це видно на рисунку 2.

```
5   // Визначення пінів і типу датчика
6   #define DHTPIN 4          // Пін, до якого підключений датчик DHT11
7   #define DHTTYPE DHT11     // Визначення типу датчика DHT11
8   #define MOTION_SENSOR_PIN 2 // Пін, до якого підключений датчик руху
9   #define LED_PIN 13        // Вбудований LED пін
10
11  // Визначення пінів для Bluetooth HC-05
12  #define BLUETOOTH_TX_PIN 0 // Пін для передачі даних Bluetooth
13  #define BLUETOOTH_RX_PIN 1 // Пін для прийому даних Bluetooth
14
```

Рис. 2. Присвоєння пінів для датчиків та адаптера Bluetooth

© Багнюк Н.В., Лавренчук С.В., Міскевич О.І., Свистун Б.В.

Функція setup() виконується один раз під час запуску програми на платі Arduino. Вона використовується для налаштування початкових параметрів та ініціалізації обладнання. Функція setup() у коді виконує наступні дії:

– Встановлює пін для зовнішнього світлодіода як вихідний. Встановлює пін для датчика руху як вхідний.

– Ініціалізує серійне з'днання для моніторингу.

– Ініціалізує датчик DHT11 для зчитування температури та вологості. Ініціалізує програмний серійний порт для Bluetooth з'єднання.

– Вимикає вбудований світлодіод.

Ці кроки забезпечують правильну роботу всіх компонентів системи перед початком виконання основного циклу програми. Функцію setup() можна побачити на рисунку 3.

```
19  void setup() {
20    pinMode(MOTION_SENSOR_PIN, INPUT);  //
21    pinMode(LED_PIN, OUTPUT);           //
22    Serial.begin(9600);                 //
23    bluetoothSerial.begin(9600);        //
24    dht.begin();                        //
25  }
```

Рис. 3. Функція setup()

Після налаштування необхідних компонентів у функції setup(), програма переходить до виконання основної логіки у циклі loop(). Під час кожного циклу виконуються кілька ключових завдань. Спочатку зчитується стан датчика руху за допомогою функції detectMotion(), яка визначає наявність руху у приміщенні. Якщо рух виявлено, вбудований світлодіод на платі Arduino вмикається, якщо ні – вимикається. Далі зчитуються значення вологості та температури з датчика DHT11 за допомогою функцій readHumidity() та readTemperature(). У разі успішного зчитування даних формується повідомлення, яке містить інформацію про стан руху, температуру та вологість, за допомогою функції createMessage(). Це повідомлення виводиться на серійний монітор для відлагодження та передається через модуль Bluetooth на мобільний додаток. Наприкінці циклу встановлюється затримка на 6 секунд перед початком нового циклу для уникнення помилок спрацювання. Функція loop() наведена на рисунку 4.

```
void loop() {
  // Зчитування значень з датчика руху
  int motionState = digitalRead(MOTION_SENSOR_PIN);

  if (motionState == HIGH) {  // Якщо датчик виявляє рух
    digitalWrite(LED_PIN, HIGH);  // Вмикаємо вбудований LED
    //Serial.println("Motion detected!");  // Виводимо повідомлення в серійний монітор
  } else {
    digitalWrite(LED_PIN, LOW);  // Вимикаємо вбудований LED
    //Serial.println("No motion.");  // Виводимо повідомлення в серійний монітор
  }

  // Затримка для стабілізації датчика DHT11
  delay(2000);

  // Зчитування значень з датчика DHT11
  float humidity = dht.readHumidity();
  float temperature = dht.readTemperature();

  // Перевірка на наявність помилок зчитування
  if (isnan(humidity) || isnan(temperature)) {
    Serial.println("Failed to read from DHT sensor!");
    return;
  }

  // Формування повідомлення для виведення в серійний монітор і передачі через Bluetooth
  String message = "Motion: ";
  if (motionState == HIGH) {
    message += "Motion detected, ";
  } else {
    message += "No motion, ";
  }
  message += "Humidity: " + String(humidity) + " %, Temperature: " + String(temperature) + " *C";
  Serial.println(message);
  bluetoothSerial.println(message);  // Відправка повідомлення через Bluetooth

  delay(6000);  // Невелика затримка для уникнення помилкових спрацьовувань
}
```

Рис. 4. Функція loop()

© Багнюк Н.В., Лавренчук С.В., Міскевич О.І., Свистун Б.В.

Після завершення написання коду прошивки для плати Arduino, на основі якої побудовано програмну систему «розумного будинку», наступним кроком є створення мобільного додатку для операційної системи Android. Для розробки мобільного додатку використано середовище Android Studio та мову програмування Java. Перед розглядом деталей процесу, визначається загальна структура мобільного додатку.

Спершу створюється конфігураційний файл для мобільного додатку, який називається «AndroidManifest.xml». Цей файл містить важливу інформацію про додаток та визначає його архітектуру, включаючи назву, іконку, дозволи для використання бібліотек, доступ до Інтернету, камери, Bluetooth тощо. Під час встановлення або роботи додатку користувачам буде показано запит на дозвіл. Маніфест також декларує всі основні робочі директорії додатку, такі як MainActivity, Service, Socket, BroadcastReceiver, BluetoothUnit, DevicesFragment, ContentProvider тощо. Це дозволяє системі Android знати, які компоненти є частиною додатку. Крім того, в маніфесті визначається мінімальна версія Android, на якій буде працювати додаток, інформація про сумісність з різними екранами та загальні метадані. Після створення конфігураційного файлу, створюються з графічних елементів додатку за допомогою вбудованої функції створення UI-елементів в Android Studio. В директорії res є файли формату xml, які описують колір, розмір, властивості UI-елементів. Це видно на рисунку 5.

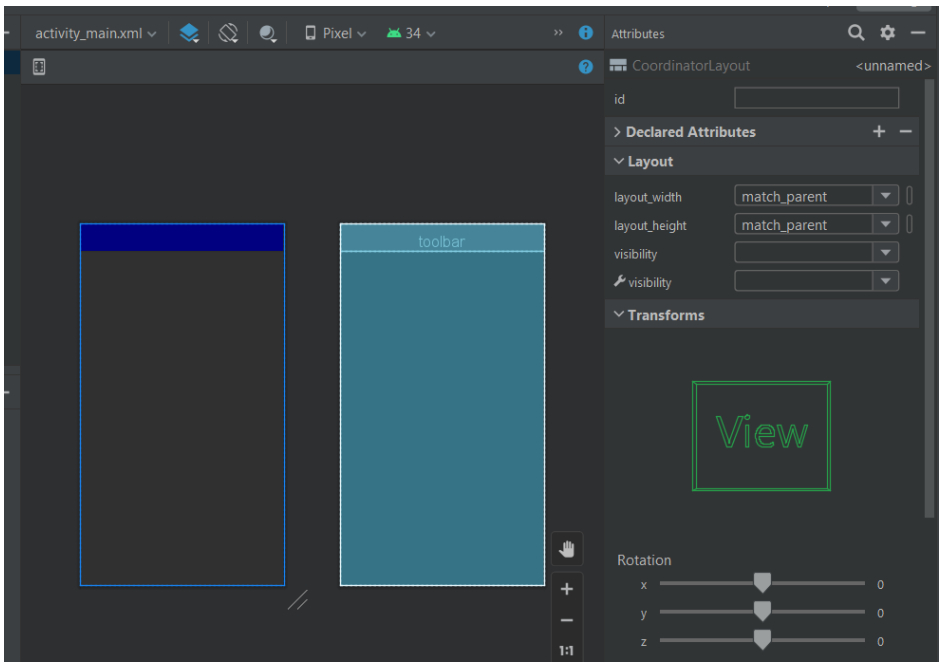


Рис. 5. Створення графічних елементів застосунку для вкладок додатку

Далі переходимо до папки MainActivity. З назви папки стає зрозуміло, що це основна папка нашого проєкту. Код в ній використовує фрагменти різного функціоналу додатку та обробляє навігацію між ними. На рисунку 6 імпортуються необхідні бібліотеки і класи для роботи з фрагментами (FragmentManager), базовою активністю (AppCompatActivity), тулбаром (Toolbar) та іншими компонентами.

```
import android.os.Bundle;
import androidx.fragment.app.FragmentManager;
import androidx.appcompat.app.AppCompatActivity;
import androidx.appcompat.widget.Toolbar;
```

Рис. 6. Імпорт необхідних бібліотек

© Багнюк Н.В., Лавренчук С.В., Міскевич О.І., Свистун Б.В.

Після імпорту бібліотек і класів прописується логіка взаємодії між елементами додатку. Реалізацію логіки для взаємодії елементів зображено на рисунку 7.

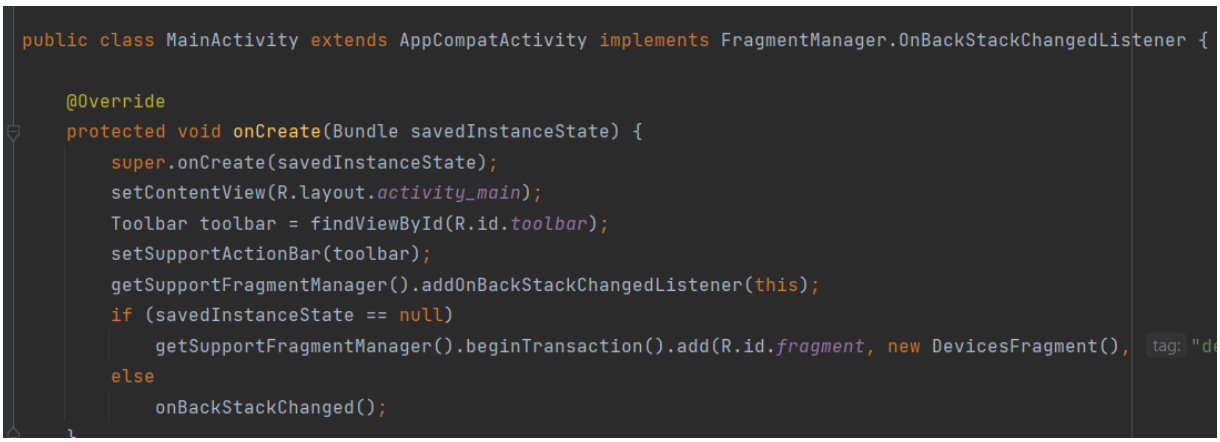


Рис. 7. Написання функції onCreate

Клас MainActivity розширює клас AppCompatActivity, забезпечуючи сумісність з попередніми версіями Android без конфліктів між старими і сучасними функціями операційної системи. Крім того, він реалізує інтерфейс FragmentManager.OnBackStackChangedListener, що дозволяє відстежувати зміни у бек-стеку фрагментів. Функція onCreate викликається при створенні активності та встановлює макет, визначений у файлі activity_main.xml, за допомогою setContentView(R.layout.activity_main).

Toolbar налаштовується як ActionBar за допомогою команди setSupportActionBar(toolbar) і додається перехоплювач змін у бек-стеку фрагментів.

Метод onBackStackChanged викликається при зміні бек-стеку фрагментів, в той час як метод onSupportNavigateUp обробляє натискання кнопки «Назад» в ActionBar. методом onBackPressed, який реалізований через булеві значення true і false. Реалізацію цих двох методів можна побачити на рисунку 8.

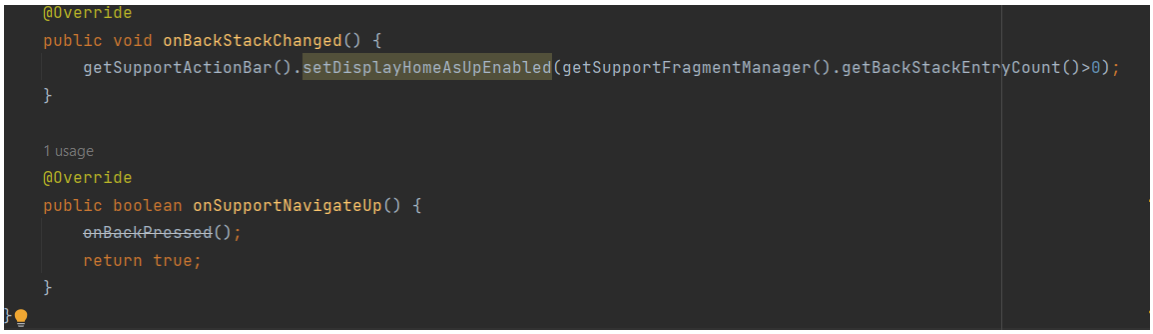


Рис. 8. Написання методів onBackStackChanged та onSupportNavigateUp

У процесі розробки системи «Розумний будинок» на базі плати Arduino Uno з використанням стандарту зв'язку Bluetooth використано два типи Bluetooth терміналів: кастомний термінал, створений власноруч, та стандартний термінал, завантажений із Google Play. Порівняльна характеристика кастомного і стандартного Bluetooth-терміналів наведена в таблиці 1.

Таблиця 1. Порівняльна характеристика кастомного і стандартного Bluetooth-терміналів

Характеристики	Кастомний Bluetooth-термінал	Стандартний Bluetooth-термінал

© Багнюк Н.В., Лавренчук С.В., Міскевич О.І., Свистун Б.В.

Інтерфейс користувача	Призначений для конкретних потреб проєкту, зручний та адаптований для керування системою	Загальний інтерфейс, призначений для роботи з різними пристроями, без можливості кастомізації під конкретний проект
Функціонал	Відображення даних з датчиків у реальному часі, можливість зворотнього зв'язку та налаштувань	Підключення до семи пристроїв одночасно, базове відображення даних, без спеціалізованих функцій
Гнучкість	Легке налаштування і розширення функціоналу відповідно до потреб проєкту	Обмежені можливості для кастомізації та розширення, складність внесення змін
Безпека	Реалізовані базові заходи безпеки, можливість додаткового налаштування під конкретні вимоги	Базові заходи безпеки, які можуть бути недостатніми для специфічних потреб
Підключення	Підключення через стандарт Bluetooth	Підключення через стандарт Bluetooth та Bluetooth Low Energy
Взаємодії з пристроями	Підключення до одного пристрою одночасно	Можливість підключення до 7 модулів Bluetooth одночасно

Візуалізацію порівняння характеристик двох терміналів можна побачити на рисунку 9.

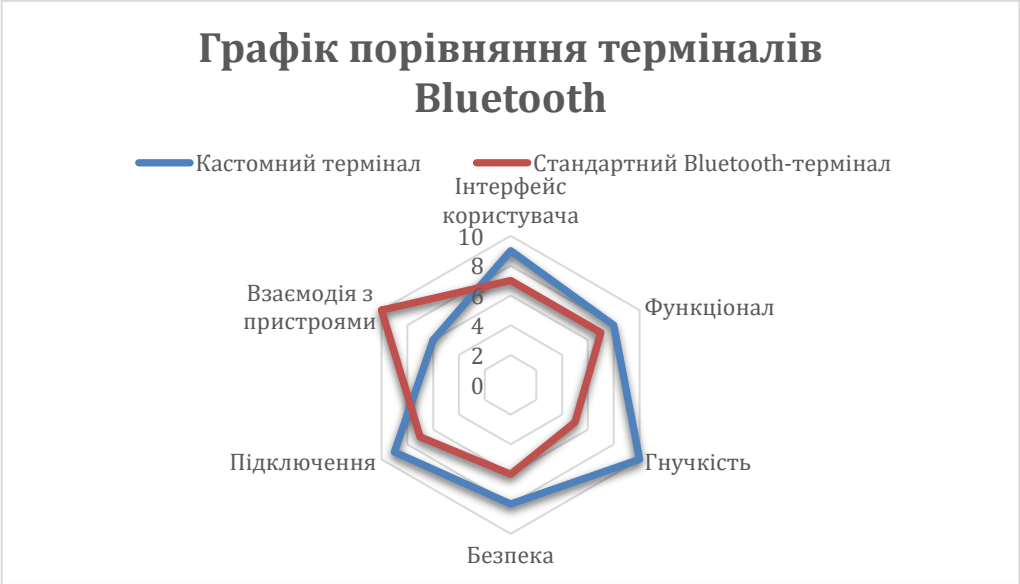


Рис. 9. Графік з порівнянням кастомного і стандартного терміналів

Висновки та перспективи подальшого дослідження. У ході виконання роботи успішно реалізовано побудову прототипу системи «Розумний будинок», що підтвердило можливість використання платформи Arduino в поєднанні з датчиками для моніторингу параметрів навколишнього середовища та передачі даних на мобільний додаток через Bluetooth-з'єднання. На основі Arduino IDE створено скетч-програму, яка забезпечує стабільну роботу датчиків і передачу

© Багнюк Н.В., Лавренчук С.В., Міскевич О.І., Свистун Б.В.

інформації в структурованому вигляді на мобільний додаток. Подальша розробка включала створення мобільного додатку для операційної системи Android з використанням Android Studio і мови програмування Java. В результаті успішно скомпільовано білд програми, який встановився на телефон і забезпечує коректну взаємодію з платою Arduino, приймаючи дані від датчиків у реальному часі. Порівняння кастомного Bluetooth терміналу, розробленого спеціально для системи «Розумний будинок», зі стандартним терміналом із Google Play показує, що для специфічних проєктів, таких як система «Розумний будинок», кастомний термінал є більш оптимальним вибором завдяки своїй гнучкості, адаптивності та можливості тонкого налаштування під конкретні завдання. Водночас, стандартний термінал може бути корисним для загальних потреб і тих користувачів, які шукають просте рішення без необхідності вносити зміни у функціонал. Перспективи подальшого дослідження включають розширення функціональності системи «Розумний будинок» шляхом інтеграції додаткових датчиків і пристроїв для автоматизації різних аспектів управління середовищем. Також варто дослідити можливості впровадження інших бездротових технологій, таких як Wi-Fi або Zigbee, для розширення радіусу дії та забезпечення більш надійного зв'язку.

Список бібліографічного опису

1.Monk S. Programming Arduino: Getting Started with Sketches, Third Edition, New York : McGraw-Hill, 2023. 176 p.
2.IoT Applications, Security Threats, and Countermeasures / ed. by P. Nayak, N. Ray, P. Ravichandran; Boca Raton : CRC Press, 2021. 277 p.
3.Кіберфізичні системи: багаторівнева організація та проектування / А. О. Мельник та ін. Київ : Магнолія, 2023. 238 с.
4. Arduino and HC-05 Bluetooth Module Complete Tutorial. *How To Mechatronics*. URL: https://howtomechatronics.com/tutorials/arduino/arduino-and-hc-05-bluetooth-module-tutorial/ (date of access: 06.07.2024).
5. DHT11–Temperature and Humidity Sensor. *Components101*. URL: https://components101.com/sensors/dht11-temperature-sensor (date of access: 15.07.2024).
6. SR312 Інфрачервоний Датчик Руху *Мій Проект*. URL: https://myproject.com.ua/hc-sr312-miniatjurnij-infrachervonij-datchik-ruhu-ua.html (дата звернення: 23.07.2024).
7. Fazio M. Kotlin and Android Development Featuring Jetpack: Build Better, Safer Android Apps. O'Reilly Media, Incorporated, 2021.

References

1.Monk S. Programming Arduino: Getting Started with Sketches, Third Edition, New York : McGraw-Hill, 2023. 176 p.
2.IoT Applications, Security Threats, and Countermeasures / ed. by P. Nayak, N. Ray, P. Ravichandran; Boca Raton : CRC Press, 2021. 277 p.
3.Кіберфізичні системи: багаторівнева організація та проектування / А. О. Мельник та ін. Київ : Магнолія, 2023. 238 с.
4. Arduino and HC-05 Bluetooth Module Complete Tutorial. *How To Mechatronics*. URL: https://howtomechatronics.com/tutorials/arduino/arduino-and-hc-05-bluetooth-module-tutorial/ (date of access: 06.07.2024).
5. DHT11–Temperature and Humidity Sensor. *Components101*. URL: https://components101.com/sensors/dht11-temperature-sensor (date of access: 15.07.2024).
6. SR312 Інфрачервоний Датчик Руху *Мій Проект*. URL: https://myproject.com.ua/hc-sr312-miniatjurnij-infrachervonij-datchik-ruhu-ua.html (дата звернення: 23.07.2024).
7. Fazio M. Kotlin and Android Development Featuring Jetpack: Build Better, Safer Android Apps. O'Reilly Media, Incorporated, 2021.

© Багнюк Н.В., Лавренчук С.В., Міскевич О.І., Свистун Б.В.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-11
УДК 004.72
Баранчук Сергій Анатолійович, студент
Бортник Катерина Яківна, к.т.н., доцент
https://orcid.org/0000-0001-5282-099x
Луцький національний технічний університет, м. Луцьк, Україна

СИСТЕМА ДЛЯ МОНІТОРИНГУ ВІДКЛЮЧЕНЬ СВІТЛА НА БАЗІ ПЛАТФОРМИ ARDUINO ТА ФРЕЙМВОРКУ REACT NATIVE

Баранчук С.А., Бортник К.Я. Система для моніторингу відключень світла на базі платформи Arduino та фреймворку React Native. У цій статті розглянуто розробку та впровадження системи для моніторингу відключень електроенергії на основі сучасних технологій Інтернету речей (IoT), яка поєднує в собі платформу Arduino та фреймворк React Native. Система складається з чотирьох ключових компонентів: мікроконтролера NodeMCU, який відповідає за збір даних про наявність електроенергії, серверної частини на базі Node.js, яка обробляє запити і керує базою даних, хмарного сервісу Firebase для збереження даних та надання функціоналу push-сповіщень, а також мобільного додатка для Android, який дозволяє користувачам отримувати актуальну інформацію про стан електропостачання. Основний функціонал системи полягає у фіксації часу та місця відключення електроенергії в автоматичному режимі, що забезпечує надійність та своєчасність передачі інформації користувачам через мобільний додаток. Завдяки інтеграції з Firebase, дані зберігаються у хмарі, що дозволяє доступ до них навіть у випадку локальних збоїв інтернет-зв'язку. Також у статті детально описані технічні аспекти розробки, включно з принципами використання необхідного ПЗ та додаткових бібліотек, а також механізми надсилання push-сповіщень через Firebase Cloud Messaging. Особливу увагу приділено використанню сучасних підходів до управління станом React компонентів у мобільному додатку, що дозволяє динамічно оновлювати інтерфейс користувача на основі зміни даних у режимі реального часу. Запропонована система є важливим кроком у розвитку технологій моніторингу електропостачання, а також має перспективи подальших досліджень, зокрема, в напрямку оптимізації споживання енергії пристроями IoT, підвищення безпеки передачі даних та розширення функціоналу.
Ключові слова: Інтернет речей, веб-технології, NodeMCU, Firebase, React Native, мобільна розробка

Baranchuk S., Bortnyk K. A system for monitoring power outages based on the Arduino platform and the React Native framework. The article discusses the development and implementation of a system for monitoring power outages based on modern Internet of Things (IoT) technologies, combining the Arduino platform and the React Native framework. The system consists of four key components: a NodeMCU microcontroller responsible for collecting data on electricity availability, a server-side component based on Node.js that processes requests and manages the database, a Firebase cloud service for data storage and push notification functionality, and a mobile Android application that allows users to receive up-to-date information about the status of power supply. The system's main functionality is to automatically record the time and location of power outages, ensuring reliability and timely transmission of information to users through the mobile app. Due to integration with Firebase, data is stored in the cloud, allowing access even in case of local internet disruptions. The article also provides a detailed description of the technical aspects of the development, including the principles of using the necessary software and additional libraries, as well as the mechanisms for sending push notifications via Firebase Cloud Messaging. Special attention is given to the use of modern state management approaches in React components in the mobile application, which enables the dynamic updating of the user interface based on real-time data changes. The proposed system represents a significant step forward in the development of power supply monitoring technologies and holds potential for further research, particularly in the areas of optimizing energy consumption by IoT devices, enhancing data transmission security, and expanding system functionality.
Keywords: Internet of Things, web technologies, NodeMCU, Firebase, React Native, mobile development

Постановка наукової проблеми. Через обстріли критичної інфраструктури в Україні, проблема відключень електроенергії стала особливо актуальною. Масштабні та непередбачувані відключення світла дуже негативно впливають на життя громадян та функціонування бізнесу, тому в умовах нестабільного енергопостачання важливо мати ефективні засоби для моніторингу ситуації та швидкого реагування на відключення.

Технології в сфері Інтернету речей (IoT) дають хороші можливості для вирішення цього питання. Застосування IoT дає змогу створити автоматизовану систему для моніторингу електропостачання в реальному часі, що дозволить відстежувати наявність або відсутність електроенергії практично в будь-якому місці.

Аналіз досліджень. Термін «Інтернет речей» (IoT) описує колективну мережу підключених пристроїв і технологій, які забезпечують зв'язок між цими пристроями та хмару, яка виступає в якості агрегатора отриманих даних.

Інтернет речей інтегрує повсякденні «речі» з інтернетом. Ще в 90-х роках інженери почали додавати сенсори та процесори до повсякденних об'єктів, однак спочатку прогрес був доволі повільним, оскільки чіпи були великими і громіздкими. З часом обчислювальні пристрої зменшилися в розмірах, а вартість інтеграції обчислювальної потужності в маленькі об'єкти також

© Баранчук С.А., Бортник К.Я.

значно знизилася. Як наслідок, виникла ціла індустрія, яка зосереджена на наповненні обчислювальними пристроями цілих будинків, бізнесів та офісів (рис. 1). Усі ці пристрої та технології, що пов'язані з ними, колективно називаються Інтернетом речей.

Рис. 1. Сфери застосування Інтернету речей

Типова IoT система працює через збір та обмін даними в режимі реального часу та, зазвичай, складається з наступних трьох компонентів:

– Мікроконтролер та додаткові периферійні пристрої, що збирають дані з навколишнього середовища, слідкують за діями користувачів, тощо, та передають отримані дані через інтернет до IoT застосунку та отримують дані у відповідь.

– IoT-застосунок – це набір сервісів і програмного забезпечення, що інтегрує та обробляє дані, отримані звід мікроконтролерів.

– Графічний інтерфейс користувача – це, фактично, «шлюз» між користувачем та IoT-системою. Зазвичай це мобільний застосунок або вебсайт, через які можна керувати мікроконтролерами та працювати з отриманими даними.

Розроблена мною IoT-система має схожу структуру, але окрім застосунку у вигляді HTTP сервера, присутня також платформа Firebase, що використовується у якості бази даних та агрегатора push-сповіщень.

Використання такої структури зумовлено тим, що система повинна повністю виконувати наступні завдання:

– Фіксувати місце та час відключення та відновлення електропостачання, надаючи детальну інформацію для аналізу.

– Сповіщати користувачів шляхом відправлення push-сповіщень на мобільні додатки, у разі виникнення проблем з електропостачанням.

– Забезпечувати надійність зібраних даних, шляхом збереження їх у хмарі, що забезпечить доступ до інформації навіть у випадку локальних перебоїв у зв'язку.

Виклад основного матеріалу й обгрунтування отриманих результатів дослідження. Як було сказано попередньо, архітектура системи складається з чотирьох основних компонентів. В якості мікроконтролеру була використана модель NodeMCU – китайський, дешевий аналог Arduino, з вбудованим модулем Wi-Fi (рис. 2).

© Баранчук С.А., Бортник К.Я.

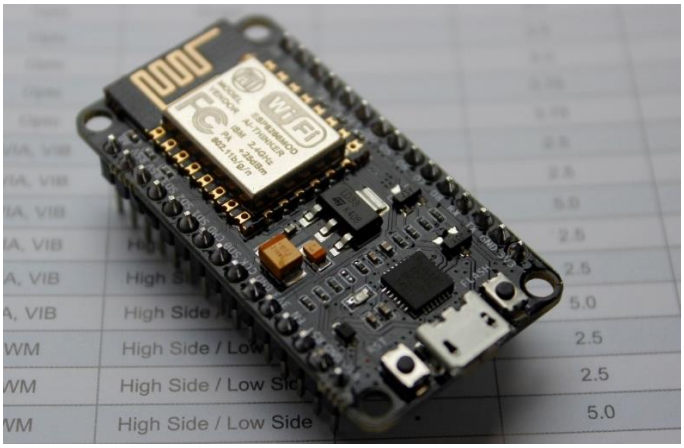


Рис. 2. Вигляд мікроконтролеру NodeMCU

Мікроконтролер відповідає за збір даних про наявність електроенергії шляхом постійного підключення до електромережі та інтернет-мережі. Кожні 7 секунд NodeMCU надсилає HTTP-запит на сервер, який в свою чергу потім оновлює поле з даними про наявність електроенергії на вказаній локації, вказуючи статус «ON». Коли електроенергія зникає, мікроконтролер відповідно вимикається і перестає надсилати запити, а сервер, не отримуючи запит від мікроконтролеру більше ніж 7 секунд, оновлює поле зі статусом про наявність світла на локації значенням «OFF».

IoT-застосунком у даному випадку буде виступати програмне забезпечення, що використовує Node.js для створення Express-сервера, який відповідає за обробку отриманих з мікроконтролера запитів, збереження інформації у базу даних платформи Firebase та відправки push-сповіщення на мобільні пристрої користувачів.

Для своєї роботи сервер використовує ряд сторонніх бібліотек та залежностей, таких як, наприклад, dotenv (для завантаження змінних середовища з файлу .env), dayjs (бібліотека для роботи з датами та часом), firebase-admin (SDK для підключення до Firebase), request (бібліотека для відправки HTTP-запитів).

При запуску файлу server.js, в якому описаний весь код застосунку, першочергово відбувається підключення до платформи Firebase за допомогою облікових даних, вказаних у файлі serviceKey.json, та ініціалізація бази даних Firestore. Далі створюється Express веб-сервер, що працює на порті 3000, та ряд необхідних масивів та змінних: collectionsList (масив для зберігання списку колекцій з Firestore), collectionsData (об'єкт для зберігання даних по кожній колекції), lastRequestTime (час останнього запиту), inactivityCheckInterval (інтервал для перевірки неактивності).

Основний функціонал сервера полягає у тому, що функція checkInactivity() перевіряє, чи минуло більше 10 секунд з часу останнього запиту для кожної колекції. Колекція в даному випадку представляє собою кожну унікальну адресу. Якщо колекція неактивна, сервер надсилає сповіщення на мобільні пристрої через Firebase Cloud Messaging (FCM) (рис. 2) про відключення електрики на відповідній адресі, а статус колекції оновлюється в Firestore до "OFF" (рис. 4), і в історії додається новий запис про відключення (рис. 5).

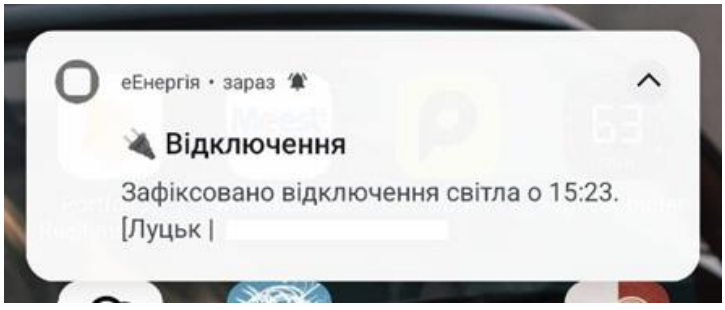


Рис. 3. Вигляд сповіщення

© Баранчук С.А., Бортник К.Я.

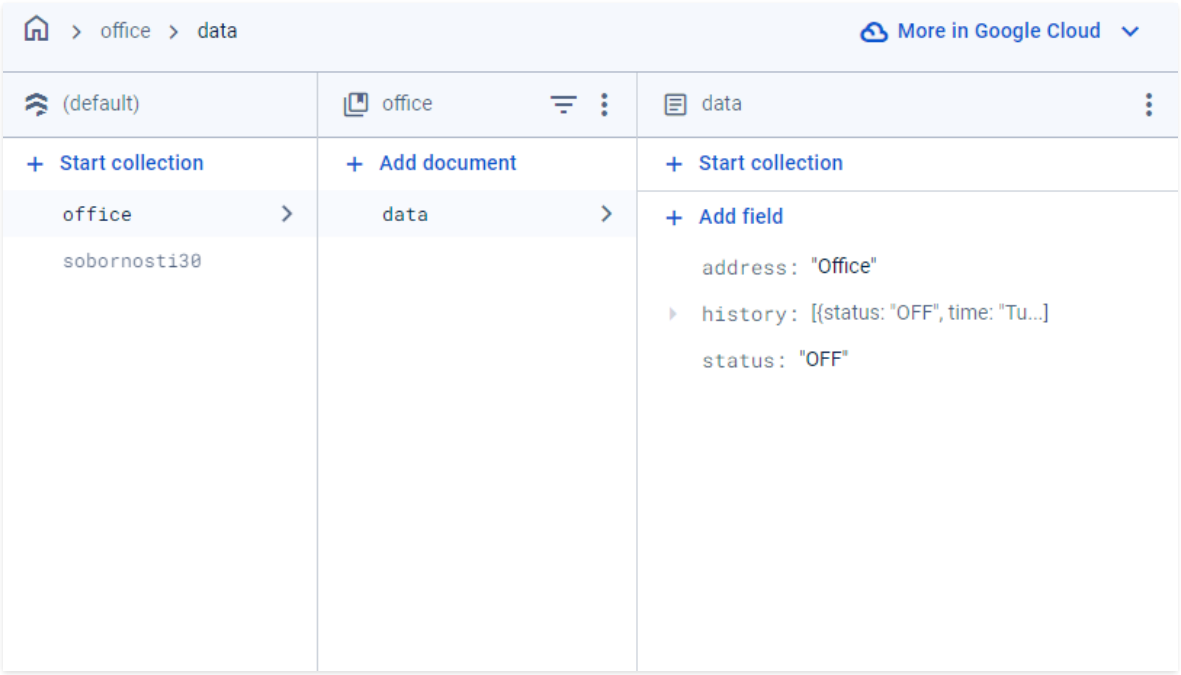


Рис. 4. Вигляд бази даних Firestore

Рис. 5. Історія записів про зміну статусу наявності електроенергії в Firestore

Мобільний додаток розроблений з використанням фреймворку React Native, є інтерфейсом для взаємодії користувачів із системою. Він дозволяє переглядати дані про стан електропостачання, отримувати сповіщення про відключення та аналізувати історію подій (рис. 6).

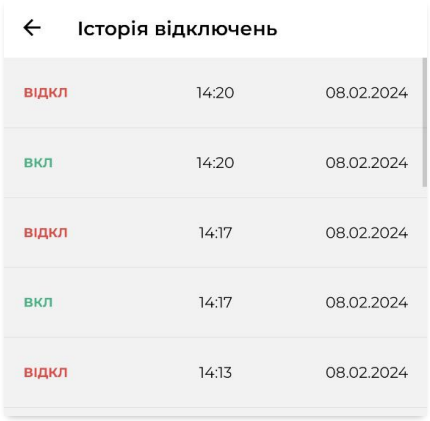


Рис. 6. Історія записів про зміну статусу наявності електроенергії в мобільному додатку

Варто також зауважити, що відображення поточного стану наявності електроенергії було відображене шляхом перемикання світлої та темної теми додатку, тобто при наявності

© Баранчук С.А., Бортник К.Я.

електроенергії тема додатку буде світлою, а при її відсутності – темною (рис. 7), та на екрані буде показаний відповідний текст.

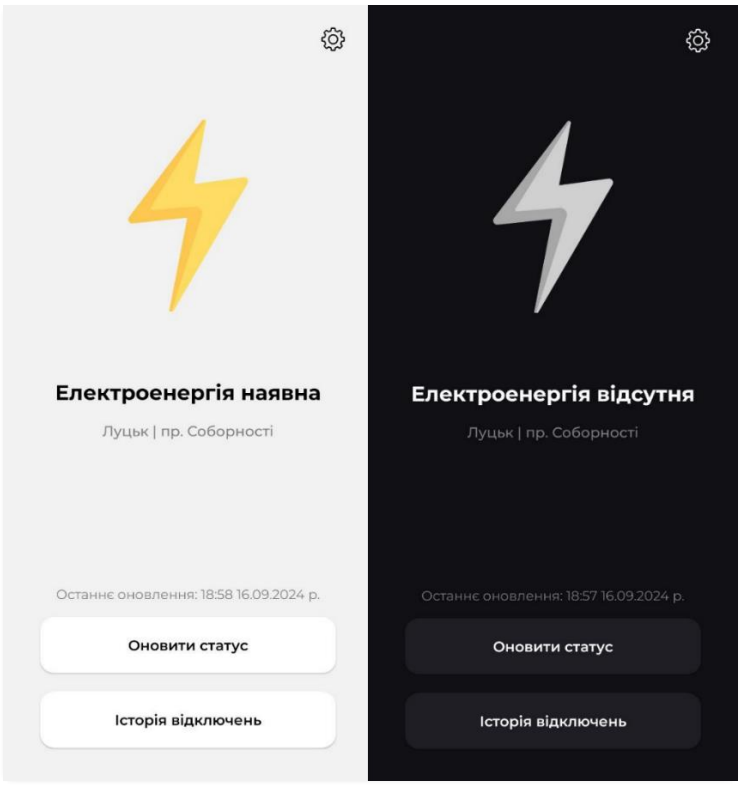


Рис. 7. Зміна теми головного екрану залежно від статусу наявності електроенергії

Під час розробки мобільного додатку були використані сучасні підходи управління станом React компонентів через хуки useState і useEffect, забезпечуючи динамічне оновлення інтерфейсу на основі змін у даних і темі додатку.

Під час завантаження показується сплеш-екран, а у фоні відбуваються асинхронні операції з отримання необхідних даних та збереження їх у локальне сховище. Після цього дані передаються в стан додатку за допомогою хуків, що дозволяє відслідковувати зміни в базі даних Firebase і оновлювати інтерфейс у реальному часі (рис. 8). Наприклад, ключ addressKey, отриманий з локального сховища AsyncStorage, використовується для доступу до конкретної колекції в Firestore, а зміни в ній відслідковуються за допомогою функції onSnapshot. Час останнього оновлення даних зберігається в стані, і цей показник регулярно оновлюється завдяки використанню бібліотеки dayjs.

```
useEffect(() => {
    // @ts-ignore
    setTheme(data.status === "OFF" ? "dark" : "light");
}, [data]);

useEffect(() => {
    if (addressKey) {
        onSnapshot(collection(firebase, addressKey), snapshot => {
            //console.log("Data updated");
            // @ts-ignore
            setData(snapshot.docs[0].data());
            setLastUpdateTime(
                dayjs(new Date()).locale("ru").format("HH:mm DD.MM.YYYY")
            );
        });
    }
}, [refreshTrigger, addressKey]);
```

Рис. 8. Лістинг коду, що відповідає за динамічне оновлення даних з Firebase

© Баранчук С.А., Бортник К.Я.

Висновки та перспективи подальшого дослідження. У ході дослідження було успішно розроблено і впроваджено систему моніторингу відключень електроенергії на основі технологій Інтернету речей (IoT). Система дозволяє оперативно фіксувати події відключення та відновлення електропостачання на локальних об'єктах, що критично важливо в умовах частих перебоїв енергопостачання. Інтеграція з платформою Firebase забезпечує надійне зберігання та обробку даних, а використання технології push-сповіщень дає змогу користувачам швидко реагувати на будь-які зміни у постачанні електроенергії.

Подальші дослідження можуть бути спрямовані на реалізацію та покращення існуючого функціоналу додатку, як наприклад: оптимізація споживання енергії самими пристроями IoT, інтеграція системи в більші інфраструктурні проекти, можливість відправки звітів або зворотного зв'язку від користувачів, а також підвищення рівня безпеки, шифрування даних і захисту від несанкціонованого доступу до системи.

Загалом, розроблена система є основою для створення масштабованого і надійного рішення для моніторингу електропостачання, що має значний потенціал для розвитку в багатьох напрямках.

Список бібліографічного опису

1. Баранов О. А. Інтернет речей (IoT) і блокчейн. Інформація і право. 2018. № 1 (24). С. 59–71.
2. Юскович-Жуковська В. І., Лотюк Ю. Г., Соловей Л. Я. Технології Інтернет речей в освіті. Дослідження інновацій та перспективи розвитку науки і техніки у XXI столітті. 2022. URL: https://doi.org/10.36059/978-966-397-278-7-25 (дата звернення: 16.09.2024).
3. «Інтернет речей» – як перспектива розвитку мобільних систем / Г. І. Сокол. 2021. № 1. С. 49–57. URL: https://doi.org/10.53920/its-2021-1-6 (дата звернення: 16.09.2024).
4. Жеребцов О.А, Іжиков А.Ю Використання функціонального програмування у JavaScript та фреймворках. Інноваційні рішення в інженерії програмного забезпечення. 2022. №9. С. 162–169.
5. Романько Р. О., Городецька О. С. Оптимізація мобільних застосунків за допомогою фреймворку React Native. Матеріали LIII науково-технічної конференції підрозділів ВНТУ, Вінниця, 20-22 березня 2024 р. Електрон. текст. дані. 2024. URI: https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2024/paper/view/19953 (дата звернення: 16.09.2024).

References

1. Baranov, O. A. (2018). Internet of Things (IoT) and Blockchain. Information and Law, (1)24, 59–71. [in Ukrainian].
2. Yuskovych-Zhukovska, V. I., Lotyuk, Y. H., & Solovey, L. Y. (2022). Internet of Things Technologies in Education. In Research on Innovations and Prospects for the Development of Science and Technology in the XXI Century. https://doi.org/10.36059/978-966-397-278-7-25 [in Ukrainian].
3. Sokol, H. I. (2021). Internet of Things as a Development Prospect for Mobile Systems. ITS Journal, (1), 49–57. https://doi.org/10.53920/its-2021-1-6 [in Ukrainian].
4. Zherebtsov, O. A., & Izhikov, A. Y. (2022). Use of Functional Programming in JavaScript and Frameworks. Innovative Solutions in Software Engineering, (9), 162–169. [in Ukrainian].
5. Romanko, R. O., & Horodetska, O. S. (2024). Optimization of Mobile Applications Using the React Native Framework. In Proceedings of the LIII Scientific and Technical Conference of VNTU Departments, Vinnytsia, March 20-22, 2024. https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2024/paper/view/19953 [in Ukrainian].

© Баранчук С.А., Бортник К.Я.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-12
УДК 004.4`2
Бегліцов Сергій Валерійович, аспірант
http://orcid.org/0009-0000-9690-9881
Державний університет інформаційно-комунікаційних технологій, м. Київ, Україна

АВТОМАТИЗАЦІЯ МІГРАЦІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ У ХМАРНУ АРХІТЕКТУРУ ІЗ ВИКОРИСТАННЯМ ІНСТРУМЕНТУ ІНФРАСТРУКТУРИ ЯК КОД TERRAFORM У СЕРЕДОВИЩІ AWS

Бегліцов С.В. Автоматизація міграції програмного забезпечення у хмарну архітектуру із використанням інструменту інфраструктури як код Terraform у середовищі AWS. У цій статті розглядається комплексний підхід до автоматизації надання інфраструктури за допомогою Terraform і AWS. Основною метою цього дослідження було оптимізувати розгортання та керування хмарними ресурсами за допомогою практик інфраструктури як коду (IaC). Стаття починається з пояснення фундаментальних принципів інфраструктури як коду, підкреслюючи її переваги з точки зору послідовності, повторюваності та співпраці. Центральне місце в цій методології займає Terraform, інструмент із відкритим кодом, розроблений для декларативного керування конфігурацією. Здатність Terraform визначати компоненти інфраструктури за допомогою простих, зрозумілих для людини файлів конфігурації забезпечує потужний механізм для надання та керування хмарними ресурсами в різних провайдерів. Ключовим моментом реалізації є використання AWS як хмарного провайдера. AWS пропонує широкий спектр послуг, від обчислювальних екземплярів (EC2) до керованих баз даних (RDS) і безсерверних обчислень (Lambda). У статті розглядаються конкретні випадки використання, коли Terraform використовується для автоматизації надання ресурсів AWS, таких як екземпляри EC2 для масштабованих веб-додатків, бакети S3 для зберігання об'єктів і конфігурації VPC для ізоляції мережі. Важливим аспектом, який обговорюється в статті, є керування станом Terraform, який відстежує поточний стан розгорнутих ресурсів і полегшує співпрацю між членами команди. Стратегії обробки файлів стану, включно з конфігураціями віддаленої серверної частини за допомогою Amazon S3 для блокування, вивчаються, щоб забезпечити узгодженість і уникнути конфліктів під час одночасного розгортання. Крім того, стаття стосується модулів Terraform, які інкапсулюють багаторазові конфігурації для різних компонентів інфраструктури, сприяючи багаторазовому використанню коду та зручності обслуговування. Завдяки модульній конфігурації команди можуть стандартизувати найкращі практики та застосовувати політики в різних середовищах, від розробки до виробництва. Інтеграція Terraform з конвеєрами CI/CD автоматизує тестування та розгортання змін інфраструктури, сприяючи гнучкості та зменшуючи ручне втручання. Ця автоматизація не тільки прискорює доставку оновлень інфраструктури, але й підвищує загальну надійність системи за допомогою автоматизованого тестування та механізмів відкату.
Ключові слова: Terraform, AWS, інфраструктура як код, керування хмарою, автоматизація, конвеєр

Behlitsov S. Software migration to a cloud architecture automation using an infrastructure as code tool Terraform AWS environment. This article explores a comprehensive approach to automating infrastructure provisioning using Terraform and AWS. The primary objective of this study was to streamline the deployment and management of cloud resources through Infrastructure as Code (IaC) practices. The article begins by elucidating the fundamental principles of Infrastructure as Code, highlighting its benefits in terms of consistency, repeatability, and collaboration. Central to this methodology is Terraform, an open-source tool designed for declarative configuration management. Terraform's ability to define infrastructure components using simple, human-readable configuration files provides a powerful mechanism for provisioning and managing cloud resources across different providers. Key to the implementation is the utilization of AWS (Amazon Web Services) as the cloud provider. AWS offers a vast array of services, from compute instances (EC2) to managed databases (RDS) and serverless computing (Lambda). The article delves into specific use cases where Terraform is leveraged to automate the provisioning of AWS resources, such as EC2 instances for scalable web applications, S3 buckets for object storage, and VPC configurations for network isolation. A critical aspect discussed in the article is the management of Terraform state, which tracks the current state of deployed resources and facilitates collaboration among team members. Strategies for handling state files, including remote backend configurations with Amazon S3 for locking, are explored to ensure consistency and avoid conflicts in concurrent deployments. Furthermore, the article addresses advanced topics such as Terraform modules, which encapsulate reusable configurations for different components of an infrastructure, promoting code reusability and maintainability. By modularizing configurations, teams can standardize best practices and enforce policies across various environments, from development to production. Integrating Terraform with CI/CD pipelines automates the testing and deployment of infrastructure changes, promoting agility and reducing manual intervention. This automation not only accelerates the delivery of infrastructure updates but also enhances overall system reliability through automated testing and rollback mechanisms.
Key words: Terraform, AWS, infrastructure as code, cloud management, automation, pipeline

Вступ та постановка проблеми. Хмарна архітектура програмного забезпечення представляє ряд важливих переваг, одними з яких є низькі капіталовкладення та швидкі можливості входу. Тим не менш, розробка алгоритму міграції певної програми на архітектуру хмарного середовища з наземної інфраструктури, представляє певний виклик, як і будь-який звичайний проект розробки програмного забезпечення. Наприклад, вибір хмарного провайдера або адаптація архітектури та вихідного коду. Інша проблема полягає в тому, щоб адаптувати програму так, щоб вона також мала специфічні для хмари функції, такі як багатокористування або масштабованість.

© Бегліцов С.В.

Аналіз останніх досліджень і публікацій. У науково-дослідницькому просторі сьогодення з'являються роботи, присвячені розробці методів та засобів автоматизації процесів переключення сервісів наземної інфраструктури в хмарне середовище.

У роботі [1] представлено результати дослідження продуктивності та доступності двох інструментів «Інфраструктура як код» (IaC) – Google Cloud Deployment Manager і Terraform. У дослідженні оцінюється розгортання та керування хмарними ресурсами за допомогою цих інструментів, а також розглядаються переваги їх інтеграції, гнучкість і мультихмарні можливості. Вона підкреслює незмінно швидший час знищення ресурсів Terraform порівняно з Google Cloud Deployment Manager, про що свідчать тести з використанням команди «time» Linux, надаючи розуміння їхньої операційної ефективності та адаптивності в хмарних середовищах. Висновки свідчать про те, що рідний Google Cloud Deployment Manager краще інтегрується з його ресурсами. На противагу цьому Terraform є універсальним, забезпечує кращу гнучкість і швидше розгортається. Його незалежність більш корисна при роботі з різними хмарними середовищами – і не обмежується GCP. Крім того, щоб отримати розуміння, у цьому дослідженні обговорюються переваги та внутрішні проблеми, пов'язані з IaC. Підкреслюється важливість технічного досвіду для ефективного використання цих інструментів для управління хмарною інфраструктурою.

Дослідження дисертації [2], проведене AIT GmbH & Co. KG, зосереджено на перенесенні їхнього власного плагіна TFS ASAP із локального середовища на хмарну платформу Microsoft, зокрема Visual Studio Online (VSO) на Microsoft Azure. Керуючись такими перевагами хмарних обчислень, як скорочення капіталовкладень і масштабованість, AIT прагне використовувати можливості хмари та інтегрувати TFS якнайшвидше у VSO, забезпечуючи модулі обробки подій із можливістю використання кількох орендарів і масштабування. У дисертації досліджуються існуючі методології хмарної міграції та адаптуються вони відповідно до міграції TFS якнайшвидше. Висвітлюються такі проблеми, як вибір хмарних провайдерів, адаптація архітектури та вихідного коду, а також забезпечення реалізації специфічних для хмари функцій, таких як багатокористування та масштабованість. Стратегічне узгодження AIT з Microsoft диктує використання служб Microsoft Azure для цієї міграції. Перехід TFS вважався успішним, незважаючи на те, що він не повністю відповідав всім вимогам до функціональності та тестування, викладеним спочатку. TFS ASAP Online успішно розгорнуто на хмарній платформі Microsoft, що робить його доступним у всьому світі. Основні функціональні вимоги були реалізовані. Протягом усього процесу міграції AIT встановив контакти зі співробітниками Microsoft, у результаті чого TFS ASAP Online було офіційно внесено до списку як розширення для Visual Studio Online. Спочатку проект у рамках дипломної роботи TFS ASAP Online перетворився на повноцінний продукт у портфоліо AIT.

Крім того, варто зазначити праці наступних науковців: Чинтапатла С. [3], Джонсон А. [4], Рахман А., Парто А., Моррісон П., Вільямс Л. [5], Сміт Л. [6], Далла Пальма С., Ді Нуччі Д., Паломба Ф., Тамбуррі Д.А. [7], Браун М. [8], Оздоган Е., Серан О., Устюндаг М. [9], Фейтоса Д., Пенча М.-Т., Берарді М., Боза Р.-Д. [10], Бату Дж. [11], Абулкішик А., Алван А., Гульзар Ю. [12], Елеракі М., Аніс Азіз Ст., Соліман Дж. [13], Мухопадьяй Н., Теварі Би. 14], Омофоєва Ю., Гребі А. [15] та інших.

Проте, беручи до уваги вище зазначену наукову документацію, питання, пов'язане з методологією по розробці методів та засобів автоматизації процесів переключення сервісів наземної інфраструктури в хмарне середовище, все ще залишається недостатньо дослідженим та потребує подальшого опрацювання.

Постановка завдання. Метою роботи є розробка методів та засобів автоматизації процесів переключення сервісів наземної інфраструктури в хмарне середовище із використанням інструменту інфраструктури як код Terraform у середовищі AWS.

Викладення основного матеріалу дослідження. Архітектура на основі хмарних обчислень зазвичай характеризується п'ятьма рисами:

– самообслуговування на вимогу, відповідно якому користувач може автоматично вимагати необхідні ресурси, коли це необхідно;

– широкий доступ до мережі, що означає здійснення доступу до ресурсів через Інтернет за допомогою стандартних механізмів;

– об'єднання ресурсів, завдяки якому ресурси управляються в пулі у віртуалізованій формі та можуть використовуватися кількома користувачами паралельно;

– швидка еластичність, яка передбачає можливість надання додаткових ресурсів користувачеві за потреби та звільнення, коли вони більше не потрібні;

© Бегліцов С.В.

– розмірене обслуговування, згідно якому споживання ресурсів контролюється та вимірюється з метою оптимізації.

Концепція хмарної інфраструктури реалізується завдяки трьом основним типам моделей, а саме програмне забезпечення як послуга (Software as a Service – SaaS), інфраструктура як послуга (Infrastructure as a Service – IaaS) і платформа як послуга (Platform as a Service – PaaS). Хмарна IaaS уможливлює впровадження платформи інфраструктури обладнання та програмного забезпечення в якості послуг. Згідно з концепцією хмарної PaaS користувачі мають можливість реалізовувати розроблюване програмне забезпечення використовувати існуючу платформу. В той же час використання концепції SaaS є можливість користуватися програмних забезпеченням лише в хмарній інфраструктурі. У своїй суті хмарні обчислення також можуть впроваджувати різні елементи програмного забезпечення в якості «коду», що здійснюється на підґрунті ідеї Інфраструктури як код (IaC – Infrastructure as Code), що є провідним об'єктом дослідження поточної роботи.

Автоматизація міграційних процесів у контексті переходу з наземної архітектури на хмарну, напряму залежить від характеру самої міграції. Найменш інвазивним типом хмарної міграції передбачає заміну окремих компонентів хмарними сервісами, в ході чого може знадобитися додаткова конфігурація та адаптація, щоб усунути потенційну несумісність компонентів, які переносяться. В ході часткової міграції здійснюється перенесення певних функцій програми. Міграція повного стеку передбачає міграцію всієї програми повністю, що є класичним випадком хмарної міграції. Зазвичай цей процес здійснюється шляхом інкапсуляції програми у віртуальну машину та запуску її в хмарному середовищі. Міграція за типом «Cloudify» передбачає повну трансформацію. У цьому відношенні, Міграція всього стека передбачає переміщення всієї наявної локальної інфраструктури в хмару з мінімальними змінами, часто використовуючи стратегію підйому та переміщення, в той час як, міграція за типом «Cloudify» означає перетворення додатків на хмарні, використовуючи специфічні функції та сервіси, що зазвичай передбачає перебудову архітектури або рефакторинг.

Визначення елементів як коду дозволяє швидко та узгоджено створювати кілька екземплярів. Це забезпечує швидкий ремонт і реконструкцію, забезпечуючи однакову структуру елементів кожного разу, роблячи поведінку системи передбачуваною, а тестування більш надійним. Це полегшує постійне тестування та доставку. Крім того, код забезпечує чіткий план того, як був побудований кожен елемент, підвищуючи прозорість.

У контексті автоматизації міграції IaC дозволяє безперебійно та узгоджено розгортати компоненти інфраструктури в хмарному середовищі. Це мінімізує ризик людської помилки та гарантує, що всі компоненти розгортаються стандартизованим способом. Безперервне тестування та виконання за допомогою IaC сприяють продуктивній інтеграції та тестуванню протягом усього процесу міграції. Цей ітеративний підхід покращує загальну ефективність і надійність міграції.

У випадку даного дослідження увага зосереджується на використанні інструменту Інфраструктури як код Terraform у автоматизації міграції програмного забезпечення, що, насамперед мотивується наданням Terraform можливості визначати хмарну інфраструктуру та керувати нею за допомогою мови декларативної конфігурації. В основі Terraform лежить його декларативний синтаксис, який дозволяє користувачам вказувати бажаний стан файлів конфігурації інфраструктури. Ці конфігураційні файли, написані мовою конфігурації HashiCorp (HCL), коротко описують ресурси, залежності та зв'язки між різними компонентами інфраструктури. Потім Terraform автоматизує надання та керування цими ресурсами, щоб переконатися, що фактичний стан збігається з бажаним станом, указаним у файлах конфігурації. Terraform автоматизує надання хмарних ресурсів шляхом взаємодії з API, наданими хмарними провайдерами. Він перетворює бажаний стан, указаний у файлах конфігурації, у виклики API для надання та налаштування необхідних ресурсів. Постачальники ресурсів Terraform абстрагуються від складнощів взаємодії з хмарними API, дозволяючи користувачам зосередитися на визначенні бажаної конфігурації інфраструктури. Зокрема, даний інструмент пропонує підтримку багатьох хмарних провайдерів, включаючи AWS, Azure, Google Cloud Platform та інших.

Зокрема, змінні Terraform дозволяють користувачам параметризувати свої конфігурації та налаштовувати розгортання на основі динамічних вхідних даних. Змінні можуть бути визначені у файлах конфігурації Terraform або надані зовні за допомогою різних методів, таких як змінні середовища, прапорці командного рядка або файли змінних. Динамічне надання змінних забезпечує

© Бегліцов С.В.

гнучке та масштабоване розгортання інфраструктури, дозволяючи організаціям плавно адаптуватися до мінливих вимог і середовищ

Terraform Cloud є розміщеною службою, яка забезпечує централізоване керування станом, функції співпраці та інтеграцію з контролем версій і пайплайнами CI/CD. Він пропонує набір функцій, призначених для спрощення надання інфраструктури, співпраці та керування. Він забезпечує централізоване розташування для зберігання та керування файлами стану Terraform, забезпечуючи послідовність і одночасність між членами команди та середовищами. Робочі простори в Terraform Cloud дають змогу командам співпрацювати над налаштуваннями інфраструктури, ділитися змінними та керувати контролем доступу, сприяючи співпраці та повторному використанню коду. До того ж він легко інтегрується з системами контролю версій, забезпечуючи керування версіями, відстеження змін і можливості відкату для конфігурацій інфраструктури. Також, ця платформа підтримує політику у вигляді коду через Sentinel, що дозволяє організаціям запроваджувати політику управління, відповідності та безпеки в усіх інфраструктурних розгортаннях.

Поточне дослідження зосереджується на впровадженні функціональних можливостей Terraform та платформи Terraform Cloud у інтеграції з хмарними сервісами AWS (Amazon Web Services) задля автоматизації процесу міграції додатку електронної комерції. Додатковими пре реквізитами є S3 бакет для зберігання файлів стану Terraform та можливість створення сценаріїв Bash для автоматизації.

Процес розпочинається визначенням необхідних ресурсів AWS за допомогою файлів конфігурації Terraform. У таблиці 1 міститься функція AWS Lambda та API Gateway.hcl. Нижче описано визначення функції AWS Lambda та шлюзу API для надання функції Lambda через кінцеву точку.

```
resource "aws_lambda_function" {
  function_name = "my-lambda-function"
  runtime       = "nodejs14.x"
  handler       = "index.handler"
  filename      = "${path.module}/lambda_function_payload.zip"
}

resource "aws_api_gateway_rest_api" {
  name        = "my-api-gateway"
  description = "My API Gateway"
}

resource "aws_api_gateway_resource" {
  rest_api_id = aws_api_gateway_rest_api.id
  parent_id   = aws_api_gateway_rest_api.root_resource_id
}

resource "aws_api_gateway_method" {
  rest_api_id   = aws_api_gateway_rest_api.id
  resource_id   = aws_api_gateway_resource.id
  http_method   = "GET"
  authorization = "NONE"
}

resource "aws_api_gateway_integration" {
  rest_api_id             = aws_api_gateway_rest_api.id
  resource_id             = aws_api_gateway_resource.id
  http_method             = aws_api_gateway_method.http_method
  integration_http_method = "POST"
  type                    = "AWS_PROXY"
  uri                     = aws_lambda_function.invoke_arn
}
```

© Бегліцов С.В.

```
resource "aws_api_gateway_deployment" " {
  depends_on  = [aws_api_gateway_integration]
  rest_api_id = aws_api_gateway_rest_api.id
  stage_name  = "dev"
```

У кінцевому випадку, функція Lambda оброблятиме запити HTTP GET, що направляються через шлюз API. Далі файли стану Terraform зберігаються у бакеті S3, щоб підтримувати узгодженість і уможливлювати співпрацю.

Наступним кроком є створення сценарію bash, щоб динамічно встановлювати конфігурацію серверної частини для Terraform, дозволяючи різні імена файлів стану на основі змінних середовища.

```
# backend.sh
cat <<EOF > main.tf
provider "aws" {
}

terraform {
  backend "s3" {
    bucket = "dev-area"
    key    = "Terraform/lambda/${state_name}.tfstate"
  }
  required_providers {
    aws = {
      version = ">= 4.67.0"
      source  = "hashicorp/aws"
    }
  }
}
EOF

cat <<EOL > lambda.tf
resource "aws_lambda_function" "${state_name}_function" {
  function_name = var.lambda_name
  filename      = "lambda.zip"
  handler       = "exports.handler"
  runtime       = "nodejs18.x"
  role          = "arn:aws:iam::${account_id}:role/LambdaRole"
}
EOL
```

Сценарій bash (backend.sh) динамічно генерує файл конфігурації серверної частини Terraform (main.tf) і файл конфігурації функції Lambda (lambda.tf) на основі змінних середовища. Це дозволяє створювати гнучкі та багаторазові конфігурації Terraform.

Потім команди сценарію Terraform і bash включаються у пайплайн CI/CD, щоб автоматизувати процес розгортання.

```
name: CI/CD Pipeline

on:
  push:
    branches:
      - main

jobs:
  deploy:
    runs-on: ubuntu-latest
```

© Бегліцов С.В.

```
    steps:
      - name: Checkout code
        uses: actions/checkout@v2

      - name: Set up Terraform
        uses: hashicorp/setup-terraform@v1
        with:
          terraform_version: 1.0.0

      - name: Set AWS credentials
        uses: aws-actions/configure-aws-credentials@v1
        with:
          aws-access-key-id: ${{ secrets.AWS_ACCESS_KEY_ID }}
          aws-secret-access-key: ${{ secrets.AWS_SECRET_ACCESS_KEY }}

      - name: Execute bash script
        run: |
          export state_name=lambda
          bash backend.sh

      - name: Initialize Terraform
        run: terraform init -reconfigure

      - name: Plan Terraform
        run: terraform plan -out out/lambda -var profile='aws1' -var
lambda_name='lambda_function_name'

      - name: Apply Terraform
        run: terraform apply "out/lambda "
```

GitHub Actions автоматизує перевіряє код, налаштовує Terraform, конфігурує облікові дані AWS, виконує сценарій bash для створення конфігурацій Terraform і запускає команди Terraform для ініціалізації, планування та застосування змін інфраструктури.

Terraform Cloud використовується для віддаленого керування станом, співпраці та розширених функцій, таких як контроль версій і журнали аудиту.

```
# terraform.tf
terraform {
  required_version = ">= 0.13.0"
  backend "remote" {
    organization = "organization_name"
    workspaces {
      name = "example-workspace"
    }
  }
}
```

Надана конфігурація вказує, що Terraform має використовувати віддалену серверну частину, надану Terraform Cloud.

У підсумку запускається пайплайн CI/CD для виконання команд Terraform, надання необхідних ресурсів AWS і розгортання програми електронної комерції.

```
export AWS_PROFILE=aws_profile
export state_name=lambda
```

© Бегліцов С.В.

```
echo $state_name
bash backend.sh

terraform init -reconfigure
terraform plan -out out/lambda-var profile='aws1' -var
lambda_name='lambda_function_name'
terraform apply "out/lambda"
terraform destroy -var profile='aws_profile' -var
lambda_name='lambda_function_name'
```

Ці команди виконують етапи міграції, включаючи експорт необхідних змінних середовища, запуск сценарію bash і виконання команд Terraform для ініціалізації, планування та застосування змін інфраструктури.

Висновки. Концепція IaC уможливлює швидке відтворення будь-якого елемента хмарної інфраструктури, оскільки все фіксується в одному сценарії. Ця можливість може усунути ризики та страхи, пов'язані зі змінами у процесі міграції програмного забезпечення. IaC вимагає, щоб системи проектувалися з припущенням, що хмарна інфраструктура буде змінюватися з часом і адаптуватися до мінливих потреб. Навіть якщо сервери видаляються або масштабуються, програми мають продовжувати працювати. Корисна функція інструментів IaC для легкого створення, видалення, заміни, масштабування та переміщення ресурсів хмарної інфраструктури полегшує вдосконалення та адаптацію поточної хмарної інфраструктури, що сприяє стійкості наявних сервісів.

Загалом, описана практика інтеграції Terraform і AWS, включаючи контроль версій коду інфраструктури, дотримання принципів безпеки та використання параметризації та змінних для динамічних конфігурацій гарантує, що розгортання інфраструктури є безпечним, доступним для перевірки та масштабованим, що відповідає вимогам сучасних хмарних програм. Завдяки такій реалізації, організації можуть досягти більшої гнучкості, зменшити операційні витрати та покращити співпрацю між командами розробки та операцій. Ідеї та методології, викладені в статті, забезпечують надійну основу для модернізації управління інфраструктурою в хмарних середовищах, що дає можливість організаціям впроваджувати інновації та масштабуватися в епоху хмарних обчислень.

Список бібліографічних джерел

1. Kishiyama B., Wang H., Lopez D., Yang J. An Overview of Infrastructure as Code (IaC) with Performance and Availability Assessment on Google Cloud Platform. 2024.
2. Sabacinski J. Der Weg in die Cloud: Entwicklung einer Migrations-Methodologie für Desktop- und Server-Anwendungen hin zu einer Software-as-a-Service Anwendung. *Diplomarbeit Nr. 3680, Institut für Architektur von Anwendungssystemen, Universität Stuttgart*. 2015.
3. Chinthapatla S. Unleashing the Power of AWS: Revolutionizing Cloud Management through Infrastructure as Code (IaC). 2024.
4. Johnson A., et al. Optimizing AWS Serverless Deployments with Terraform: ACaseStudy. *Journal of Cloud Computing.* 2021. № 10(2),
5. Rahman A., Partho A., Morrison P., Williams L. What questions do programmers ask about configuration as code? Proceedings of the 4th International Workshop on Rapid Continuous Software Engineering. 2018. pp 16–22, DOI:10.1145/3194760.3194769.
6. Smith L., et al. Automating Serverless Infrastructure on AWS with Terraform. *IEEE Transactions on Cloud Computing.* 2021. № 9(4), P. 789-802.
7. Dalla Palma S., Di Nucci D., Palomba F., Tamburri D. A. Within- project defect prediction of Infrastructure-as-Code using product and process metrics. IEEE Transactions on Software Engineering. 2020. № 14(8). P. 1-8.
8. Brown M., et al. Scalable Serverless Deployments on AWS using Terraform. *ACMTransactions on Cloud Computing.* 2021. № 7(3). P. 123-136.
9. Özdoğan E., Ceran O., Üstündağ M. Systematic Analysis of Infrastructure as Code Technologies. Gazi University Journal of Science Part A: Engineering and Innovation. 2023. № 10. DOI:10.54287/gujsa.1373305.
10. Feitosa D., Penca M.-T., Berardi M., Boza R.-D., Andrikopoulos V. Mining for Cost Awareness in the Infrastructure as Code Artifacts of Cloud-Based Applications: An Exploratory Study. 2023. DOI:10.2139/ssrn.4436874.
11. Batu J. Implementing Infrastructure-as-Code with Cloud Disaster Recovery Strategies. International Journal of Computer Trends and Technology. 2024. № 72. P. 41-45. DOI:10.14445/22312803/IJCTT-V72I2P108.
12. Abualkishik A., Alwan A., Gulzar Y. DisasterRecovery in Cloud Computing Systems: An Overview. *InternationalJournal of Advanced Computer Science and Applications.* 2020. № 11. DOI: 10.14569/IJACSA.2020.0110984.
13. Eleraky, Mohammed & Anis Aziz, Wagdy & Soliman, John. (2023). Using Cloud Infrastructure as a Code IaC to

© Бегліцов С.В.

Set Up Web Applications. International Journal of Simulation: Systems, Science & Technology. 24.
14. Mukhopadhyay N., Tewari B. Efficient IaC-Based Resource Allocation for Virtualized Cloud Platforms. 2022. DOI: 10.1007/978-3-030-96040-7_16.
15. Omofoyewa Y., Grebe A., Leusmann P. IaC reusability for Hybrid Cloud Environment. 2021.

References

1. Kishiyama B., Wang H., Lopez D., & Yang J. (2024). An Overview of Infrastructure as Code (IaC) with Performance and Availability Assessment on Google Cloud Platform.
2. Sabacinski J. (2015). Der Weg in die Cloud: Entwicklung einer Migrations-Methodologie für Desktop- und Server-Anwendungen hin zu einer Software-as-a-Service Anwendung. Diplomarbeit Nr. 3680, Institut für Architektur von Anwendungssystemen, Universität Stuttgart.
3. Chinthapatla S. (2024). Unleashing the Power of AWS: Revolutionizing Cloud Management through Infrastructure as Code (IaC).
4. Johnson A., et al. (2021). Optimizing AWS Serverless Deployments with Terraform: A Case Study. Journal of Cloud Computing, 10(2).
5. Rahman A., Partho A., Morrison P., & Williams L. (2018). What questions do programmers ask about configuration as code? Proceedings of the 4th International Workshop on Rapid Continuous Software Engineering, 16–22. https://doi.org/10.1145/3194760.3194769
6. Smith L., et al. (2021). Automating Serverless Infrastructure on AWS with Terraform. IEEE Transactions on Cloud Computing, 9(4), 789-802.
7. Dalla Palma S., Di Nucci D., Palomba F., & Tamburri D. A. (2020). Within-project defect prediction of Infrastructure-as-Code using product and process metrics. IEEE Transactions on Software Engineering, 14(8), 1-8.
8. Brown M., et al. (2021). Scalable Serverless Deployments on AWS using Terraform. ACM Transactions on Cloud Computing, 7(3), 123-136.
9. Özdoğan E., Ceran O., & Üstündağ M. (2023). Systematic Analysis of Infrastructure as Code Technologies. Gazi University Journal of Science Part A: Engineering and Innovation, 10. https://doi.org/10.54287/gujsa.1373305
10. Feitosa D., Penca M.-T., Berardi M., Boza R.-D., & Andrikopoulos V. (2023). Mining for Cost Awareness in the Infrastructure as Code Artifacts of Cloud-Based Applications: An Exploratory Study. https://doi.org/10.2139/ssrn.4436874
11. Batu J. (2024). Implementing Infrastructure-as-Code with Cloud Disaster Recovery Strategies. International Journal of Computer Trends and Technology, 72, 41-45. https://doi.org/10.14445/22312803/IJCTT-V72I2P108
12. Abualkishik A., Alwan A., & Gulzar Y. (2020). Disaster Recovery in Cloud Computing Systems: An Overview. International Journal of Advanced Computer Science and Applications, 11. https://doi.org/10.14569/IJACSA.2020.0110984
13. Eleraky, M., Aziz, W. A., & Soliman, J. (2023). Using Cloud Infrastructure as a Code IaC to Set Up Web Applications. International Journal of Simulation: Systems, Science & Technology, 24.
14. Mukhopadhyay N., & Tewari B. (2022). Efficient IaC-Based Resource Allocation for Virtualized Cloud Platforms. https://doi.org/10.1007/978-3-030-96040-7_16
15. Omofoyewa Y., Grebe A., & Leusmann P. (2021). IaC reusability for Hybrid Cloud Environment.

© Бегліцов С.В.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-13
УДК 004.652
Бондарчук Олег Іванович[1], магістр
http://orcid.org/0009-0003-9626-1124
Козуб Владислав Юрійович[2], д.ф. з комп.н., асистент
http://orcid.org/0000-0003-2710-7206
Козуб Юрій Гордійович[2], д.т.н., професор
http://orcid.org/0000-0002-3366-6031
[1] Akademia WSB/Університет WSB, м. Домброва Гурнича, Польща
[2]Луганський національний університет імені Тараса Шевченка, м. Полтава, Україна

АНАЛІЗ ЕФЕКТИВНОСТІ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ В ОБРОБЦІ ВЕЛИКИХ ДАНИХ

Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г. Аналіз ефективності алгоритмів машинного навчання в обробці великих даних. Метою статті є аналіз ефективності алгоритмів машинного навчання та пошук оптимальних підходів для їх використання в умовах високого навантаження і великих обсягів інформації. Особливу увагу приділено завданням, які потребують високої точності та швидкості, зокрема фінансові прогнози, медична діагностика та аналіз поведінкових даних. Методологія дослідження включала порівняльний аналіз різних алгоритмів машинного навчання, таких як лінійна регресія, дерева рішень, метод опорних векторів та глибокі нейронні мережі. Оцінювалися ключові фактори, що впливають на швидкість та точність обробки даних, зокрема розмір даних, складність моделей, обчислювальні ресурси та якість вхідних даних. Було проведено серію експериментів на реальних наборах даних, щоб оцінити результативність кожного алгоритму з погляду точності, часу навчання та потреби в обчислювальних ресурсах. У результатах дослідження встановлено, що глибокі нейронні мережі забезпечують високу точність на неструктурованих даних, але мають значні вимоги до обчислювальних ресурсів і часу навчання. Алгоритми, такі як лінійна регресія та дерева рішень, показали високу швидкість обробки на простіших наборах даних, проте їх точність знижується зі збільшенням складності завдань. Метод опорних векторів продемонстрував ефективність у завданнях класифікації та прогнозування, зокрема у фінансових та медичних застосуваннях. Випадкові ліси виявилися ефективними для текстової класифікації та забезпечили баланс між швидкістю і точністю. Висновки свідчать, що вибір алгоритму залежить від специфіки завдання, розміру даних та вимог до точності. Оптимізація моделей для роботи в умовах розподілених обчислень є ключовим напрямом для підвищення продуктивності, оскільки дозволяє паралелізувати процеси й зменшити час навчання.
Ключові слова: штучний інтелект, предиктивна аналітика, обробка даних, алгоритми класифікації, оптимізація моделі

Bondarchuk O., Kozub V., Kozub Yu. Analysis of the effectiveness of machine learning algorithms in big data processing. The aim of this article is to analyze the effectiveness of machine learning algorithms and identify optimal approaches for their use under heavy loads and large volumes of information. Special attention is paid to tasks requiring high accuracy and speed, such as financial forecasting, medical diagnostics, and behavioral data analysis. The research methodology included a comparative analysis of various machine learning algorithms, such as linear regression, decision trees, support vector machines, and deep neural networks. Key factors influencing the speed and accuracy of data processing were evaluated, including data size, model complexity, computational resources, and data quality. A series of experiments on real-world datasets was conducted to assess each algorithm's performance in terms of accuracy, training time, and computational resource demands. The results showed that deep neural networks provide high accuracy on unstructured data, but require significant computational resources and training time. Algorithms such as linear regression and decision trees demonstrated high processing speed on simpler datasets, though their accuracy decreases as task complexity increases. The support vector machine method proved effective for classification and prediction tasks, particularly in financial and medical applications. Random forests were found to be efficient for text classification, providing a balance between speed and accuracy. The conclusions indicate that the choice of algorithm depends on the task's specific characteristics, data size, and accuracy requirements. Optimizing models for distributed computing environments is a key direction for improving productivity, as it allows for parallelization and reduces training time.
Keywords: artificial intelligence, predictive analytics, data processing, classification algorithms, model optimization

Постановка наукової проблеми. Обробка великих даних стала однією з ключових проблем сучасної науки й техніки через стрімке збільшення обсягу інформації, що генерується в різних галузях, таких як медицина, економіка, соціальні мережі та наукові дослідження. Алгоритми машинного навчання забезпечують ефективні інструменти для аналізу та інтерпретації цих даних, що відкриває нові можливості для прийняття рішень та прогнозування. Проте ефективність цих алгоритмів залишається під питанням через складність їх налаштування для великих і неоднорідних наборів даних. Основним викликом є необхідність розробки таких методів обробки, які дозволять оптимізувати час обчислень, підвищити точність прогнозів та адаптувати алгоритми до умов розподілених обчислень. Це завдання має важливе практичне значення для підвищення продуктивності великих систем аналізу даних, зокрема у сфері бізнес-аналітики, фінансових ринків, охорони здоров'я та наукових досліджень.

© Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г

Аналіз останніх досліджень і публікацій. Аналіз наукових досліджень, присвячених ефективності алгоритмів машинного навчання в обробці великих даних, демонструє різноманітність підходів і технологій, що застосовуються для покращення продуктивності й точності в різних галузях.

О. Данчак і М. Войтюк [1] акцентують на ефективності сховищ даних, спеціально налаштованих для алгоритмів машинного навчання, які забезпечують швидкий доступ і обробку великих масивів інформації. Вони пропонують рішення для підвищення продуктивності таких алгоритмів у практичних сценаріях.

В. Нестеров [2] досліджує вплив аналітики великих даних на бізнес у цифрову епоху, наголошуючи на важливості машинного навчання для покращення бізнес-процесів і прийняття рішень.

К. Лозовська [3] аналізує застосування методів машинного навчання для оцінки показників інтернет-ресурсів. На думку дослідниці, алгоритми машинного навчання є ключовими інструментами для аналізу великих масивів даних, що дозволяє автоматизувати процеси й забезпечити точні прогнози.

K. Ngiam та I. Khor [4] зосереджуються на використанні великих даних та алгоритмів машинного навчання в охороні здоров'я. Їхнє дослідження демонструє значення алгоритмів для покращення якості надання медичних послуг і прогнозування ризиків.

V. Koliesetty і D. Rajput [5] та M. Jan і співавтори [6] здійснюють огляди застосування машинного навчання для аналізу великих даних, підкреслюючи роль глибокого навчання в підвищенні ефективності обробки даних у різних галузях. Їхні дослідження підтверджують важливість розробки спеціалізованих моделей для адаптації під конкретні потреби.

Колектив авторів на чолі з G. Saranya [7] та G. Saranya і S. Asha [8] аналізують продуктивність алгоритмів машинного навчання для систем виявлення вторгнень у систему охорони здоров'я. Вони наголошують на важливості вибору оптимальних моделей для забезпечення безпеки та надання точних прогнозів.

J. Shang і Z. You [9] вивчають вплив машинного навчання на розвиток розумного виробництва, підкреслюючи, що успішна інтеграція алгоритмів машинного навчання в процеси виробництва дозволяє оптимізувати управління та підвищити ефективність роботи.

A. Adadi [10] робить огляд алгоритмів, які забезпечують ефективну обробку даних у великих масштабах, наголошуючи на важливості економічності моделей для роботи з обмеженими ресурсами. Це узгоджується з висновками колективу авторів під керівництвом M. Amanullah [11], який досліджує глибоке навчання та технології великих даних у сфері безпеки IoT.

O. Faker та E. Dogdu [12] продовжують тему безпеки, зосереджуючись на виявленні вторгнень за допомогою глибокого навчання, підкреслюючи значення адаптації моделей під конкретні потреби індустрій, де безпека є пріоритетом.

Дослідження в цій галузі демонструють важливість удосконалення алгоритмів машинного навчання та оптимізації сховищ даних для забезпечення ефективної обробки великих обсягів інформації.

Виділення раніше невирішених частин проблеми. Попри різноманітність досліджень, залишаються нерозв'язаними питання дослідження наявних алгоритмів машинного навчання з погляду їхньої ефективності при роботі з великими даними; оцінки ключових факторів, що впливають на швидкість і точність обробки даних; порівняння результативності різних алгоритмів на основі практичних експериментів і наукових досліджень.

Метою статті є аналіз ефективності алгоритмів машинного навчання для обробки великих даних та виявлення оптимальних підходів до їх використання в умовах високого навантаження та великих обсягів інформації.

Завдання статті:

1)дослідити наявні алгоритми машинного навчання з погляду їхньої ефективності при роботі з великими даними;

2) оцінити ключові фактори, що впливають на швидкість і точність обробки даних;

3) порівняти результативність різних алгоритмів на основі практичних експериментів і наукових досліджень;

4) вивчити проблеми та обмеження, пов'язані з масштабуванням алгоритмів для великих обсягів даних;

© Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г

5) розробити рекомендації щодо оптимізації алгоритмів для роботи в умовах розподілених обчислень.

Виклад основного матеріалу. Дослідження наявних алгоритмів машинного навчання з погляду їхньої ефективності при роботі з великими даними є важливим аспектом аналізу продуктивності сучасних систем обробки інформації. Алгоритми, такі як лінійна регресія, дерева рішень, метод опорних векторів та глибокі нейронні мережі, мають різні підходи до обробки великих обсягів даних, що впливає на їх точність та швидкість виконання. Наприклад, алгоритми глибокого навчання, хоча й забезпечують високу точність, вимагають значних обчислювальних ресурсів та часу на тренування, особливо при збільшенні розмірів даних. З іншого боку, простіші алгоритми, такі як лінійна регресія, можуть бути швидкими та менш ресурсомісткими, але їх ефективність може знижуватися при складних і великомасштабних наборах даних. У таблиці 1 наведено порівняння основних алгоритмів за кількома критеріями, такими як швидкість навчання, масштабованість та точність прогнозів.

Табл. 1. Порівняння основних алгоритмів машинного навчання

Алгоритм	Опис алгоритму	Швидкість навчання	Точність прогнозів	Масштабованість	Обчислювальні вимоги
Лінійна регресія	Статистичний метод для моделювання залежності між змінними	Висока	Низька в складних завданнях	Висока	Низькі
Дерева рішень	Деревоподібна структура для прийняття рішень на основі правил	Середня	Середня	Обмежена у великих наборах	Середні
Метод опорних векторів	Алгоритм для класифікації та регресії, що знаходить оптимальну гіперплощину	Низька	Висока	Середня	Високі
Глибокі нейронні мережі	Алгоритм із великою кількістю шарів нейронів для виявлення складних закономірностей.	Низька	Висока	Висока	Дуже високі

Лінійна регресія є одним із найстаріших алгоритмів і використовується переважно для простих моделей передбачення в таких сферах, як економіка, аналітика продажів або прогнозування клімату. Вона дозволяє швидко отримати результат, однак її точність може бути низькою в умовах складних нелінійних даних. Дерева рішень широко застосовуються в галузях, де необхідно пояснити процес прийняття рішень, наприклад, в медицині для діагностичних завдань або у фінансах для оцінки ризиків. Вони надають візуально зрозумілу структуру, що легко інтерпретується, але їх продуктивність знижується при великій кількості ознак. Метод опорних векторів, завдяки своїй здатності працювати з високовимірними даними, застосовується в біоінформатиці, обробці зображень і машинному перекладі, але має високі вимоги до обчислювальних потужностей. Глибокі нейронні мережі, які здатні вивчати надзвичайно складні зв'язки в даних, сьогодні є основою таких технологій, як розпізнавання мови, зображень та прогнозування поведінки в реальному часі [2], [5]. Вони широко використовуються в таких галузях, як штучний інтелект, автономні транспортні засоби та персоналізована медицина.

У сучасних умовах штучний інтелект є важливим інструментом для оптимізації процесів обробки великих обсягів даних, особливо в складних середовищах, де швидкість та точність рішень є критично важливими. Інтеграція методів машинного навчання та штучного інтелекту дає змогу використовувати алгоритми класифікації для прогнозування й автоматизації процесів прийняття рішень у реальному часі.

Наприклад, штучний інтелект може використовувати алгоритми класифікації, такі як підтримуючі векторні машини (SVM), або глибокі нейронні мережі (DNN), для аналізу поведінкових патернів користувачів на онлайн-платформах. Такі моделі здатні ефективно опрацьовувати великі обсяги даних, виявляти ключові закономірності та на основі цих даних

© Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г

проводити прогнозування, як-от передбачення потенційних ризиків або визначення ключових трендів (табл. 2).

Табл. 2. Порівняння алгоритмів штучного інтелекту для класифікації та прогнозування великих даних

Алгоритм ШІ	Призначення	Переваги	Недоліки
Глибокі нейронні мережі (DNN)	Обробка складних і неструктурованих даних	Висока точність, можливість глибокого навчання	Потребують великих обчислювальних ресурсів
Підтримуючі векторні машини (SVM)	Класифікація та прогнозування на основі великих даних	Висока точність, добре працює з лінійно роздільними даними	Менша ефективність для великих неструктурованих даних
Багатошарова персептронна модель (MLP)	Прогнозування та класифікація в реальному часі	Здатність обробляти дані в реальному часі	Чутливість до перенавчання моделі

Глибокі нейронні мережі застосовуються для обробки великих наборів даних, таких як зображення або текст, де необхідно знайти складні закономірності. Вони забезпечують високу точність, але потребують значних обчислювальних потужностей для навчання й прогнозування. Алгоритми підтримуючих векторних машин (SVM) підходять для розв'язання завдань класифікації, де дані можуть бути лінійно розділеними, що дозволяє досягти високої точності прогнозів. Мультишарова персептронна модель (MLP) використовується для обробки великих потоків даних у реальному часі, що дозволяє виконувати прогнозування на основі даних, що надходять в режимі реального часу.

На практиці такі алгоритми можуть використовуватися для підвищення ефективності систем штучного інтелекту у сфері фінансів, медицини та промисловості.

Процес вибору алгоритму залежно від обсягу даних та вимог до точності прогнозування представлено на рис. 1

Рис. 1. Процес вибору алгоритму залежно від обсягу даних та вимог до точності прогнозування

У сучасних умовах алгоритми машинного навчання відіграють ключову роль в ефективному обробленні величезних масштабів даних. Прості алгоритми, такі як лінійна регресія, використовуються при невеликому обсязі даних для швидких прогнозів з обмеженими вимогами до ресурсів. Для середніх обсягів даних та потреби в інтерпретації результатів краще підходять дерева рішень, особливо в ситуаціях, коли важлива наочність процесу прийняття рішень. Якщо ж дані

складні та їх обсяг значний, метод опорних векторів забезпечить високу точність, але за умови наявності достатніх обчислювальних ресурсів. Нарешті, для роботи з величезними масивами даних найкраще підходять глибокі нейронні мережі, здатні виявляти складні взаємозв'язки й використовувати сучасні обчислювальні системи для масштабованого навчання й прогнозування.

Обчислювальні ресурси є ще одним визначальним чинником, оскільки алгоритми потребують значної кількості ресурсів для роботи з великими наборами даних. Алгоритми з високими вимогами до ресурсів, такі як метод опорних векторів або нейронні мережі, можуть уповільнювати процес навчання та прогнозування в умовах обмеженого доступу до обчислювальних ресурсів. Окрім цього, якість вхідних даних також впливає на точність моделі. Алгоритми можуть забезпечити кращі результати, якщо дані попередньо оброблені та очищені від шуму або пропущених значень.

У таблиці 3 наведено порівняння основних факторів, що впливають на швидкість і точність обробки даних різними алгоритмами машинного навчання.

Табл. 3. Порівняння основних факторів, що впливають на швидкість і точність обробки даних різними алгоритмами машинного навчання.

Фактор	Вплив на швидкість обробки	Вплив на точність обробки	Пояснення
Розмір даних	Знижує швидкість	Залежить від алгоритму	Великий обсяг даних уповільнює роботу складніших алгоритмів, зокрема нейронних мереж
Складність моделі	Знижує швидкість	Підвищує точність	Більша кількість параметрів або шарів вимагає більше часу для навчання та налаштувань.
Обчислювальні ресурси	Підвищує швидкість	Не впливає безпосередньо	Доступ до потужних серверів або розподілених систем прискорює обробку даних.
Якість вхідних даних	Не впливає значно	Підвищує точність	Попередня обробка та очищення даних знижують ймовірність помилок у прогнозах.

У сучасних умовах ключові фактори, що впливають на швидкість і точність обробки даних, значно еволюціонували завдяки розвитку обчислювальних технологій та новим підходам до управління даними [15]. Обсяг даних постійно збільшується через широке застосування Інтернету речей (IoT), соціальних медіа та великих комерційних і наукових систем збору даних. Це вимагає використання більш складних алгоритмів і потужніших обчислювальних ресурсів. У зв'язку з цим масштабованість обчислень і можливості хмарних інфраструктур стали критичними для збереження швидкості обробки навіть при значних обсягах даних.

Складність моделей зросла разом із розвитком алгоритмів глибокого навчання, які тепер можуть виконувати завдання, що раніше були недоступні через обмеження в обчислювальних потужностях. Однак збільшення кількості параметрів і шарів у нейронних мережах призвело до підвищення вимог до ресурсів, що робить необхідною інтеграцію розподілених систем обробки. Водночас якість даних залишається важливим аспектом, оскільки навіть найкращі алгоритми не можуть дати точних результатів за відсутності належної попередньої обробки даних. Використання сучасних методів очищення та нормалізації даних дозволяє значно підвищити точність моделей, мінімізуючи вплив шумів або пропущених значень.

Порівняння результативності алгоритмів машинного навчання на основі практичних експериментів є одним із ключових способів визначення найбільш ефективних підходів до обробки великих даних. У ході експериментів було проведено тестування декількох популярних алгоритмів на різних наборах даних із реальних застосувань, таких як аналіз медичних зображень, прогнозування фінансових ринків і класифікація текстових документів. Метою експериментів було оцінити ефективність кожного алгоритму з погляду точності, швидкості навчання та вимог до обчислювальних ресурсів, а також визначити їхню придатність для різних типів завдань.

У таблиці 4 наведено результати експериментів, що включають основні показники роботи алгоритмів при обробці наборів даних різного розміру та складності.

© Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г

Табл. 4. Основні показники роботи алгоритмів при обробці наборів даних різного розміру та складності

Алгоритм	Набір даних	Точність (%)	Час навчання (с)	Обчислювальні ресурси (CPU)	Обсяг даних (млн записів)	Використання на практиці
Нейронні мережі	Медичні зображення	94	600	Високі	2.5	Діагностика захворювань
Метод опорних векторів	Фінансові ринки	89	450	Середні	1.2	Прогнозування трендів
Випадкові ліси	Текстова класифікація	87	120	Середні	3.1	Аналіз документів
Лінійна регресія	Економічні показники	72	30	Низькі	0.5	Просте моделювання
K-ближчих сусідів	Маркетингові дані	78	300	Середні	0.7	Сегментація клієнтів

Експерименти проводилися з різними наборами даних, щоб оцінити ефективність алгоритмів в умовах, наближених до реальних застосувань. Наприклад, для медичних зображень використовувалися нейронні мережі, які продемонстрували найвищу точність (94%), але час навчання був значно тривалішим порівняно з іншими алгоритмами, що пояснюється великою кількістю параметрів, необхідних для обробки зображень. Метод опорних векторів був успішно застосований для прогнозування на фінансових ринках, досягаючи точності 89% за середнього часу навчання (450 секунд) та помірного використання ресурсів.

Випадкові ліси виявилися ефективними для класифікації текстів, забезпечуючи точність 87% з відносно швидким навчанням (120 секунд), що робить їх оптимальними для завдань із великими наборами текстових даних. Лінійна регресія використовувалася для економічних показників, де точність 72% виявилася прийнятною для простих завдань прогнозування, але її низькі вимоги до обчислювальних ресурсів і час навчання всього в 30 секунд роблять цей алгоритм ефективним у простих моделях. Алгоритм K-ближчих сусідів був використаний для сегментації маркетингових даних, де показав середню точність (78%) та відносно високий час навчання (300 секунд), що обмежує його використання на великих наборах даних.

Оцінка результатів показує, що вибір алгоритму залежить від конкретного завдання, його складності, обсягу даних та вимог до точності. Для складних завдань, таких як медична діагностика, нейронні мережі забезпечують високу результативність, тоді як для швидкого аналізу економічних показників краще підходять простіші моделі, такі як лінійна регресія. У таблиці 5 проаналізовано взаємозв'язок між обсягом даних та часом навчання алгоритмів.

Табл. 5. Взаємозв'язок між обсягом даних та часом навчання алгоритмів

Алгоритм	Обсяг даних (млн записів)	Час навчання (с)	Залежність часу від обсягу даних
Нейронні мережі	1.0	300	Лінійне збільшення
Нейронні мережі	2.5	600	Лінійне збільшення
Метод опорних векторів	0.6	200	Лінійне збільшення
Метод опорних векторів	1.2	450	Лінійне збільшення
Випадкові ліси	1.5	70	Нелінійне збільшення
Випадкові ліси	3.1	120	Нелінійне збільшення

© Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г

Результати експериментів показують, що для нейронних мереж та методу опорних векторів час навчання збільшується лінійно зі збільшенням обсягу даних. Це пояснюється складністю цих моделей, які потребують більше ресурсів із кожним новим записом у наборі даних. Водночас випадкові ліси показують нелінійне збільшення часу навчання, що свідчить про їхню здатність ефективніше обробляти великі обсяги даних завдяки розподіленню обчислень між різними деревами прийняття рішень.

Масштабування алгоритмів машинного навчання для обробки великих обсягів даних є складним завданням, яке супроводжується низкою проблем і обмежень. Однією з ключових проблем є потреба в значних обчислювальних ресурсах, які стають усе важливішими зі збільшенням обсягу даних. Зростання складності даних, що обробляються, призводить до того, що навіть найбільш оптимізовані алгоритми можуть не справлятися з вимогами щодо швидкості навчання та точності результатів. Це створює проблеми не лише для наукових досліджень, але й для практичних застосувань у сфері фінансів, медицини та інших галузях, де рішення необхідно приймати оперативно.

Ще однією важливою проблемою є низька масштабованість деяких алгоритмів. Наприклад, традиційні моделі, такі як лінійна регресія чи K-ближчих сусідів, мають обмежену здатність ефективно працювати з великими наборами даних через зростання обчислювальних витрат або зниження точності результатів. Крім того, багато сучасних алгоритмів, таких як нейронні мережі, демонструють високу точність на малих обсягах даних, але їх масштабування на великі набори потребує спеціалізованого обладнання, зокрема графічних процесорів (GPU) або розподілених систем, що збільшує вартість і складність обчислень.

У таблиці 6 представлено основні проблеми та обмеження, пов'язані з масштабуванням різних алгоритмів для великих обсягів даних.

Табл. 6. Основні проблеми та обмеження, пов'язані з масштабуванням різних алгоритмів для великих обсягів даних

Алгоритм	Проблеми масштабування	Обмеження	Приклад впливу на практику
Нейронні мережі	Значні вимоги до обчислювальних ресурсів	Низька швидкість навчання	Потреба в розподілених обчислювальних системах
Метод опорних векторів	Збільшення часу навчання	Високі вимоги до пам'яті	Труднощі при обробці великих наборів даних
Випадкові ліси	Ускладнене паралельне обчислення	Лінійне зростання часу навчання	Обмежена масштабованість для масивних даних
Лінійна регресія	Низька точність на великих обсягах	Обмежена складність моделі	Використовується тільки для простих завдань
K-ближчих сусідів	Зростання обчислювальних витрат	Зниження швидкості зі збільшенням обсягу даних	Низька ефективність для великих наборів

Отже, нейронні мережі, хоча й забезпечують високу точність на малих наборах даних, мають значні вимоги до обчислювальних ресурсів при масштабуванні. Для їхнього ефективного використання потрібні потужні апаратні засоби, такі як GPU або розподілені системи, що робить процес навчання дорогим і повільним. Метод опорних векторів також характеризується збільшенням часу навчання і вимогами до пам'яті при обробці великих наборів даних, що обмежує його ефективність у масштабованих задачах.

Випадкові ліси, хоча і є відносно ефективними для невеликих наборів даних, демонструють лінійне зростання часу навчання зі збільшенням обсягу даних, що обмежує їхню масштабованість у великих системах. Лінійна регресія як проста модель має обмежену точність і складність, тому її застосування обмежене тільки простими завданнями, що не потребують глибокого аналізу великих обсягів даних. Алгоритм K-ближчих сусідів демонструє значне зростання обчислювальних витрат зі збільшенням обсягу даних, що негативно впливає на його ефективність при масштабуванні.

Для наукових досліджень і реальних застосувань стає важливим пошук компромісу між точністю, швидкістю та ресурсами, щоб забезпечити оптимальне масштабування алгоритмів на

© Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г

великих обсягах даних. Наприклад, гібридні підходи або використання оптимізованих обчислювальних платформ можуть дозволити розв'язати деякі з цих проблем. Таблиця 7 демонструє вплив об'єму даних на продуктивність алгоритмів.

Табл. 7. Вплив обсягу даних на продуктивність алгоритмів

Алгоритм	Обсяг даних (млн записів)	Точність (%)	Час навчання (с)	Обчислювальні витрати (CPU/GPU)
Нейронні мережі	1.0	92	300	Високі (GPU)
Нейронні мережі	5.0	90	1000	Дуже високі (GPU)
Метод опорних векторів	0.5	85	200	Середні (CPU)
Метод опорних векторів	2.0	83	700	Високі (CPU)
Випадкові ліси	1.5	80	150	Середні (CPU)
Випадкові ліси	4.0	78	400	Високі (CPU)

Так, нейронні мережі демонструють високу точність на мільйоні записів, але їхня продуктивність значно погіршується зі збільшенням обсягу даних до п'яти мільйонів записів, що проявляється в значному збільшенні часу навчання та витрат на GPU. Метод опорних векторів демонструє схожу динаміку, де збільшення обсягу даних зменшує точність і збільшує час навчання, роблячи його менш ефективним для великих наборів.

Ці експерименти вказують на необхідність пошуку балансів між точністю й ефективністю алгоритмів для обробки великих обсягів даних, а також на можливість оптимізації цих алгоритмів шляхом використання розподілених обчислювальних систем або гібридних моделей, які можуть комбінувати переваги різних підходів.

Оптимізація алгоритмів для роботи в умовах розподілених обчислень є актуальним завданням в епоху постійного збільшення обсягів даних і ускладнення процесів їхньої обробки. Сучасні алгоритми штучного інтелекту, що використовуються для аналізу великих даних, стикаються з проблемами ефективного використання ресурсів при розподілі обчислень між різними вузлами обчислювальної системи. Розподілені обчислення дозволяють прискорити обробку даних через паралелізацію процесів, проте алгоритми потребують спеціальних підходів до оптимізації для забезпечення ефективної роботи в такому середовищі.

Основними аспектами, що потребують оптимізації, є зменшення затримок при передачі даних між вузлами, забезпечення узгодженості результатів паралельних обчислень, а також максимізація використання доступних обчислювальних потужностей. Для цього необхідно впроваджувати інтелектуальні стратегії розподілу завдань та обчислювальних ресурсів між вузлами системи.

Наприклад, глибокі нейронні мережі можуть бути оптимізовані через розподіл обчислень між кількома графічними процесорами (GPU) або обчислювальними кластерами. Алгоритми, що використовують ансамблеве навчання, також можуть бути адаптовані для паралельного виконання, що дає змогу обробляти великі обсяги даних у різних частинах системи одночасно, зменшуючи загальний час обробки (табл. 8).

Табл. 8 Оптимізація алгоритмів штучного інтелекту для розподілених обчислень

Алгоритм	Підхід до оптимізації в умовах розподілених обчислень	Переваги	Недоліки
Глибокі нейронні мережі	Паралельне навчання на декількох GPU або CPU	Швидкість обробки, масштабованість	Складність синхронізації між вузлами
Випадкові ліси	Розподілена обробка дерев рішень	Можливість ефективної паралелізації	Ресурсомісткість при роботі з великими даними
Градієнтний бустинг	Паралелізація обчислень між вузлами	Покращення швидкості навчання	Високі вимоги до пам'яті та передачі даних

У випадку глибоких нейронних мереж паралельне навчання на декількох GPU дозволяє значно скоротити час, необхідний для навчання моделі на великих обсягах даних. Водночас виникає потреба в синхронізації результатів між різними вузлами, що може бути складним завданням. Алгоритми ансамблевого навчання, такі як випадкові ліси або градієнтний бустинг, дозволяють

розподілити процеси між різними частинами системи, що дає можливість паралельно виконувати незалежні обчислення для кожного дерева рішень або послідовної моделі. Однак це вимагає значних ресурсів, особливо для великих наборів даних.

На практиці такі підходи використовуються для оптимізації роботи систем штучного інтелекту в хмарних обчислювальних середовищах, де обробка великих даних потребує ефективного використання розподілених обчислень для досягнення максимальної швидкості та точності результатів.

Висновки та перспективи подальших досліджень. У результаті здійсненого дослідження встановлено, що ефективність алгоритмів машинного навчання для обробки великих даних залежить від кількох ключових факторів, таких як розмір і складність даних, обчислювальні ресурси та якість вхідних даних. Основними проблемами, які виникають під час роботи з великими наборами даних, є складність масштабування алгоритмів, що потребує значних обчислювальних ресурсів, а також час навчання моделей, який збільшується пропорційно обсягам даних. Зокрема, алгоритми глибокого навчання, хоча і забезпечують високу точність, потребують значних ресурсів, що робить їх менш ефективними у випадках обмежених обчислювальних потужностей.

Рекомендується застосовувати оптимізаційні підходи, такі як розподілені обчислення та паралелізація процесів, для прискорення навчання алгоритмів. Це дозволить зменшити затримки при обробці великих обсягів даних і покращити узгодженість результатів у різних обчислювальних вузлах. Крім того, для підвищення ефективності алгоритмів необхідно впроваджувати методи очищення та нормалізації даних, що дозволить підвищити точність моделей і знизити їхню чутливість до пропущених значень або шуму в даних.

Перспективи подальших досліджень полягають у розвитку нових методів оптимізації алгоритмів для роботи з великими даними в умовах хмарних та розподілених обчислень. Також необхідним є дослідження гібридних моделей, які поєднують різні алгоритми для забезпечення більш високої ефективності й точності прогнозів у складних умовах обробки даних.

Список бібліографічного опису

1. Данчак О., Войтюк М. Ефективні сховища даних для рішень машинного навчання. *Herald of Khmelnytskyi National University. Technical sciences.* 2024. Вип. 337. № 3(2). С. 57–63. DOI: https://doi.org/10.31891/2307-5732-2024-337-3-8
2. Нестеров В. Дослідження впливу аналітики великих даних на ефективність бізнесу в цифрову епоху. *Інформаційні технології та суспільство*. 2024. Вип. 1 (12). С. 70–76. DOI: https://doi.org/10.32689/maup.it.2024.1.1
3. Лозовська К. Аналіз використання методів машинного навчання в аналітиці показників інтернет-ресурсів. *Сталий розвиток економіки.* 2023. Вип. 2 (47). С. 65–69. DOI: https://doi.org/10.32782/2308-1988/2023-47-9
4. Ngiam K. Y., Khor W. Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology.* 2019. Vol. 20. № 5. P. e262–e273. URL: https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(19)30149-4/abstract (дата звернення: 02.09.2024)
5. Koliesetty V. V., Rajput D. S. A review on the significance of machine learning for data analysis in big data. *Jordanian Journal of Computers and Information Technology*. 2020. Vol. 6. № 1. P. 155–171. URL: https://research.vit.ac.in/publication/a-review-on-the-significance-of-machine-learning-for-data/pdf/publisher-pdf-fulltext-a-review-on-the-significance-of-machine-learning-for-data.pdf (дата звернення: 02.09.2024)
6. Bilal J., Haleem F., Murad K., Muhammad I., Ihtesham U., Awais A., Shaukat A., Gwanggil J. Deep learning in big data analytics: a comparative study. *Computers & Electrical Engineering.* 2019. Vol. 75. P. 275–287. DOI: https://doi.org/10.1016/j.compeleceng.2017.12.009
7. Saranya T., Sridevi S., Deisy C., Tran Duc Chung, Ahamed Khan M. Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science*. 2020. Vol. 171. P. 1251–1260. DOI: https://doi.org/10.1016/j.procs.2020.04.133
8. Saranya P., Asha P. Survey on big data analytics in health care. *In: 2019 International conference on smart systems and inventive technology (ICSSIT).* (27–29 November 2019) (pp. 46–51). DOI: https://doi.org/10.1109/ICSSIT46314.2019.8987882
9. Shang J., You Z. Data analytics and machine learning for smart process manufacturing: Recent advances and perspectives in the big data era. *Engineering.* 2019. Vol. 5. № 6. P. 1010–1016. DOI: https://doi.org/10.1016/j.eng.2019.01.019
10. Adadi A. A survey on data-efficient algorithms in big data era. *Journal of Big Data.* 2021. Vol. 8. № 1. URL: https://link.springer.com/article/10.1186/S40537-021-00419-9
11. Amanullah M., Ariyaluran Habeeb R, Fariza R., Gani A, Ahmed E., Abdul Nainar A., Akim N., Imran M. Deep learning and big data technologies for IoT security. *Computer Communications*. 2020. Vol. 151. P. 495–517. DOI: https://doi.org/10.1016/j.comcom.2020.01.016
12. Faker O., Dogdu E. Intrusion detection using big data and deep learning techniques. *In: Proceedings of the 2019 ACM Southeast conference* (pp. 86–93). DOI: https://doi.org/10.1145/3299815.3314439

© Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г

13. Habeeb R, Nasaruddin R., Gani A., Abaker Targio Hashem I., Ahmed E., Imran M. Real-time big data processing for anomaly detection: A survey. *International Journal of Information Management.* 2019. Vol. 45. P. 289–307. DOI: https://doi.org/10.1016/j.ijinfomgt.2018.08.006
14. Sarker I. H. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science.* 2021. Vol. 2. № 3. DOI: https://doi.org/10.1007/s42979-021-00592-x
15. Deng A. Database task processing optimization based on performance evaluation and machine learning algorithm. *Soft Computing – A Fusion of Foundations, Methodologies & Applications*. 2023. Vol. 27. № 10. DOI: https://doi.org/10.1007/s00500-023-08111-1

References

1. Danchak, O., & Voitiuk, M. (2024). Efficient data warehouses for machine learning solutions. *Herald of Khmelnytskyi National University. Technical sciences*, 337(3–2), 57–63. DOI: https://doi.org/10.31891/2307-5732-2024-337-3-8
2. Nesterov, V. (2024). Study of the impact of big data analytics on business efficiency in the digital era. *Informatsiyni technolohiyi i suspilstvo – Information Technology and Society*, 1(12), 70–76. DOI: https://doi.org/10.32689/maup.it.2024.1.10
3. Lozovska, K. (2023). Analysis of the use of machine learning methods in the analytics of internet resource indicators. *Stalyi rozvytok ekonomiky – Sustainable Economic Developme*nt, 2(47), 65–69. DOI: https://doi.org/10.32782/2308-1988/2023-47-9
4. Ngiam, K., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, 20(5), e262–e273. Retrieved from https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(19)30149-4/abstract
5. Koliesetty, V., & Rajput, D. (2020). A review on the significance of machine learning for data analysis in big data. *Jordanian Journal of Computers and Information Technology*, 6(1), 155–171. Retrieved from https://research.vit.ac.in/publication/a-review-on-the-significance-of-machine-learning-for-data/pdf/publisher-pdf-fulltext-a-review-on-the-significance-of-machine-learning-for-data.pdf
6. Bilal, J., Haleem, F., Murad, K., Muhammad, I., Ihtesham, U., Awais, A., Shaukat, A., & Gwanggil, J. (2019). Deep learning in big data analytics: A comparative study. Computers & *Electrical Engineering*, 75, 275–287. DOI: https://doi.org/10.1016/j.compeleceng.2017.12.009.
7. Saranya, T., Sridevi, S., Deisy, C., Tran, D. C., & Ahamed, K. M. (2020). Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science*, 171, 1251–1260. DOI: https://doi.org/10.1016/j.procs.2020.04.133
8. Saranya, P., & Asha, P. (2019). Survey on big data analytics in health care. *In 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)* (27–29 November 2019) (pp. 46–51). DOI: https://doi.org/10.1109/ICSSIT46314.2019.8987882
9. Shang, J., & You, F. (2019). Data analytics and machine learning for smart process manufacturing: Recent advances and perspectives in the big data era. *Engineering*, 5(6), 1010–1016. DOI: https://doi.org/10.1016/j.eng.2019.01.019
10. Adadi, A. (2021). Survey on data-efficient algorithms in big data era. *Journal of Big Data*, 8(1). Retrieved from https://link.springer.com/article/10.1186/S40537-021-00419-9
11. Amanullah, M., Ariyaluran Habeeb, R., Fariza, R., Gani, A., Ahmed, E., Abdul Nainar, A., Akim, N., & Imran, M. (2020). Deep learning and big data technologies for IoT security. *Computer Communications*, 151, 495–517. DOI: https://doi.org/10.1016/j.comcom.2020.01.016
12. Faker, O., & Dogdu, E. (2019). Intrusion detection using big data and deep learning techniques. *In Proceedings of the 2019 ACM Southeast Conference*, 86–93. DOI: https://doi.org/10.1145/3299815.3314439.
13. Habeeb, R., Nasaruddin, R., Gani, A., Abaker Targio Hashem, I., Ahmed, E., & Imran, M. (2019). Real-time big data processing for anomaly detection: A survey. *International Journal of Information Management*, 45, 289–307. DOI: https://doi.org/10.1016/j.ijinfomgt.2018.08.006
14. Sarker, I. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2(3). DOI: https://doi.org/10.1007/s42979-021-00592-x
15. Deng, A. (2023). Database task processing optimization based on performance evaluation and machine learning algorithm. *Soft Computing – A Fusion of Foundations, Methodologies & Applications*, 27(10). DOI: https://doi.org/10.1007/s00500-023-08111-1

© Бондарчук О.І., Козуб В.Ю., Козуб Ю.Г

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-14
УДК 62-23
Борисов Олег Володимирович, к.т.н.
https://orcid.org/0000-0002-9460-2605
Артабаєв Юрій Зуберович, к.т.н.
https://orcid.org/0000-0001-9446-3011
Сурма Анатолій Іванович,
https://orcid.org/0009-0000-1778-9229
Воєнна академія імені Євгенія Березняка, м. Київ

КІБЕРБЕЗПЕКА БЕЗПІЛОТНИХ ВІЙСЬКОВИХ АПАРАТІВ: МЕТОДИ ЗАХИСТУ ВІД ПЕРЕХОПЛЕННЯ ТА ДИСТАНЦІЙНОГО УПРАВЛІННЯ

Борисов О.В., Артабаєв Ю.З., Сурма А.І. Кібербезпека безпілотних військових апаратів: методи захисту від перехоплення та дистанційного управління. У статті розглянуто актуальні питання кібербезпеки безпілотних військових апаратів, зокрема загрози перехоплення сигналів управління та несанкціонованого дистанційного керування. Проаналізовано методи виявлення безпілотних військових апаратів кібератаки, спрямовані на військові безпілотники, та наведено приклади успішних атак на ці системи. Основну увагу приділено методам захисту, таким як криптографічний захист сигналів, багаторівнева аутентифікація, захист програмного забезпечення та сенсорних систем, а також фізичний захист комунікацій. Стаття також містить аналіз реальних кейсів впровадження цих методів та розглядає виклики, що виникають під час їх реалізації. Запропоновано перспективи подальших досліджень у цій галузі.
Ключові слова: кібербезпека, безпілотні військові апарати, перехоплення сигналів, дистанційне управління, криптографічний захист, аутентифікація, захист програмного забезпечення

Borysov O., Artabaiev Y., Surma A. Cybersecurity of unmanned military aerial vehicles: methods of protection against signal interception and remote control. The article discusses current issues of cybersecurity for unmanned military aerial vehicles (UAVs), particularly the threats of signal interception and unauthorized remote control. It analyzes typical cyberattacks targeting military drones and provides examples of successful attacks on these systems. The main focus is on protection methods such as cryptographic protection of signals, multi-level authentication, software and sensor system protection, as well as physical protection of communications. The article also includes an analysis of real cases of implementing these methods and examines the challenges encountered during their implementation. Prospects for further research in this field are proposed.
Keywords: cybersecurity, unmanned military aerial vehicles, signal interception, remote control, cryptographic protection, authentication, software pr

Постановка завдання. У сучасних військових конфліктах безпілотні військові апарати відіграють важливу роль у розвідці, спостереженні, а також у бойових операціях. Проте з розвитком технологій зростає й кількість кіберзагроз, які можуть суттєво вплинути на ефективність і безпеку цих систем. Перехоплення сигналів управління та несанкціоноване дистанційне керування безпілотниками становлять серйозну загрозу, що може призвести до втрати контролю над апаратом, компрометації місії та загрози безпеці військових підрозділів. Основним завданням цієї статті є аналіз поточних кіберзагроз, спрямованих на безпілотні військові апарати, а також розробка та оцінка методів захисту від цих загроз.

Аналіз досліджень. Останні дослідження у сфері кібербезпеки безпілотних військових апаратів демонструють значний інтерес до проблеми захисту цих систем від різноманітних загроз. Зокрема, багато науковців звертають увагу на зростання кількості атак, спрямованих на перехоплення та маніпуляцію сигналами управління безпілотниками, що може призвести до критичних наслідків для військових операцій.

Метою статті є дослідити ефективність криптографічного захисту сигналів управління, аутентифікації та контролю доступу, захисту програмного забезпечення та сенсорних систем, а також фізичного захисту комунікаційних каналів.

Виклад основного матеріалу. У сучасних військових конфліктах безпілотні апарати відіграють критично важливу роль, забезпечуючи розвідку, спостереження та навіть виконання бойових завдань. Однак їх широке використання робить їх привабливою мішенню для кіберзагроз. Нижче розглянуто основні типи атак, спрямованих на безпілотні військові апарати, а також їх можливі наслідки.

Для того, щоб перехопити чи знищити БПЛА їх на першому етапі потрібно виявити тому розглянемо методи виявлення БПЛА. В даний час існує безліч систем і комплексів для виявлення та протидії дронам і безпілотним літальним апаратам.

Основні методи виявлення та пошуку безпілотних військових апаратів :

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

1. Акустичні (звукові) рис.1. системи виявлення дронів контролюють характерні спектри звукових частот, що випромінюються безпілотними літальними апаратами. Вони здатні виявляти БПЛА на відстані кількох сотень метрів у безпосередній близькості. Однак, ефективність таких детекторів значно знижується в міських умовах через високий рівень фонового шуму. Більшість мікрофонів здатні виявляти дрони лише на відстані до 25-50 метрів, а в шумних міських районах аудіовиявлення стає практично марним. Крім того, зміна характеристик БПЛА, наприклад, заміна гвинтів або внесення інших модифікацій, може ускладнити їх акустичну ідентифікацію.

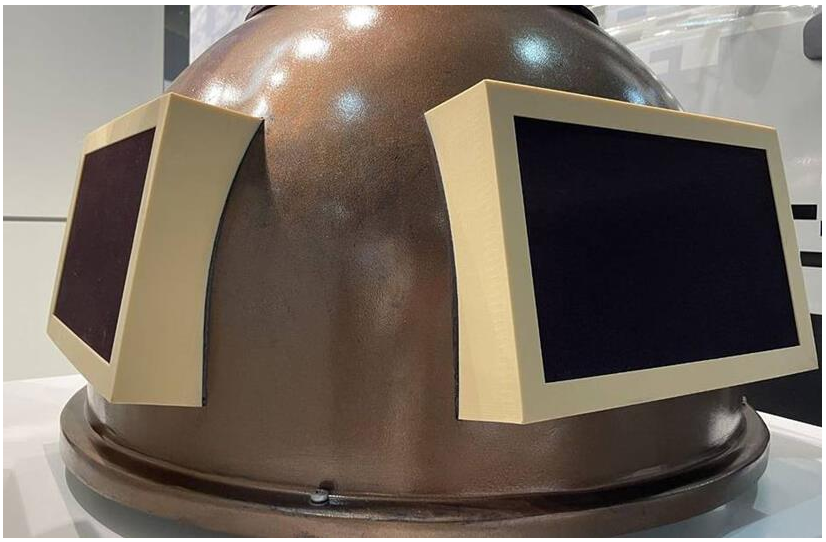


Рис.1. Акустичні (звукові) системи виявлення дронів

2. Візуальні (оптичні) рис.2. системи виявлення дронів використовують камери спостереження для пошуку рухомих повітряних об'єктів. Вони намагаються диференціювати БПЛА та птахів на основі розміру, траєкторії польоту та характеру руху. Однак, ефективна дальність дії таких камер обмежується відстанню не більше 350 метрів. Навіть застосовуючи спеціалізовані комп'ютерні алгоритми для відстеження польоту об'єктів, складно з впевненістю визначити, чи є спостережуваний об'єкт БПЛА або птахом. Деякі ключові характеристики польоту безпілотників, такі як ширяння, притаманні також багатьом видам птахів. Яскравим прикладом є чайки, які демонструють схожі польотні характеристики з деякими типами БПЛА.

Рис.2. Візуальні (оптичні) системи виявлення дронів

3. Теплові системи (рис.3) виявлення дронів ґрунтуються на аналізі теплових сигнатур безпілотних літальних апаратів. Вони здатні виявляти БПЛА на ефективній дальності близько 350 метрів. Однак, тепловізійні системи є менш ефективними порівняно з акустичними датчиками. Більшість сучасних БПЛА виготовлені з пластику та оснащені електричними двигунами, які випромінюють мінімальну кількість тепла. У деяких випадках теплова сигнатура БПЛА навіть нижча, ніж у птахів. Це значно знижує ефективність теплового методу виявлення безпілотників.

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

Рис.3. Теплові системи виявлення дронів

4. Радіочастотні методи (рис.4). виявлення дронів аналізують радіочастотні сигнали в діапазонах, які використовуються для керування безпілотними літальними апаратами. Це включає моніторинг частот 2,4 та 5,8 ГГц, а також виявлення SSID та MAC-адрес Wi-Fi, які передаються недорогими комерційними БПЛА. Такі системи мають більшу дальність дії порівняно з іншими методами. Однак, їм складно визначити напрямок (азимут та кут місця), звідки прилітає дрон, оскільки для цього потрібні додаткові технології пеленгації.

Рис.4. Радіочастотні методи виявлення дронів

5. Радарні (рис. 5) системи виявлення дронів використовують технології ближньої радіолокації для більш точного визначення місцезнаходження та характеристик безпілотних літальних апаратів. Основні розробки таких просунутих радарних рішень ведуться в Ізраїлі (MAGOS), США (ICX Technologies Inc., Rockwell Collins), Європі (Prime Consulting and Technologies), Росії («ЄНОТ», «Радескан-Антидрон», «Купол-М») та Україні (КП «НПК «Іскра», ДП НДІ РС «Квант-Радіолокація»), однак поки що вони застосовуються переважно військовими.

Технології ближньої радіолокації дозволяють отримувати детальну інформацію про рух відстежуваного об'єкта: місце виявлення, траєкторію, швидкість, розміри. Якщо радарна система інтегрована з поворотною платформою, на якій встановлені відеокамера або тепловізор, то можна також фіксувати відеозображення.Проте радіолокаційні системи виявлення дронів мають певні обмеження. Вони ефективно працюють на відкритих

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

просторах, таких як відкриті ділянки місцевості без високої рослинності чи будівель, акваторії, аеропортові території тощо. Водночас, такі радари практично не здатні виявляти малогабаритні БПЛА, що запускаються з близької відстані або летять на малих висотах. Камери, інтегровані з радаром, також не можуть виявити такі «малюки» через обмежений кут огляду.

Рис.5. Радарні системи виявлення дронів

Розглянемо перехоплення сигналів управління. Опис типових атак на безпілотні апарати через перехоплення радіосигналів Безпілотні апарати керуються дистанційно за допомогою радіосигналів, які можуть бути вразливими до перехоплення. Кібервійська противника можуть здійснити атаку шляхом прослуховування та підміни цих сигналів, що дозволяє їм перехопити управління апаратом або внести дезінформацію в його роботу. Зазвичай такі атаки використовують радіоелектронну боротьбу (рЕБ) також пристрої для перехоплення радіочастот або спеціалізовані комплекси для аналізу та модифікації сигналів управління. Деякі з найвідоміших систем включають:

Красуха-4 – комплекс радіоелектронної боротьби, який може глушити та перехоплювати сигнали управління дронами на великій відстані. Ця система здатна створювати перешкоди для радіозв'язку та управління, що ускладнює роботу дронів і робить їх вразливими для перехоплення або знищення (рис.6).

Рис.6. Комплекс радіоелектронної боротьби Красуха-4

Леєр-3 – система, яка використовує спеціалізовані дрони для придушення сигналів GSM, що може включати і дрони, які працюють на основі цих сигналів. Цей комплекс може глушити та спотворювати сигнали управління, що ускладнює контроль над апаратами (рис.7).

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

Рис.7. Система Леєр-3

Автобаза 1Л222 – система радіорозвідки і РЕБ, що може перехоплювати сигнали управління дронами та створювати перешкоди для їх нормального функціонування. «Автобаза» може виявляти джерела радіосигналів, аналізувати їх і впливати на роботу дронів (рис 8.).

Рис.8. Автобаза 1Л222

Приклади успішних атак у світовіті на військові безпілотники. Історія кібербезпеки безпілотних апаратів налічує кілька резонансних випадків перехоплення сигналів управління. Один із найвідоміших випадків стався в 2011 році, коли іранські військові заявили про захоплення американського безпілотного літального апарату RQ-170 Sentinel (рис 9.)

Рис.9 RQ-170 Sentinel

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

Іранці стверджували, що їм вдалося взяти під контроль безпілотник, перехопивши та підмінивши сигнали управління. Це показало потенційні вразливості навіть у високотехнологічних військових системах.

Дистанційне управління та взлом. Захоплення контролю над безпілотними апаратами є однією з найсерйозніших кіберзагроз, оскільки воно може призвести до повної втрати апарату та компрометації місії. Одним із методів злому є атака типу «людина посередині» (Man-in-the-Middle)(рис 10.), коли атакуючий перехоплює сигнали між оператором і безпілотником та підмінює їх власними командами.

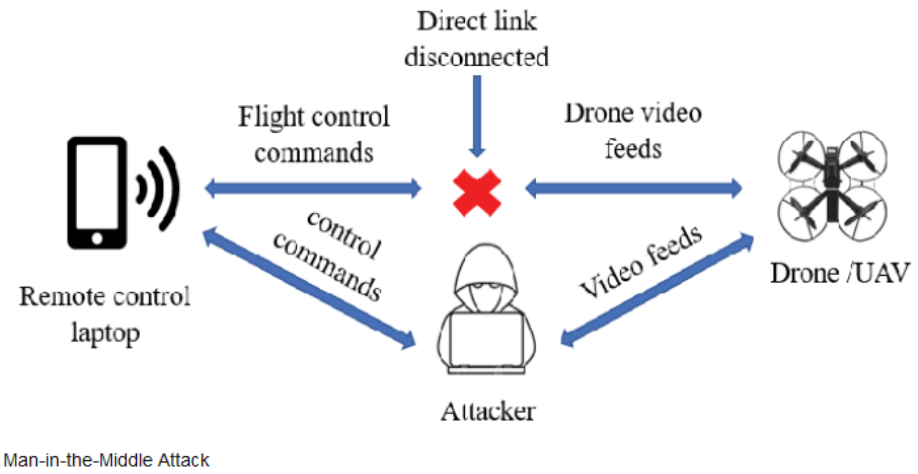


Рис.10 Man-in-the-Middle

Іншим методом є використання експлойтів, які атакують вразливості програмного забезпечення безпілотника, що дозволяє зловмисникам отримати доступ до системи управління.

Огляд відомих кейсів атак на військові дрони окрім випадку з RQ-170, відомий випадок у 2009 році, коли іракські повстанці використовували прості комерційні програмні засоби для перехоплення відеопотоку з американських безпілотників Predator (рис.11).

Рис.11 MQ-1 Predator

Це стало можливим через відсутність шифрування у відеопотоках, що було серйозною вразливістю в тодішніх системах управління дронами.

Інші вразливості вразливості в програмному забезпеченні та сенсорах безпілотних апаратів. Програмне забезпечення безпілотних апаратів також є важливою точкою атаки. Недоліки у коді, помилки в логіці або недостатньо захищені інтерфейси можуть бути використані для несанкціонованого доступу або для виведення апарату з ладу (рис.12).

Зокрема, атаки на сенсори можуть призвести до того, що безпілотник буде отримувати хибні дані про своє оточення, що може призвести до аварій або виконання некоректних дій. Атаки типу

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

«відмова в обслуговуванні» (DoS) та їх вплив на функціональність апаратів. Атаки типу «відмова в обслуговуванні» (DoS) спрямовані на перевантаження системи, що призводить до її неспроможності обробляти запити в реальному часі.

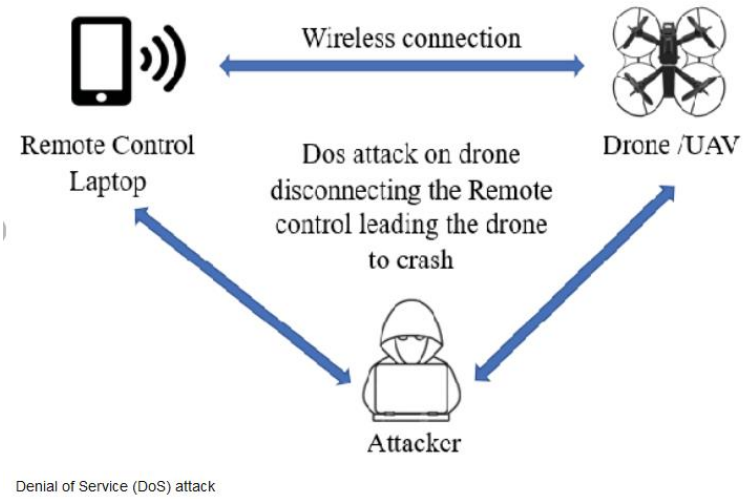


Рис.12 Відмова в обслуговуванні (DoS)

У контексті безпілотних апаратів такі атаки можуть перешкодити отриманню сигналів управління або даних з сенсорів, що призведе до втрати контролю над апаратом або його некоректної роботи. Наприклад, атака DoS може змусити дрон зависнути на місці або аварійно приземлитися, що робить його вразливим для подальших атак або фізичного захоплення.

Ці кіберзагрози підкреслюють необхідність постійного розвитку методів захисту та вдосконалення систем управління безпілотними військовими апаратами для забезпечення їх надійної роботи в умовах зростаючої складності сучасних кіберзагроз.

Розглянемо основні методи захисту від перехоплення та дистанційного управління.

Криптографічний захист сигналів управління:

1. Використання сучасних методів шифрування для захисту радіокомунікацій: У військових безпілотних апаратах критично важливо забезпечити надійний захист радіосигналів, щоб уникнути їхнього перехоплення або підробки. Сучасні криптографічні методи, такі як AES (Advanced Encryption Standard) або ECC (Elliptic Curve Cryptography), можуть використовуватись для шифрування сигналів, що передаються між оператором та безпілотником. Це дозволяє гарантувати, що тільки авторизовані сторони можуть інтерпретувати передані дані.

Огляд існуючих стандартів криптографії для військових безпілотників: Існують різноманітні криптографічні стандарти, які використовуються для захисту комунікацій військових безпілотників. До таких стандартів належать MIL-STD-188-125 та STANAG 4586, які регулюють безпеку зв'язку і захист від електромагнітних імпульсів. Ці стандарти визначають вимоги до стійкості шифрування та забезпечують відповідність вимогам національної і міжнародної безпеки.

2. Аутентифікація та контроль доступу. Впровадження багаторівневої аутентифікації для захисту від несанкціонованого доступу: Для забезпечення надійного захисту від спроб несанкціонованого доступу до безпілотних апаратів використовується багаторівнева аутентифікація. Це може включати комбінацію криптографічних ключів, біометричних даних, одноразових паролів та інших методів, які ускладнюють злом системи. Така аутентифікація дозволяє переконатися, що управління апаратом здійснюється тільки авторизованими особами.

Системи управління доступом та моніторинг активності: Системи управління доступом до безпілотних апаратів повинні включати засоби постійного моніторингу активності, щоб виявляти та запобігати спробам несанкціонованого доступу. Наприклад, системи SIEM (Security Information and Event Management) можуть використовуватися для виявлення підозрілої активності та автоматичного блокування несанкціонованих спроб доступу.

3. Захист програмного забезпечення та сенсорних систем. Методи захисту програмного забезпечення від вразливостей та експлойтів: Програмне забезпечення, що використовується у безпілотних апаратах, повинно бути захищеним від вразливостей та експлойтів. Це включає регулярні оновлення та патчі, аналіз коду на предмет можливих вразливостей, а також

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

впровадження механізмів захисту, таких як DEP (Data Execution Prevention) та ASLR (Address Space Layout Randomization). Також можуть бути застосовані методи машинного навчання для виявлення та нейтралізації аномальної поведінки в реальному часі.

Інтеграція механізмів безпеки у сенсорні системи для запобігання хибним даним: Сенсорні системи безпілотних апаратів є критичною частиною їхньої інфраструктури і повинні бути захищеними від фальсифікації даних. Це можна досягти через використання криптографічних методів для перевірки цілісності даних, отриманих із сенсорів, а також через застосування технологій машинного навчання для виявлення аномалій та хибних даних, що можуть бути наслідком кібернападу.

4. Фізичний захист та безпека комунікацій. Технічні рішення для захисту від фізичних атак та перехоплення даних: Безпілотні апарати також потребують фізичного захисту від спроб їхнього перехоплення або прямого втручання. Це може включати встановлення захищених корпусів, самознищення у разі захоплення, а також використання технологій для виявлення і нейтралізації загроз під час польоту.

Використання стійких до перешкод та завад комунікаційних протоколів: Щоб захистити безпілотні апарати від перешкод та глушіння, використовуються стійкі комунікаційні протоколи, такі як FHSS (Frequency Hopping Spread Spectrum) та DSSS (Direct Sequence Spread Spectrum). Ці технології дозволяють ефективно знижувати вплив радіоелектронної боротьби на управління апаратом, зберігаючи стабільність та безпеку зв'язку навіть у складних умовах.

Ці методи дозволяють суттєво знизити ризики перехоплення та дистанційного управління військовими безпілотниками, забезпечуючи їхню надійність та ефективність у бойових умовах.

Практичні приклади та впровадження методів захисту.

Реальні кейси успішної реалізації методів захисту аналіз випадків, де були успішно застосовані методи захисту безпілотних апаратів: Розглянемо конкретні випадки, коли методи захисту безпілотних апаратів довели свою ефективність на практиці. Наприклад, аналіз кейсів, де використання криптографічних методів захисту сигналів управління та багаторівневої аутентифікації запобігло спробам перехоплення або взлому безпілотників. Особлива увага приділяється випадкам застосування FHSS для захисту зв'язку та стійкості проти радіоелектронної боротьби.

Оцінка ефективності використаних рішень: проводиться оцінка ефективності застосованих рішень у реальних бойових умовах. Оцінюється, наскільки успішно вдалося забезпечити захист від перехоплення сигналів, дистанційного управління та інших кіберзагроз, а також яким чином вдалося підвищити надійність та стійкість систем управління безпілотними апаратами.

Проблеми та виклики впровадження та виявлення проблем, що виникають під час впровадження методів кібербезпеки: Незважаючи на те, що методи захисту безпілотних апаратів мають високу ефективність, існує ряд проблем та викликів, що виникають під час їхнього впровадження. До таких проблем належать складність інтеграції криптографічних рішень у наявні системи, витрати на модернізацію апаратного та програмного забезпечення, а також проблеми з забезпеченням сумісності різних систем безпеки. Окрім цього, можуть виникати питання щодо продуктивності системи після впровадження додаткових захисних механізмів.

Аналіз можливих шляхів подолання цих проблем розглядаються потенційні шляхи подолання виявлених проблем. Пропонується впровадження нових підходів до проектування безпілотних систем, які включають безпеку на ранніх етапах розробки (security by design), адаптивне шифрування, що мінімізує вплив на продуктивність, а також використання машинного навчання для автоматизації процесів виявлення та реагування на загрози. Розглядаються також організаційні заходи, такі як підвищення кваліфікації персоналу та удосконалення процедур безпеки. Цей розділ акцентує увагу на практичних аспектах захисту безпілотних військових апаратів, що дозволяє зрозуміти, як теоретичні методи захисту втілюються в реальному світі, а також які виклики можуть виникнути під час їхньої реалізації.

Висновки та перспективи подальшого дослідження.

Дослідження показало, що безпілотні військові апарати піддаються значним кіберзагрозам, включаючи перехоплення сигналів управління, дистанційне захоплення контролю, вразливості програмного забезпечення та атаки на сенсорні системи. Найбільш ефективними методами нейтралізації цих загроз є використання криптографічного захисту, багаторівневої аутентифікації, інтеграції механізмів безпеки на рівні програмного забезпечення та фізичного захисту комунікацій.

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

Проведений аналіз підтверджує, що впровадження запропонованих рішень суттєво підвищує рівень захисту безпілотних апаратів від кіберзагроз. Реальні кейси показали ефективність застосування криптографії та контролю доступу в реальних умовах, що дозволяє рекомендувати ці рішення для широкого впровадження в сучасних військових системах.

Перспективи подальших досліджень незважаючи на досягнуті результати, залишається низка питань, що потребують подальшого дослідження. Це стосується розробки нових методів шифрування, які будуть стійкими до майбутніх загроз, таких як квантові обчислення, а також дослідження ефективності методів захисту в умовах масованих кібер- та радіоелектронних атак.

Пропозиції щодо вдосконалення існуючих методів захисту можуть зосередитися на вдосконаленні існуючих методів захисту, таких як адаптація системи аутентифікації до змінних умов бойових операцій, впровадження машинного навчання для проактивного виявлення загроз, та розробка нових комунікаційних протоколів, стійких до перешкод і спроб перехоплення. Також важливим напрямом є розробка інтегрованих систем захисту, які об'єднують криптографію, контроль доступу та фізичний захист у єдину платформу.

Список бібліографічного опису

1. Kurtz, J. (2021). Cybersecurity in Unmanned Aerial Systems: Challenges and Solutions. Springer.ISBN: 978-3030598654.
Wang, X., & Liu, S. (2019). Cryptographic Techniques for Secure Communication in UAV Networks. IEEE Transactions on Information Forensics and Security, 14(5), 1427-1438.DOI: 10.1109/TIFS.2019.2892214.
2 Bertino, E., & Sandhu, R. (2020). Database Security: Concepts, Approaches, and Challenges. Springer. ISBN: 978-3030201717.
3 Kumar, A., & Arora, A. (2022). Authentication Mechanisms for Military UAVs: A Comparative Study. Journal of Cyber Security and Privacy, 3(2), 145-160.DOI: 10.1002/cem.3589.
4. Chen, Z., & Zhang, Y. (2018). Secure Communication Protocols for UAVs: A Survey. IEEE Access, 6, 20415-20428.DOI: 10.1109/ACCESS.2018.2824724.
5. Srinivas, S., & Nair, S. (2019). Software Vulnerabilities in UAV Systems: Identifying and Mitigating Risks. International Journal of Information Security, 18(1), 1-12.DOI: 10.1007/s10207-018-0414-4.
6 National Institute of Standards and Technology (NIST) (2020). Guidelines on Securing Unmanned Aircraft Systems. NIST Special Publication 800-213.URL: nist.gov.
7 Meyer, R., & Barnum, S. (2021). Physical and Cyber Security of UAV Systems: A Comprehensive Review. ACM Computing Surveys, 54(4), 1-35.DOI: 10.1145/3457658.
8. Zhang, L., & Li, Q. (2022). Emerging Threats and Defensive Measures for UAV Communication Systems. IEEE Transactions on Aerospace and Electronic Systems, 58(1), 369-382.DOI: 10.1109/TAES.2021.3075270.

References

1. Kurtz, J. (2021). Cybersecurity in Unmanned Aerial Systems: Challenges and Solutions. Springer.ISBN: 978-3030598654.
Wang, X., & Liu, S. (2019). Cryptographic Techniques for Secure Communication in UAV Networks. IEEE Transactions on Information Forensics and Security, 14(5), 1427-1438.DOI: 10.1109/TIFS.2019.2892214.
2 Bertino, E., & Sandhu, R. (2020). Database Security: Concepts, Approaches, and Challenges. Springer. ISBN: 978-3030201717.
3 Kumar, A., & Arora, A. (2022). Authentication Mechanisms for Military UAVs: A Comparative Study. Journal of Cyber Security and Privacy, 3(2), 145-160.DOI: 10.1002/cem.3589.
4. Chen, Z., & Zhang, Y. (2018). Secure Communication Protocols for UAVs: A Survey. IEEE Access, 6, 20415-20428.DOI: 10.1109/ACCESS.2018.2824724.
5. Srinivas, S., & Nair, S. (2019). Software Vulnerabilities in UAV Systems: Identifying and Mitigating Risks. International Journal of Information Security, 18(1), 1-12.DOI: 10.1007/s10207-018-0414-4.
6 National Institute of Standards and Technology (NIST) (2020). Guidelines on Securing Unmanned Aircraft Systems. NIST Special Publication 800-213.URL: nist.gov.
7 Meyer, R., & Barnum, S. (2021). Physical and Cyber Security of UAV Systems: A Comprehensive Review. ACM Computing Surveys, 54(4), 1-35.DOI: 10.1145/3457658.
8. Zhang, L., & Li, Q. (2022). Emerging Threats and Defensive Measures for UAV Communication Systems. IEEE Transactions on Aerospace and Electronic Systems, 58(1), 369-382.DOI: 10.1109/TAES.2021.3075270.

© Борисов О.В., Артабаєв Ю.З., Сурма А.І.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-15
УДК 004.62
Веремеєнко Андрій Андрійович, магістрант
https://orcid.org/0009-0000-1916-3254
Національний університет «Львівська політехніка», м. Львів, Україна

ДОСЛІДЖЕННЯ ВИКОРИСТАННЯ ФРАКТАЛЬНИХ АЛГОРИТМІВ КОМПРЕСІЇ В ІНФОРМАЦІЙНИХ СИСТЕМАХ, ЩО ПРАЦЮЮТЬ З ВЕЛИКИМИ ОБСЯГАМИ ДАНИХ ТА ВИМАГАЮТЬ ВИСОКИХ КОЕФІЦІЄНТІВ СТИСНЕННЯ ПРИ МАКСИМАЛЬНОМУ ЗБЕРЕЖЕННІ ЯКОСТІ ЗОБРАЖЕНЬ

Веремеєнко А.А. Дослідження використання фрактальних алгоритмів компресії в інформаційних системах, що працюють з великими обсягами даних та вимагають високих коефіцієнтів стиснення при максимальному збереженні якості зображень. В роботі досліджено ефективність застосування фрактальних алгоритмів компресії в інформаційних системах, що обробляють великі обсяги даних та вимагають високих коефіцієнтів стиснення при збереженні якості зображень. Описано принципи роботи фрактальних алгоритмів, які використовують математичні моделі для виявлення самоподібних структур у даних, забезпечуючи високе співвідношення стискання без значної втрати якості. Проведено порівняння фрактальних методів з іншими популярними методами стиснення, такими як JPEG та Wavelet Transform, оцінено їх ефективність на основі продуктивності, коефіцієнтів стиснення та якості відновлених зображень. Виявлено, що фрактальні алгоритми є перспективними для систем, що потребують збереження високоякісних зображень при мінімальній втраті якості, особливо в умовах обмежених ресурсів зберігання та пропускної здатності мереж. Розроблено рекомендації щодо впровадження фрактальних алгоритмів у інформаційні системи для підвищення ефективності управління великими обсягами даних.
Ключові слова: фрактальний алгоритм, компресія зображень, інформаційна система, афінні перетворення, квадратичне розбиття.

Veremeenko A. Study of the usage of fractal compression algorithms for information systems that handle large volumes of data and require high compression ratios while preserving image quality. This work investigates the effectiveness of applying fractal compression algorithms in information systems that process large volumes of data and require high compression ratios while maintaining image quality. The principles of fractal algorithms, which utilize mathematical models to detect self-similar structures in data, are described, providing high compression ratios without significant loss of quality. A comparison of fractal methods with other popular compression methods, such as JPEG and Wavelet Transform, is conducted, evaluating their effectiveness based on performance, compression ratios, and the quality of restored images. It is found that fractal algorithms are promising for systems that need to preserve high-quality images with minimal quality loss, especially under conditions of limited storage resources and network bandwidth. Recommendations are developed for implementing fractal algorithms in information systems to improve the efficiency of managing large volumes of data.
Keywords: fractal algorithm, image compression, information system, affine transformations, quadratic partitioning.

Постановка наукової проблеми. Сучасний світ стикається з експоненціальним зростанням обсягів даних, що створюються та зберігаються в різних галузях, таких як телекомунікації, наука, медицина, мультимедіа та індустрія розваг. Ефективне стиснення даних стає критично важливим для зниження вимог до зберігання та покращення швидкості передачі даних. Стиснення даних дозволяє значно зменшити обсяг, що займає інформація на дисках та інших носіях. Це знижує витрати на обладнання для зберігання даних і покращує ефективність використання існуючих ресурсів. Також зменшується навантаження на мережеві канали при передачі великих обсягів даних. Зменшення обсягу даних за допомогою стиснення дозволяє прискорити процес передачі інформації через мережі, особливо в умовах обмеженої пропускної здатності.

Будь-яке зображення може бути збережене як сукупність пікселів. Зображення, збережене в комп'ютері, може мати певну кількість надлишковості. Якщо зображення має високу надлишковість – воно вимагає не тільки більше місця для зберігання, але й високошвидкісних каналів для передачі цього зображення з однієї мережі до іншої. Для мінімізації часу передачі та вартості зберігання необхідне стиснення.

Фрактальні алгоритми стиснення даних можуть забезпечити високу якість відтворення зі значним зменшенням розміру файлі. Фрактальні алгоритми компресії даних базуються на математичних принципах, що дозволяють ефективно виявляти самоподібні структури у даних. Це забезпечує високе співвідношення стискання без значних втрат якості. Фрактальні алгоритми також мають здатність до адаптивного стискання, що дозволяє більш ефективно стискати дані з різними характеристиками.

Фрактальні алгоритми, з їх здатністю до високоякісного стискання, можуть бути інтегровані в системи, що потребують збереження високої роздільної здатності зображень. В ході розробки однієї

© Веремеєнко А.А.

з таких систем, а саме – універсальної web-платформи для астрономів аматорів, виникла потреба в збереженні високоякісних світлин космосу, при як найменшій втраті якості. Оскільки така платформа передбачає роботу з великим об'ємом зображень, завантажуваних користувачами, вартість зберігання цих даних можу бути суттєвою. Також, це може вплинути на швидкодію платформи.

Таким чином, актуальність теми обумовлена необхідністю ефективного управління великими обсягами даних, зниженням витрат на зберігання та передачу, покращенням якості обслуговування користувачів даної платформи. В результаті аналізу та порівння існуючих алгоритмів компресії – фрактальний алгоритм був обраний як найбільш оптимальний та відповідний поставленим задачам і цілям.

Предметом дослідження є комплекс методів та алгоритмів фрактального стиснення даних, зокрема зображень, з метою досягнення високого рівня стиснення без значної втрати якості для збереження в розробленій інформаційній системі. Особлива увага приділяється порівнянню фрактальних методів з іншими популярними методами стиснення, такими як JPEG та Wavelet Transform, для оцінки їх ефективності та застосування у різних галузях.

Об'єктом дослідження є процеси стиснення та декомпресії даних за допомогою фрактальних алгоритмів, які використано у розробленій інформаційній системі. Дослідження охоплює аналіз ефективності фрактальних методів у зменшенні обсягів даних та впливу на якість відновлених зображень, а також порівняння цих методів з іншими технологіями стиснення.

Завданням даного дослідження є аналіз існуючих методів фрактального стиснення зображень та їх основних принципів; порівняння ефективності фрактальних алгоритмів з іншими популярними методами стиснення, такими як JPEG та Wavelet Transform, на основі показників продуктивності, коефіцієнтів стиснення та якості відновлених зображень; аргументація доцільності використання фрактальних алгоритмів компресії для систем, що вимагають збереження великих обсягів зображень з мінімальною втратою якості, формування вимог до інформаційної системи, яка ґрунтується на методах фрактального стиснення зображень.

Аналіз досліджень. Поняття фракталу було описане математиком IBM Бенуа Мандельбротом. Саме він виявив, що традиційна геометрія не може повною мірою використовуватися для опису структури природних об'єктів, які є складними та хаотичними, наприклад, гори, хмари, берегові лінії, дерева, тощо.

Неевклідова геометрія або фрактальна геометрія займається нерегулярними та фрагментними візерунками. Фрактали є складними об'єктами, які мають властивість самоподібності, тобто частина фрактального об'єкта подібна до всього об'єкта, отже, фрактали є повторенням тієї ж самої структурної форми.

Для створення будь-якого фрактала потрібні три речі: набір перетворень (IFS - Iterated Function System), база, з якої починається ітерація, та множина конденсації (можливо, порожня множина).

У фрактальному стисненні зображень застосовуються де-які математичні функції та афінні перетворення, які також називаються контрактними афінними перетвореннями. Афінне перетворення, застосоване до фрактального стиснення зображень, є комбінацією перетворень, тобто: обертання, зміна масштабу, нахил або переміщення. Контрактне афінне перетворення грає значну роль, оскільки відстань між двома точками в оригінальному зображенні стає все ближче і ближче до фіксованої точки, і ця фіксована точка після афінного перетворення є не що інше, як зображення. Зображення, отримане після застосування контрактного афінного перетворення, є наближенням до оригінального зображення.

В основі ітераційної системи функцій лежить обрання точки або фігури та її заміна на декілька інших, ідентичних. Ітераційна система функцій – це колекція контрактних афінних перетворень c1, c2 і так далі до cn.

Застосовування ітераційної системи функцій (IFS) відбувається до тих пір, доки вона не збігається до певної точки, яка є атрактором IFS. М. Барнслі вивів спеціальну форму контрактного відображення перетворення (CMT - Contractive Mapping Transform), що застосовується до IFS та називається теоремою Коледжу. Вона дістала таку назву через те, що М. Барнслі вперше представив свою ідею під час лекцій у коледжі. Ця теорема стверджує, що відстань Гаусдорфа, що є математичною мірою для визначення максимальної відстані між точками двох множин у метричному просторі, зокрема, між двома зображеннями має бути якомога меншою, тобто атрактор IFS має бути близьким наближенням до оригінального зображення. Відстань між оригінальним зображенням та атрактором називається помилкою коледжу, і повинна бути якомога меншою.

© Веремеєнко А.А.

На основі напрацювань М. Барнслі, інший відомий дослідник в галузі фракталів, М.Ф. Жакен, запропонував деталізований підхід. У методі Жакена зображення розділяється на підзображення, які називаються "блоками діапазону", і PIFS (Partitioned Iterated Function System, або ж розподілена ітераційна система функцій) застосовується до підзображень, а не до всього зображення, щоб знайти найбільш підходящий блок домену.

У розділеній IFS замість застосування афінного перетворення до всього зображення, контрактні афінні перетворення застосовуються до його частин, і об'єднання афінних перетворень створює кінцеве зображення.

Алгоритм фрактального стиснення зображень зображений на рис. 1.

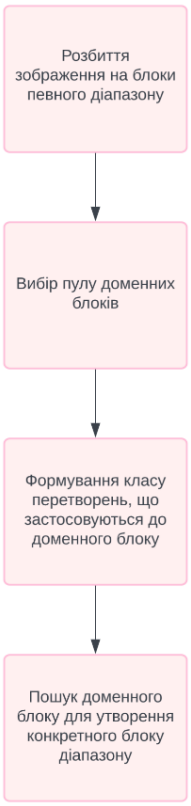


Рис. 1. Алгоритм фрактального стиснення зображень

Основною метою є збільшення коефіцієнта стиснення та мінімізація часу стиснення. Іншим важливим аспектом є розмір стисненого файлу у порівнянні з якістю декомпресованого зображення. Існує два типи стиснення: стиснення без втрат та стиснення з втратами. Стиснення з втратами завжди створює зображення, які не є точною копією оригінальних даних, воно усуває надлишковість і створює наближення до оригіналу. Фрактальний алгоритм компресії зображень є методом стиснення з втратами.

Мета роботи. Метою даної статті є аналіз існуючих методів фрактального стиснення зображень, порівняння їх з іншими популярними методами стиснення, такими як JPEG і Wavelet Transform, та аргументація доцільності використання фрактального алгоритму компресії для розроблення інформаційної системи, що сприяє збереженню великих обсягів зображень, з мінімальною втратою якості. Також розглядається проблематика пошуку самоподібних блоків у зображеннях, вплив різних схем розбиття зображення на ефективність стиснення та основні показники якості відновлених зображень.

1. Порівняння алгоритмів та результати

Існують різні методи компресії зображень, такі як Joint Photographic Experts Group (JPEG), JPEG2000, Wavelet Transform та фрактальне стиснення зображень. Ці методи були протестовані для реальних зображень розміром *512×512* та *1024×1024*.

Результати продуктивності наведені в таблицях 1, 2 та 3.

© Веремеєнко А.А.

Табл. 1. Продуктивність алгоритмів компресії для зображення 512×512

Алгоритм	**Значення PSNR(дБ)**	**Компресія**	**Декомпресія**
Wavelet	37.65	9.89 сек	6.7 сек
JPEG	33.72	0.49 сек	0.29 сек
JPEG2000	38.27	15.20 сек	10.27 сек
Фрактальний	31.03	1201.45 сек	4.92 сек

Табл. 2. Продуктивність алгоритмів компресії для зображення відбитків пальців 1024×1024

Алгоритм	**Значення PSNR(дБ)**	**Компресія**	**Декомпресія**
Wavelet	39.0	24.5 сек	14.9 сек
JPEG	35.2	1.03 сек	0.57 сек
JPEG2000	40.0	31.2 сек	20.1 сек
Фрактальний	30.2	3019.21 сек	9.7 сек

Табл. 3. Продуктивність на основі коефіцієнта стиснення різних алгоритмів кодування

Алгоритм	**Коефіцієнт стиснення**
Wavelet	13
JPEG	4
JPEG2000	16
Фрактальний	8

Можна трактувати результати наведені в таблиці 3, наступним чином: фрактальне стиснення забезпечує коефіцієнти стиснення не менше 8:1. Хоча цей метод не досягає таких високих коефіцієнтів, як Wavelet або JPEG2000 і тривалість компресії є значно більшою, проте, він має перевагу у збереженні якості зображення при високих ступенях стиснення та забезпечує відносно швидку декомпресію, що є основною вимогою в розробці платформи.

Результати стиснення за допомогою фрактального алгоритму зображені на рис. 2-5 для прикладів світлини «Кіт» та «Жук» та демонструють високе збереження якості навіть при великому ступені компресії (розмір рангового блоку до 64 пікселів, що забезпечує компресію в більш ніж 1000 разів). Проте, висока якість досягається найкраще для зображень, що містять елементи самоподібності, що добре видно в різниці між зображенням «Кіт» та «Жук».

Рис. 2. Оригінальні зображення

Рис. 3. Декомпресійовані зображення при розмірі рангового блоку r=2

© Веремеєнко А.А.

Рис. 4. Декомпресійовані зображення при розмірі рангового блоку r=8

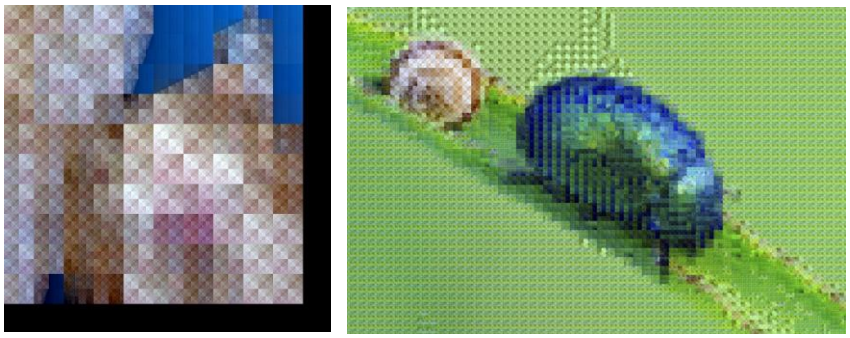


Рис. 5. Декомпресійовані зображення при розмірі рангового блоку r=64

1.1 Переваги фрактального стиснення

1. Незалежність від роздільної здатності та пристроїв.
2. Декомпресія в 6 разів швидша, ніж у JPEG2000.
3. Вища якість при високих коефіцієнтах стиснення.
4. Коефіцієнти стиснення можуть бути покращені шляхом збільшення часу стиснення без збільшення часу декомпресії або зниження якості зображення.

1.2 Недоліки фрактального стиснення

1. Час стиснення в значно довший, ніж у інших методів.
2. Непідтверджена технологія.
3. Висока вартість для звичайних користувачів.

1.3 Переваги JPEG

1. JPEG підходить тоді, коли потрібні низькі коефіцієнти стиснення.
2. Оскільки це промисловий стандарт, він є безкоштовним.

1.4 Недоліки JPEG

1. JPEG не працює з бінарними зображеннями.
2. JPEG не працює добре з нереалістичними зображеннями.
3. Якість зображень значно погіршується при високих коефіцієнтах стиснення.

2. Перетворення

2.1 Афінне перетворення

Алгоритм фрактальної компресії зображень базується на ідеї самоподібності, тобто частини зображення повторюються в різних масштабах. Це дозволяє представити зображення як набір афінних перетворень, що значно зменшує обсяг даних, необхідних для його збереження.

Афінні перетворення – це лінійні перетворення, які є композицією обертання, трансляцій, масштабувань та зсувів. Воно не зберігає кути або довжини. Два або більше послідовних перетворень можуть бути застосовані до зображення за допомогою афінного перетворення.

Перетворення $\omega: R^{\mathbb{2}} \rightarrow R^{\mathbb{2}}$ має вигляд:

$$w(x, y) = (ax + by + e, cx + dy + f),$$

де a, b, c, d, e і f – дійсні числа, це називається (двовимірним) афінним перетворенням. Афінні перетворення є ключовим поняттям для фрактальних алгоритмів.

2.2 Теорема про контрактні відображення Банака

Перетворення $f: X \rightarrow X$ в метричному просторі (X, d)) називається контрактним або контрактним відображенням, якщо існує константа $0 \leq s < 1$ така, що:

© Веремеєнко А.А.

$$d\big(f(x), f(y)\big) \le s \cdot d(x, y), \quad \forall x, y \in X$$

Будь-яке таке число s називається коефіцієнтом контрактності для f.

2.3 Простір Гаусдорфа

Нехай (X, d) є повним метричним простором. Тоді відстань Гаусдорфа між точками A і B у $H(X)$ визначається як:

$$h(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} d\,(a, b), \sup_{b \in B} \inf_{a \in A} d\,(a, b)\}$$

Це також називається метричним простором Гаусдорфа.

3. Ітераційні системи функцій (IFS)

IFS стали дуже популярними у середині 1980-х років. М. Барнслі та його колеги з Технологічного інституту в Джорджії першими помітили потенціал IFS для застосування в комп'ютерній графіці. М. Барнслі і А. Слоан представили в наукових журналах здатність IFS до стиснення кольорових зображень із відношенням понад 10 000 до 1.

Ітераційна система функцій з ймовірностями складається з IFS разом з упорядкованим набором чисел $\{p_1, p_2, \dots, p_N\}$, таких що $p_1 + p_2 + p_3 + \cdots + p_N = 1$ і $p_i > 0$ для $(i = 1, 2, \dots, N)$.

Ймовірності пов'язані з теорією міри атракторів IFS і відіграють роль у обчисленні зображень атрактора IFS за допомогою алгоритму випадкової ітерації.

3.1 Недоліки IFS

Основний недолік IFS полягає в складності процесу отримання набору афінних перетворень для представлення даного зображення. Для повністю автоматизованого фрактального стиснення зображень необхідний алгоритм, який не вимагає "користувацького втручання", такого як налаштування коефіцієнтів в IFS.

4. Методи під IFS

4.1 Детерміністичний метод

У цьому методі, де-які афінні перетворення застосовуються до кожної підмножини зображення і таким чином, шукається наступне зображення, яке повинно бути повною підмножиною простору (R^2), де знаходиться зображення. Після багаторазового застосування афінних перетворень ми отримуємо послідовність зображень, які повинні зближуватися до деякої точки, яка буде граничною точкою. Ця гранична точка – і є зображення. Для використання цього підходу в генерації фракталів потрібна велика кількість пам'яті, оскільки на кожній ітерації ми генеруємо деяке зображення, яке повинно бути тимчасово збереженим.

Нехай $\{X; w_1, w_2, \dots, w_N\}$ є гіперболічним IFS. Виберіть компактну множину $A_0 \subseteq R^{\mathbb{2}}$. Потім обчислюйте послідовно $A_n = W^n(A)$ відповідно до:

$$A_{n+1} = \bigcup_{j=1}^{N} w_j(A_n) \quad \text{для } n = 1,2, \dots$$

Таким чином, створюється послідовність $\{A_n : n = 0,1,2,3, \dots\} \subseteq H(X)$. Згідно з теоремою IFS, послідовність $\{A_n\}$ збігається до атрактора IFS у метриці Гаусдорфа. Фрактали можуть бути згенеровані за допомогою ітерації Ішікави. Метою ітерації Ішікави, в свою чергу, є визначенням різних типів орбіт для генерації фракталів.

4.2 Алгоритм випадкової ітерації

Підхід випадкової ітерації відрізняється від детерміністичного підходу тим, що початкова множина є одиничною точкою, і на кожному рівні ітерації використовується лише одне з визначальних афінних перетворень для обчислення наступного рівня. На кожному рівні афінне перетворення, що застосовується, обирається випадковим чином. Усі точки, крім ранніх, наносяться на графік і відкидаються після використання для обчислення наступного значення. Випадковий алгоритм уникає потреби у великій пам'яті комп'ютера, та найкраще підходить для простих комп'ютерів, на яких можна обчислювати та відображати одну точку за раз. З іншого боку, для створення результуючого зображення цим способом потрібно тисячі точок, щоб воно не було недостатнім .

Нехай $\{X; w_1, w_2, \dots, w_N; p_1, p_2, \dots, p_N\}$ є IFS з ймовірностями. Виберіть $(x_0 \in X)$ і далі вибирайте рекурсивно, незалежно,

$$x_n \in \{w_1(x_{n-1}), w_2(x_{n-1}), \dots, w_N(x_{n-1})\} \quad \text{для } n = 1,2,3, \dots,$$

© Веремеєнко А.А.

де ймовірність того, що $x_n = w_i(x_{n-1})$ дорівнює p_i. Таким чином, створюється послідовність$\{x_n : n = 0,1,2,3, \dots\} \subseteq X$.

5. Фрактальна компресія зображень

Фрактальна компресія зображень є методом стиснення з втратами, що використовує фрактальну теорію. Основна ідея фрактального стиснення полягає в тому, щоб використовувати властивість самоподібності зображень для їх стиснення. Збереження зображення як набору пікселів вимагає більшої кількості пам'яті. Зберігаючи лише коефіцієнти афінних перетворень, потреба в пам'яті може бути зменшена. Стиснення є більш обчислювально витратним, оскільки інтенсивний пошук використовується для знаходження самоподібних блоків. При високому коефіцієнті компресії, фрактальне стиснення може забезпечити вищу якість.

Стандартні методи стиснення зображень можуть оцінюватися за їх коефіцієнтом стиснення: відношенням пам'яті, необхідної для збереження оригінального зображення, до пам'яті, необхідної для збереження зображення в скомпресійованій формі. Після застосування алгорритму фрактальної компресії, зображення стають незалежними від роздільної здатності. Це означає, що при генерації збільшених зображень стає видимою більша кількість деталей порівняно зі звичайним збільшенням, яке поступово розмивається. Фрактальне стиснення зображень є дуже ефективним для компресії зображень з високим ступенем самоподібності. Проте, процес пошуку відповідних блоків та застосування афінних перетворень може бути дуже обчислювально важким і займати значний час.

Аналіз цих методів дозволив зробити висновок, що для стиснення зображень в інформаційній системі найбільш доцільно використовувати фрактальну компресію зображень з декількох причин. Детерміністичний метод не забезпечує такої високої якості компресії, як фрактальна компресія. Він підходить більше для математичних та теоретичних задач, ніж для практичного використання в інформаційних системах для стиснення зображень. Хоча метод випадкової ітерації може бути корисним у генерації фракталів та деяких графічних застосуваннях, він менш ефективний для стиснення зображень через високу обчислювальну складність і менш передбачувані результати.

Фрактальна компресія здатна забезпечити високий ступінь стиснення без значної втрати якості. Це особливо важливо для великих зображень або для інформаційної системи, яка зберігатиме велику кількість графічних даних. Цей метод дозволяє зберегти дрібні деталі та текстури, що робить його ідеальним для фотографій та складних зображень. Фрактальна компресія добре підходить для зображень, які часто піддаються масштабуванню, оскільки дозволяє зберігати високий рівень деталізації навіть при збільшенні зображення. Отже, для розроблення інформаційної системи, яка займається стисненням зображень, фрактальна компресія є найкращим вибором, забезпечуючи баланс між високим ступенем стиснення та збереженням якості зображень.

5.1 Алгоритм фрактальної компресії зображення

Для компресії зображення за допомогою фрактального алгоритму компресії, слід виконати наступні кроки:

1. Розділити зображення на блоки діапазону.

Розділити вихідне зображення на неперекривні блоки однакового розміру, які називаються блоками діапазону (range blocks).

2. Розділити зображення на доменні блоки.

Розділити зображення на перекривні блоки, які є більшими за блоки діапазону, і називаються доменними блоками (domain blocks). Розмір доменних блоків зазвичай вдвічі більший за розмір блоків діапазону.

3. Застосування афінних перетворень.

Для кожного доменного блоку виконується наступне:

- Масштабування доменного блоку до розміру блоку діапазону.
- Обчислення середнього значення яскравості та контрасту.

4. Пошук найкращого співпадіння.

Для кожного блоку діапазону слід знайти найкраще співпадіння серед масштабованих доменних блоків з використанням афінних перетворень. Це здійснюється за допомогою мінімізації різниці між блоками діапазону та доменними блоками.

5. Збереження параметрів перетворення.

Збереження параметрів найкращого афінного перетворення (масштаб s і зсув o) для кожного блоку діапазону.

© Веремеєнко А.А.

Для декомпресії зображення необхідно виконати ітераційний процес, що полягає в наступному:

1. Ініціалізація буферу зображення випадковими значеннями.
2. Для кожного блоку діапазону застосовуються збережені афінні перетворення до відповідного доменного блоку, оновлюючи значення пікселів у буфері зображення.
3. Процес повторюється, доки буфер зображення не збіжиться до стабільного стану.

5.2 Приклади використання фрактальної компресії зображень

Алгоритми з даного сімейства не є популярними в побуті, через свої унікальні особливості та специфічні вимоги до використання. Проте, вони успішно застосовуються в сферах, що працюють з великими обсягами зображень та вимагають максимальної збереження якості та можливості маштабування.

1. Медична візуалізація - у медичній візуалізації важливо зберігати високоякісні зображення для діагностики, але обсяг даних може бути дуже великим. Фрактальна компресія дозволяє зберігати медичні зображення (наприклад, МРТ, КТ) з високою якістю, забезпечуючи при цьому значне зменшення розміру файлів.
2. Астрономічні зображення - як було зазначено в статті, фрактальна компресія використовується для збереження високоякісних астрономічних зображень на універсальній web-платформі для астрономів-аматорів. Вона дозволяє зберігати великий обсяг зображень космічних тіл, зменшуючи витрати на зберігання і забезпечуючи швидку передачу даних.
3. Збереження супутникових зображень - супутникові зображення мають високу роздільну здатність і часто використовуються для моніторингу земної поверхні, прогнозування погоди та спостереження за екологічними змінами. Фрактальна компресія допомагає зберігати ці зображення з високою деталізацією, зменшуючи обсяг даних, що передаються та зберігаються.
4. Застосування у наукових дослідженнях - у наукових дослідженнях часто використовуються високоякісні зображення, наприклад, у біології (зображення клітин, тканин) або фізиці (зображення результатів експериментів). Фрактальна компресія дозволяє зберігати такі зображення з мінімальними втратами якості, що є критичним для точності наукових висновків.
5. Географічні інформаційні системи (ГІС) - ГІС використовують великі обсяги географічних даних, включаючи карти, аерофотознімки та супутникові зображення. Фрактальна компресія допомагає зменшити обсяг даних, необхідних для зберігання та обробки в ГІС, забезпечуючи при цьому високу якість зображень для точного аналізу та прийняття рішень.

6. Схеми розбиття

При розробці схеми фрактального кодування необхідно приймати конкретне рішення щодо типу розбиття зображення, яке використовується для формування блоку діапазону. Існує широкий вибір схем розбиття.

6.1 Розбиття на квадрати фіксованого розміру

Це найпростіший метод розбиття зображення. Такий тип розбиття успішний у перетворювальному кодуванні окремих блоків зображення, наприклад, у реалізації стандарту JPEG.

6.2 Квадродерево

Цей тип розбиття базується на рекурсивному поділі вибраних квадрантів зображення. Після розбиття зображення буде виглядати як деревоподібна структура, в якій кожен нетермінальний вузол має чотири нащадки.

6.3 Горизонтально-вертикальне розбиття

У цій схемі розбиття зображення поділяється на дві частини горизонтальною або вертикальною лінією, щоб утворити два нові прямокутники. Розбиття повторюється рекурсивно, доки не буде задоволено критерій покриття, як у схемі квадродерева.

6.4 Трикутне розбиття

Інший спосіб розбиття зображення базується на трикутниках. Прямокутне зображення ділиться по діагоналі на два трикутники. Кожен з цих трикутників поділяється на 4 трикутники шляхом сегментації трикутника вздовж лінії, яка з'єднує три точки поділу вздовж сторін трикутника, що є краще та гнучкіше, порівняно зі схемою горизонтально-вертикального розбиття.

Для стисненням зображень в інформаційній системі для астрономів-аматорів, найбільш доцільним є використання квадродерева з декількох причин. Метод розбиття на квадрати фіксованого розміру простий у реалізації, але не враховує різноманітність вмісту зображення. Для

© Веремеєнко А.А.

областей з високою деталізацією він може бути неефективним, оскільки використовує однаковий розмір блоків незалежно від складності. Метод горизонтально-вертикального розбиття менш гнучкий порівняно з квадродеревом. Він може створювати неприродні блоки, що можуть не відповідати вмісту зображення, що погіршує якість стиснення.

6.5 Вибір оптимального розбиття для інформаційної системи

Хоча метод трикутного розбиття може бути ефективним для певних типів зображень (наприклад, з багатьма діагональними лініями), він складніший в реалізації та обчисленнях. Також він може призводити до створення надлишкових трикутників в однорідних областях, що знижує ефективність стиснення. Квадродерево дозволяє адаптивно розділяти зображення на блоки різних розмірів залежно від вмісту зображення. Це означає, що однорідні області можуть бути представлені великими блоками, тоді як складніші області — меншими. Це підвищує ефективність стиснення. Завдяки фрагментації зображення до необхідного рівня деталізації, Квадродерево добре зберігає важливі деталі зображення, що важливо для високої якості результату після декомпресії. Квадродерево також є досить ефективним з точки зору обчислювальної складності, оскільки дозволяє швидко ідентифікувати однорідні області та оптимізувати процес стиснення. Враховуючи вищезазначені фактори, метод квадродерево є найкращим вибором для стиснення зображень в інформаційній системі, забезпечуючи баланс між ефективністю стиснення, якістю зображення та обчислювальною складністю.

Алгоритм квадродерева на прикладі компресії зображень:

1. Ініціалізація:
- Починаємо з кореня дерева, який представляє все зображення.
2. Рекурсивний поділ:
- Для кожного блоку перевіряємо, чи є він однорідним.
- Якщо блок однорідний, залишаємо його як є.
- Якщо блок не однорідний, ділимо його на чотири квадранти і повторюємо процес для кожного квадранту.
3. Завершення:
- Процес продовжується, поки не будуть досягнуті однорідні блоки або мінімальний розмір блоків.

6.6 Приклади застосування квадродерева

Тип розбиття квадродерево найбільше підходить для наступних цілей

1. Зберігання картографічних даних - картографічні дані можуть бути збережені у вигляді квадродерев, де кожен вузол представляє певну область карти. Однорідні області (наприклад, великі водойми або пустелі) можуть бути представлені великими блоками, тоді як більш складні області (міста, дороги) розбиваються на дрібніші блоки.
2. Адаптивне зберігання зображень - аерофотознімки та супутникові зображення можна також зберігати у вигляді квадродерев, що дозволяє адаптивно розділяти зображення залежно від його вмісту. Це зменшує обсяг даних, що зберігаються, і забезпечує швидкий доступ до різних частин зображення.
3. Обробка даних - квадродерево дозволяє швидко знаходити та обробляти необхідні області зображення або карти. Наприклад, для визначення ділянок лісу або водних ресурсів можна швидко знаходити відповідні блоки в квадродереві.

7. Показники якості у фрактальній компресії

Якість стиснутих зображень стає важливим фактором разом із значним коефіцієнтом стиснення, що визначає ефективне стиснення зображень. Стиснення зображення зменшує кількість надлишкових даних у ньому. Мета показників якості полягає в тому, щоб відобразити якість декодованих зображень. Об'єктивний показник якості повинен добре відображати спотворення зображення, наприклад, через розмитість, шум або стиснення.

7.1 Середньоквадратична помилка (MSE - Mean squared error)

MSE або відношення сигналу до шуму (SNR - Signal-to-noise ratio) є показником якості, який часто використовується для вимірювання відхилення між оригінальним і закодованим зображенням.

7.2 Пікове відношення сигналу до шуму (PSNR - Peak signal-to-noise ratio)

Більші значення SNR і PSNR вказують на меншу різницю між оригінальним і відновленим зображенням. Це найпоширеніший об'єктивний показник якості/спотворення зображень. Основною

© Веремеєнко А.А.

перевагою цього показника є легкість обчислення, але він не відображає сприйняття якості. Важливою властивістю PSNR є те, що невеликий просторовий зсув зображення може спричинити значне числове спотворення без візуального спотворення і навпаки, невелике середнє спотворення може призвести до значного візуального артефакту.

7.3 Середня різниця (AD – Average difference)

Нижче значення середньої різниці дає "чистіше" зображення, оскільки зменшується кількість шуму.

7.4 Максимальна різниця (MD – maximum difference)

Велике значення MD означає, що якість зображення є низькою.

7.5 Нормалізована кореляція

Ці показники вимірюють схожість між двома зображеннями. Всі кореляційні показники наближаються до 1, оскільки різниця між двома зображеннями наближається до нуля. Так як показники різниці та кореляції доповнюють один одного, мінімізація показників відстані означає максимізацію показників кореляції.

7.6 Середня абсолютна помилка (MAE – Mean absolute error)

MAE оцінює середню різницю між відповідними пікселями оригінального та відновленого зображень, що дозволяє визначити рівень спотворення. Велике значення MAE означає, що якість зображення є низькою.

7.7 Нормалізована абсолютна помилка (NAE – Normalized absolute error)

За цим показником оцінюється, наскільки відновлене зображення відрізняється від оригінального зображення, причому значення нуль є ідеальною відповідністю.

7.8 Вибір оптимального показника якості для інформаційної системи

Для оцінювання якості зображення в інформаційній системі найдоцільніше використовувати пікове відношення сигналу до шуму з кількох причин.

Середньоквадратична помилка є основою для розрахунку, але сама по собі не так інтуїтивно зрозуміла і менш зручна для інтерпретації. Високе значення середньоквадратичної помилки вказує на значну помилку, але воно не враховує динамічний діапазон зображення.

Середня різниця є дуже простим методом, який враховує лише середню різницю між відповідними пікселями, що не відображає загальну якість зображення так точно, як інші методи.

Максимальна різниця фокусується на максимальній різниці між відповідними пікселями, що може бути корисним у виявленні суттєвих спотворень, але не дає повної картини про загальну якість зображення.

Метод нормалізованої кореляції більше підходить для оцінки подібності зображень, ніж для оцінки якості після стиснення.

Середня абсолютна помилка є більш інтуїтивно зрозумілим показником, ніж середньоквадратична помилка, але не враховує варіацій яскравості та контрастності, що може бути важливим для візуального сприйняття якості зображення.

Нормалізована абсолютна помилка нормалізує помилки відносно амплітуди зображення, але також не враховує всі аспекти людського сприйняття.

Метод пікового відношення сигналу до шуму добре враховує різні види шуму, які можуть виникнути в процесі стиснення та декомпресії зображення. Він враховує максимальні амплітуди сигналу та шуму, що робить його чутливим до суттєвих спотворень. Значення пікового відношення сигналу до шуму виражається в децибелах, що робить його зручним для інтерпретації. Високе значення пікового відношення сигналу до шуму вказує на високу якість зображення, що полегшує порівняння різних методів стиснення або налаштувань параметрів.

Хоча пікове відношення сигналу до шуму не є ідеальним з точки зору моделювання людського сприйняття, він все ж корелює з тим, як люди сприймають якість зображення. Це робить його практичним вибором у багатьох реальних додатках. Таким чином, пікове відношення сигналу до шуму є найбільш доцільним методом для оцінки якості зображення в інформаційній системі, оскільки він поєднує в собі інтуїтивну зрозумілість, зручність інтерпретації та достатню точність для оцінки різних видів спотворень.

Середньоквадратична помилка та пікове відношення сигналу до шуму є ключовими метриками для оцінки якості зображення після стиснення. Середньоквадратична помилка обчислює середнє значення квадрату різниць між відповідними пікселями двох зображень: оригінального (I) та зображення після стиснення (K). Формула для MSE виглядає наступним чином

© Веремеєнко А.А.

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j],$$

де:

I(i,j) - значення пікселя в координатах (i, j) оригінального зображення,

K(i,j) - значення пікселя в координатах (i, j) зображення після стиснення,

m та n - кількість рядків та стовпців у зображенні відповідно.

Пікове відношення сигналу до шуму виражається в децибелах і обчислюється за формулою:

$$PSNR = 10 \cdot \log_{10}\frac{MAX^2}{MSE}$$

де:

MAX - максимальне можливе значення пікселя в зображенні. Для 8-бітних зображень, це зазвичай 255.

MSE - середньоквадратична помилка, обчислена за попередньою формулою.

Наведемо приклад обчислення. Припустимо, у нас є 2 маленькі зображення розміром 2x2 пікселі:

Оригінальне зображення: $\begin{bmatrix}100 & 150\\200 & 250\end{bmatrix}$

Зображення після стиснення: $\begin{bmatrix}98 & 152\\203 & 247\end{bmatrix}$

Обчислимо MSE:

$$MSE = \frac{1}{4}((100-98)^2 + (150-152)^2 + (200-203)^2 + (250-247)^2)$$

$$MSE = \frac{1}{4}(4+4+9+9)$$

$$MSE = \frac{26}{4}=6{,}5$$

Наведемо приклад обчислення PSNR. Припустимо, що максимальне можливе значення пікселя (MAX) для наших зображень - 255. Використаємо обчислене значення MSE для обчислення PSNR:

$$PSNR = 10 \cdot \log_{10}\frac{255^2}{6{,}5}$$

$$PSNR = 10 \cdot \log_{10}\frac{65025}{6{,}5}$$

$PSNR = 10 \cdot \log_{10}(10003.846)$, PSNR≈10·4, PSNR≈40dB

Таким чином, для цього прикладу PSNR становить приблизно 40 дБ.

8. Процес фрактального стиснення зображення

Щоб застосувати фрактальне стиснення для зображення, інформаційна система реалізує такий алгоритм:

1. Зчитати файл з фрактальним кодом
2. Створити буфери пам'яті для доменного і діапазонного екранів
3. Ініціалізувати буфер діапазонного екрана довільним початковим станом
4. Повторити:

- Для кожного домену:
- Замінити цей домен відповідним перетвореним блоком діапазону
- Наступний домен
- Скопіювати вміст доменного екрана на діапазонний екран
- Доки не буде досягнуто необхідної кількості ітерацій

5. Зберегти кінцевий файл

8.1 Проблема швидкості стиснення

Час, витрачений на стиснення будь-якого зображення, є основним недоліком будь-якого алгоритму фрактального стиснення. Основною важливою частиною алгоритму є зіставлення блоку діапазону з відповідним доменним блоком таким чином, щоб відстань Гаусдорфа між оригінальним зображенням і його атрактором була мінімальною. Алгоритм вимагає багато часу на пошук, коли точний збіг не знайдено. Для великого зображення час стиснення може вимірюватися в годинах, а

© Веремеєнко А.А.

не в секундах або навіть хвилина. Однак можна стверджувати, що час декомпресії є важливішим, ніж час стиснення.

Висновки. Фрактальне стиснення зображень, забезпечує високу якість відновлених зображень і є незалежним від роздільної здатності та пристроїв. Його головною перевагою є швидка декомпресія, що робить його застосування перспективним для медичних та наукових сфер. В свою чергу, основний недолік методу – тривалий час стиснення, який обмежує практичне застосування в умовах, де важлива швидкість обробки.

Порівняння з методами JPEG, JPEG2000 та Wavelet Transform демонструє, що фрактальне стиснення перевершує інші методи при високих коефіцієнтах стиснення, але поступається JPEG при низьких коефіцієнтах через простоту та доступність останнього. Фрактальне стиснення має перспективи для систем, що вимагають високої якості зображень, і працюють з великим обсягом даних. Особлива ефективність фрактального алгоритму досягається у випадках, коли компресія зображення відбувається всього один раз, а доступ до зображення здійснюється численну кількість разів. Таким чином, фрактальне стиснення зображень є перспективною технологією, яка потребує подальших вдосконалень для зменшення часу обробки, що дозволить його ширше впровадження в різні галузі.

Список бібліографічного опису

1. Працьовитий М.В. Фрактальний підхід у дослідженнях сингулярних розподілів. Київ: НПУ імені М.П. Драгоманова, 1998. 296 с.
2. Працьовитий М.В., Свинчук О.В. Розсіювання значень однієї фрактальної неперервної немонотонної функції канторівського типу // Нелінійні коливання. 2018. Том 21, № 1. С. 116 – 130.
3. Лапчик В. В., Олексієнко С. О. Фрактальний аналіз та його застосування в інформаційних технологіях. Київ: Видавничий дім "Слово", 2007. 218 с.
4. Тарасенко Ф. П., Бойченко В. В. Математичні моделі фрактальної геометрії в інформаційних системах. Харків: ХНУРЕ, 2010. 272 с.
5. Зінченко А. В., Білозерцев Є. І. Застосування фрактальної геометрії в обробці та стисненні зображень // Вісник КНУ. Серія: Фізико-математичні науки. 2015. № 3. С. 45-54.
6. Барнслі, М. Ф. Фрактали всюди // Академічна преса, 1988. 394 с.
7. Фішер, Й. Фрактальна компресія зображень: теорія та застосування. Спрінгер, 1995. 316 с.
8. Жакен, А. Е. Кодування зображень на основі фрактальної теорії ітераційних контрактних перетворень зображень // IEEE Transactions on Image Processing, 1992. 1(1), 18-30.
9. Слоан, К. Р. Фрактальна компресія зображень: огляд // Proceedings of the IEEE, 1994. 82(4), 647-655.
10. Девіс, Г. М. Вейвлет-аналіз фрактальної компресії зображень // IEEE Transactions on Image Processing, 1994. 3(2), 213-220.
11. Гартенштейн, Г. Оцінка продуктивності алгоритмів фрактальної компресії зображень // Proceedings of the Data Compression Conference, 1996. сс. 123-132. IEEE.
12. Врскай, Е. Ітераційні функціональні системи: теорія, застосування та автоматичне створення фракталів // Journal of Mathematical Imaging and Vision, 1994. 4, 75-100.

References

1. Pratsiovytyi M.V. Fractal Approach in the Study of Singular Distributions. Kyiv: NPU named after M.P. Drahomanov, 1998. 296 p.
2. Pratsiovytyi M.V., Svynchuk O.V. Scattering of Values of a Fractal Continuous Non-monotonic Function of Cantor Type // Nonlinear Oscillations. 2018. Vol. 21, No. 1. pp. 116-130.
3. Lapchyk V. V., Oleksiienko S. O. Fractal Analysis and Its Application in Information Technologies. Kyiv: Publishing House "Slovo", 2007. 218 p.
4. Tarasenko F. P., Boichenko V. V. Mathematical Models of Fractal Geometry in Information Systems. Kharkiv: KhNURE, 2010. 272 p.
5. Zinchenko A. V., Bilozercev Y. I. Application of Fractal Geometry in Image Processing and Compression // Bulletin of KNU. Series: Physical and Mathematical Sciences. 2015. No. 3. pp. 45-54.
6. Barnsley, M. F. Fractals Everywhere // Academic Press, 1988. 394 p.
7. Fisher, Y. Fractal Image Compression: Theory and Application. Springer, 1995. 316 p.
8. Jacquin, A. E. Image Coding Based on a Fractal Theory of Iterated Contractive Image Transformations // IEEE Transactions on Image Processing, 1992. 1(1), 18-30.
9. Sloan, K. R. Fractal Image Compression: A Review // Proceedings of the IEEE, 1994. 82(4), 647-655.
10. Davis, G. M. A Wavelet-Based Analysis of Fractal Image Compression // IEEE Transactions on Image Processing, 1994. 3(2), 213-220.
11. Hartenstein, H. Performance Evaluation of Fractal Image Compression Algorithms // Proceedings of the Data Compression Conference, 1996. pp. 123-132. IEEE.
12. Vrscay, R. E. Iterated Function Systems: Theory, Applications, and the Automatic Generation of Fractals // Journal of Mathematical Imaging and Vision, 1994. 4, 75-100.

© Веремеєнко А.А.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-16
УДК 004:657
Вознюк Анастасія Вадимівна, асистент
https://orcid.org/0009-0000-0168-3235
Сачук Вікторія Олегівна, асистент
https://orcid.org/0009-0009-9532-7586
Повстяна Соломія Олександрівна, здобувач вищої освіти
Луцький національний технічний університет, м.Луцьк, Україна

ІНТЕРАКТИВНІ МЕТОДИ НАВЧАННЯ ВЕБ-РОЗРОБЦІ: ЯК ЗАЛУЧИТИ СТУДЕНТІВ ДО АКТИВНОГО НАВЧАННЯ

Вознюк А.В., Сачук В.О., Повстяна С.О. Інтерактивні методи навчання веб-розробці: як залучити студентів до активного навчання. У статті висвітлено основні інтерактивні методи навчання веб-розробці, які покращують навченість студентів. До основних із них належать метод проєктів, кейс-метод, кодування в режимі реального часу, метод спаринг-партнерства, диспути, дискусії, інтерактивні ігри, віртуальні лабораторії і симуляції. Представлені у статті інтерактивні методи допомагають студентам реалізувати свої фахові компетентності завдяки активній діяльності. У роботі висвітлюються способи, як інтерактивні методи забезпечують практичний досвід, стимулюють критичне мислення і творчість, а також сприяють розвитку навичок співпраці та проблемного вирішення. Підкреслено, що використання цих підходів робить навчання більш захоплюючим і продуктивним, відповідно до сучасних вимог до освіти в галузі веб-розробки. Метод проєктів і кейс-метод надають можливість працювати над реальними або змодельованими ситуаціями, що розвиває навички критичного мислення і вирішення проблем. Кодування в режимі реального часу дозволяє студентам бачити вплив змін у коді безпосередньо, що сприяє глибшому розумінню процесів. Метод спаринг-партнерства і групові дискусії заохочують студентів до активної комунікації і співпраці, а інтерактивні ігри і віртуальні лабораторії створюють мотивуюче навчальне середовище. Стаття демонструє, як ці методи разом допомагають студентам краще засвоювати матеріал, розвивати професійні компетенції і бути готовими до реальних викликів у сфері веб-розробки. У роботі вказано системи та засоби, які мають освоїти студенти здійснюючи веб-розробку веб-сайтів, веб-сторінок, веб-додатків. Аналіз наукової літератури з теми дав можливість відобразити сучасні тенденції розвитку IT сфери, з акцентуалізацією на веб-розробки.
Ключові слова: веб-розробка, інтерактивні методи навчання, активне навчання, метод проєктів, кейс-методи, віртуальні лабораторії.

Vozniuk A., Sachuk V., Povstiana S. Interactive methods of teaching web development: how to engage students in active learning. The article highlights the main interactive web development teaching methods that improve student learning. The main ones include the project method, case study method, real-time coding, sparring partnership method, debates, discussions, interactive games, virtual laboratories, and simulations. The interactive methods presented in the article help students to realise their professional competences through active work. The paper highlights the ways in which interactive methods provide practical experience, stimulate critical thinking and creativity, and promote the development of collaboration and problem-solving skills. It is emphasised that the use of these approaches makes learning more engaging and productive, in line with the modern requirements for web development education. The project method and the case method provide an opportunity to work on real or simulated situations, which develops critical thinking and problem-solving skills. Real-time coding allows students to see the impact of changes in the code directly, which contributes to a deeper understanding of the processes. Sparring partnerships and group discussions encourage students to actively communicate and collaborate, while interactive games and virtual labs create a motivating learning environment. The article demonstrates how these methods together help students to learn better, develop professional competences and be ready for real-world challenges in web development. The paper describes the systems and tools that students should master when developing websites, web pages, and web applications. The analysis of scientific literature on the topic made it possible to reflect the current trends in the development of the IT sphere, with an emphasis on web development.
Keywords: web development, interactive teaching methods, active learning, project-based learning, case studies, virtual laboratories.

Постановка проблеми. Спостереження за діяльністю студентів під час освітнього процесу дає можливість побачити тенденцію до необхідності збагачення арсеналу методів та прийомів навчання від активних до інтерактивних. Сучасне покоління здобувачів вищої освіти потребує нових підходів до подачі матеріалу, спрямоване на усвідомлення ними пропонованого змісту як теоретичного, так і практичного. Розглядаючи проблему залучення студентів до активного навчання визначено ефективність саме інтерактивних методів навчання веб-розробці, які сприятимуть цілковитій залученості студентів до освітнього процесу.

Веб-розробка є однією з найдинамічніших і найпопулярніших галузей в IT-індустрії. В умовах сучасного освітнього процесу традиційні лекції і конспекти вже не є достатніми для досягнення високих результатів. Тому інтерактивні методи навчання набирають популярність як ефективний засіб залучення студентів і покращення їхнього розуміння матеріалу.

Аналіз останніх досліджень і публікацій. Здійснюючи аналіз наукової літератури дійшли висновку, що висвітлена проблематика є мало дослідженою. Вивченням впливу інтерактивних

@ Вознюк А.В., Сачук В.О., Повстяна С.О.

методів навчання на діяльність студентів займалися науковці Алієва О.Г., Дьяковська Г.О., Скляр О.Г., Скляр Р.В., Тимошенко Н. Є. Вони розглядали метод діалогу, рольові ігри, групові презентації, метод проблемного навчання, метод проєктів, метод мозкового штурму, групові дискусії та ін. як ефективний інструмент подачі та опрацювання матеріалу під час якого студенти максимально включені в освітній процес [6, С.732].

У роботах вітчизняних науковців розкрито технології веб-розробок, застосування та розроблення адаптивних веб-сайтів (Хорошевська І.О.), які допомагають бізнесу реагувати на виклики сьогодення [4]. Широкого попиту зазнає використання нейромереж під час веб-розробки сайтів. Цієї думки дотримується вчена Бережна О. Б. [3, С. 67-68.]. Вченою досліджено переваги нейромереж та їхню ефективність для здійснення аналітичної та дослідницької роботи. Однією із переваг визначено вміння здійснювати швидку аналітичну роботу щодо відвідувачів сайтів, їхні дії та реакцію на пропонований контент. Такими нейромережами є RNN (Recurrent Neural Networks), LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), CNN та ін. [7].

Проблематика є сучасною та потребує поглибленого вивчення, адже IT сфера постійно удосконалюється, тому і методи роботи з майбутніми фахівцями мають бути виключно інноваційними та інтерактивними, зорієнтовані на практичне застосування.

Виділення невирішених раніше частин загальної проблеми. Методологія нинішньої вищої освіти потребує суттєвих змін, які анонсовані викликами технологічного прогресу та реаліями сьогодення. Оскільки більшість досліджень пов'язані з використанням ІКТ під час освітнього процесу у закладах вищої освіти, то питання активної залученості студентів в навчальний процес завдяки інтерактивним методам навчання є мало дослідженим.

Мета статті. Основною метою представленої на розгляд статті є висвітлення значущості та дієвості інтерактивних методів навчання веб-розробці в освітньому процесі закладів вищої освіти.

Виклад основного матеріалу дослідження. Інтерактивні методи навчання – це методи, які активно залучають студентів до процесу навчання. Замість пасивного сприйняття інформації, вони беруть участь у створенні знань, спільному вирішенні задач і практичному застосуванні отриманих знань. Це може включати в себе різноманітні методи, такі як проєкти, симуляції, обговорення, ролеві ігри та ін. Саме їх застосування допомагає краще усвідомити зміст та сутність веб-розробок [2].

На сьогоднішній день веб-розробка є тим інструментом, який допомагає розширювати уявлення бізнесу про можливості сучасних інформаційних технологій. Загалом вона стосується розробки веб-сайтів, веб-додатків та веб-сторінок, які допомагають здійснювати ефективну комунікацію з об'єктами впливу.

Інтерактивні методи навчання веб-розробки дозволяють створювати динамічні та захоплюючі користувацькі інтерфейси, що підвищують залученість і забезпечують кращий користувацький досвід. Серед сучасних трендів та інструментів, які використовуються для створення інтерактивних веб-додатків є односторінкові додатки (SPA), що стали стандартом у сучасній веб-розробці завдяки своїй здатності забезпечувати плавний і швидкий досвід користувача. SPA завантажує один HTML-документ і динамічно оновлює контент на основі взаємодії користувача, без повторних запитів до сервера для кожної зміни. React, Angular та Vue.js є популярними фреймворками для створення SPA. Вони пропонують потужні інструменти для розробки інтерактивних інтерфейсів та ефективного управління станом додатка. Інтерактивні елементи та анімації допомагають зробити веб-додатки більш привабливими та функціональними. Вони можуть використовуватися для підвищення зрозумілості інтерфейсу, а також для привернення уваги до важливих елементів. Для створення анімацій часто використовуються бібліотеки як GreenSock (GSAP) та Anime.js. CSS-анімації та переходи також є потужними інструментами для досягнення плавності анімацій.

Інтерактивні методи веб-розробки продовжують еволюціонувати, забезпечуючи нові можливості для створення динамічних і захоплюючих веб-додатків. Односторінкові додатки, адаптивний і респонсивний дизайн, інтерактивні елементи та анімації, WebAssembly, API та прогресивні веб-додатки – всі ці технології і підходи допомагають розробникам створювати інтерфейси, які відповідають сучасним вимогам і очікуванням користувачів. Використання цих методів дозволяє створювати не лише функціональні, але й інтуїтивно зрозумілі та захоплюючі веб-додатки.

Оскільки сучасне суспільство орієнтоване на соціальні мережі, на взаємодію в мережі Інтернет, то бізнесова структура стрімко переходить у це середовище. Окрім створення сайтів,

@ Вознюк А.В., Сачук В.О., Повстяна С.О.

додатків, активно застосовується використання адаптивних їх форм. Серед фундаментальних засобів розробки, які допомагають реалізувати адаптивність є HTML (Hypertext Markup Language), CSS (Cascading Style Sheets), Java Script, PHP, CSS-фреймворки (Bootstrap, Foundation, Bulma), Media Queries та Responsive Frameworks (Foundation, Bootstrapі Semantic UI) [4, С. 65].

Їх використання в освітньому процесі вищої школи надає можливість студентові стати конкурентоспроможним на ринку праці. Серед основних напрямків веб-розробки, потрібно розрізняти їх поділ на застосування веб-редактора, якими є VSCode, SublimeText та систем, які допомагають вправлятися з контентом для наповнення веб-сайтів, ними є Joomla, WordPress та ін.

Інформування студентів щодо зазначних веб-розробок є можливим завдяки наступним інтерактивним методам: метод проєктів, метод спаринг-партнерства, диспути, дискусії, інтерактивні ігри, кейс-методи, віртуальні лабораторії і симуляції тощо. Одним із ефективних методів є метод проєктів, під час реалізації якого студенти мають можливість проявити не лише вміння програмування, а й критичне та креативне мислення. Завдання проєктів мають бути орієнтовані на створення веб-сайтів, веб-додатків відповідно до пропонованої викладачем тематики. Під час створення продукту студенти вчаться налагоджувати активну комунікацію один з одним, генерують цікаві ідеї, обирають дієві форми презентації та визначають оратора, який буде розкривати сутність і значимість проєкту.

Проєктні роботи є однією з найефективніших інтерактивних методик у навчанні веб-розробці. Студенти можуть працювати над реальними проєктами, створювати веб-сайти, додатки чи інші інтерактивні продукти. Це дозволяє їм застосовувати теоретичні знання на практиці, вдосконалювати свої навички програмування і дизайну, а також отримувати зворотний зв'язок від викладачів і однокурсників.

Цікавим та дієвим є використання методу спаринг-партнерства, що стимулює студентів до самовідповідальності та саморефлексії. Оскільки сучасний веб-розробник має вміти конкурувати та постійно прагнути до підвищення свого професійного рівня, то даний метод є ефективним. Під час його реалізації студенти вчаться демонструвати свої навички та переваги, накопичують досвід один одного та проводять змістовний аналіз власної діяльності. Основна роль викладача полягає у підборі партнерів для спарингу, адже вони мають мати орієнтовно однакові знання та навички для того, щоб мати можливість ними ділитися та покращувати.

Наступним методом, який можна використовувати в освітньому процесі є кодування в режимі реального часу, також відоме як «live coding», дозволяє студентам бачити, як код впливає на результати безпосередньо під час написання [6]. Це може бути організовано через спільні сесії кодування, де викладач або досвідчений розробник працює разом зі студентами. Такі сесії можуть бути проведені як онлайн, так і офлайн, що дозволяє знижувати бар'єри для участі і підвищувати рівень взаємодії.

Традиційні методи навчання, такі як лекції та індивідуальні завдання, залишаються важливими складовими освітнього процесу. Проте з розвитком технологій та новими педагогічними підходами все більше навчальних закладів та викладачів переходять до інтерактивних ігрових та комунікативних методів. Використання інтерактивних вправ та ігор може значно підвищити мотивацію студентів. Ігрові елементи, такі як квести, вікторини або кодинг-ігри, дозволяють студентам змагатися між собою або виконувати завдання в ігровій формі. Платформи, такі як Codewars, HackerRank або FreeCodeCamp, пропонують різноманітні задачі та ігри для покращення навичок програмування [1].

Групові дискусії є потужним інструментом для стимулювання активної участі студентів у навчальному процесі. Вони не лише сприяють розвитку критичного мислення, але й допомагають формувати навички спілкування та командної роботи. Групові дискусії дозволяють студентам висловлювати свої думки, аргументувати їх та розглядати різні точки зору. Це стимулює розвиток критичного мислення і вміння аналізувати інформацію. Спільна робота над завданнями та обговореннями сприяє розвитку навичок командної роботи. Студенти навчаються делегувати завдання, знаходити спільні рішення та вирішувати конфлікти.

Групові дискусії та інтерактивні ігри стали популярними інструментами, які активно заохочують студентів до спільної роботи, розвитку комунікативних навичок та створення мотивуючого навчального середовища.

Зважаючи на специфіку програмування є необхідність демонструвати студентам ефективні та цікаві кейси відомих сучасних вітчизняних та зарубіжних веб-розробників. Так, студенти зможуть надихнутися твореннями професіоналів і розробити власну веб-розробку, що стане

продуктивним та високоякісним кейсом. Кейс-метод, або метод аналізу конкретних випадків, є одним з найефективніших підходів до навчання, який активно використовують у різних освітніх установах. Це метод, при якому навчальний процес організовується навколо детального вивчення реальних або змодельованих ситуацій. Кейс-метод передбачає аналіз і обговорення конкретних випадків або сценаріїв, які мають на меті відобразити реальні проблеми та ситуації. Це може бути реальний бізнес-кейс, ситуація, яка вимагає критичного мислення і прийняття рішень. Цей метод є потужним інструментом в освітній сфері, що забезпечує практичне навчання і розвиток критичного мислення. Він дозволяє студентам працювати над реальними проблемами, розвивати навички комунікації та прийняття рішень.

Цікавими методами роботи зі студентами є віртуальні лабораторії і симуляції. Вони дозволяють студентам експериментувати з кодом в безпечному середовищі, можуть тестувати різні техніки і методи без ризику пошкодження реального проєкту. До прикладу, студенти можуть використовувати онлайн-редактори, такі як CodePen або JSFiddle, щоб миттєво бачити результати змін у коді.

Окрім вказаних інтерактивних методів навчання, вагомий вплив на залученість студентів до професійної діяльності мають інтерактивні онлайн-платформи та інструменти що забезпечують студентам можливість вивчати веб-розробку у зручний час і в інтерактивному форматі. Платформи Codecademy, freeCodeCamp і Udacity пропонують інтерактивні курси з веб-розробки, які включають завдання, проєкти і зворотний зв'язок. Онлайн-платформи дозволяють студентам навчатись у своєму темпі, отримувати миттєвий зворотний зв'язок і практикувати навички на реальних завданнях.

Сучасним методом організації пошукової діяльності є хакатони, які сприяють мотивації студентів до професійної діяльності, допомагають розвивати пізнавальний інтерес, вчать швидко приймати рішення, сприяють творчому розвиткові та навчають практикуватися у створенні веб-додатків під час командної взаємодії. Змагання створюють азарт і стимулюють студентів до поглибленого вивчення матеріалу, а також дозволяють їм працювати в команді та взаємодіяти з іншими учасниками.

Варіативність інтерактивних методів навчання дозволяє урізноманітнювати навчальну діяльність студентів та стимулювати їх до реалізації власного потенціалу.

Висновки та перспективи подальших досліджень. Веб-розробка – це динамічна і швидко розвиваюча галузь, що вимагає постійного оновлення знань та навичок. Традиційні методи навчання часто не відповідають сучасним вимогам студентів, тому інтерактивні підходи стають дедалі популярнішими. Вони дозволяють не тільки краще засвоїти матеріал, але й стимулювати зацікавленість і мотивацію здобувачів вищої освіти.

Завдяки проєктному навчанню, кодуванню в режимі реального часу, інтерактивним вправам, віртуальним лабораторіям, спаринг-партнерству можна створити ефективне і захоплююче навчальне середовище. Залучення студентів до активного навчання не тільки покращує їх навички, але й готує їх до реальних викликів у світі веб-розробки.

Сфера веб-розробки постійно еволюціонує, відзначаючи нові технології та тенденції, які визначають майбутнє вебу. Перспективи подальших розробок у веб-розробці вказують на постійне вдосконалення технологій та нові можливості для створення інтерактивних, безпечних та інноваційних веб-додатків. Штучний інтелект, прогресивні веб-додатки, AR/VR, веб-3.0, голосові інтерфейси та інші тенденції формують майбутнє веб-розробки. Розробники, що слідкують за цими тенденціями, зможуть впроваджувати найновіші рішення і підтримувати конкурентоспроможність у динамічному світі технологій.

Список бібліографічного опису.

1. Алієва О.Г., Дьяковська Г.О. Методика викладання філософії з залученням інтерактивних методів. *Духовність особистості: методологія, теорія і практика*. 2023. № 1(105). С. 14–21.
2. Антюшко Д.П., Володавчик В.С., Сєногонова Л.І. Інтерактивні методи навчання у вищій школі: монографія. Харків: Видавництво Іванченка І. С., 2022. 189 с.
3. Бережна О. Б. Тенденції використання нейромереж у розробці веб-сайтів. *Поліграфічні, мультимедійні та web-технології* : *тези доп. IX Міжнар. наук.-техн. конф., 14-18 травня 2024 р.* Т. 1. Харків: ТОВ «Друкарня Мадрид», 2024. С. 67-68.
4. Кисельов Т., Хорошевська І. Огляд засобів розроблення адаптивних веб-сайтів. *Матеріали конференцій МЦНД*, (29.03.2024; Херсон, Україна), 2024. С. 64–66. URL : https://archive.mcnd.org.ua/index.php/conference-proceeding/article/view/1116

@ Вознюк А.В., Сачук В.О., Повстяна С.О.

5. Скляр О.Г., Скляр Р.В. Використання в освітньому процесі інтерактивних методів навчання. *Удосконалення освітньо-виховного процесу в закладі вищої освіти: збірник наук.-метод. праць*. 2023. Вип. 26. С. 250–25.
6. Тимошенко Н.Є. Використання інтерактивних методів навчання у викладацькій діяльності. *Перспективи та інновації науки*. № 4(38). 2024. С. 730 – 736.
7. Medium. Deep Learning Based Recommender Systems. 2021.URL: https://medium.com/sciforce/deeplearning-based-recommender-systems-b61a5ddd5456.

References

1. Aliieva, O.H., Diakovska, H.O. (2023). Metodyka vykladannia filosofii z zaluchenniam interaktyvnykh metodiv *[Methods of teaching philosophy with the use of interactive methods]*. Dukhovnist osobystosti: metodolohiia, teoriia i praktyka.1(105), 14-21 [in Ukrainian]
2. Antiushko D.P., Volodavchyk V.S., Sienohonova L.I. (2022). Interaktyvni metody navchannia u vyshchii shkoli: monohrafiia. *[Interactive teaching methods in higher education : a monograph]*. Kharkiv: Vydavnytstvo Ivanchenka I. S., 189 s. [in Ukrainian].
3. Berezhna O. B. (2024). Tendentsii vykorystannia neiromerezh u rozrobtsi veb-saitiv. *[Trends in the use of neural networks in website development]*. Polihrafichni, multymediini ta web-tekhnolohii : tezy dop. IX Mizhnar. nauk.-tekhn. konf., 14-18 travnia 2024 r. T. 1. Kharkiv: TOV «Drukarnia Madryd». S. 67-68. [in Ukrainian].
4. Kyselov T., Khoroshevska I. (2024). Ohliad zasobiv rozroblennia adaptyvnykh veb-saitiv. *[An overview of responsive website development tools.]*. Materialy konferentsii MTsND, (29.03.2024; Kherson, Ukraina). S. 64–66. URL: https://archive.mcnd.org.ua/index.php/conference-proceeding/article/view/1116 [in Ukrainian].
5. Skliar O.H., Skliar R.V. (2023). Vykorystannia v osvitnomu protsesi interaktyvnykh metodiv navchannia. *[The use of interactive teaching methods in the educational process]*. Udoskonalennia osvitno-vykhovnoho protsesu v zakladi vyshchoi osvity: zbirnyk nauk.-metod. prats. Vyp. 26. S. 250–25. [in Ukrainian].
6. Tymoshenko N.Ie. (2024). Vykorystannia interaktyvnykh metodiv navchannia u vykladatskii diialnosti. *[Using interactive teaching methods in teaching activities]*.Perspektyvy ta innovatsii nauky. № 4(38).S. 730 – 736. [in Ukrainian].
7. Medium. (2021). Deep Learning Based Recommender Systems.URL: https://medium.com/sciforce/deeplearning-based-recommender-systems-b61a5ddd5456.[in English].

@ Вознюк А.В., Сачук В.О., Повстяна С.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-17
УДК 37:004.9(045)
Інна Олександрівна Гулівата, к. пед. н., доцент
http://orcid.org/0000-0003-4752-535X
Вінницький торговельно-економічний інститут Державного торговельно-економічного університету, м. Вінниця, Україна

КОНЦЕПТ ЦИФРОВОЇ ОСВІТНЬОЇ ДОСТУПНОСТІ: ЄВРОПЕЙСЬКИЙ ДОСВІД

Гулівата І.О. Концепт цифрової освітньої доступності: європейський досвід. Здійснено огляд сучасних інтерактивних та персоналізованих цифрових рішень для підвищення доступності освіти. Серед освітніх ресурсів, що сприяють розбудові безбар'єрного середовища для всіх громадян, включаючи людей з інвалідністю, визначено наступні: онлайн-курси та платформи; віртуальні класи; освітні мобільні застосунки; масові відкриті онлайн-курси з різних дисциплін; інтерактивні навчальні ресурси для навчання програмуванню; електронні підручники та цифрові ресурси з різних предметів; доповнена та віртуальна реальність; навчальні платформи з інтеграцією штучного інтелекту; навчальні симулятори та ігри. Проаналізовано проблеми та виклики пов'язані з ними: цифровий розрив (доступ до технологій, рівень цифрової грамотності); якість контенту (надійність джерел в Internet, адаптація цифрового контенту до навчальних програм); педагогічні виклики (багато вчителів потребують додаткового навчання для ефективного використання цифрових інструментів та розробка нових методик навчання, які враховують можливості та обмеження цифрових технологій); мотивація та залученість учнів (самодисципліна, відсутність соціальної взаємодії); безпека в мережі та приватність; захист даних (забезпечення конфіденційності особистих даних учасників освітнього процесу, Internet-безпека); технічні проблеми; економічні та організаційні виклики (фінансові витрати та організаційні зміни пов'язані з перебудовою освітніх процесів та структур для інтеграції цифрових технологій). Розглянуто європейські практики та висвітлено європейський підхід щодо цифрової освітньої доступності, який ґрунтується на вимірах: технології, учасники освітнього процесу, заклад освіти, національний/регіональний рівень.
Ключові слова: концепт освітньої доступності, цифрова безбар'єрність, цифрова освіта, європейський досвід

Hulivata I. The concept of digital accessibility in education: European experience. This paper carries out an overview of modern interactive and personalized digital solutions in order to increase the accessibility of education. Among the educational resources that contribute to the development of a barrier-free environment for all citizens, including people with disabilities, are the following: online courses and platforms; virtual classrooms; educational mobile applications; massive open online courses in various disciplines; interactive learning resources for learning to program; e-textbooks and digital resources in various subjects; augmented and virtual reality; learning platforms with artificial intelligence integration; training simulations and games. It was analysed the problems and challenges associated with them: digital divide (access to technology, level of digital literacy); content quality (reliability of sources on the Internet, adaptation of digital content to curricula); pedagogical challenges (many teachers need additional training to use digital tools effectively and to develop new teaching methods that take into account the possibilities and limitations of digital technologies); student motivation and engagement (self-discipline, lack of social interaction); online security and privacy; data protection (ensuring the confidentiality of personal; network security and privacy; data protection (ensuring the confidentiality of personal data of participants in the educational process, Internet security); technical problems; economic and organizational challenges (financial costs and organizational changes to restructure educational processes and structures to integrate digital technologies). The main focus presented in this paper is describing the European practices in the light of the European approach to digital accessibility in education, which is based on such dimensions: technology, participants in the educational process, educational institution, national/regional level.
Keywords: the concept of accessibility of education, digital accessibility, digital education, European experience

Постановка наукової проблеми. Безбар'єрне середовище є важливим аспектом сучасності, оскільки дозволяє людям з різними фізичними, психічними та іншими обмеженнями вільно взаємодіяти з оточуючим світом. Розбудова безбар'єрного середовища є питанням соціальної справедливості для забезпечення рівних можливостей і прав для всіх громадян, що особливо актуально для українського суспільства. За даними Міністерства соціальної політики в Україні нараховується 3 мільйони людей з інвалідністю, а за останні півтора року їх кількість збільшилася приблизно на 300 тисяч [1]. Для їх інтеграції в суспільне життя необхідно створити безбар'єрне середовище, що включає доступ до будівель, транспорту, інформаційних та освітніх ресурсів.

Разом з тим, Україна прагне стати частиною Європейського Союзу, де питання безбар'єрності є важливою складовою соціальної політики, а впровадження європейських стандартів у цій сфері є необхідним для наближення до членства в ЄС. Європейські країни мають розвинену систему стандартів та практик у галузі безбар'єрності, яка включає законодавство, нормативні документи, технічні рішення та культурні підходи. Вивчення та імплементація європейських стандартів може допомогти Україні покращити свої підходи розбудови безбар'єрного середовища.

Аналіз останніх досліджень і публікацій. Потенціал інформаційно-комунікаційних технологій (ІКТ) в освіті і виклики, пов'язані з ним, у центрі уваги Європейської агенції з питань

@ Гулівата І.О.

спеціальних потреб та інклюзивної освіти [2, 8, 10] та ЮНЕСКО [7]. Європейські практики щодо розбудови цифрової доступності в освіті реалізовані у проєктах [3, 4, 5, 6]. Політичні ініціативи Європейського Союзу визначені планом дій у сфері цифрової освіти (2021-2027) [2]. У дослідженні [10] проаналізовано нові пріоритети та вимоги щодо інклюзивної цифрової освіти та змішаного навчання. Аналіз цифрових інструментів у реалізації безбар'єрного освітнього середовища висвітлені у працях Близнюк Т.О., Кривонос О.М., Котенко О.Д., Ніколіної І.І. [12] та інших.

Метою статті є дослідження Європейського досвіду у реалізації цифрової доступності в освіті.

Виклад основного матеріалу дослідження. Європа активно використовує сучасні технології та інновації для забезпечення безбар'єрного доступу для всіх груп користувачів. Одним із засобів виступає цифровізація суспільства, яка має значний вплив на розбудову безбар'єрного середовища, сприяючи більшій інклюзивності та доступності для всіх громадян, включаючи людей з інвалідністю. Особливо вагомим для нинішнього і майбутнього поколінь є впровадження нових інформаційних технологій в освіту.

Цифрова освіта охоплює широкий спектр навчальних та освітніх технологій, які використовують цифрові інструменти та ресурси, серед яких наступні:

1. Онлайн-курси та платформи, що пропонують курси від провідних університетів та організацій: Coursera (https://www.coursera.org/courses?query=free), edX (https://www.edx.org/courses?q=free+online+courses); Khan Academy (https://uk.khanacademy.org/).

2. Віртуальні класи, як інструменти для проведення онлайн-уроків та зустрічей: Zoom; Microsoft Teams; Google Classroom.

3. Освітні мобільні застосунки: Photomath (https://photomath.com/) – застосунок для вирішення математичних задач за допомогою камери смартфона; Kahoot! (https://kahoot.com/) – платформа для створення ігор і вікторин для навчання; Duolingo (https://www.duolingo.com/learn) – застосунок для вивчення мов.

4. Масові відкриті онлайн-курси (MOOCs) з різних дисциплін: Udacity (https://www.udacity.com/); FutureLearn (https://www.futurelearn.com/).

5. Інтерактивні навчальні ресурси для навчання програмуванню дітей: Scratch (https://scratch.mit.edu/); Code.org (https://code.org/); mate.academy (https://mate.academy).

6. Електронні підручники та ресурси з різних предметів: Amazon Kindle, Google Books (https://books.google.com.ua/).

7. Доповнена та віртуальна реальність (AR/VR): Google Expeditions (https://sites.google.com/view/virtuarealities/vr-resources/google-expeditions) – застосунок для віртуальних екскурсій; Nearpod (https://nearpod.com/) – інтерактивні уроки з використанням VR.

8. Навчальні платформи з інтеграцією штучного інтелекту: Socratic by Google (https://socratic.org/) – застосунок, що використовує штучний інтелект у навчанні.

9. Навчальні симулятори та ігри: Minecraft: Education Edition, SimCityEDU.

Такі цифрові рішення є ефективними засобами, які можуть покращити доступ до освіти, зробити навчання інтерактивним та персоналізованим, а також сприяти розвитку навичок, необхідних у сучасному світі.

Проте, поряд з новими технологічними можливостями для навчання та розвитку освіта стикається з численними проблемами та викликами.

1. Цифровий розрив:

- Доступ до технологій: не всі учні та студенти мають доступ до необхідних пристроїв (комп'ютерів, планшетів, смартфонів) та швидкого інтернету.
- Рівень цифрової грамотності: вчителі та учні можуть мати різний рівень знань і навичок використання цифрових технологій.

2. Якість контенту:

- Надійність джерел: в Інтернеті є багато неякісного або неперевіреного навчального матеріалу.
- Адаптація до навчальних програм: цифровий контент не завжди відповідає вимогам національних навчальних програм.

3. Педагогічні виклики:

- Підготовка вчителів: багато вчителів потребують додаткового навчання для ефективного використання цифрових інструментів.

- Методологічна адаптація: потрібно розробляти нові методики навчання, які враховують можливості та обмеження цифрових технологій.

4. Мотивація та залученість учнів:

- Самодисципліна: онлайн-навчання вимагає від здобувачів освіти більшої самодисципліни та мотивації.
- Відсутність соціальної взаємодії: брак особистого спілкування може негативно вплинути на соціальні навички учнів.

5. Проблеми з безпекою та приватністю:

- Захист даних: Забезпечення конфіденційності особистих даних учасників освітнього процесу.
- Інтернет-безпека: захист від кіберзагроз, таких як зломи, віруси, та інтернет-булінг.

6. Технічні проблеми:

- Технічні збої: перебої в роботі платформ, проблеми з підключенням до інтернету та інші технічні несправності можуть впливати на процес навчання.
- Обслуговування та підтримка: потрібні ресурси для постійного оновлення та підтримки технічної інфраструктури.

7. Економічні та організаційні виклики:

- Фінансові витрати: впровадження цифрових технологій в освіту вимагає значних фінансових інвестицій.
- Організаційні зміни: необхідність перебудови освітніх процесів та структур для інтеграції цифрових технологій.

Розв'язання цих викликів потребує комплексного підходу, включаючи співпрацю між урядами, освітніми інституціями, технологічними компаніями та громадськими організаціями.

Європейські країни мають значний досвід у розробці та впровадженні політик і практик, що сприяють інклюзивності та доступності в освіті. Вивчення і впровадження цього досвіду може суттєво допомогти Україні у створенні більш доступного та інклюзивного цифрового освітнього середовища.

Європейський досвід говорить про те, що потенціал інформаційно-комунікаційних технологій (ІКТ) в освіті був у центрі уваги Європейської агенції з питань спеціальних потреб та інклюзивної освіти (Агенція) [2] ще з 1999 року. На початку своєї діяльності Агенція досліджувала як ІКТ можуть сприяти інклюзії учнів з особливими потребами, тоді як подальші дослідження розширили фокус на те, як ІКТ можуть сприяти інклюзивній освіті для всіх учасників освітнього процесу.

ІКТ для інклюзії було головним питанням у роботі Агенції. Зокрема, проекти «ІКТ для інклюзії» (ICT4I) [3] та «ІКТ для доступу до інформації в навчанні» (ICT4IAL) [4], а також співпраця з ЮНЕСКО [5] та G3ict [6] спрямовані на реалізацію політики інклюзивних ІКТ в освіті для осіб з обмеженими можливостями.

Прискорення технологічного прогресу та зміни у ставленні й політиці спрямовані на забезпечення більш справедливих освітніх можливостей для всіх учасників освітнього процесу призвели до появи нових пріоритетів і вимог у сфері змішаного навчання. Це лягло в основу політичної ініціативи Європейського Союзу «План дій у сфері цифрової освіти» (Digital Education Action Plan (2021-2027) [7], який передбачає довгострокове стратегічне бачення високоякісної, інклюзивної та доступної європейської цифрової освіти. Його перший пріоритетний напрям зосереджений на сприянні розвитку високоефективної цифрової освітньої екосистеми.

Глобальна пандемія COVID-19 і нинішня війна в Україні висвітлюють слабкі місця та виклики для систем освіти з точки зору їхньої здатності забезпечувати безперервне навчання для людей у цілому, і особливо для учнів та студентів. Системи освіти адаптувалися до безпрецедентних обставин під час пандемії. Звіт Агенції «Вплив COVID-19 на інклюзивну освіту на європейському рівні» [8] вказує, що доступ до навчання стає проблематичним, особливо для осіб з особливими потребами.

У ньому [8] також підкреслюється, що не можна ігнорувати потенційну роль ІКТ у подоланні цього виклику. Проте, розглядаючи роль ІКТ у забезпеченні доступу до навчання та встановлення комунікації вчитель-учень, відзначаються численні складні проблеми, пов'язані з використанням ІКТ та можливостями, які вони пропонують для побудови більш стійких освітніх систем, що забезпечують усім учасникам освітнього процесу безперервні та справедливі навчальні можливості.

@ Гулівата І.О.

У всесвітній доповіді UNESCO [9, с.2] з моніторингу освіти у 2023 році, яка присвячена технологіям та освіті, зазначається, що якісна освіта залежить від можливостей і викликів, які створюють технології.

Освітні системи в усьому світі адаптувалися до безпрецедентних обставин під час пандемії COVID-19. Незважаючи на швидкі та комплексні кроки в напрямку цифровізації освіти, доступ до навчання, особливо для вразливих осіб, залишається проблемою. У дослідженні [10] ретельно проаналізовано нові пріоритети та вимоги щодо інклюзивної цифрової освіти та змішаного навчання.

Постійний процес цифрової трансформації у суспільному житті, де цифрові медіа використовуються все ширше, а також зростання використання цифрових технологій у навчанні та викладанні, відкриває нові можливості для подолання соціального виключення. Водночас з'являються нові бар'єри на шляху до участі в житті суспільства та в освіті. Вони можуть посилити існуючу вразливість окремих осіб або груп, а також створити нові форми вразливості до соціального та освітнього виключення.

Актуальність впровадження сучасних технологій зумовлена їх використанням різними суб'єктами на різних рівнях системи освіти та залежить від умов, що сприяють або перешкоджають їх використанню. Дослідження Агенції щодо визначення цих рівнів керується моделлю екосистеми інклюзивної освіти.

На індивідуальному рівні у дослідженні виділяють учнів та вчителів, оскільки вони можуть використовувати різні або однакові цифрові технології в контексті свого навчання та викладання як самостійно, так і спільно. Це означає, що кожна з цих груп має свої унікальні потреби та способи використання технологій, проте інколи їхні цілі та інструменти можуть перетинатися.

Роль освітнього закладу розглядається як ключовий компонент інклюзивної цифрової освіти, який співвідносить індивідуальний рівень з політичними та громадськими аспектами. Нарешті, виокремлено питання національного/регіонального рівня, оскільки він торкається законодавчих аспектів інклюзивної освіти.

У дослідженні [10] визначено концепт цифрової освітньої доступності, який ґрунтується на вимірах: технології, учні, вчителі, рівень закладу освіти та національний/регіональний рівень.

На технологічному рівні підкреслена необхідність розробки інклюзивних технологій зі збалансованими технологічними та педагогічними підходами, де першість надається педагогіці, визначаючи пріоритетність над усіма іншими аспектами. Особливість технологічного підходу побудована на застосуванні концепції універсального дизайну, орієнтованої на користувача. «Універсальний дизайн - це проектування продуктів, середовищ, програм і послуг, які максимально придатні для використання всіма, без потреби в адаптації або спеціалізованому дизайні» [11, с. 420].

Допоміжні технології рекомендовано використовувати як компенсаторний засіб лише там, де універсально розроблені технології недостатньо задовольняють потреби всіх користувачів. Більш цілеспрямоване використання ІКТ для створення можливостей для взаємного навчання на різних рівнях має великий невикористаний потенціал. Відкриті освітні ресурси однозначно спрямовані на поліпшення доступу до навчальних матеріалів. До таких належать масові відкриті онлайн-курси, застосунки зі штучним інтелектом доступні для підтримки навчання дітей з аутизмом, з проблемами у навчанні або з вадами зору чи слуху. У деяких галузях технологій з'являються розробки з потенційним майбутнім впливом на освіту, наприклад, домашня робототехніка, системи мобільної телеприсутності, чат-боти і «розумні» колонки.

Інклюзія в цифровій освіті - це складне явище, на яке впливають різноманітні фактори, такі як суспільство, технічні засоби, освітні заклади, навчальна ситуація та окремі учні. Вразливість до соціального виключення в цифровій освіті може бути викликана явищами в навчанні, що тісно пов'язані з соціальними механізмами, де на людину, групу людей, організацію чи соціальну проблему впливає низка тисків, сил, важелів, дискримінації та несприятливих умов.

Для окремих осіб інклюзія в цифровій освіті відображається в технічній доступності, присутності та видимості, активному соціальному залученні, взаємодії та співпраці один з одним, а також відчутті того, що їх цінують і залучають до навчальної спільноти.

Існує широкий спектр контекстів, в яких інклюзія є надзвичайно актуальною темою для навчання в цифровій освіті. Цифрові компетентності учасників освітнього процесу відіграють важливу роль, особливо у спілкуванні, співпраці та безпеці, шанобливій та вдячній соціальній взаємодії, розвитку та розширенні можливостей себе як цифрової особистості, вираженні власного

голосу, критичному осмисленні цифрових медіа та самозахисті від насильства в цифровому середовищі.

При розробці умов, сприятливих для інклюзії в цифровому та аналоговому освітньому середовищі, необхідно поєднувати інформацію, отриману від різних осіб або груп, вразливих до соціального виключення, щоб розробити заходи для цілісної перспективи інклюзії для забезпечення якісної освіти для всіх учнів. Крім того, дистанційне навчання може призвести до посилення нерівності та відсіву учнів.

На рівні вчителів також існує ряд проблем та викликів. Вчителі потребують підтримки у виборі інклюзивних навчальних матеріалів, які не створюють або створюють незначні бар'єри та підходять для всіх учнів.

Досвід показав, що під час пандемії використання цифрових медіа не здатне компенсувати брак контакту та необхідної навчальної підтримки з боку вчителів, особливо для соціально та економічно незахищених учнів. А дистанційна освіта може бути ефективною короткостроковою заміною, у разі неминучих обмежень особистого контакту. Однак, у контексті інклюзивного цифрового навчання важливими є такі компетенції, як медіаграмотність, грамотність у роботі з даними та прийняття рішень на основі даних.

Ключовим компонентом у розбудові інклюзивного простору є заклад освіти, оскільки освітні організації, які охоплюють процес цифровізації з точки зору контенту та фінансування, можуть допомогти зменшити соціальну ізоляцію.

Розширення можливостей вчителів відіграє основну роль і має супроводжуватися заходами організаційної підтримки, підвищенням кваліфікації та врахуванням індивідуальних потреб вчителів.

Через недостатній зв'язок між інклюзивною освітою та цифровізацією освітні організації повинні здійснювати цифрову трансформацію самостійно та на власний ризик. Це може пояснити нерішучість у її впровадженні.

Успішна інклюзивна цифрова освіта передбачає створення сприятливих умов на національному та регіональному рівнях. Готовність до інклюзії з точки зору контенту та технологій свідчить про стійкість організації у випадку кризи. Пандемія COVID-19 поглибила освітню нерівність, але вона також створила унікальну можливість реформувати освіту.

Комп'ютерні та інформаційні технології набувають значення не тільки в контексті предмета навчання, а й як досить дієвий результативний засіб корекційного впливу у побудові безбар'єрного освітнього середовища [12, с. 38].

Досі освітній сектор недостатньо залучений до розробки та розвитку технологій, а також до обговорення етичних наслідків використання цифрових медіа та технологій для проактивного реагування на вимоги інклюзивної освіти.

Висновки та перспективи подальших досліджень. Цифровізація освіти у реалізації політики безбарєрності потребує набагато більше, ніж застосування належним чином розроблених цифрових технологій в освіті. Вона вимагає залучення всіх рівнів – від окремої людини, закладу освіти до регіонального чи національного рівнів, де інклюзія та цифровізація є наскрізними питаннями. Ця участь має вирішальне значення, якщо інклюзивна цифрова освіта не просто запроваджуватиметься від випадку до випадку, а буде назавжди закріплена в структурах системи освіти. Однак існує дуже мало прикладів зі сфери освіти, які показують, що передбачає успішний процес трансформації та які конкретні кроки повинні зробити окремі особи, організації та політики.

Багатий європейський досвід з безбар'єрності, співпраця з європейськими партнерами, обмін досвідом та навчання на основі кращих практик може стати цінним джерелом знань для України, що сприятиме прискоренню процесу розбудови безбар'єрного середовища.

Список бібліографічного опису

1.Міністерство соціальної політики України : вебсайт. URL: https://www.msp.gov.ua/timeline/invalidnist.html.
2.European Agency for Special Needs and Inclusive Education: вебсайт. URL: European Agency for Special Needs and Inclusive Education website (european-agency.org).
3.ICT for Inclusion: вебсайт. URL: https://www.european-agency.org/activities/ICT4I.
4.ICT for Information Accessibility in Learning: вебсайт. URL: https://www.european-agency.org/activities/ict-information-accessibility-learning.
5.UNESCO: вебсайт. URL: https://www.unesco.org/en.
6. G3ict The Global Initiative for Inclusive ICTs: вебсайт. URL: https://g3ict.org/.
7. European Education Area: вебсайт. URL: https://education.ec.europa.eu/focus-topics/digital-education/action-plan.

@ Гулівата І.О.

8. European Agency for Special Needs and Inclusive Education, 2021. *The Impact of COVID-19 on Inclusive Education at the European Level: Literature Review*. (C. Popescu, ed.). Odense, Denmark. URL: https://www.european-agency.org/sites/default/files/COVID-19-Impact-Literature-Review.pdf.
9. UNESCO Digital Library: вебсайт. URL: https://unesdoc.unesco.org/ark:/48223/pf0000378950/.
10. European Agency for Special Needs and Inclusive Education, 2022. *Inclusive Digital Education.* (H. Weber, A. Elsner, D. Wolf, M. Rohs and M. Turner-Cmuchal, eds.). Odense, Denmark. URL: https://www.european-agency.org/sites/default/files/Inclusive_Digital_Education.pdf
11. UNESCO Digital Library: вебсайт. URL: https://unesdoc.unesco.org/ark:/48223/pf0000373718.
12. Гулівата, І., Ніколіна, І. Цифрові інструменти у реалізації безбар'єрного освітнього середовища. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво.* 2023, № 51. С. 37-42. URL: https://doi.org/10.36910/6775-2524-0560-2023-51-05.

References

1. Sait «Ministerstvo sotsialnoi polityky Ukrainy» [Ministry of Social Policy of Ukraine website]. *msp.gov.ua.* Retrieved from https://www.msp.gov.ua/timeline/invalidnist.html [in Ukraine].
2. Sait «European Agency for Special Needs and Inclusive Education» [European Agency for Special Needs and Inclusive Education website]. *european-agency.org.* Retrieved from https://www.european-agency.org/ [in English].
3. Sait «ICT for Inclusion» [ICT for Inclusion website]. *european-agency.org.* Retrieved from https://www.european-agency.org/activities/ICT4I [in English].
4. Sait «ICT for Information Accessibility in Learning» [ICT for Information Accessibility in Learning website]. *european-agency.org.* Retrieved from https://www.european-agency.org/activities/ict-information-accessibility-learning [in English].
5. Sait «UNESCO» [UNESCO website]. *unesco.org.* Retrieved from https://www.unesco.org/en [in English].
6. Sait «G3ict The Global Initiative for Inclusive ICTs» [G3ict The Global Initiative for Inclusive ICTs website]. *g3ict.org* Retrieved from https://g3ict.org/ [in English].
7. Sait «European Education Area» [European Education Area website]. *education.ec.europa.eu*. Retrieved from https://education.ec.europa.eu/focus-topics/digital-education/action-plan [in English].
8. Popescu, C. ed. (2021). *The Impact of COVID-19 on Inclusive Education at the European Level: Literature Review*. European Agency for Special Needs and Inclusive Education. Odense, Denmark. https://www.european-agency.org/sites/default/files/COVID-19-Impact-Literature-Review.pdf.
9. Sait «UNESCO Digital Library» [UNESCO Digital Library]. *unesdoc.unesco.org*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000378950/ [in English].
10. Weber, H., Elsner, A., Wolf, D., Rohs, M. & Turner-Cmuchal, M. (2022). *Inclusive Digital Education.* European Agency for Special Needs and Inclusive Education. Odense, Denmark. https://www.european-agency.org/sites/default/files/Inclusive_Digital_Education.pdf
11. Sait «UNESCO Digital Library» [UNESCO Digital Library]. *unesdoc.unesco.org*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000373718 [in English].
12. Hulivata, I. & Nikolina, I. (2023). Tsyfrovi instrumenty u realizatsii bezbariernoho osvitnoho seredovyshcha [Digital tools in the implementation of a barrier-free educational environment]. *Kompiuterno-intehrovani tekhnolohii: osvita, nauka, vyrobnytstvo -Computer-integrated technologies: education, science, production*, (51), 37-42. https://doi.org/10.36910/6775-2524-0560-2023-51-05 [in Ukraine].

@ Гулівата І.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-18
УДК 004.05:512.64:514.12(045)
Гусак Людмила Петрівна, канд. пед. наук, доцент
https://orcid.org/ 0000-0002-0022-9644
Радзіховська Лариса Миколаївна, канд. пед. наук, доцент
https://orcid.org/ 0000-0003-0185-8036
Вінницький торговельно-економічний інститут Державного торговельно-економічного університету, м. Вінниця, Україна

ВИКОРИСТАННЯ КОМП'ЮТЕРНИХ ПРОГРАМ ПІД ЧАС ВИКЛАДАННЯ КУРСУ «ЛІНІЙНА АЛГЕБРА ТА АНАЛІТИЧНА ГЕОМЕТРІЯ»

Гусак Л.П., Раздзіховська Л.М. Використання комп'ютерних програм під час викладання курсу «Лінійна алгебра та аналітична геометрія». В статті розглядаються питання навчання здобувачів першого курсу освітнього ступеня «бакалавр» під час вивчення дисципліни «Лінійна алгебра та аналітична геометрія» із застосуванням комп'ютерних програм. Аналізуються можливості застосування як спеціалізованих математичних пакетів, так і найбільших офісних програм. Перераховано переваги та наведено приклади використання офісної програми MS Excel та програмних продуктів Mathcad, Mathematica, MatLAB, GeoGebra у процесі вивчення окремих тем з лінійної алгебри та аналітичної геометрії. Встановлено, що впровадження розглянутих комп'ютерних програм у навчальний процес дало позитивну динаміку результатів вивчення студентами курсу «Лінійна алгебра та аналітична геометрія». Студенти, звичні до використання комп'ютерних технологій, перебувають у комфортному для себе психологічному середовищі, що покращує результат навчальної діяльності. Завдяки комп'ютерним програмам освітнього призначення здійснюється візуалізація абстрактних математичних понять, що сприяє якнайшвидшому сприйняттю матеріалу, більш глибокому його розумінню, прискорює процес розв'язання задач, спрощує обчислення і підвищує інтерес до дисципліни. Крім того, використання у навчальній діяльності різних комп'ютерних пакетів дозволяє індивідуалізувати навчальну діяльність студентів, з першого курсу відчути досвід наукової роботи під час розв'язування задач з лінійної алгебри та аналітичної геометрії різними способами. Безліч можливостей досягнення мети формує цілісніше бачення постановки навчальної проблеми, а також формує можливість широкого спектру самостійної діяльності студента у науковій сфері.
Ключові слова: комп'ютерні математичні пакети, інформаційні технології, MS Excel, MathCad, MatLab, GeoGebra

Husak L., Radzihovska L. The usage of computer programs during the teaching course «Linear Algebra and Analytical Geometry». The article showes the issue of teaching first-year bachelor's degree holders while studying the discipline «Linear Algebra and Analytical Geometry» with the usage of computer programs. The advantages are listed and examples of using the MS Excel office program and software products Mathcad, Mathematica, MatLAB, GeoGebra in the process of studying individual topics in linear algebra and analytical geometry are given. It was established that the introduction of the considered computer programs into the educational process gave positive dynamics to the results of the students' study of the course «Linear Algebra and Analytical Geometry». Students who are accustomed to using computer technologies are in a psychological environment that is comfortable for them, which improves the results of educational activities. Thanks to educational computer programs, visualization of abstract mathematical concepts is carried out, which contributes to the fastest possible perception of the material, deeper understanding of it, accelerates the process of solving problems, simplifies calculations and increases interest in the discipline. In addition, the use of various computer packages in educational activities allows to individualize the educational activities of students, from the first year to feel the experience of scientific work while solving linear algebra and analytical geometry problems in different ways. Many opportunities to achieve the goal form a more holistic vision of the educational problem, and also form the possibility of a wide range of independent activity of the student in the scientific field.
Key words: computer mathematical packages, information technologies, MS Excel, MathCad, MatLab, GeoGebra

Постановка наукової проблеми. В останні роки в процесі викладання різних дисциплін у закладах вищої освіти все частіше використовуються програмні продукти, як професійного, так і освітнього призначення. Особливо актуальним стало використання різних пакетів програм у ході вивчення математичних дисциплін.

Існує досить велика кількість програмних продуктів, здатних активізувати освітній процес. Наприклад, такі програми, як: Maple, Mathcad, Mathematica, MatLAB, GeoGebra, GeoNext, C.a.R. та інші. Застосування програмних засобів є особливо актуальним на заняттях з математичних дисциплін. Перераховані вище пакети програм сприяють виконанню розрахунків за короткий час, дозволяють виконувати алгебраїчне обчислення, геометричні креслення математичних об'єктів, проводити маніпуляції з готовим кресленням, змінюючи значення параметрів, мають можливості анімації і т. д.

Аналіз досліджень. Аналіз науково-методичної літератури щодо удосконалення математичної освіти дозволяє стверджувати, що питанню використання у навчальному процесі різного програмного забезпечення присвячено значну кількість науково-методичних праць. Вибір

© Гусак Л.П., Раздзіховська Л.М.

того чи іншого програмного засобу має відповідати завданням математичної дисципліни, в якій передбачається його використання, оскільки кожен засіб має сильні та слабкі сторони.

Зокрема, автори Дащенко А.Ф., Жалдак М.І., Зюков М.Є., Лазарєв Ю.Ф., Кравченко І.В., Микитенко В.І. [5], Левчук О.В. [6], Хоцкіна В.Б., Вдовиченко І.Н. [9], зазначали, що системи Mathematica, MatLAB Mathcad дозволяють здійснити широкий спектр символьних перетворень, які включають операції математичного аналізу, лінійної та векторної алгебри, теорії ймовірностей та математичної статистики.

Використання інтерактивних геометричних систем, зокрема GeoGebra, у навчальному процесі є одним із актуальних напрямів дослідницької діяльності як вітчизняних так і закордонних вчених та викладачів: Гуліватої І.О.[3], Дубовика В.В. [4], Samura A. O., Darhim, Juandi D. [8], Yuliardi R. [10] та інші. У своїх дослідженнях вони відмічали, що GeoGebra має такі якісно нові дидактичні можливості як наочність, моделювання, динаміка, що дозволяють змінювати традиційні підходи до вивчення багатьох розділів математики, розвиваючи пізнавальний інтерес студентів та навички дослідницької діяльності.

Мета статті. Розглянути можливості використання комп'ютерних програм освітнього призначення у процесі вивчення лінійної алгебри та аналітичної геометрії.

Виклад основного матеріалу. Сучасна індустрія пропонує широкий вибір програм, які можна використовувати для математичних обчислень. З усього різноманіття програмних засобів необхідно виділити для використання на комп'ютерних практикумах з математичних навчальних дисциплін наступні групи:

− математичні пакети (MatLab, MathCAD, Maple та ін.);

− статистичні пакети (Статистика 5.0, STADIA та ін.).

Вищезазначені програми є зручним інструментарієм для вирішення різних прикладних завдань. Проте працювати з матрицями можна у спеціалізованих математичних пакетах типу Matematica, MatLab, MathCad, а й у звичних всім офісних електронних таблицях.

Вивчення освітнього компоненту «Лінійна алгебра та аналітична геометрія» студентами спеціальності «Інформаційні системи та технології» у Вінницькому торговельно-економічного інституту ДТЕУ традиційно припадає на перший семестр. У цей час колишні абітурієнти ще тільки знайомляться з навчальним закладом, правилами та особливостями розкладу. Тому цей час не є зручним для вивчення нових великих математичних пакетів.

Під час вивчення тем «Матриці, лінійні операції над ними», «Визначники та їх властивості», «Системи лінійних алгебраїчних рівнянь» зручно користуватися стандартними електронними таблицями. Наведемо приклади розрахунку обернених матриць, добутку матриць, систем лінійних алгебраїчних рівнянь в офісній програмі Excel (рис. 1). Електронні таблиці набули широкого поширення, вивчаються в більшості шкіл, часто встановлюються і на планшети, і на сучасні телефони, тобто студенти мають доступ до них. Студенти не відчувають труднощів під час роботи в електронних таблицях, знаходяться в комфортному психологічному середовищі і можуть повністю зосередитися на новому матеріалі з вищої математики.

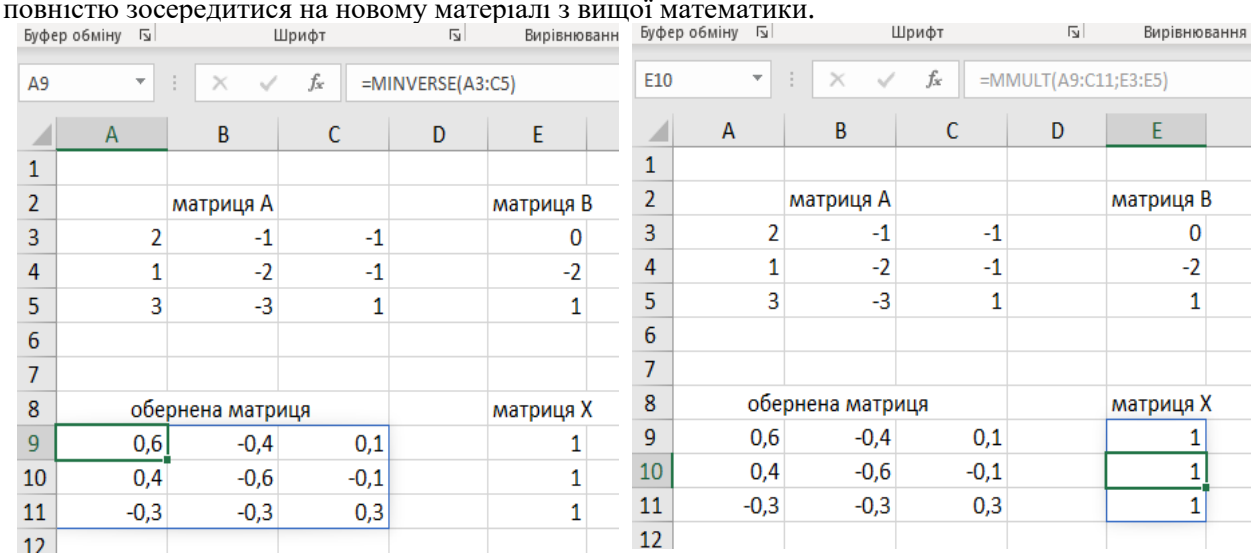


Рис. 1. Приклад розв'язання в Excel

© Гусак Л.П., Раздзіховська Л.М.

Для виконання розрахунків у цій програмі студентам необхідно знати дві функції: обчислення оберненої матриці та множення матриць. В офісній програмі Excel для цього використовується вбудований редактор формул, що містить різні категорії функцій. При виборі конкретної категорії з'являється список функцій та опис дій, що виконуються. Для обчислення оберненої матриці в Excel є функція MINVERSE (). Для знаходження відповіді необхідно провести обчислення добутку оберненої матриці на стовпець вільних членів за допомогою вбудованої функції MMULT(). Оскільки поряд з обраною функцією з'являється опис її дії, студентам необхідно хоча б у загальних рисах уявляти собі перебіг розв'язання задачі. При роботі в електронних таблицях у студентів у будь-який момент часу є можливість змінити обрану функцію на іншу, якщо це не суперечить вихідним даним.

Середовище спеціалізованого пакету MathCad умовно розуміє оточення написаних формул. Для обчислення в даному математичному пакеті достатньо задати матриці вихідних даних та вказати формулу для розрахунку. Пакет не вимагає спеціальної підготовки для виконання простих обчислень і може бути використаний при вивченні лінійної алгебри на першому курсі. Важливою особливістю електронних таблиць та пакету MathCad є миттєвий перерахунок всіх формул та результатів обчислень за зміни вихідних даних. Так, при вивченні лінійної алгебри зручно наочно демонструвати студентам, як від зміни на невелику величину однієї цифри у вихідних даних змінюється значення визначників, розв'язання системи лінійних рівнянь, як змінює вигляд і положення графіку функції. Маючи можливість багаторазово змінювати вихідні дані, студенти починають краще розуміти взаємозв'язок різних факторів.

Розглянемо алгоритм знаходження розв'язків системи лінійних рівнянь за допомогою оберненої матриці, використовуючи MathCad. Дано систему лінійних рівнянь:

$$
\begin{cases}
x_1 + 2x_2 + 3x_3 = -4 \\
2x_1 - x_2 - 3x_3 = 4 \\
3x_1 - x_2 + 2x_3 = 2
\end{cases}
$$

1. Визначаємо матрицю A і вектор-стовпець B.

```mathcad
A :=
\begin{pmatrix}
   1 & 2 & 3 \\
   2 & -1 & -3 \\
   3 & -1 & 2
   \end{pmatrix}
```

```mathcad
B := \begin{pmatrix}
   -4 \\
   4 \\
   2
   \end{pmatrix}
```

2. Знайдемо обернену матрицю A^{-1}.

```mathcad
A^{-1} := inverse(A)
```

3. Знаходимо розв'язок X.

```mathcad
X := A^{-1} B
```

В результаті отримаємо:

```mathcad
X = \begin{pmatrix}
   -2 \\
   1 \\
   0
   \end{pmatrix}
```

Отже, розв'язок СЛР $x_1 = -2$, $x_2 = 1$, $x_3 = 0$.

У сучасній інженерній діяльності широко використовуються потужні комп'ютерні пакети, спеціально призначені для математичних обчислень, такі як Mathematica і MatLab. Приклад розв'язання системи лінійних рівнянь у математичному пакеті Mathematica:

```
Solve[{x^2-1=0, x^2-3x+2=0},x].
```

Відповіддю буде запис $x \to 1$.

Приклад розв'язання системи лінійних алгебраїчних рівнянь алгебри в математичному пакеті MatLab.

© Гусак Л.П., Раздзіховська Л.М.

```
A = [4 1 2
     3 7 1
     2 2 8];
f = [7; 11; 12];

x = A\f
x =
     1
     1
     1
```

$$\begin{cases} 4x_1 + x_2 + 2x_3 = 7 \\ 3x_1 + 7x_2 + x_3 = 11 \\ 2x_1 + 2x_2 + 8x_3 = 12 \end{cases}$$

Використання інтерактивного геометричного середовища GeoGebra актуальне щодо таких розділів курсу «Лінійна алгебра та аналітична геометрія», у яких крім обчислень необхідно виконувати і геометричні побудови математичних об'єктів, що вивчаються.

Наприклад, у процесі вивчення теми «Лінії другого порядку, їх канонічні рівняння та зображення» виконуються геометричні побудови наступних кривих: еліпса, гіперболи, параболи. А також розглядаються різні форми цих кривих та їхнє розташування на координатній площині залежно від заданого рівняння. Використовуючи динамічність образів, що створюються з використанням програми GeoGebra, можна показати студентам зміну форми еліпса в залежності від відношення його великої та малої півосей (a та b). На рисунку 2 зображено еліпс для випадку, коли його велика вісь лежить на осі Ox, тобто $a>b$. Використовуючи динамічність створеного образу, досягнемо рівності великої і малої півосей еліпса, тобто $a=b$. Графік еліпса, як показано на рисунку 3, перетворюється на коло.

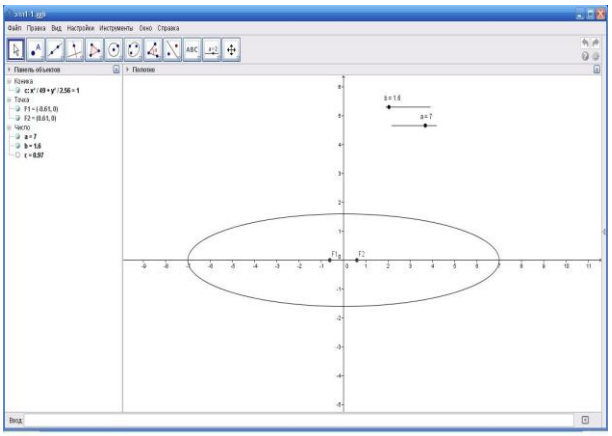


Рис. 2. Побудова еліпса в GeoGebra

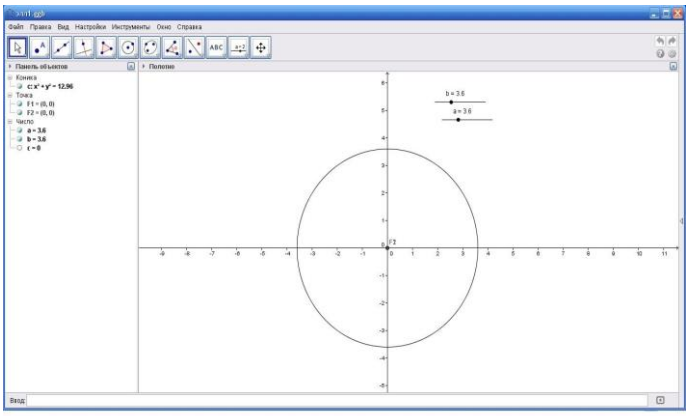


Рис. 3. Побудова кола в GeoGebra

Інтерактивне геометричне середовище GeoGebra також можна використовувати в процесі вивчення гіперболи та параболи.

Знайти площу трикутника, утвореного асимптотами гіперболи $9x^2-4y^2=36$ і прямої $9x+2y-12=0$.

Використовуючи програму GeoGebra, побудуємо гіперболу та її асимптоти, потім задану пряму. Таким чином, отримуємо трикутник ABC, після чого знаходимо його площу. Побудова трикутника, а також усі обчислення, одержані у процесі розв'язання задачі, показані на рисунку 4.

© Гусак Л.П., Раздзіховська Л.М.

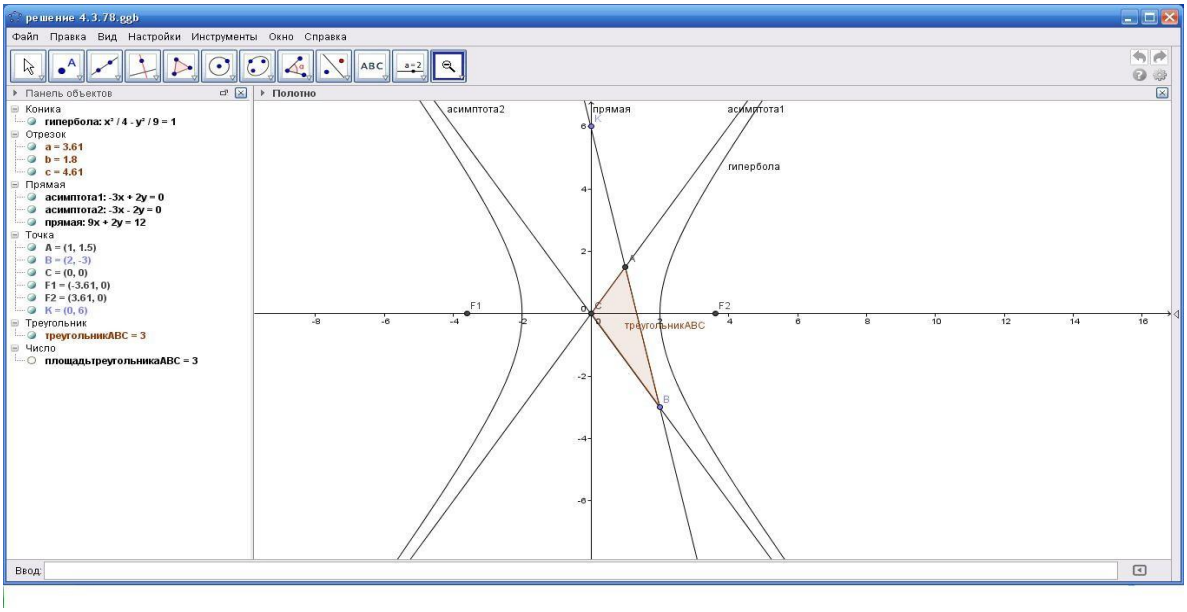


Рис.4. Приклад розв'язання в GeoGebra

Застосування інтерактивного геометричного середовища GeoGebra в ході розв'язування завдань, а також при вивченні лекційного матеріалу дозволяє виконати наочне зображення всіх математичних об'єктів, що сприяє кращому розумінню нового матеріалу, прискорює процес розв'язання задач, спрощує обчислення.

Висновки та перспективи подальшого дослідження. З погляду педагогічних технології вкрай важливо дозволити студентам випробувати різні методи вирішення тих самих навчальних завдань. Відчувши вплив мінімальної зміни вихідних даних на отриманий результат, студенти починають уважніше ставитися до чисельних значень у завданнях, поважати деталі, звертати більше уваги на дрібниці, навчаються аналізувати зміни і під час вирішення математичних завдань, й у житті, й у вчинках. Уважне вивчення математики сприяє формуванню відповідальнішого ставлення до життя, виховує любов до продуманого, послідовного виконання різних дій.

Використання у навчальній діяльності різних комп'ютерних пакетів дозволяє індивідуалізувати навчальну діяльність студентів, з першого курсу відчути досвід наукової роботи під час розв'язування задач з лінійної алгебри та аналітичної геометрії різними способами. Безліч можливостей досягнення мети формує цілісніше бачення постановки навчальної проблеми, а також формує можливість широкого спектру самостійної діяльності студента у науковій сфері. Використання комп'ютерних програм заощаджує час, наприклад, дозволяє використовувати матриці більшої розмірності та вирішувати виробничо-орієнтовані задачі. Рекомендуємо зіставляти можливість використання спеціалізованих комп'ютерних пакетів під час навчання студентів першого курсу з навичками роботи студентів у програмному середовищі.

Впровадження розглянутих комп'ютерних програм у навчальний процес дало позитивну динаміку результатів вивчення студентами курсу «Лінійна алгебра та аналітична геометрія». Студенти, звичні до використання комп'ютерних технологій, перебувають у комфортному для себе психологічному середовищі, що покращує результат навчальної діяльності. Можливості даних математичних програм можна використовувати і при вивченні дисциплін «Математичний аналіз», «Вища та прикладна математика», «Теорія ймовірностей та математична статистика». Вищезазначені комп'ютерні програми освітнього призначення дозволяють здійснити візуалізацію абстрактних математичних понять, що сприяє якнайшвидшому сприйняттю матеріалу, більш глибокому його розумінню і підвищує інтерес до дисципліни.

Список бібліографічних посилань

1. Балик Н.Р. (2017). Підходи та особливості сучасної STEM-освіти. *Фізико-математична освіта,* 2(12), 26-30.
2. ГуліватаІ.О. (2019). Інноваційні методики забезпечення змісту навчання математики. Журнал інформаційних технологій в освіті (ITE), 41, 64-74.
3. Гулівата І.О., Гусак Л.П., Радзіховська Л.М. (2018). Вища та прикладна математика: теорія ймовірностей: навч. посібник. 208 с.

© Гусак Л.П., Раздзіховська Л.М.

4. Дубовик В.В. (2021). Використання середовища GeoGebra на практичних заняттях з лінійної алгебри під час навчання студентів педагогічних університетів. *Scientific Collection «InterConf».* Вип. 52. Brighton, Great Britain: A.C.M. Webb Publishing Co Ltd., С. 167–174.
5. Кравченко І. В., Микитенко В. І. (2018). Інформаційні технології: Системи комп'ютерної математики: навч. посібник. 243 с.
6. Левчук О.В. (2019). Математичне моделювання на базі Mathcad як засіб формування професійної компетентності майбутніх економістів. Економіка, фінанси, менеджмент: актуальні питання науки і практики. № 5. С. 73-83.
7. Радзіховська Л. М., Гусак Л.П. (2019). Використання табличного процесора MS EXCEL у викладанні дисциплін професійного спрямування в економічних ЗВО. *Сучасні інформаційні технології та інноваційні методики навчання в підготовці фахівців: теорія, досвід, проблеми*. Вип. 54. с. 136-140.
8. Samura A. O., Darhim, Juandi D. (2021). Improving the Creative Thinking Ability of Junior High School Students Through GeoGebra Assisted Learning Community in Mathematics. International Journal of Interactive Mobile Technologies 15(22), 84–98. URL: https://online-journals.org/index.php/i-jim/article/ view/24797/10255 (дата звернення: 12.09.2024).
9. Хоцкіна В.Б., Вдовиченко І.Н. (2023). Робота в пакеті MATLAB: навч. посібник. 130 с.
10. Yuliardi R. (2017). Mathematics Learning Assisted Geogebra using Technologically Aligned Classroom (TAC) to Improve Communication Skills of Vocasional High School Student. Journal of Physics: Conference Series, 895(1), 012156. URL: https://iopscience.iop.org/article/10.1088/1742-6596/895/1/012156/pdf (дата звернення: 11.09.2024).

References

1. Balyk N.R. (2017). Pidkhody ta osoblyvosti suchasnoyi STEM-osvity. Fizyko-matematychna osvita, 2(12), 26-30.
2. HulivataI.O. (2019). Innovatsiyni metodyky zabezpechennya zmistu navchannya matematyky. Zhurnal informatsiynykh tekhnolohiy v osviti (ITE), 41, 64-74.
3. Hulivata I.O., Husak L.P., Radzikhovs′ka L.M. Vyshcha ta prykladna matematyka: teoriya ymovirnostey: navch. posibnyk (2018), 208 s.
4. Dubovyk V.V. (2021). Vykorystannya seredovyshcha GeoGebra na praktychnykh zanyattyakh z liniynoyi alhebry pid chas navchannya studentiv pedahohichnykh universytetiv. Scientific Collection «InterConf». Vyp. 52. Brighton, Great Britain: A.C.M. Webb Publishing Co Ltd., S. 167–174.
5. Kravchenko I. V., Mykytenko V. I. (2018) Informatsiyni tekhnolohiyi: Systemy komp″yuternoyi matematyky: navch. posibnyk. 243s.
6. Levchuk O.V. (2019). Matematychne modelyuvannya na bazi Mathcad yak zasib formuvannya profesiynoyi kompetentnosti maybutnikh ekonomistiv. Ekonomika, finansy, menedzhment: aktual′ni pytannya nauky i praktyky.. № 5. S. 73-83.
7. Radzikhovs′ka L. M., Husak L.P. (2019). Vykorystannya tablychnoho protsesora MS EXCEL u vykladanni dystsyplin profesiynoho spryamuvannya v ekonomichnykh ZVO. Suchasni informatsiyni tekhnolohiyi ta innovatsiyni metodyky navchannya v pidhotovtsi fakhivtsiv: teoriya, dosvid, problemy. Vyp. 54. s. 136-140.
8. Samura, A.O. & Darhim, Juandi, D. (2021). Improving the Creative Thinking Ability of Junior High School Students Through GeoGebra Assisted Learning Community in Mathematics. International Journal of Interactive Mobile Technologies 15(22), 84-98. Retrieved from: https://online-journals.org/index.php/i-jim/article/ view/24797/10255 (accessed 12.09.2024).
9. Khotskina V.B., Vdovychenko I.N. (2023). Robota v paketi MATLAB: navch. posibnyk. 130 s.
10. Yuliardi R. (2017). Mathematics Learning Assisted Geogebra using Technologically Aligned Classroom (TAC) to Improve Communication Skills of Vocasional High School Student. Journal of Physics: Conference Series, 895(1), 012156. URL: https://iopscience.iop.org/article/10.1088/1742-6596/895/1/012156/pdf (accessed 11.09.2024).

© Гусак Л.П., Раздзіховська Л.М.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-19
УДК 004.9
Гусєв Олександр Юрійович, к.ф.-м.н., професор
https://orcid.org/0000-0002-0548-728X
Рибальченко Юрій Петрович, асистент
https://orcid.org/0000-0003-1363-9885
Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ КОНТРОЛЮ ТЕПЛОВОГО СТАНУ НИЖНЬОЇ ЧАСТИНИ ДОМЕННОЇ ПЕЧІ

Гусєв О.Ю., Рибальченко Ю.П. Інформаційна технологія підвищення ефективності контролю теплового стану нижньої частини доменної печі. У роботі проведено аналіз особливостей доменного процесу, зокрема мінливості вхідних параметрів, періодичності завантаження матеріалів та циклів накопичення і випуску рідких продуктів плавки. Це впливає на тепловий стан доменної печі, що ускладнює підтримку стабільного ходу процесу. Для вирішення цієї проблеми запропоновано інформаційну технологію контролю теплового стану доменної печі, яка дозволяє здійснювати моніторинг температури футерування металоприймача в режимі реального часу. Такий підхід забезпечує швидке реагування на відхилення та підтримку стабільного процесу плавки. Особливу увагу приділено циклу "накопичення-випуск" рідких продуктів плавки, що дозволяє контролювати рівень наповнення металоприймача і своєчасно організовувати випуск продукції. На основі цієї технології розроблено програмне середовище, яке автоматично надає оператору кількісні дані про тепловий стан горна та прогнозує вміст кремнію в чавуні. Використання цієї інформаційної технології підвищує ефективність управління процесом плавки, знижує витрати енергоресурсів, покращує техніко-економічні показники виробництва та забезпечує стабільну якість продукції.
Ключові слова: інформаційна технологія, контроль, доменний процес, тепловий стан, монолітна архітектура.

Gusiev A., Rybalchenko Y. Information technology to improve the efficiency of thermal control of the bottom part of the blast furnace. The paper analyzes the characteristics of the blast furnace process, particularly the variability of input parameters, the periodicity of material charging, and the accumulation and discharge cycles of molten products. These factors affect the thermal state of the blast furnace, complicating the maintenance of a stable process. To address this issue, an information technology for controlling the thermal state of the blast furnace is proposed, allowing real-time monitoring of the lining temperature of the hearth. This approach ensures quick responses to deviations and helps maintain a stable smelting process. Special attention is given to the "accumulation-discharge" cycle of molten products, enabling control of the hearth's filling level and timely discharge of products. Based on this technology, a software environment has been developed to automatically provide the operator with quantitative data on the thermal state of the hearth and forecast the silicon content in the pig iron. The use of this information technology enhances the efficiency of managing the smelting process, reduces energy consumption, improves the technical and economic indicators of production, and ensures the stable quality of the final product.
Keywords: information technology, control, blast furnace process, thermal state, monolithic architecture.

Постановка проблеми. Доменний процес, незважаючи на його інерційність і стійкість, унаслідок значного часу перебування шихтових матеріалів у печі, відносять до квазістаціонарного. Нестаціонарність процесу спричинена мінливістю значень вхідних параметрів процесу доменної плавки, періодичністю завантаження шихтових матеріалів, накопиченням і випуском рідких продуктів плавки і, як наслідок, відповідною зміною умов відновлювальної роботи горнових газів у шахті та теплової роботи в горні. Усе це проявляється і в циклічному характері зміни ступеня використання газів-відновлювачів, а також фізичного і хімічного нагріву чавуну в міру накопичення та спорожнення горна в процесі плавки. У зв'язку з мінливістю значень вхідних параметрів процесу доменної плавки, періодичністю завантаження шихтових матеріалів, накопиченням і випуском рідких продуктів плавки виникають проблеми з управлінням процесом і забезпеченням якості продукції. Для розв'язання цієї проблеми пропонується використання інформаційної технології контролю теплового стану доменної печі, що дає змогу проводити моніторинг у режимі реального часу та швидко реагувати на будь-які проблеми, які можуть вплинути на процес плавки.

Мета роботи - підвищення ефективності моніторингу теплового стану нижньої частини доменної печі за допомогою інформаційної технології контролю у режимі реального часу, що дозволить швидко реагувати на будь-які проблеми, які можуть вплинути на процес плавки.

Виклад основного матеріалу. У процесі доменної плавки відбувається циклічна зміна теплового стану горна, пов'язана з накопиченням і випуском рідких продуктів плавки в горні. Накопичення рідких продуктів плавки супроводжується накопиченням кількості тепла, що вноситься чавуном і шлаком у горн, а випуск рідких продуктів - виносом тепла з горна [1].

© Гусєв О.Ю., Рибальченко Ю.П.

Тепловий стан горна безпосередньо пов'язаний з температурою його футерування. Таким чином, контролюючи температуру футерування металоприймача, можна здійснювати контроль теплової роботи горна доменної печі.

Якісні техніко-економічні показники доменної плавки - висока продуктивність, низька питома витрата коксу і потрібний хімічний склад чавуну, можна отримати тільки за "нормальної роботи" печі. Поняття про нормальну роботу пов'язується з так званим "рівним ходом" печі.

Таким чином, оптимальним режимом доменної плавки, що дає змогу мінімізувати енергоресурси і забезпечити отримання чавуну на випуску із заданим хімічним складом, є "рівний хід" доменної печі [2], який характеризується:

1. Плавним і стійким сходом шихтових матеріалів з постійною швидкістю опускання шматків шихти в поперечному перерізі печі. Тут важливо зазначити, що швидкість опускання матеріалів в окремих точках перерізу печі може бути різною, але за рівного перебігу плавки ця швидкість має бути постійною.
2. Сталістю кількості продуктів плавки, що випускаються з печі на кожному випуску.
3. Стійким характером розподілу шихтових матеріалів і газового потоку в перерізі доменної печі, за якого досягається максимальний ступінь використання хімічної і теплової енергії газу з мінімальною для даних умов питомою витратою коксу.
4. Постійними параметрами дуттьового режиму, тобто за умови стабільності витрати дуття, його температури і вологості, а також сталості витрат природного газу (пиловугільного палива або мазуту) і технологічного кисню.
5. Стійким тепловим станом плавки, коли має місце отримання чавуну і шлаку заданого складу з відповідними фізико-хімічними властивостями.
6. Наявністю стійкого гарнісажу, що забезпечує сталість робочого профілю доменної печі та його збереження в процесі експлуатації.

Рівний хід плавки досягається суворим дотриманням усіх технологічних інструкцій і положень, починаючи від приймання та складування шихтових матеріалів, завантаження їх у піч, виконання технології плавки і закінчуючи випуском продуктів плавки.

Контроль перебігу доменної плавки і коригування режиму здійснюється за результатами візуального спостереження, за допомогою контрольно-вимірювальних приладів (КВП) і сучасних експертних систем, що реалізують математичні моделі розвитку фізико-хімічних явищ. Метою контролю є виявлення можливих відхилень основних параметрів процесів від заданих.

При класифікації того чи іншого виду відхилення ходу плавки від нормального режиму аналізується весь комплекс вимірюваних параметрів у сукупності з результатами візуального спостереження [2].

Виходячи з викладеного, можна зробити висновок, що для ефективного контролю теплового режиму доменної печі необхідно виробити якийсь інтегральний показник, який відображав би максимально повно всю інформацію про рівний хід.

Як такий показник пропонується використовувати інформацію про процес "накопичення-випуск" рідких продуктів доменної плавки, що безпосередньо відображають рівень рідких продуктів плавки в металоприймачі, тобто ступінь його наповнення і спорожнення.

Для реалізації запропонованого методу використовувався універсальний пристрій контролю теплових процесів у горні доменної печі (УКТП), (рис. 1) [3].

Рис.1. Пристрій контролю теплового режиму горна і циклів "накопичення-випуск"

© Гусєв О.Ю., Рибальченко Ю.П.

Схему встановлення УКТП у футеровку металоприймача доменної печі наведено на рис. 2.

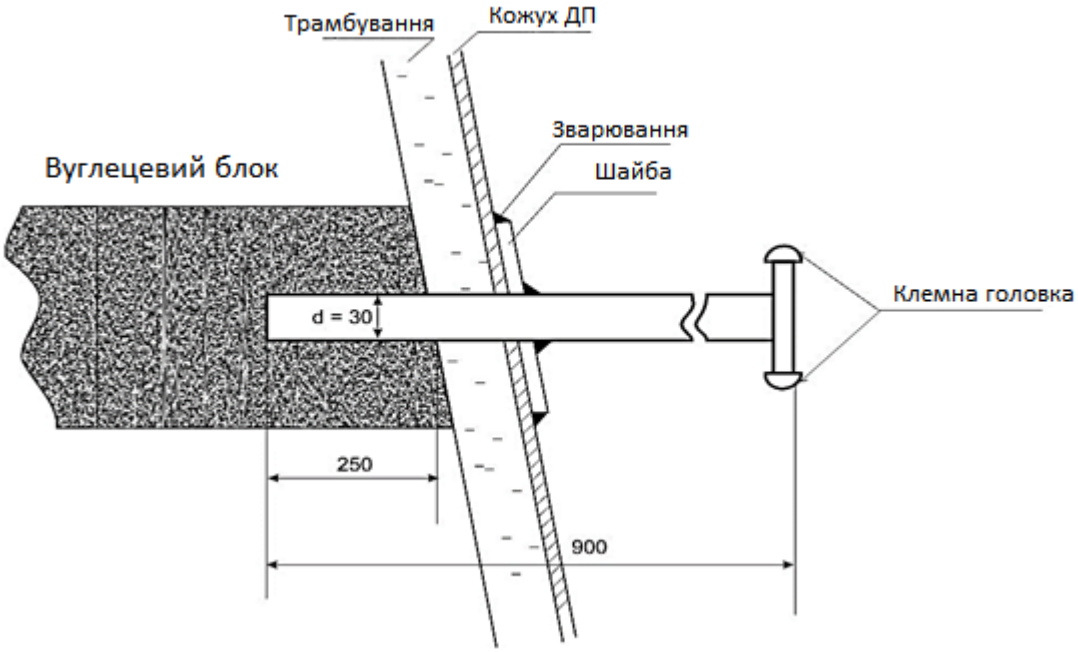


Рис. 2. Встановлення УКТП у футеровку металоприймача

УКТП дає змогу контролювати температуру футеровки, яка, як уже згадувалося, безпосередньо пов'язана з рівнем рідких продуктів плавки.

На рис. 3 наведено графік, що відображає поведінку процесу "накопичення-випуск".

З рисунка видно, що показання УКТП мають вигляд полігармонійних коливань. Висхідні лінії відображають процес "накопичення" металоприймачем рідких продуктів плавки (прихід тепла), низхідні - "випуск" (витрата тепла) продуктів плавки з доменної печі.

На верхніх перегинах спостерігаються порівняно "рівні" майданчики тривалістю за часом 5-10 хв., що свідчить про наповнення металоприймача рідкими продуктами плавки та "готовність" доменної печі до випуску. При цьому вважається, що рівень розплаву досяг 100%.

У разі затримок випусків тривалість за часом "рівного майданчика" досягає 30-50 хв. і більше.

Рис. 3. Графік поведінки процесу "накопичення-випуск"

Таким чином, інформація про цикли "накопичення - випуск" дає змогу візуально визначати ступінь наповнення металоприймача рідкими продуктами плавки й організувати своєчасний випуск продуктів плавки.

© Гусєв О.Ю., Рибальченко Ю.П.

Візуальна оцінка трендів записів показань УДТП-2 дає змогу спостерігати тенденцію тренду теплового режиму горна - "розігрів", "похолодання", "норма" (рис. 4, 5, 6 відповідно).

Рис. 4. Реєстрація УДТП-2 режиму "розігрів" і гістограма прогнозу вмісту кремнію в чавуні (зелений колір)

Рис. 5. Реєстрація УДТП-2 режиму "похолодання" і гістограма прогнозу вмісту кремнію в чавуні (зелений колір)

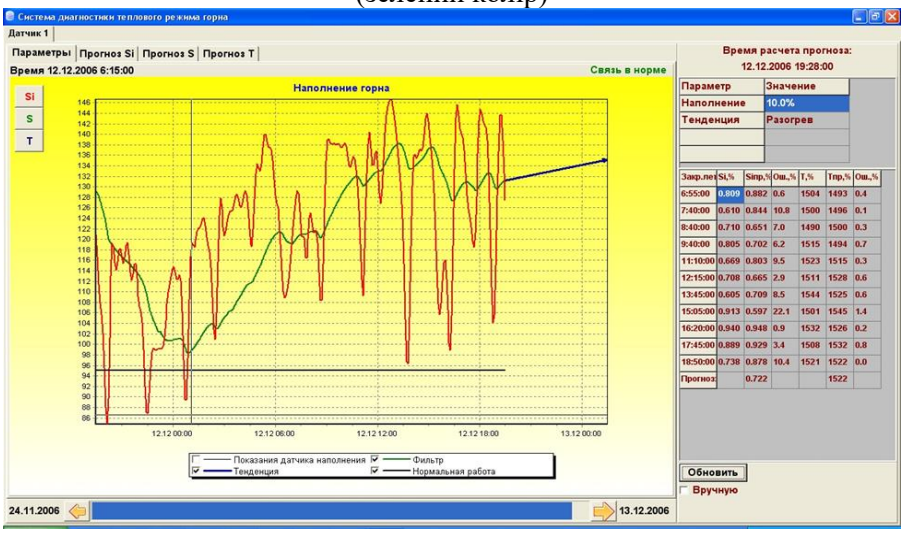


Рис. 6. Реєстрація УДТП-2 режиму "нормальної" роботи горна доменної печі

Важливою перевагою використання запропонованої інформаційної технології контролю теплового стану доменної печі є можливість проводити моніторинг у режимі реального часу, що дає

© Гусєв О.Ю., Рибальченко Ю.П.

змогу операторам швидко реагувати на будь-які інші проблеми, що можуть вплинути на процес плавки, та скоригувати виробничі параметри для забезпечення ефективності та якості продукції.

На основі запропонованої інформаційної технології було розроблено програмне середовище, що дає змогу оператору отримувати кількісну оцінку інформації, яка автоматично відображає тренд теплового стану горна доменної печі та прогноз вмісту кремнію в чавуні на випуску.

Висновки. Використання інформаційної технології для контролю теплового стану доменної печі є необхідним компонентом сучасного виробництва чавуну. Вона дає змогу операторам-технологам контролювати і коригувати виробничі параметри в режимі реального часу, використовуючи, в перспективі, технології штучного інтелекту і машинного навчання. Крім того, контроль теплового стану доменної печі дають змогу підвищити безпеку виробничого процесу шляхом запобігання можливим аваріям, пов'язаним з перегрівом печі, а також підвищити якість продукції.

Список бібліографічного опису:

1. Корнієнко В.І. Комплексна оцінка, ідентифікація та прогнозування складних нелінійних процесів / В.І. Корнієнко, І.Г. Гуліна, Л.В. Будкова, // Науковий вісник НГУ. – 2013. – № 6. – С. 124-131. – ISSN 2071-2227.
2. Гулина И.Г. Идентификация, прогнозирование и управление сложным многосвязным объектом управления / И.Г. Гулина, В.И. Корниенко, А.Ю. Гусев, В.Г. Макиенко // Системы обработки информации. – 2012. – № 9 (107). – С. 31-35. – ISSN 1681-7700.
3. Боковиков Б.А. Моделирование тепловых процессов в горне доменной печи с учётом периодичности накопления и выпуска продуктов плавки / Б.А. Боковиков, В.Г. Макиенко // Известия вузов. Чёрная металлургия, 1986, №4, с.126-130.

References:

1. Korniyenko V.I. Comprehensive evaluation, identification, and forecasting of complex nonlinear processes / V.I. Korniyenko, I.G. Hulina, L.V. Budkova // Scientific Bulletin of the National Mining University. – 2013. – No. 6. – P. 124-131. – ISSN 2071-2227.
2 Hulina I.G. Identification, forecasting, and control of a complex multivariable control object / I.G. Hulina, V.I. Korniyenko, A.Yu. Gusev, V.G. Makiyenko // Information Processing Systems. – 2012. – No. 9 (107). – P. 31-35. – ISSN 1681-7700.
3. Bokovikov B.A. Modeling of heat processes in the hearth of blast furnace, taking into account the periodicity of accumulation and release of melt products / B.A. Bokovikov, V.G. Makienko // Izvestia vuzov. Ferrous metallurgy, 1986, №4, p.126-130.

© Гусєв О.Ю., Рибальченко Ю.П.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-20
УДК 534.6:004.52
Загинайло Євгеній Олександрович, аспірант
https://orcid.org/0009-0004-5109-2053
Приходченко Сергій Дмитрович, к.т.н.
https://orcid.org/0000-0002-6562-0601
Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна

РОЗВИТОК ЗАСОБІВ АКУСТИЧНОЇ ЛОКАЛІЗАЦІЇ ТА АНАЛІЗ ЇЇ БАЗОВИХ МЕТОДІВ

Загинайло Є.О., Приходченко С.Д. Розвиток засобів локалізації та розгляд базових методів акустичної локалізації. У статті розглядається шлях розвитку галузі просторової локалізації. Визначаються основні види просторових локалізацій за способом виявлення джерел сигналу та приклади використання різних локаторів. Розглядається принцип радіолокації та як її приклад - радар, задля більш повного аналізу існуючих локаторів акустичного простору. Локалізація джерел акустичних сигналів розглядається з моменту створення першого пристрою, який поклав початок розвитку галузі. Також розглядається головний принцип за яким стояли перші локатори, а саме інтерференції хвиль. Розглядаються інші локатори за видом отримання сигналу (активність, пасивність) та приводяться приклади та сфери в яких тій чи інший вид локатора може бути застосований. В статті розглядаються різні методи визначення джерел звуку, такі як: визначення джерел за часом затримки (Time delay estimation) або метод формування променя (Beamforming). Приводять приклад використання масиву мікрофонів, як основного пристрою який дає можливість застосовувати наявні методи акустичної локації. Детально описується метод розрахунку локації за затримкою в часі та приводяться базові формули представлення сигналу відносно кожного мікрофону, формули розрахунку крос-кореляції та узагальненої крос-кореляції. Виділення особливості локалізації в середах з високим рівнем відношення сигналу до шуму (SNR) та варіанту вирішення даної проблеми. Детально описується процес акустичної локалізації методом формування променю та процес що стоїть за ним. Приводяться можливі варіанти використання нейронних мереж задля покращення ефективності наявних алгоритмів чи можливості створення новітніх алгоритмів на базі нейронних мереж.
Ключові слова: акустична локалізація, TOA, TDOA, SNR, Beamforming

Zahynailo Y., Prihodchenko S. Development of localization techniques and review of basic methods of acoustic localization. The article shows the development of spatial localization. The main types of spatial localization are determined by the method of identifying signals and the use of different locators. The principle of radio location and radar as its application, is examined for a better analysis of existing locators of the acoustic space. The acoustic signal sources locators are observed from the moment of creation of the first device, which is the beginning of sphere development. The main principle, which was used in first locators (waves interference) is also considered. Other locators are examined based on the type of signal interception (active, passive) examples are added with spheres in which different locator types can be used. Different methods of sources location are considered: time delay estimation or beamforming methods. There is an example of microphone array usage, as the main device, which makes methods of acoustic location work. Location defining method, which uses time delay is described in details and basic formulas are demonstrated like: a formula to define the signal relative to the microphone, cross correlation formula or generalized cross correlation formula. Highlighting the features of localization in environments with a high signal-to-noise ratio with different solutions for this problem. The process of acoustic localization by the method of beam formation and the process behind it are described in detail. Possible options for using neural networks to improve the efficiency of existing algorithms or the possibility of creating new algorithms based on neural networks are presented.
Keywords: acoustic localization, TOA, TDOA, SNR, Beamforming

Вступ. Розвиток сучасної науки глибоко пов'язаний із цифровізацією, та переносом більшості рутинних завдань до контролю та виконання в автоматичному режимі програмованими та машинно-навченими системами. Нажаль, до сьогодні не всі властивості, притаманні людині можливо автоматизувати, чи цифровізувати з багатьох причин, які стосуються швидкодії, рівнів чутливості і багатьох інших факторів. Але наукові дослідження невпинно намагаються перекласти важкі і рутинні завдання від людини до штучного агенту, який в ідеалі дає менше помилок, ніж людина, і ніколи не втомлюється.

Одним з таких завдань є просторова локалізація джерела звуку у тривимірному просторі. Локалізація може поділятись на різні види за способом виявлення об'єктів, або джерела сигналу. Це може бути акустична або радіо локація, сейсмічна, лазерна, гравітаційна локація тощо. Кожен з перелічених видів локалізації використовується у своїй специфічній сфері для розв'язання задачі розміщення тих чи інших об'єктів в просторі.

Постановка задачі. Акустична локалізація об'єктів у тривимірному просторі є важливою технологією, яка знаходить застосування в багатьох галузях, від робототехніки до систем безпеки. Вона дозволяє визначати положення джерела звуку, використовуючи різницю в часі прибуття звукових хвиль до різних мікрофонів. Це особливо корисно для автономних систем, таких як дрони або роботи, які повинні орієнтуватися в просторі без допомоги людини. Завдяки акустичній

@ Загинайло Є.О., Приходченко С.Д.

локалізації, ці системи можуть точно визначати місцезнаходження об'єктів, уникати перешкод і виконувати складні завдання в реальному часі.

Крім того, акустична локалізація має велике значення для розвитку технологій віртуальної та доповненої реальності. Вона дозволяє створювати більш реалістичні та інтерактивні середовища, де користувачі можуть взаємодіяти з віртуальними об'єктами, які точно відповідають їхнім фізичним положенням у просторі. Це відкриває нові можливості для навчання, розваг та професійної підготовки, забезпечуючи більш глибоке занурення та ефективність. Таким чином, акустична локалізація є ключовою технологією, яка сприяє розвитку інноваційних рішень у багатьох сферах.

Тому, актуальною є розробка систем акустичної локалізації для пошуку та встановлення актуальних координат об'єкту, що породжує звукові сигнали.

Метою статті є аналіз попередніх досліджень в галузі акустичної локалізації об'єктів та визначення напрямків подальших досліджень в галузі акустичного контролю простору та локалізації об'єктів, що породжують звукові коливання.

Аналіз попередніх досліджень і публікацій. Акустична локалізація, є достатньо дослідженим методом, дослідження якого почалось ще на початку 1880 років, а саме з першого локалізаційного пристрою Мейєра (топофону, рисунок 1) [1], та наукових робіт, які були породжені натомість.

Рис. 1. Топофон Мейєра

Перші акустичні локатори були побудовані на принципі пасивності. Тобто в конструкції таких локаторів були тільки приймачі акустичного сигналу і тому їх область застосування була достатньо обмежена, а принцип роботи дуже не досконалим.

Принципом тогочасних акустичних пристроїв була звичайна інтерференція хвиль, у випадку, якщо уловлювачі звуку отримували сигнал одночасно, або різниця була не більше ніж в довжину отриманої хвилі, то отриманий сигнал посилювався, в іншому випадку сигнал послаблювався. Саме цей принцип давав можливість визначати напрямок джерела акустичного сигналу.

Надалі, даний принцип використовувався в різних пристроях аж до появи радарів. Прикладом розповсюдженого використання таких пристроїв слугує "військові туби", або "звукові труби", які використовувались в першій світовій війні Францією, Британією для локації літаків і дирижаблів.

З появою радарів (RADAR - акронім, який означає "Radio Detection and Ranging") люди почали більше переходити з пасивних локаторів до активних локаторів.

Активні локатори, такі як радар, мають у своїй конструкції джерело сигналу та пристрій, що отримує випущений сигнал. Принципом радара є принцип дуже схожий на ехолокацію. Радар випускає короткі радіохвилі, які рухаються зі швидкістю світла і покладається на відбиття хвилі об'єктом в напрямку приймача. Завдяки цьому принципу можна встановлювати дальність до об'єкта, його матеріал тощо. [2]. Цей принцип також можна використовувати не тільки в стаціонарних радарних установках, а й для побудови карти місцевості шляхом вимірювання відстані до перешкод, що може гарно слугувати у ситуаціях де неможливо користуватись камерою, або використання камери є недостатньо інформативним.

Також, застосування радара не обмежується локалізацією об'єктів, або побудовою карти місцевості, радари також використовують у передбачені погодних умов. Таким радаром, який вже довгі роки допомагає поратись з цією задачею є радар Доплера. Принцип радара збігається з

@ Загинайло Є.О., Приходченко С.Д.

принципом звичайного радара за винятком того що радар Доплера може аналізувати ефект Доплера, який відбувається завдяки краплям води в повітрі, що рухаються в атмосфері, або використовувати технологію двійної поляризації, що дасть можливість визначити тип опадів, які очікуються та інше [3]

Зазвичай, можливість користуватись активними локаторами обмежена або правовим полем, або фінансами людини, тому в якості компромісу зазвичай використовуються пасивні системи локації.

В сучасних системах акустичної локації часто застосовується метод визначення джерела звуку завдяки розрахунку часу затримки (Time delay estimation метод), метод формування променю, бінауральний метод тощо.

В рамках задачі визначення локації джерела сигналу виділяються така низка підзадач:

- Визначення TOA (часу отримання сигналу мікрофоном).
- Визначення TDOA (різниця в часі отримання сигналу між окремими мікрофонами).

Кожен з вищезазначених параметрів зазвичай зберігається в наносекундах (10-9), або пікосекундах (10-12) для зменшення можливих втрат точності при розрахунках.

Основою усіх методів розрахунку затримки часу є масив мікрофонів, який робить заміри TOA та TDOA від звукової хвилі, що випущена джерелом сигналу. Припускається що масив мікрофонів був направлений під кутом до джерела і кожен окремий мікрофон отримує хвилю з затримкою в часі n (0<n<N-1), де N - кількість мікрофонів. Сигнал може бути представлений таким чином:

$$r_1(t) = aS(t) + n_1(t) \tag{1}$$

$$r_2(t) = aS(t - \tau_1) + n_2(t) \tag{2}$$

Де $r_1(t)$ та $r_2(t)$ це сигнали, що уловлюються мікрофоном, S(t) - це сигнал який був спродукований джерелом, $n_x(t)$ - шум на проміжку між джерелом та приймачем та a - коефіцієнт ослаблення, який показує те з якою легкістю промінь, або хвиля може перетнути довільний обсяг матеріалу (0<t<T, де T - час споглядання) [4].

Далі, застосовуються алгоритми, що дають визначити локацію джерела. Базовим алгоритмом для визначення локації вважається алгоритм крос кореляції, який може виразити формулою: [4]

$$R_{r1r2}(\tau) = E(r_1(t)r_2(t - \tau)) \tag{3}$$

Або ж згідно з властивостями, за формулою 4:

$$R_{r1r2}(\tau) = E(r_1(t + \tau)r_2(t)) \tag{4}$$

В якості покращеного варіанту простого алгоритму крос кореляції є алгоритм узагальненої крос кореляції (Generalized cross correlation), що був створений Кнапом та Картером в їх роботах 1976 року [5]. Принципом даного методу є визначення часу затримки шляхом максимізації значень крос кореляції сигналів окремих мікрофонів. Даний тип алгоритму має гарні показники точності при середніх значеннях SNR (signal to noise ratio) та час роботи. Розрахунок алгоритму узагальненої крос кореляції проводиться за формулою [4]:

$$R_{r1r2}(\tau) = \int_{-\infty}^{\infty} \psi(f) G_{r1r2}(f) e^{j2\pi f\tau} df \tag{5}$$

Щодо роботи в середах з високими значеннями SNR є можливість модифікації даного алгоритму, де буде додана специфічна вагова функція, яка нормалізує спектральну щільність сигналу за величиною спектра, що призводить до узагальненого методу фазового перетворення крос-кореляції (GCC-PHAT) [6].

@ Загинайло Є.О., Приходченко С.Д.

Також хочеться виділити метод формування променю (Beamforming), як один з можливих методів локації акустичного сигналу. В основі методу формування променю знаходиться алгоритм "Суми та затримки" [7], а сам метод, являє собою так званий просторово-часовий фільтр сигналу. Тобто завдяки цьому методу можна фільтрувати сигнали з визначеного напрямку, в той час, як сигнали з інших напрямків будуть пригнічуватись. Це працює так, що якщо при надходженні сигналу під кутом , кожен окремий мікрофон в масиві мікрофонів буде зчитувати сигнал в різні моменти, якийсь раніше, якийсь пізніше і тільки у випадку коли сигнал буде розповсюджуватись перпендикулярно площині розміщення масиву мікрофонів (визначений напрямок масиву буде правильний) - при складанні сигналів з усіх мікрофонів - на виході буде отриманий сигнал з найбільшою амплітудою (Рис. 2).

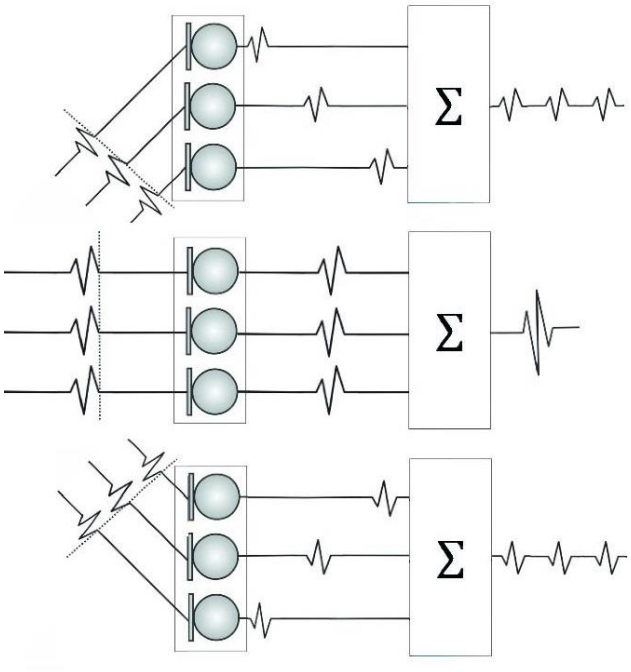


Рис. 2. Приклад складання сигналу з масиву мікрофонів під різними кутами

Таким чином, якщо звукова хвиля буде розповсюджуватись в напрямку відмінному від напрямку, що визначений в методі формування променю, в результаті утвориться хвиля з довгим сигналом маленької амплітуди, або хвилі можуть компенсувати одна одну інтерференцією хвиль.

Також, хотілося б виділити ефективність використання нейронних мереж для розв'язання задач акустичної локації, які з часом набирають все більше популярності.

Нейромережі можуть використовуватись задля покращення точності визначення джерела сигналу, зменшення рівня навколишнього шуму, надання адаптивності статичним алгоритмам, або навіть для перевірки можливих нових алгоритмів акустичної локації.

Одним з прикладів можливого покращення наявних алгоритмів може бути нейромережа, яка завдяки згортковим шарам зможе виділяти походження, або окремі характеристики звукової хвилі, таким чином інші шари нейромережі зможуть використовувати дану інформацію для більш точного налаштування чутливостей мікрофонів і напрямку масиву мікрофонів (у разі можливості його регулювання) що дасть можливість створити систему визначення акустичного сигналу ще більшої точності.

Висновки та перспективи подальшого дослідження. Таким чином, беручи до уваги шлях розвитку акустичної локації та наявність класичних методів просторової локалізації, кожен з яких має свої переваги і недоліки, користуючись новітніми технологіями з створення та впровадження нейромереж, можна передбачити можливість подальшого покращення методів акустичної локації і створення нових методів з вже більшим покладанням на новітні технології.

Одним із ключових аспектів цих досліджень є використання багатоканальних мікрофонних масивів та обробки сигналів. Завдяки цьому можна отримати більш точні дані про напрямок та відстань до джерела звуку. Сучасні методи обробки сигналів, такі як алгоритми на основі машинного навчання, дозволяють значно підвищити точність та швидкість локалізації. Крім того, використання

@ Загинайло Є.О., Приходченко С.Д.

тривимірних моделей середовища допомагає враховувати відбиття та поглинання звуку, що робить результати більш надійними.

Перспективи розвитку цієї галузі включають інтеграцію з іншими технологіями, такими як комп'ютерний зір та штучний інтелект. Це дозволить створювати більш комплексні системи, здатні не лише визначати місцезнаходження джерела звуку, але й аналізувати його характеристики та контекст. Такі системи можуть бути використані для автоматичного розпізнавання мовлення в шумних умовах або для створення інтелектуальних систем спостереження, які реагують на звукові події в реальному часі.

Список бібліографічного опису:

1. Топофон : патент 224,199 Сполучені Штати Америки. URL: https://patents.google.com/patent/US224199A.
2. Боле А., Волл А., Норріс А., Базові принципи радару. *Посібник по радару та ARPA*. 2014. P. 1–28. URL: https://doi.org/10.1016/b978-0-08-097752-2.00001-5 (дата звернення: 19.08.2024).
3. Довіак Р.Дж., Жрнік Д.С., Сірманс Д.С. Погодній радар Доплера. *Робота з IEEE*. 1979. Том. 67, час. 11. с. 1522–1553. URL: https://doi.org/10.1109/proc.1979.11511 (дата звернення: 19.08.2024).
4. Ріту К., Санджів К., Огляд методів локалізації джерела звуку. *Європейський журнал досягнень в сфері інженерії та технологій,*. URL: https://ejaet.com/review-on-acoustic-source-localization-techniques/ (дата звернення: 19.08.2024).
5. Узагальнена крос кореляція з частотним ковзанням: підхід до оцінки затримки під діапазону / Кобос М., Антоначі Ф., Командуччі Л., Сарті А. *arxiv*. 2019. 19 жовтня. URL: https://doi.org/10.48550/arXiv.1910.08838 (дата звернення: 19.08.2024).
6. Кнапп С., Картер Г. Узагальнений кореляційний метод для оцінки затримки часу. *IEEE Операції з акустики, мови та обробки сигналів*. 1976. Том. 24, час. 4. с. 320–327. URL: https://doi.org/10.1109/tassp.1976.1162830 (дата звернення: 19.08.2024).
7. Валенсія-Пальма А., Кордова-Еспарза Д.-М.. Локалізація джерела звуку за допомогою формування променю та його представлення у вбудованому пристрої змішаної реальності. *Конспект лекцій з комп'ютерних наук*: книга. 2019. URL: https://doi.org/10.1007/978-3-030-21077-9_35 (дата звернення: 19.08.2024).

References:

1. Topophone : patent 224,199 United States. URL: https://patents.google.com/patent/US224199A. (date of access: 19.08.2024).
2. Bole A., Wall A., Norris A. Basic Radar Principles. Radar and ARPA Manual. 2014. P. 1–28. URL: https://doi.org/10.1016/b978-0-08-097752-2.00001-5 (date of access: 19.08.2024).
3. Doviak R. J., Zrnic D. S., Sirmans D. S. Doppler weather radar. Proceedings of the IEEE. 1979. Vol. 67, no. 11. P. 1522–1553. URL: https://doi.org/10.1109/proc.1979.11511 (date of access: 19.08.2024).
4. Ritu K., Sanjeev K. Review on Acoustic Source Localization Techniques. European Journal of Advances in Engineering and Technology,. URL: https://ejaet.com/review-on-acoustic-source-localization-techniques/ (date of access: 19.08.2024).
5. Frequency-Sliding Generalized Cross-Correlation: A Sub-band Time Delay Estimation Approach / Cobos M., Antonacci F., Comanducci L., Sarti A., arxiv. 2019. 19 October. URL: https://doi.org/10.48550/arXiv.1910.08838 (date of access: 19.08.2024).
6. Knapp C., Carter G. The generalized correlation method for estimation of time delay. IEEE Transactions on Acoustics, Speech, and Signal Processing. 1976. Vol. 24, no. 4. P. 320–327. URL: https://doi.org/10.1109/tassp.1976.1162830 (date of access: 19.08.2024).
7. Valencia-Palma A., Cordova-Esparza D.-M. Sound Source Localization Using Beamforming and Its Representation in a Mixed Reality Embedded Device. Lecture Notes in Computer Science : book. 2019. URL: https://doi.org/10.1007/978-3-030-21077-9_35 (date of access: 19.08.2024).

@ Загинайло Є.О., Приходченко С.Д.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-21
УДК 004.4`4:004.85
Іваненко Антон Романович, аспірант,
https://orcid.org/0000-0002-9846-537X
Марченко Олександр Іванович, к.т.н., доцент,
https://orcid.org/0000-0002-4537-3420
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

МЕТОД КОМПІЛЯЦІЇ ОГОЛОШЕННЯ КЛАСІВ МОВ WEB-ПРОГРАМУВАННЯ У МОВИ БАЙТ-КОДОВОГО ТИПУ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Іваненко А.Р., Марченко О.І. Метод компіляції оголошення класів мов WEB-програмування у мови байт-кодового типу на основі машинного навчання. У даній статті пропонується метод, який дозволяє компілювати декларування класів мов WEB-програмування у мови байт-кодового типу на основі машинного навчання. Запропонований метод базується на ідеї використання двох видів штучного інтелекту. Для виконання перевірки введеної програми користувачем, запропоновано використати нейронну мережу, яка навчена для виконання задачі бінарної класифікації введеної програми на коректну і не коректну. Для навчання мережі було обрано алгоритм SDCA, що добре зарекомендував себе для розв'язку задач бінарної класифікації. З метою генерації інструкцій цільової мови на основі введеної програми у даній статті було використано генеративний штучний інтелект на основі LLM від компанії OpenAI. Для розв'язку задачі модель ChatGPT була донавчена на відповідних прикладах методом fine-tuning. Дослідження було апробоване на прикладі створення тестового компілятора, що перевіряє на коректність введену програму, що написана мовою TypeScript, та генерує відповідний код мовою CIL з достатньо високою точністю. Отриманий результат доводить, що використання методів машинного навчання для створення компіляторів є можливим. Такий підхід дозволить звести розробку компілятора тільки до підготовки правильного набору даних для донавчання відповідних моделей, що є значно простішим і менш часовитратним у порівнянні з класичним підходом, коли для кожного способу чи методу компіляції підмножини мови, що компілюється, необхідно вносити зміни і доопрацьовувати лексичний, синтаксичний, семантичний аналізатори та генератор коду для кожної нової чи зміненої конструкції вхідної мови програмування.
Ключові слова: компілятор, машинне навчання, генеративний штучний інтелект, бай-код, CIL, .NET, TypeScript, LLM, ChatGPT, ML.NET

Ivanenko A., Marchenko O. Method of compiling declarations of classes of WEB-programming languages into byte-code languages based on machine learning. This article proposes a method that allows you to compile class declarations of WEB programming languages into byte-code languages based on machine learning. The proposed method is based on the idea of using two types of artificial intelligence. In order to check the entered program by a user, it is proposed to use a neural network trained to perform the task of binary classification of the entered program into correct and incorrect. The SDCA algorithm was chosen for network training, which has proven itself well for solving binary classification problems. In order to generate CIL instructions based on the entered program, this article used generative artificial intelligence based on LLM from the OpenAI company. To solve the problem, the ChatGPT model was retrained on relevant examples using the fine-tuning method. The study was tested on a developed test compiler that checks the entered program written in TypeScript for correctness and generates the corresponding code in the Common Intermediate Language (CIL) with sufficiently high accuracy. The obtained result proves that the use of machine learning methods to create compilers is possible. This approach will reduce the development of the compiler only to the preparation of the correct data set for retraining the corresponding models, which is much simpler and less time-consuming compared to the classical approach, when for each method or method of compiling a subset of the language being compiled, it is necessary to make changes and refine the lexical, syntactic, semantic analyzers and a code generator for each new or changed construct of the input programming language.
Keywords: compiler, machine learning, generative artificial intelligence, byte-code CIL, .NET, TypeScript, ChatGPT, LLM, ML.NET

Постановка наукової проблеми.

Розробка компілятора з однієї мови програмування у іншу – досить складна задача, що складається з багатьох підзадач, зокрема і розробка компілятора з мов WEB-програмування у мови байт-кодового типу. У попередніх дослідженнях, [1] та інших, ми дослідили способи і методи компіляції певних конструкцій мови TypeScript у проміжну мову CIL використовуючи класичний підхід при розробці компілятора, що складається з лексичного, синтаксичного та семантичного аналізу, а також з написання генератора, який перетворює проміжне представлення програми у CIL інструкції.

Сьогодні все більшу популярність набирає використання різних засобів штучного інтелекту для розв'язання прикладних задач, в тому числі і генеративний штучний інтелект, який використовується для створення різних текстів, зображень, відео та музичних композицій за запитами користувача. Виходячи з цього, виникла ідея використання машинного навчання для створення компілятора мови WEB-програмування у мову байт-кодового типу. Такий підхід дозволив

© Іваненко А.Р., Марченко О.І.

би узагальнити і спростити вирішення проблеми розширення функціональності компілятора лише проводячи донавчання відповідної моделі штучного інтелекту.

Для апробації запропонованого методу і проведення експериментів у цьому напрямку було вирішено обрати невелику підмножину мови TypeScript, а саме оголошення класів, та розробити компілятор за запропонованим методом для цієї підмножини конструкцій мови у проміжну мову CIL платформи .NET. Зазначимо, що цей метод є придатним не тільки для мови Typescript та мови CIL, а й для компіляції оголошення класів будь-яких мов WEB-програмування у будь-які мови байт-кодового типу.

Аналіз останніх досліджень і публікацій.

На сьогодні є декілька різних досліджень та розробок, що пов'язані з використанням штучного інтелекту, а саме методів машинного навчання, для вирішення різних задач програмування. Розглянемо декілька найпопулярніших рішень використання машинного навчання для розв'язання задач генерації коду та трансляції коду з однієї мови в іншу.

У своїй статті англійські дослідники Юджіа Лі, Девід Чої та інші [2] дослідили, яким чином можливо навчити AlphaCode [3] вирішувати різні прикладні задачі шляхом генерації відповідного коду. З метою навчання штучного інтелекту було проаналізовано мільйони рішень різних задач. Автори статті використали підхід навчання, що складався з 3 етапів (рис. 1) [2]:

1. Pre-training (попереднє навчання) – на цьому етапі було проаналізовано сотні тисяч рішень на платформі Github на різних мовах програмування шляхом розділення кожного файлу на 2 частини, які використовувались як вхідні дані для енкодера для декодера відповідно, щоб навчити модель прогнозувати та генерувати продовження коду.

2. Fine-tuning (точне налаштування) – на цьому етапі на вхід енкодера був поданий опис задачі англійською мовою, а на вхід декодера – її рішення на відповідній мові програмування.

3. Sampling & Evaluation (семплінг та розрахунок) – на цьому етапі AlphaCode генерує багато зразків рішення для кожного опису проблеми, потім виконує їх, щоб відфільтрувати погані зразки та згрупувати ті, що залишилися. Після чого залишає невеликий набір кандидатів на правильне рішення заданої проблеми.

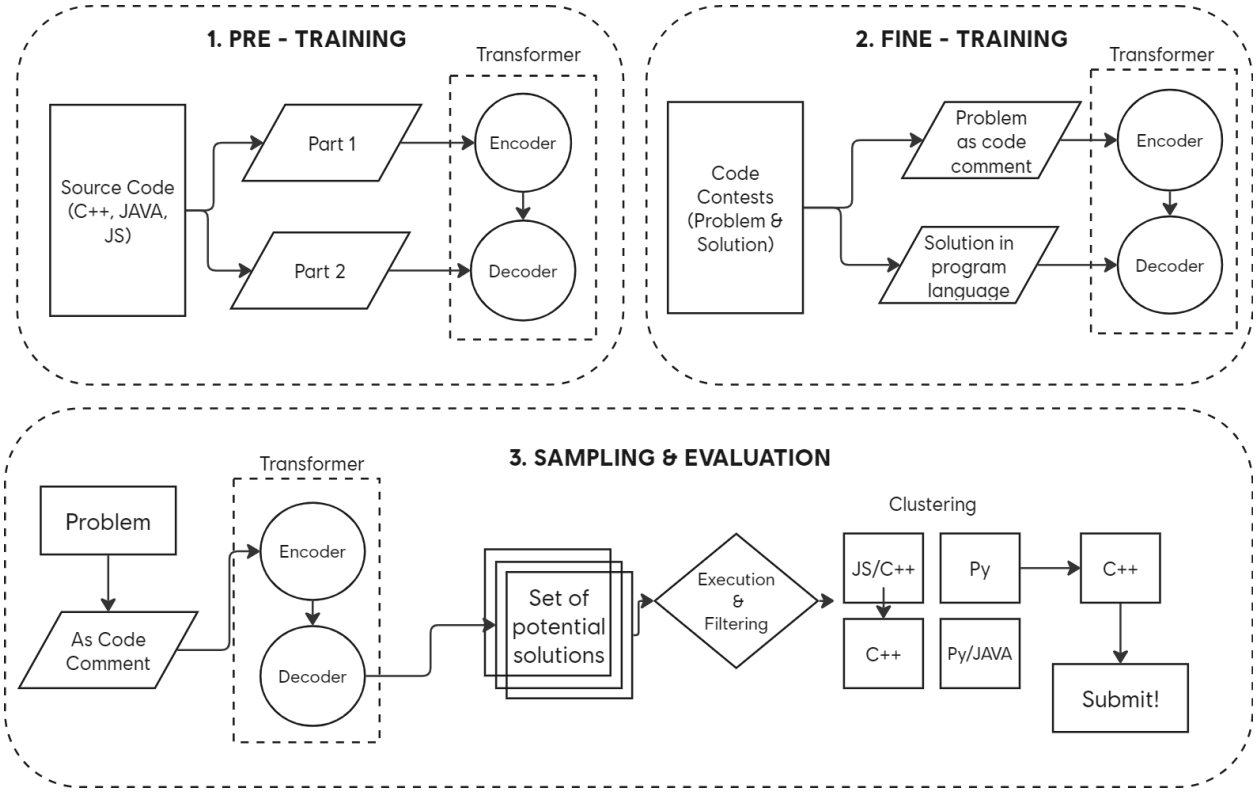


Рис. 1. Метод навчання AlphaCode для розв'язання задач програмування

Завдяки такому методу навчання, AlphaCode став спроможним розв'язувати задачі з

© Іваненко А.Р., Марченко О.І.

програмування краще, ніж 56% користувачів платформи Codeforces.

Розглянемо статтю [4], у якій китайські вчені Кьюші Сун, Нуо Чен та інші дослідили трансляцію програми з однієї мови програмування високого рівня в іншу. Вони назвали своє рішення TransCoder, що скорочено від Transferable Code Representation Learning. Цей метод базується на використанні Code Pre-Trained Models (далі CodePTM) та складається з двох етапів (рис. 2) [4]:

1. Source Task Training (навчання на вхідній задачі). На цьому етапі створюються універсальні префікси знань випадковим чином та додаються до CodePTM, що використовується для класифікації можливих розв'язків задачі певною мовою програмування.

2. Target Task Specification (специфікація цільової задачі). На цьому етапі універсальні префікси знання додаються до нової CodePTM, що відповідає розв'язку поставленої задачі іншими мовами програмування високого рівня.

Цей підхід дозволив реалізувати підтримку компіляції функцій написаних такими мовами програмування як Python, C++, JavaScript, Java. Наразі TransCoder є частиною загального рішення CodeGen від дослідницького центру компанії Meta.

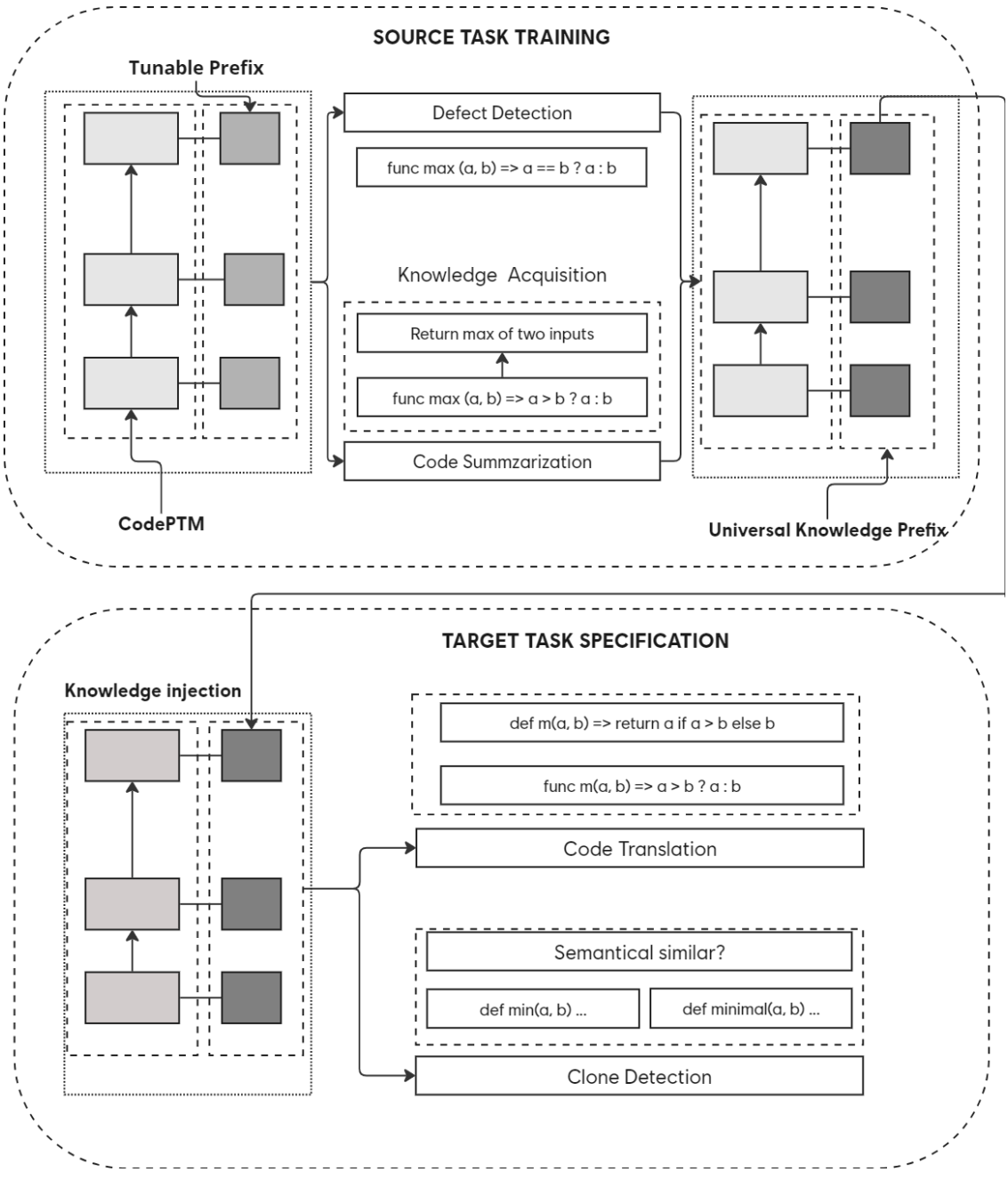


Рис. 2. Метод навчання TransCoder для компіляції програми з однієї мови високого рівня в іншу

© Іваненко А.Р., Марченко О.І.

В результаті проведеного аналізу існуючих досліджень, можна зробити два висновки.

Перший висновок – використання штучного інтелекту для вирішення різних задач програмування, починаючи від генерації коду для рішення задачі, описаної звичайною мовою, до трансляції коду, написаного однією з мов високого рівня в іншу, є можливим.

Другий висновок – дослідження компіляції мов програмування високого рівня у мови програмування низького рівня на основі машинного навчання наразі відсутні, і тому дослідження методів компіляції мов WEB-програмування у мови байт-кодового типу на основі машинного навчання матиме наукову новизну і практичну цінність.

Метою даної статті є розробка методу компіляції оголошення класів мов WEB-програмування у мови байт-кодового типу на основі машинного навчання.

Термінологія.

Large Language Model (LLM, Велика мовна модель) – це обчислювальна модель, що була навчена на великій кількості немаркованого тексту та здатна виконувати різні завдання обробки природної мови, наприклад розуміння і генерація текстів.

Fine-tuning (точне налаштування) – це підхід глибокого машинного навчання, за якого параметри попередньо навченої моделі додатково навчаються на нових даних.

Запропонований метод компіляції на основі машинного навчання.

Для апробації запропонованого методу, який є загальним і придатним для будь-яких мов WEB-програмування, було вирішено обрати мову TypeScript і виконати її компіляцію у проміжну мову CIL платформи .NET.

Розглянемо опис підграматики декларації класів мови TypeScript [5], яку було обрано для дослідження методу компіляції:

<class-declaration> --> class <class-name> <class-block>

<class-block> --> { <property-declaration-list>? }

<property-declaration-list> --> <class-property-declaration>|<class-property-declaration> <class-property-declaration-list>

<property-declaration> --> <access-modifier-keyword><property-name>:<attribute>;

<access-modifier-keyword> -> private | protected | public

<attribute> -> number | boolean | string | any

<class-name> -> <identifier>

<property-name> -> <identifier>

Приклад оголошення класу, що містить всі зазначені правила обраної підграматики, зображено на рис. 3, а очікуваний результат компіляції у проміжну мову CIL платформи .NET – на рис. 4.

```
class Test {
        private privateNumber: number;
        protected protectedString: string;
        public publicBoolean: boolean;
        public publicAny: any;
}
```

Рис. 3. Приклад оголошення класу мовою TypeScript

```
.class public auto ansi beforefieldinit Test

    extends [System.Private.CoreLib]System.Object
{
    .field private int32 privateNumber
    .field family string protectedString
    .field public bool publicBoolean
    .field public object publicAny

    .method public hidebysig specialname rtspecialname
        instance void .ctor () cil managed
    {
        .maxstack 8

        IL_0000: ldarg.0
        IL_0001: call instance void [System.Private.CoreLib]System.Object::.ctor()
        IL_0006: nop
        IL_0007: ret
    }
}
```

© Іваненко А.Р., Марченко О.І.

Рис. 4. Очікувана програма мовою CIL, що містить приклад оголошення класу з рис. 3

У мові CIL ключове слово .class відповідає класу. Ключове слово beforefieldinit означає, що тип можна ініціалізувати в будь-який момент до того, як будуть посилатися на будь-які його статичні поля (компіляція статичних класів у даній статті не розглядається). За оголошення поля відповідає ключове слово .field. Відповідність області видимості полів мови TypeScript до мови CIL наступна:

- private – private
- protected – family
- public – public.

А відповідність типів даних є такою:

- number – int32
- string – string
- boolean – bool
- any – object.

У проміжній мові CIL кожен тип даних повинен успадковуватись від загального типу-посилання System.Object (якщо дані типу зберігаються у кучі) чи типу-значення System.Value (якщо дані зберігаються на стеку) та мати конструктор за замовчуванням без параметрів. У мові TypeScript клас є типом-посиланням, тому реалізуємо успадкування від типу System.Object.

Як відомо, основний процес компіляції виконується у 4 етапи:

1) лексичний аналіз;
2) синтаксичний аналіз;
3) семантичний аналіз;
4) генерація коду.

У підмножині мови TypeScript, що розглядається в експерименті, відсутні правила, які потрібно перевіряти на семантичну некоректність, і таким чином компіляцію можна звести то задачі лексичного і синтаксичного аналізу та задачі генерації коду. Реалізацію перших двох етапів можна виконати одним об'єднаним етапом перевірки на коректність введеної програми. Таким чином, будемо розглядати дві підзадачі машинного навчання – перевірку введеної програми на помилки та генерацію вихідної програми мово CIL. Перша підзадача є задачею класифікації, де головною метою є класифікація введеного тексту на коректний і некоректний. Друга підзадача є задачею генерації тексту, для чого можна використати генеративний штучний інтелект на основі LLM (наприклад ChatGPT), який донавчено виконувати таку підзадачу. Схема компілятора, що пропонується для рішення цих підзадач, зображена на рис. 3.

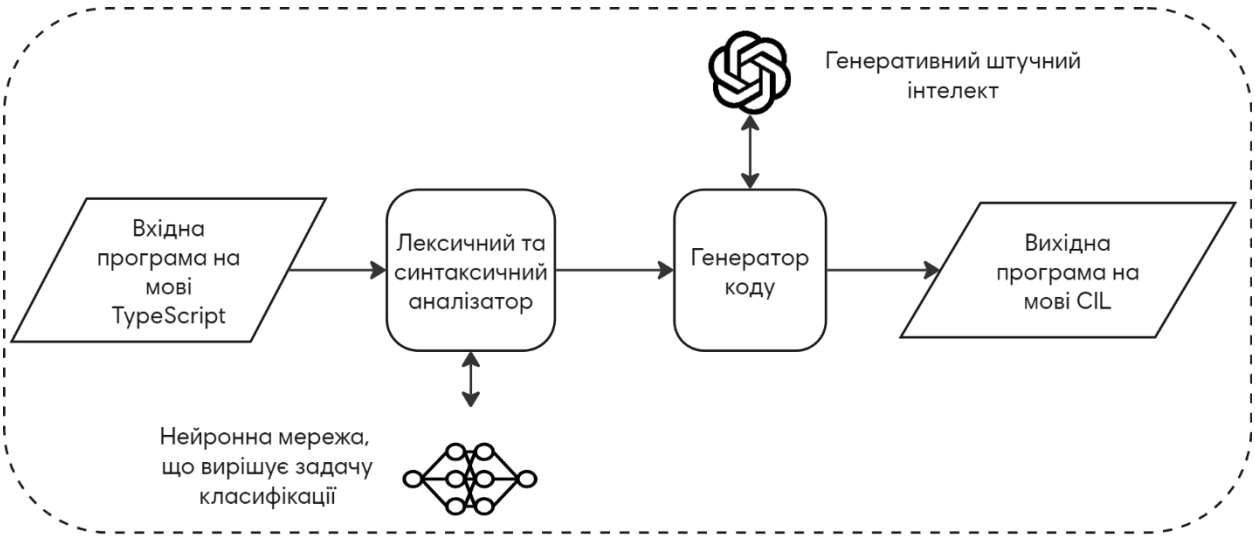


Рис. 3. Схема запропонованого компілятора

Для написання модуля, що відповідає за аналіз програми шляхом класифікації, використаємо платформу ML.NET [6].

© Іваненко А.Р., Марченко О.І.

Для початку напишемо консольну програму, що буде створювати та тренувати нашу модель на тестових даних. Ця програма має виконувати наступні етапи:

1) завантаження тестових даних з файлу та їх розподіл на дані для навчання та дані для тестування;
2) побудова і навчання моделі на відповідних даних з пункту 1;
3) тестування точності моделі на тестових даних;
4) збереження моделі в локальний файл.

Для навчання моделі було вирішено обрати алгоритм Stochastic Dual Coordinate Ascent (далі SDCA). Цей алгоритм навчання добре зарекомендував себе для вирішення задач бінарної класифікації, що висвітлено у відповідних дослідженнях [7, 8]. Схема запропонованого методу навчання зображена на рис. 4.

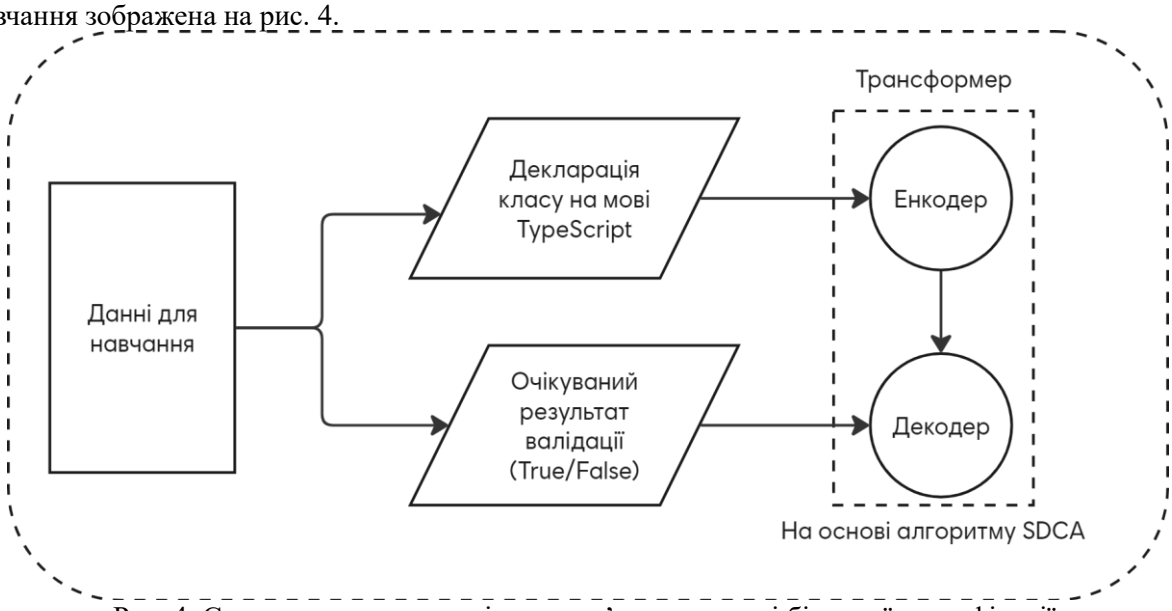


Рис. 4. Схема навчання моделі для розв'язання задачі бінарної класифікації.

Після того, як модель для виконання задачі класифікації коректного тексту налаштована і збережена у файл, у самій реалізації компілятора достатньо лише завантажити її та виконувати перевірку вхідної програми.

Наступним етапом після аналізу на коректність вхідної програми є генерація коду мовою CIL. Для виконання задачі використаємо генеративний штучний інтелект на основі LLM для генерації текстів ChatGPT від компанії Open AI. Оскільки відповідна модель ніколи не навчалась виконувати поставлену нами задачу, необхідно виконати налаштування шляхом донавчання цієї моделі за методом fine-tuning. Open AI надає відповідне API [9] для налаштування власної моделі, використовуючи підготовлені тестові дані й відповідні запити (promts). За основу власної моделі візьмемо найновішу версію gpt-4o-mini.

Створимо відповідну консольну програму, що буде завантажувати підготовлені тестові дані, та надсилати їх за протоколом HTTP для налаштування власної моделі. Ця програма виконуватиме наступні кроки:

1) завантаження даних у форматі json, що містить інформацію, як має поводити себе модель, приклад запиту та очікуваний результат на цей запит (рис. 7);
2) надсилання тестових даних у вигляді спеціальних запитів (promts);
3) перевірку результату, отриманого на тестових даних.

```
{
  "messages": [
      {
          "role": "system",
          "content": "Compiler is a chatbot that returns only generated code."
      },
      {
          "role": "user",
          "content": "Compile typescript code to common intermediate language: class
Test { private privateNumber: number…"
      },
```

© Іваненко А.Р., Марченко О.І.

```
        {
            "role": "assistant",
            "content": ".class public auto ansi beforefieldinit Test...."
        }
    ]
}
```

Рис. 7. Приклад контенту json файлу, що містить дані для навчання за методом fine-tuning

Після того, як модель налаштована і протестована, її можливо інтегрувати у наш тестовий компілятор. Для цього достатньо створити клієнт чат-бота, використовуючи той самий арі ключ, що було використано для донавчання власної моделі і зробити запит до чат-бота відповідним запитом (promt), формат якого було використано на етапі навчання (рис. 7, повідомлення role user).

Аналіз отриманих результатів.

Оскільки для виконання компіляції шляхом використання машинного навчання була взята досить мала підграматика мови TypeScript, то аналіз отриманих результатів слід проводити не у порівнянні з існуючими аналогами (як було проаналізовано у попередніх дослідженнях авторів), а дослідити точність роботи запропонованого методу на тестових даних.

За метрику точності роботи алгоритму було обрано відношення кількості коректних результатів роботи тестового компілятора до загальної кількості тестових даних (у %), в залежності від загальної кількості даних, що були використані для навчання і налаштування моделей.

Для моделі, що виконує задачу класифікації, діапазон даних для навчання був обраний від 100 до 1000000, а для моделі, що відповідає за генерацію CIL інструкцій, – від 100 до 10000. Така розбіжність діапазонів пояснюється далі в аналізі.

Для тестування було згенеровано 1000 декларацій класів з параметрами, які відрізняються за такими параметрами: кількість полів, різноманітність типів полів, імен полів та імен класів. Половина з цих даних містила помилки, які повинні були бути відсіяні на етапі лексичного і синтаксичного аналізу. Перевірка коректності згенерованих CIL інструкцій була виконана компілятором ILasm[10], який компілює проміжну мову CIL у мову асемблера. Результати тестування продемонстровані на рис. 5 та рис. 6.

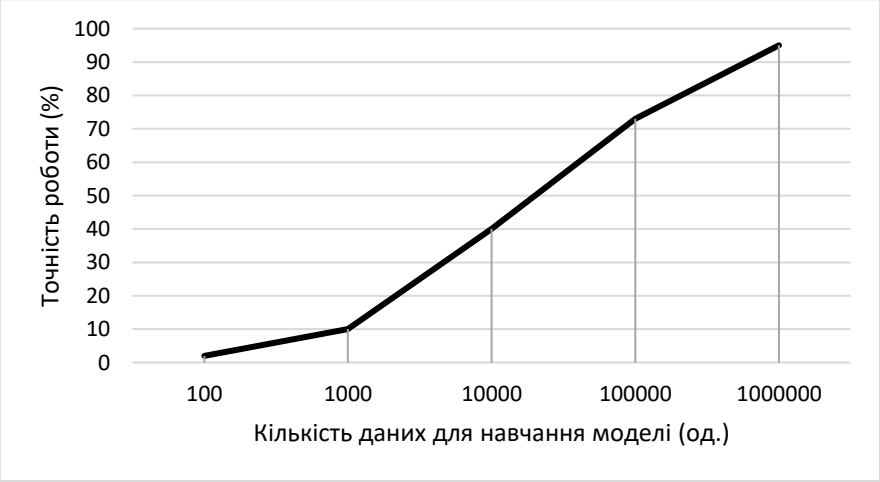


Рис. 5. Графік залежності коректності роботи лексичного та синтаксичного аналізу від загальної кількості даних для навчання моделі

© Іваненко А.Р., Марченко О.І.

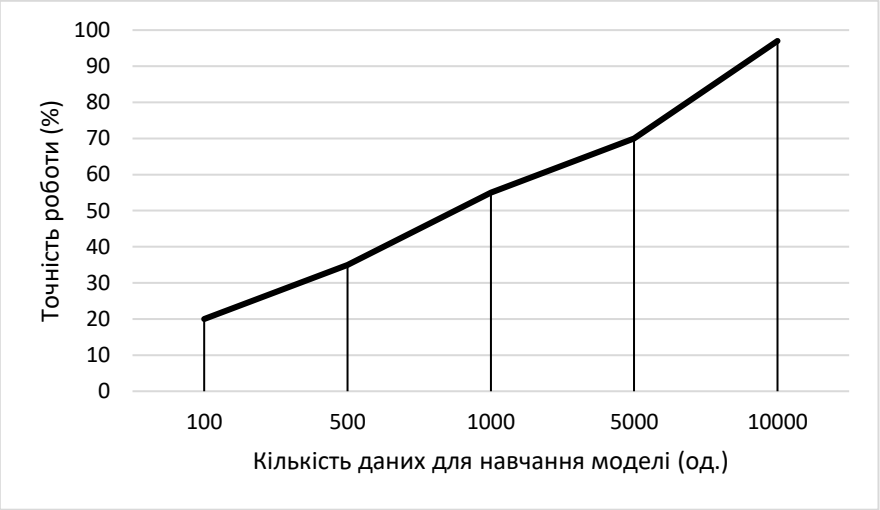


Рис. 6. Графік залежності коректності згенерованого коду від загальної кількості даних для навчання моделі

Як бачимо з отриманих результатів, для підвищення точності роботи компілятора необхідно збільшувати кількість даних для навчання. Чим ближче точність наближається до 100%, тим в рази більше даних необхідно використати для навчання.

Варто зазначити, що у випадку навчання моделі для виконання задачі класифікації (рис. 8) необхідно в рази більше тестових даних у порівнянні з донавчанням генеративної мережі (рис.9). Така розбіжність тестових даних пояснюється тим, що генеративна мережа (у нашому випадку ChatGPT) вже навчена виконувати задачу генерації тексту на великій кількості різноманітних даних і в ході експерименту ми лише донавчали її на різних прикладах, щоб відкоригувати роботу для виконання поставленої нами задачі.

Висновки та перспективи подальшого дослідження.

В цій статті отримало подальший розвиток дослідження методів компіляції мов WEB-програмування у мов байт-кодового типу. Зокрема було розроблено і досліджено метод компіляції оголошення класів на основі машинного навчання, який був апробований на прикладі компіляції оголошення класів мови TypeScript у проміжну мову CIL платформи .NET, яка є мовою байт-кодового типу.

Для реалізації тестового компілятора етапи лексичного і синтаксичного аналізу процесу компіляції були об'єднані в одну задачу машинного навчання, яка виконувала аналіз програми на коректність, а генерація вихідного коду вирішувалася як друга задача машинного навчання. Для розв'язку першої задачі було вирішено використати нейронну мережу, що була навчена на основі алгоритму SDCA виконувати бінарну класифікацію тексту введеної користувачем програми на коректну і некоректну, а для розв'язку другої задачі – використати LLM модель (ChatGPT), яку було донавчено за методом fine-tuning на відповідних прикладах генерувати CIL інструкції.

Отримані результати показують, що запропонований метод може бути застосований на практиці, але для реалізації компілятора навіть достатньо малої підмножини мови WEB-програмування необхідно мати велику кількість тестових даних для навчання. Крім того, у порівнянні з класичними підходами трансляції запропонований метод не завжди показує достатню точність роботи. Для підвищення точності роботи необхідно збільшувати кількість тестових даних, а для реалізації компілятора усієї мови кількість таких тестових даних повинна бути збільшена в рази.

Отриманий результат доводить, що використання методів машинного навчання для створення компіляторів є можливим. Такий підхід дозволить звести розробку компілятора тільки до підготовки правильного набору даних для донавчання відповідних моделей, що є значно простішим і менш часовитратним у порівнянні з класичним підходом, коли для кожного способу чи методу компіляції підмножини мови, що компілюється, необхідно вносити зміни і доопрацьовувати лексичний, синтаксичний, семантичний аналізатори та генератор коду для кожної нової чи зміненої конструкції вхідної мови програмування.

Напрямком подальших досліджень може бути розвиток запропонованого методу для інших конструкцій мов WEB-програмування, його апробація на інших вхідних мовах та мовах байт-

© Іваненко А.Р., Марченко О.І.

кодового типу, а також узагальнення цього методу компіляції на основі інших видів машинного навчання.

Список бібліографічного опису

1. Іваненко А.Р., Марченко О.І. «Метод компіляції типів об'єднання мови TypeScript у проміжну мову CIL платформи .NET», *Комп'ютерно-інтегровані технології: освіта, наука, виробництво,* 2023, № 52, с.77-84. Режим доступу: http://surl.li/mxrufv. Дата звернення: 11.09.24.
2. Yujia Li, David Choi, , Junyoung Chung and others. «Competition-level code generation with AlphaCode», *Science*, 2022, Vol 378, pp. 1092-1097. Available: http://surl.li/jbchsl. Accessed on: 11.09.24.
3. AlphaCode Attention Visualization, Available: https://alphacode.deepmind.com. Accessed on 11.09.24
4. Qiushi Sun, Nuo Chen, Jianing Wang, Xiang Li, Ming Gao. «TransCoder: Towards Unified Transferable Code Representation Learning Inspired by Human Skills», *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, 2024, pp. 16713–16726, Available: http://surl.li/sxmmcw. Accessed on: 11.09.24.
5. Документація TypeScript. Режим доступу: https://www.typescriptlang.org/docs/handbook/unions-and-intersections.html. Дата звернення 11.09.24.
6. Документація ML.NET. Режим доступу: https://learn.microsoft.com/en-us/dotnet/machine-learning. Дата звернення: 11.09.24.
7. Shai Shalev-Shwartz, Tong Zhang. «Stochastic Dual Coordinate Ascent Methods for Regularized Loss Minimization», *Journal of Machine Learning Research*, 2013, №14, pp. 567-599. Available: http://surl.li/vcaisp. Accessed on: 11.09.24.
8. Kenneth Tran, Saghar Hosseini, Lin Xiao, Thomas Finley, Mikhail Bilenko. «Scaling Up Stochastic Dual Coordinate Ascent», *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 1185 – 1194. Available: http://surl.li/cdxvze. Accessed on: 11.09.24.
9. Документація OpenAI Platform. Режим доступу: http://surl.li/rnmkpo. Дата звернення: 11.09.24.
10. Документація ILAsm. Режим доступу: http://surl.li/dtqbcv. Дата звернення: 11.09.24.

References

1. Ivanenko A.R., Marchenko O.I. «Method of compilation union types of TypeScript into Common Intermediate Language of .NET platform», *Computer-Integrated Technologies: Education, Science, Production*, 2023, №52, pp. 77-84. Available: http://surl.li/mxrufv. Accessed on: 11.09.24.
2. Yujia Li, David Choi, , Junyoung Chung and others. «Competition-level code generation with AlphaCode», *Science*, 2022, Vol 378, pp. 1092-1097. Available: http://surl.li/jbchsl. Accessed on: 11.09.24.
3. AlphaCode Attention Visualization, Available: https://alphacode.deepmind.com/. Accessed on 11.09.24.
4. Qiushi Sun, Nuo Chen, Jianing Wang, Xiang Li, Ming Gao. «TransCoder: Towards Unified Transferable Code Representation Learning Inspired by Human Skills», *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, 2024, pp. 16713–16726, Available: http://surl.li/sxmmcw. Accessed on: 11.09.24.
5. TypeScript documentation. Available: https://www.typescriptlang.org/docs/handbook/unions-and-intersections.html. Accessed on: 11.09.24.
6. ML.NET documentation. Available: https://learn.microsoft.com/en-us/dotnet/machine-learning/. Accessed on: 11.09.24.
7. Shai Shalev-Shwartz, Tong Zhang. «Stochastic Dual Coordinate Ascent Methods for Regularized Loss Minimization», *Journal of Machine Learning Research*, 2013, №14, pp. 567-599. Available: http://surl.li/vcaisp. Accessed on: 11.09.24.
8. Kenneth Tran, Saghar Hosseini, Lin Xiao, Thomas Finley, Mikhail Bilenko. «Scaling Up Stochastic Dual Coordinate Ascent», *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 1185 – 1194. Available: http://surl.li/cdxvze. Accessed on: 11.09.24.
9. OpenAI Platform documentation. Available: http://surl.li/rnmkpo. Accessed on: 11.09.24.
10. ILAsm documentation. Available: http://surl.li/dtqbcv. Accessed on: 11.09.24

© Іваненко А.Р., Марченко О.І.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-22
УДК 004.4`2
Касянчук Дмитро Павлович, аспірант
https://orcid.org/0009-0004-2824-8232
Марченко Олександр Іванович, к.т.н., доцент
https://orcid.org/0000-0002-4537-3420
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

ЗАСІБ СИНТАКСИЧНОГО РОЗБОРУ НА ОСНОВІ ГЕНЕРАТОРА СИНТАКСИЧНИХ АНАЛІЗАТОРІВ ANTLR ТА МОВИ CLOJURE

Касянчук Д.П., Марченко О.І. Засіб синтаксичного розбору на основі генератора синтаксичних аналізаторів ANTLR та мови Clojure. Ця стаття присвячена процесу синтаксичного розбору програмного коду. В результаті виконаного аналізу низки публікацій, які містять порівняння існуючих генераторів синтаксичних аналізаторів, було визначено, що генератор синтаксичних аналізаторів ANTLR користується популярністю у розробників завдяки своїй простоті, широкому спектру функцій, а також якості і ефективності згенерованого коду. В статті представлено засіб синтаксичного розбору програмного коду, що базується на використанні генератора синтаксичних аналізаторів ANTLR, а також сучасної функційної мови програмування Clojure, характеристики якої дозволяють значно легше створювати деревоподібні структури даних, а також аналізувати та змінювати їх, якщо порівнювати з нефункційними мовами програмування. Крім того, широкий спектр конструкцій, які має мова Clojure, дозволяють робити код лаконічнішим та зрозумілішим, ніж код записаний нефункційними мовами програмування. Це є великою перевагою при обробці складних структур даних, таких як синтаксичні дерева. На відміну від існуючих бібліотек мови Clojure, що надають деякі засоби синтаксичного розбору, які використовують генератор ANTLR у неефективному режимі інтерпретатора, розроблений засіб містить функції, які надають розробнику повноцінні можливості виконання синтаксичного розбору програмного коду за допомогою ANTLR. Множина цих функцій включає: функцію генерації класів лексичного та синтаксичного аналізаторів, яка має використовуватися в якості окремого етапу збірки проєкту; функцію реєстрації, мета якої полягає у аналізі класів згенерованих аналізаторів і створенні структур даних мови Clojure для реалізації інтерфейсу взаємодії між програмами, написаними мовою Clojure і згенерованими класами аналізаторів, написаними мовою Java; функцію синтаксичного розбору програмного коду, що використовує зареєстровані структури даних. Завдяки тому, що програмний код, написаний мовою Clojure, працює на основі Java Virtual Machine (JVM), в цих функціях вдалося використати класи бібліотек генератора ANTLR, що написані мовою Java.
Ключові слова: ANTLR, Clojure, синтаксичний аналіз, синтаксичне дерево, лексичний аналіз, токен, рефлексія

Kasianchuk D., Marchenko O. A parsing tool based on the ANTLR parser generator and the Clojure language. This article is devoted to the process of parsing program code. As a result of the analysis of a number of publications comparing existing parser generators, it has been determined that the ANTLR parser generator is popular among developers due to its simplicity, wide range of functions, as well as the quality and efficiency of the generated code. The article presents a tool for parsing program code based on the use of the ANTLR parser generator and the modern functional programming language Clojure, whose characteristics make it much easier to create tree-like data structures, as well as analyze and modify them, when compared to non-functional programming languages. In addition, the wide range of constructs available in Clojure makes it possible to make code more concise and understandable than code written in non-functional programming languages. This is a great advantage when processing complex data structures such as syntax trees. Unlike the existing Clojure libraries that provide some parsing tools that use the ANTLR generator in an inefficient interpreter mode, the developed tool contains functions that provide the developer with full-fledged capabilities to perform parsing of program code using ANTLR. The set of these functions includes: the function of generating classes of lexical and syntax analyzers, which should be used as a separate stage of project build; the registration function, the purpose of which is to analyze the classes of generated analyzers and create Clojure data structures to implement an interface between programs written in Clojure and generated classes of analyzers written in Java; the function of parsing program code using registered data structures. Due to the fact that the program code written in Clojure runs on the Java Virtual Machine (JVM), these functions were able to use the classes of the ANTLR generator libraries written in Java.
Keywords: ANTLR, Clojure, parsing, parse tree, lexical analysis, token, reflection

Постановка наукової проблеми

Основою засобів статичного аналізу програмного коду є спеціальне внутрішнє представлення цього коду, зазвичай, синтаксичне дерево, яке є результатом синтаксичного розбору рядка термінальних символів, що відповідає правилам формальної граматики. Така граматика може описувати, як природну мову (напр. українську або англійську), так і мову програмування або формат даних. Процес виділення лексем (або токенів) з початкового програмного коду, називають лексичним аналізом. Синтаксичні аналізатори – це програмні засоби, які на основі результату лексичного аналізу, перевіряють вхідний текст на відповідність синтаксису та будують синтаксичне дерево, що представляє вхідну програму деревоподібною структурою даних. Хоча певний аналіз вхідного коду програми можна зробити і в оригінальній текстовій формі, використання ієрархічної деревовидної структури даних для представлення цього коду є набагато ефективнішим і зручнішим,

© Касянчук Д.П., Марченко О.І.

ніж виконання перетворень безпосередньо початкового текстового представлення програмного коду.

З вищесказаного можна зробити висновок, що розробка нових ефективних і зручних для використання засобів обробки тексту на основі синтаксичних дерев є актуальною.

Аналіз досліджень

Широке використання синтаксичного аналізу в сфері програмного забезпечення породжує попит у створенні засобів, мета яких полягає у автоматичному створенні лексичних та синтаксичних аналізаторів. В таких засобах розробник описує синтаксис мови, якою написаний програмний код, що має бути розпізнаним, а в результаті отримує аналізатори для заданої граматики.

На сьогоднішній день, існує багато генераторів синтаксичних аналізаторів, які також називають ще компіляторами компіляторів (compiler-compiler), наприклад:

1) ANTLR [1];
2) Lex/Yacc [2];
3) LISA [3].

В роботі [4] дослідники здійснили якісне порівняння генераторів синтаксичних аналізаторів LISA , Lex/Yacc та ANTLR 3. Для виконання поставленої задачі вони визначили граматику предметно-орієнтованої мови Lavanda, метою якої є опис мішків для білизни, які компанія відправляє на прання.

Якісну оцінку цих трьох генераторів синтаксичних аналізаторів автори виконують на основі наступних критеріїв:

1) специфікація мови;
2) процес генерації синтаксичних аналізаторів;
3) згенерований код.

Дослідники стверджують, що за показниками зручності використання, наявності додаткових засобів та якості кінцевого продукту Lex/Yacc є найбіднішим і застарілим засобом, в той час як інші два генератори є достатньо ефективними, а також дуже схожими між собою.

Незважаючи на те, що ці засоби відрізняються такими своїми властивостями як стратегія синтаксичного аналізу або мова програмування згенерованих аналізаторів, переважна більшість розробників обирають або класичний засіб Lex/Yacc, або засіб ANTLR.

В роботі [5] здійснено емпіричне дослідження, метою якого є аналіз успішності студентів, що розроблюють компілятор, в залежності від використання ними засобів Lex/Yacc або ANTLR. Дослідники резюмують, що генератор ANTLR значно переважає Lex/Yacc. Студенти, які використовували ANTLR, показали значно кращі результати при виконанні як початкових завдань з розробки лексичного та синтаксичного аналізаторів, так і завдань з подальших модифікацій цих аналізаторів. Також ці студенти мали вищі показники за кількістю виконаних завдань, окремих оцінок за лабораторні роботи, а також підсумкових оцінок, порівняно зі студентами, що виконували розробку з використанням Lex/Yacc. За відгуками студентів ANTLR є більш інтуїтивним та простим засобом.

В якості мови програмування для розробки нового засобу для підтримки процесів лексичного і синтаксичного аналізу, що пропонується в цій статті, був обраний сучасний діалект мови Lisp – Clojure [6]. Проаналізуємо властивості цієї мови та її перспективність для роботи з ієрархічними структурами даних.

Оскільки мова Clojure є функційною мовою програмування, то її основними властивостями є, в першу чергу, загальні властивості функційних мов:

1) незмінність даних (варто відзначити, що в мові Clojure також існує багато засобів для реалізації змінного стану);

2) функції також є структурами даних нарівні з іншими структурами даних (тобто, вони можуть передаватися в інші функції, як аргументи, або повертатися з них, як результат);

3) функції вищого порядку;

4) чисті функції;

5) спискові структури даних.

Такі властивості дозволяють значно легше створювати деревоподібні структури даних, а також аналізувати та змінювати їх, якщо порівнювати з нефункційними мовами програмування. Крім того, широкий спектр конструкцій, які має мова Clojure, дозволяють робити код коротшим та зрозумілішим, що є великою перевагою при обробці складних структур даних, таких як синтаксичні дерева. Мова Clojure має широкий набір конструкцій для реалізації паралельних та розподілених обчислень. Ці конструкції можуть бути використанні у певних задачах статичного аналізу коду, які

© Касянчук Д.П., Марченко О.І.

можливо розпаралелити, з метою покращення продуктивності.

Програмний код, написаний мовою Clojure, працює на основі віртуальної машини Java (Java Virtual Machine – JVM), що дозволяє використовувати широкий спектр бібліотек, які існують у середовищі JVM [7]. Яскравим прикладом цього є використання класів бібліотек генератора ANTLR, а також класів лексичних та синтаксичних аналізаторів, згенерованих за його допомогою. Варто також зазначити, що ця мова програмування має систему макросів, що дозволяє додавати нові високорівневі синтаксичні конструкції, які орієнтовані на певні прикладні задачі.

Мова Clojure підтримує, так-званий, інтерактивний процес розробки, коли взаємодія з програмою відбувається вже під час її написання. Робота при такому процесі розробки здійснюється через Read-Eval-Print Loop (REPL), що дозволяє швидко визначати і тестувати різні функції без необхідності перекомпіляції всього вихідного коду проєкту.

Розглянемо дві бібліотеки мови Clojure для розробки лексичних і синтаксичних аналізаторів: clj-antlr [8] та instaparse [9].

Бібліотека clj-antlr місить функції для синтаксичного аналізу програмного коду з використанням генератора ANTLR у режимі інтерпретатора. Цей режим дозволяє спростити інтеграцію ANTLR у власний проєкт завдяки тому, що не потрібно виконувати етапи генерації лексичного і синтаксичного аналізаторів, а також компіляції згенерованого коду. В документації ANTLR зазначено, що такий режим зручно використовувати для невеликих завдань розбору програмного коду. Розглянемо переваги та недоліки режиму інтерпретатора.

Переваги:

1) цей режим дозволяє швидше виконувати розробку і тестування граматики (розробник може без затримки бачити результати змін, які він вносить у граматику, без необхідності генерування синтаксичного аналізатора і компілювання його коду);

2) простота інтеграції генератора ANTLR у проєкт.

Недоліки:

1) інтерпретатор працює не так швидко, як згенеровані засоби, особливо для вхідних програм великого розміру або для складних граматик;

2) деякі варіанти оптимізації вихідного коду можуть бути недоступними в цьому режимі.

Незважаючи на те, що розробники clj-antlr використовують ANTLR у неефективному режимі інтерпретатора для виконання задачі синтаксичного розбору, пропонований засіб, за їх словами, працює швидше, ніж засоби іншої бібліотеки instaparse, що демонструється ними на прикладі синтаксичного розбору JSON документу.

Розглянемо тепер більш детально бібліотеку instaparse.

Почнемо з головної відмінності, яка полягає у тому, що при використанні цієї бібліотеки для виконання процесу синтаксичного аналізу не потрібне використання ні генератора ANTLR, ні інших генераторів, а генерування синтаксичного аналізатора за заданою граматикою відбувається за допомогою функцій, написаних мовою Clojure. Розглянемо переваги та недоліки бібліотеки instaparse.

Переваги:

1) може бути використана для будь-якого типу граматики: ліво-рекурсивна, право-рекурсивна або двозначна;

2) для представлення вихідного дерева використовуються найпопулярніші формати мови Clojure, такі як вектор або хеш-таблиця;

3) має детальне звітування про помилки;

4) легке впровадження у проєкт.

Недоліки:

1) для великої кількості вхідних даних або у випадку складної граматики, процес синтаксичного розбору з використанням бібліотеки instaparse може виконуватися повільніше порівняно з синтаксичними аналізаторами, написаними вручну або згенерованими спеціалізованими засобами;

2) звіт про помилки може бути занадто перевантаженим, що ускладнює виявлення проблем у складних граматиках;

3) на відміну від ANTLR, не має напрацьованої бази готових граматик для великої кількості мов програмування [10].

Підсумовуючи вище описане, обидві бібліотеки мови Clojure містять функції для підтримки процесів лексичного і синтаксичного аналізу, які є ефективними тільки для нескладних граматик, а

© Касянчук Д.П., Марченко О.І.

також при аналізі невеликих програм. Переваги цих бібліотек полягають, швидше, у надані зручного інтерфейсу для розробника, і легкій інтеграції засобів синтаксичного розбору у проєкт, а не у швидкодії виконання процесів лексичного і синтаксичного аналізу.

Постановка завдання

Метою цієї статті є розробка нового ефективного засобу реалізації синтаксичного розбору вхідної програми, призначеного для розробників, які, з однієї сторони хочуть використовувати функційну мову програмування Clojure, а з другої сторони – популярний генератор синтаксичних аналізаторів ANTLR, який написаний мовою Java. Основною задачею розробки такого засобу є реалізація інтерфейсу взаємодії між програмами, написаними мовою Clojure, і засобами генератора ANTLR, написаними мовою Java.

Виклад основного матеріалу

Розроблений авторами засіб, який пропонується в цій статті, містить наступні функції, що написані мовою Clojure:

1) функція antlr4, призначення якої полягає у генерації коду лексичного і синтаксичного аналізаторів;

2) функція register-lang, призначення якої полягає у створені структур даних мови Clojure, які реалізують інтерфейс взаємодії між кодом програм, написаних мовою Clojure, і кодом згенерованих аналізаторів, написаними мовою Java;

3) функція parse-source, яка використовує зареєстровані структури даних і призначення якої полягає у виконанні власне синтаксичного розбору вхідного програмного коду.

Ці функції реалізують доступ до класів та методів генератора синтаксичних аналізаторів ANTLR. Розглянемо більш детально ці функції. В якості засобу для збірки та розповсюдження розробленого проєкту був обраний відомий засіб Leiningen [11]. Це один з найпоширеніших засобів для розробки проєктів мовою Clojure, який здобув свою популярність завдяки простоті у використанні, а також широкому спектру можливостей, які він надає. Конфігурація проєкту визначається у спеціальному файлі – project.clj, приклад якого зображено на рис.1.

```
(defproject <project-name> <version>
  :plugins [[lein-pprint "1.1.1"] ...]
  :source-paths ["src"]
  :java-source-paths ["src/java"]
  :prep-tasks ["javac" "compile"]
  :dependencies [[org.clojure/clojure "1.11.1"] ...]
  ...)
```

Рис. 1. Приклад project.clj файлу

Цей файл містить опис налаштувань, які необхідні для роботи функцій розробленого засобу:

1. :plugins – містить список плагінів. Кожен плагін містить функції, які мають доступ до налаштувань файлу project.clj. Ці функції можуть бути використані під час збірки проєкту, як додаткові етапи.

2. :source-paths – список, кожен елемент якого є відносним шляхом до директорії, яка містить файли з програмним кодом, що написаний мовою Clojure.

3. :java-source-paths – список, кожен елемент якого є відносним шляхом до директорії, яка містить файли з програмним кодом, що написаний мовою Java.

4. :prep-tasks – список етапів збірки проєкту, в який можна додати функції з п. 1. За замовчуванням, цей список містить наступні етапи:

1) javac – виконує компіляцію файлів, що знаходяться в java-source-paths (п. 3);

2) compile – виконує компіляцію файлів, що знаходяться в source-paths (п. 2).

5. :dependencies – список залежностей проєкту. Містить назви бібліотек та їх версії, які автоматично завантажуються з центрального репозиторію Maven або інших джерел.

Розглянемо детально функцію antlr4, з якої починається інтеграція генератора ANTLR у проєкт, що розроблюється. Мета цієї функції полягає у генерації коду лексичних і синтаксичних аналізаторів для певної вхідної мови. Функція antlr4 була розроблена і реалізована в рамках окремого плагіну засобу Leiningen. Цей плагін отримав ім'я lein-clj-parsing. Це було зроблено для того, щоб автоматизувати процес генерації коду аналізаторів шляхом додавання цієї функції у етапи збірки проєкту.

© Касянчук Д.П., Марченко О.І.

Параметри функції antlr4:

1) project – конфігурація project.clj файлу, яка має містити налаштування, необхідні для генерації коду аналізаторів; ці налаштування представлені хеш-таблицею, що пов'язана з ключовим словом :antlr, і має містити наступні ключі:

a) :grammars-dir – шлях до директорії, піддиректорії якої містять необхідні для генерації коду лексичного та синтаксичного аналізаторів файли, а ім'я кожної піддиректорії відповідає назві граматики;

b) :package – ім'я Java пакету, в якому будуть знаходитися згенеровані класи; варто зазначити, що це ім'я доповнюється іменем граматики;

c) :out-dir – шлях до директорії, яка міститиме згенеровані класи, що написані мовою Java;

2) grammars – (необов'язковий) список імен граматик (тобто відповідних директорій), для яких треба згенерувати аналізатори; за замовчуванням, аналізатори генеруються для всіх наявних граматик.

В список залежностей розробленого плагіну була додана бібліотека org.antlr/antlr4 генератора ANTLR, що містить клас Tool, який призначений для генерації коду лексичного і синтаксичного аналізаторів. Також, варто відзначити, що генерація аналізаторів для певної граматики відбувається тільки у випадку відсутніх згенерованих Java файлів або якщо відбулися зміни у директорії граматики. Перевірка таких змін базується на порівнянні часу останньої зміни директорії граматики і директорії зі згенерованими Java файлами для цієї граматики.

Для використання функції antlr4 у проєкті розробник має зробити наступні налаштування у файлі project.clj:

1) додати lein-clj-parsing у список плагінів (:plugins);
2) додати налаштування, що необхідні для генерації коду аналізаторів (:antlr);
3) додати функцію antlr4 у список етапів збірки проєкту перед javac етапом (:prep-tasks);
4) додати значення :out-dir налаштування у список Java директорій (:java-source-paths) для того, щоб згенеровані аналізатори компілювалися під час збірки проєкту.

Приклад project.clj файлу з описаними вище налаштуваннями, зображено на рис. 2.

```
(defproject <project-name> <version>
  :plugins [[lein-clj-parsing "0.1.0"] ...]
  :antlr {:grammars-dir "resources/grammars"
          :out-dir "src/java"
          :package "parser.generated"}
  :prep-tasks ["antlr4" "javac" "compile"]
  :java-source-paths ["src/java"]
  ...)
```

Рис. 2. Приклад project.clj файлу з налаштуваннями для генерації коду аналізаторів

Функції запропонованого авторами засобу, що описуються далі, були розроблені і реалізовані в рамках окремого проєкту засобу Leiningen, що отримав ім'я clj-parsing. Тобто, для їх використання, розробник має додати clj-parsing в список залежностей проєкту (:dependencies).

В список залежностей проєкту clj-parsing була також додана бібліотека org.antlr/antlr4-runtime генератора ANTLR, оскільки вона містить функції, що необхідні для створення і обробки синтаксичних дерев, отриманих за допомогою згенерованих аналізаторів.

Розглянемо основні функції проєкту clj-parsing.

Після виконання генерації, створені класи аналізаторів мають бути обов'язково зареєстровані за допомогою функції register-lang. Суть реєстрації полягає у аналізі класів згенерованих аналізаторів за допомогою Java рефлексії [12] і у створені структур даних мови Clojure, які реалізують інтерфейс взаємодії між кодом програм, написаних мовою Clojure, і кодом згенерованих аналізаторів, написаних мовою Java. Рефлексія допомагає досліджувати вміст класу, а також викликати його методи, незалежно від специфікатора доступу, який з ними використовується.

Опишемо структури даних мови Clojure, які будуть створені в результаті роботи функції register-lang (нагадаємо, що у функційних мовах функції також є структурами даних):

1. Функція make-lexer. Ця функція в якості параметра приймає потік символів вхідної програми, для якої буде виконуватися лексичний аналіз. Цей потік символів мусить мати

© Касянчук Д.П., Марченко О.І.

спеціальний тип CharStream з бібліотеки генератора ANTLR. Функція make-lexer повертає об'єкт лексичного аналізатора для потоку символів певної вхідної програми.

2. Функція make-parser. Ця функція через параметр приймає потік токенів, який є результатом роботи лексичного аналізатора і для якого буде виконаний синтаксичний аналіз. Цей потік має бути об'єктом класу, що реалізує інтерфейс TokenStream з бібліотеки генератора ANTLR. Функція make-parser повертає об'єкт синтаксичного аналізатора для конкретного потоку токенів перетвореної вхідної програми.

3. Хеш-таблиця rule-parse-fn. Ключем цієї таблиці є ім'я правила граматики, а значенням – функція, яка в якості параметра приймає об'єкт синтаксичного аналізатора і викликає його метод. Цей метод синтаксичного аналізатора повертає дерево, яке є результатом синтаксичного розбору, що починається з правила, з ім'ям якого ця функція пов'язана у хеш-таблиці.

Параметри функції register-lang:

1) lang-name – ім'я граматики, з яким будуть пов'язані створені структури даних;
2) lexer-class – згенерований клас лексичного аналізатора;
3) parser-class – згенерований клас синтаксичного аналізатора.

На рис. 3 зображено реалізацію функції реєстрації register-lang. В цій функції описані вище структури даних зберігаються у глобальній хеш-таблиці languages, в якій ключем є значення параметра lang-name (тобто, імені граматики), а значенням є вкладена хеш-таблиця з наступними ключами:

1) :make-lexer, що пов'язується з функцією make-lexer;
2) :make-parser, що пов'язується з функцією make-parser;
3) :rule-parse-fn, що пов'язується з хеш-таблицею rule-parse-fn.

```
(defonce languages (atom {}))

(defn register-lang [lang-name ^Class lexer-class ^Class parser-class]
  (let [lang-name (name lang-name)]
    (check-lang-input lang-name lexer-class parser-class)
    (let [lexer-cnstr (.getDeclaredConstructor lexer-class (class-array [CharStream]))
          parser-cnstr (.getDeclaredConstructor parser-class (class-array [TokenStream]))
          rule-names (-> (.getField parser-class "ruleNames") (.get nil))]
      (->>
       (for [rule-name rule-names
             :let [method (.getMethod parser-class rule-name (class-array []))]]
         [rule-name (fn [parser] (.invoke method parser (object-array [])))])
       (into {})
       (hash-map :make-lexer (fn [stream] (.newInstance lexer-cnstr (object-array [stream])))
                 :make-parser (fn [tokens] (.newInstance parser-cnstr (object-array [tokens])))
                 :rule-parse-fn)
       (swap! languages assoc lang-name)))))
```

Рис. 3. Програмна реалізація функції реєстрації register-lang

Оскільки код вхідної програми може зберігатися у різних структурах даних (файл, потік, рядок тощо), а функція створення об'єкта лексичного аналізатора приймає в якості параметра тільки потік символів (див. register-lang), то для формування потоку символів з різних структур даних було вирішено створити протокол мови Clojure (протокол – це набір іменованих методів з сигнатурами без конкретної реалізації). У цьому протоколі, що отримав ідентифікатор CharStreamBuilder, був реалізований тільки один метод make-stream, який власне і виконує таке формування, приймаючи в якості параметра структуру даних з кодом вхідної програми і формує із символів цього коду потік символів. Існуючі реалізації методу make-stream використовують методи класу CharStreams з бібліотеки генератора ANTLR, призначення яких полягає у формуванні потоку символів із символів вхідної програми, яка може знаходитися в наступних структурах даних (рис. 4): 1) файл; 2) потік символів; 3) потік байтів; 4) рядок.

© Касянчук Д.П., Марченко О.І.

```
(defprotocol CharStreamBuilder
  (^CharStream make-stream [source] "Create CharStream from `source`"))

(extend-protocol CharStreamBuilder
  File
  (make-stream [source] (CharStreams/fromFileName (.getPath source)))
  Path
  (make-stream [source] (CharStreams/fromPath source))
  Reader
  (make-stream [source] (CharStreams/fromReader source))
  InputStream
  (make-stream [source] (CharStreams/fromStream source))
  String
  (make-stream [source] (CharStreams/fromString source)))
```

Рис. 4. Протокол переписування програмного коду із заданої початкової структури даних у потік символів

При необхідності, метод make-stream може бути легко розширений реалізаціями для інших структур даних.

Функція parse-source створює об'єкти лексичного і синтаксичного аналізаторів і виконує власне синтаксичний розбір програмного коду.

Параметри функції parse-source:

1) lang-name – ім'я зареєстрованої граматики;
2) rule – ім'я правила граматики, з якого треба починати розбір програмного коду;
3) source – структура даних, що містить програмний код.

На рис. 5 схематично зображено алгоритм роботи цієї функції, а також використання зареєстрованих структур даних і методу створеного протоколу CharStreamBuilder.

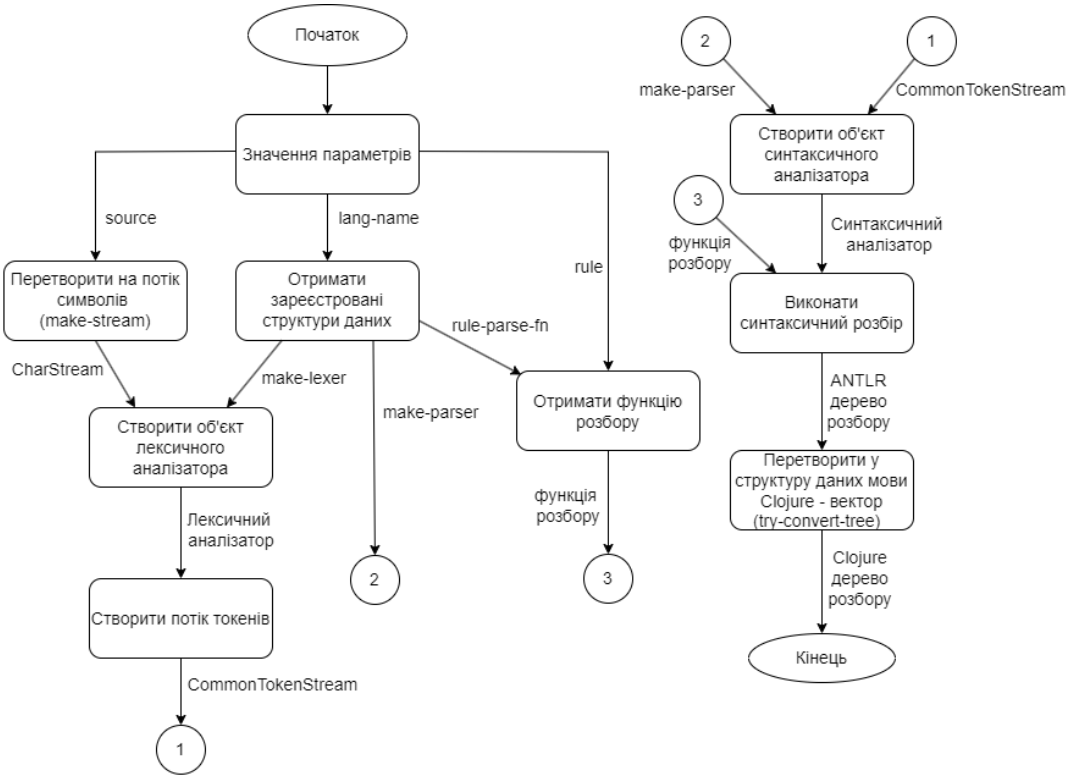


Рис. 5. Алгоритм роботи функції розбору parse-source

На рис. 6 зображено програмну реалізацію функції parse-source. В якості класу для створення потоку токенів був обраний клас CommonTokenStream з бібліотеки генератора ANTLR. Цей клас використовується у більшості випадків для отримання потоку токенів. Для створення об'єкта класу CommonTokenStream, в його конструктор необхідно передати об'єкт лексичного аналізатора в якості

© Касянчук Д.П., Марченко О.І.

аргументу.

```
(defn parse-source [lang-name rule source]
  (if-let [{:keys [make-lexer make-parser rule-parse-fn]} (@languages lang-name)]
    (let [lexer (-> source make-stream make-lexer)
          tokens (CommonTokenStream. lexer)
          parser (make-parser tokens)]
      (if-let [parse-fn (rule-parse-fn rule)]
        (try-convert-tree (parse-fn parser) parser)
        (throw (IllegalArgumentException.
                (str "Invalid rule name - " rule "."))))))
    (throw (IllegalArgumentException.
            (str "Language '" lang-name "' isn't registered.")))))
```

Рис. 6. Функція parse-source

Функція try-convert-tree перетворює синтаксичне дерево, що представлене об'єктом класу ParserRuleContext з бібліотеки генератора ANTLR, у синтаксичне дерево, що представлене структурою даних «вектор», яка є у мові Clojure.

Параметри функції try-convert-tree:

1) node – початкова вершина (корінь) дерева ANTLR;
2) parser – об'єкт синтаксичного аналізатора.

Перетворення відбувається тільки у випадку, якщо:

- не знайдено жодної синтаксичної помилки;
- процес розбору використав всі наявні токени; ця перевірка була додана через те, що синтаксичний аналізатор, згенерований за допомогою ANTLR, вважає розбір успішним, незважаючи на токени, які залишилися нерозібраними; ця перевірка виконується шляхом перевірки наступного (у потоці токенів) токену на рівність спеціальному значенню, що представляє кінець тексту.

Опишемо більш детально алгоритм перетворення дерева у форматі ANTLR у формат типу «вектор» мови Clojure.

У випадку, якщо вершина є термінальною (TerminalNode), то об'єкт токена цієї вершини перетворюється у хеш-таблицю, що містить значення типу і тексту цього токена, які пов'язуються з ключовими словами :type і :text відповідно.

Наприклад, термінальна вершина, яка зображена у вигляді UML діаграми об'єктів, на рис. 7, перетворюється у хеш-таблицю, зображену на рис. 8.

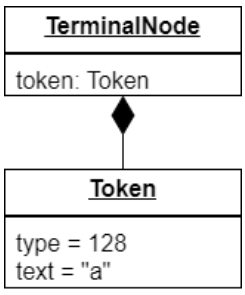


Рис. 7. Термінальна вершина

```
{:type 128
 :text "a"}
```

Рис. 8. Перетворена термінальна вершина

У випадку, якщо вершина є нетермінальною (ParserRuleContext), то створюється вектор. Першим його елементом є ім'я правила граматики, яке отримується за допомогою функції, зображеної на рис. 9.

```
(defn- rule-name [^Parser parser ^ParserRuleContext node]
  (->> node .getRuleIndex (get (.getRuleNames parser)) keyword))
```

Рис. 9. Функція отримання імені правила граматики з нетермінальної вершини

Решта елементів вектора – це результати перетворення дочірніх вершин цієї нетермінальної вершини. Наприклад, нетермінальна вершина з дочірніми вершинами, які зображені на рис. 10,

© Касянчук Д.П., Марченко О.І.

перетворюється у вектор, зображений на рис. 11.

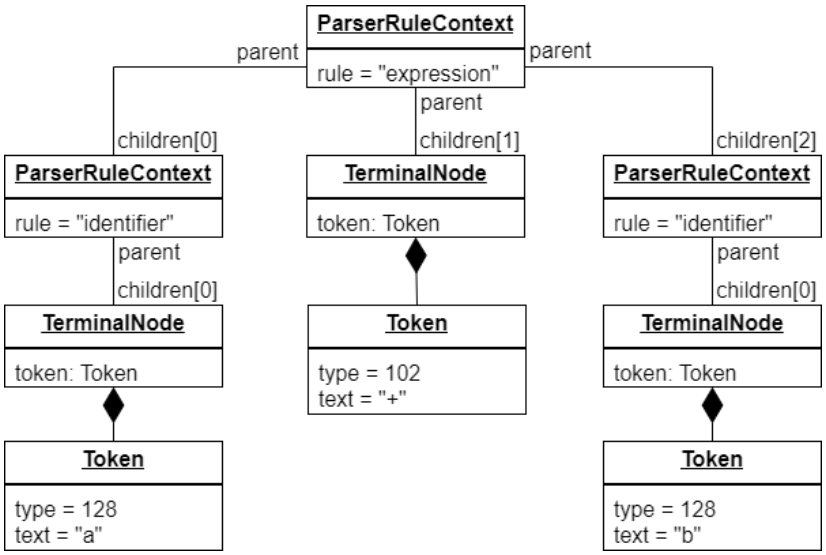


Рис. 10. Нетермінальна вершина

```
[:expression
 [:identifier {:type 128, :text "a"}]
 {:type 102, :text "+"}
 [:identifier {:type 128, :text "b"}]]
```

Рис. 11. Перетворена нетермінальна вершина

На рис. 12 зображено програмну реалізацію функції try-convert-tree, яка виконує описане перетворення.

```
(defn try-convert-tree [node ^Parser parser]
  (letfn [(helper [node]
            (condp instance? node
              TerminalNode
              (let [token (.getSymbol node)]
                {:type (.getType token)
                 :text (.getText token)})
              ParserRuleContext
              (into [(rule-name parser node)] (map helper) (get-children node))
              :else (throw (IllegalArgumentException.
                            (str "Unrecognized tree type - " (class node) ".")))))]
    (let [errors-count (.getNumberOfSyntaxErrors parser)]
      (cond
        (pos? errors-count)
        (throw (IllegalStateException.
                (str "Parsing has failed. Number of errors - " errors-count ".")))
        (= (.. parser getInputStream (LA 1)) (Token/EOF))
        (helper node)
        :else (throw
               (IllegalStateException.
                (str "Rule '" (name (rule-name parser node)) "' hasn't consumed all input."))))))))
```

Рис. 12. Функція try-convert-tree

На рис. 13 зображено виклик функції parse-source для випадку виконання синтаксичного розбору програмного коду, написаного мовою Java. Розбір цього програмного коду починається з правила граматики, яке має ім'я statement. Також, припускаємо, що граматика java, була попередньо зареєстрована за допомогою функції register-lang, як показано на рис. 14.

```
(parse-source :java :statement "a = b;")
```

Рис. 13. Виклик функції parse-source

© Касянчук Д.П., Марченко О.І.

```
(register-lang :java
               parser.generated.java.JavaLexer
               parser.generated.java.JavaParser)
```

Рис. 14. Реєстрація граматики java за допомогою функції register-lang

Розглянемо приклад перетворення синтаксичного дерева з формату бібліотеки генератора ANTLR у формат типу «вектор» мови Clojure. На рис. 15 показана UML діаграма об'єктів синтаксичного дерева, яке було отримане за допомогою синтаксичного аналізатора, що згенерований за допомогою ANTLR, а на рис. 16 показане те ж саме дерево після його перетворення функцією try-convert-tree у вектор мови Clojure.

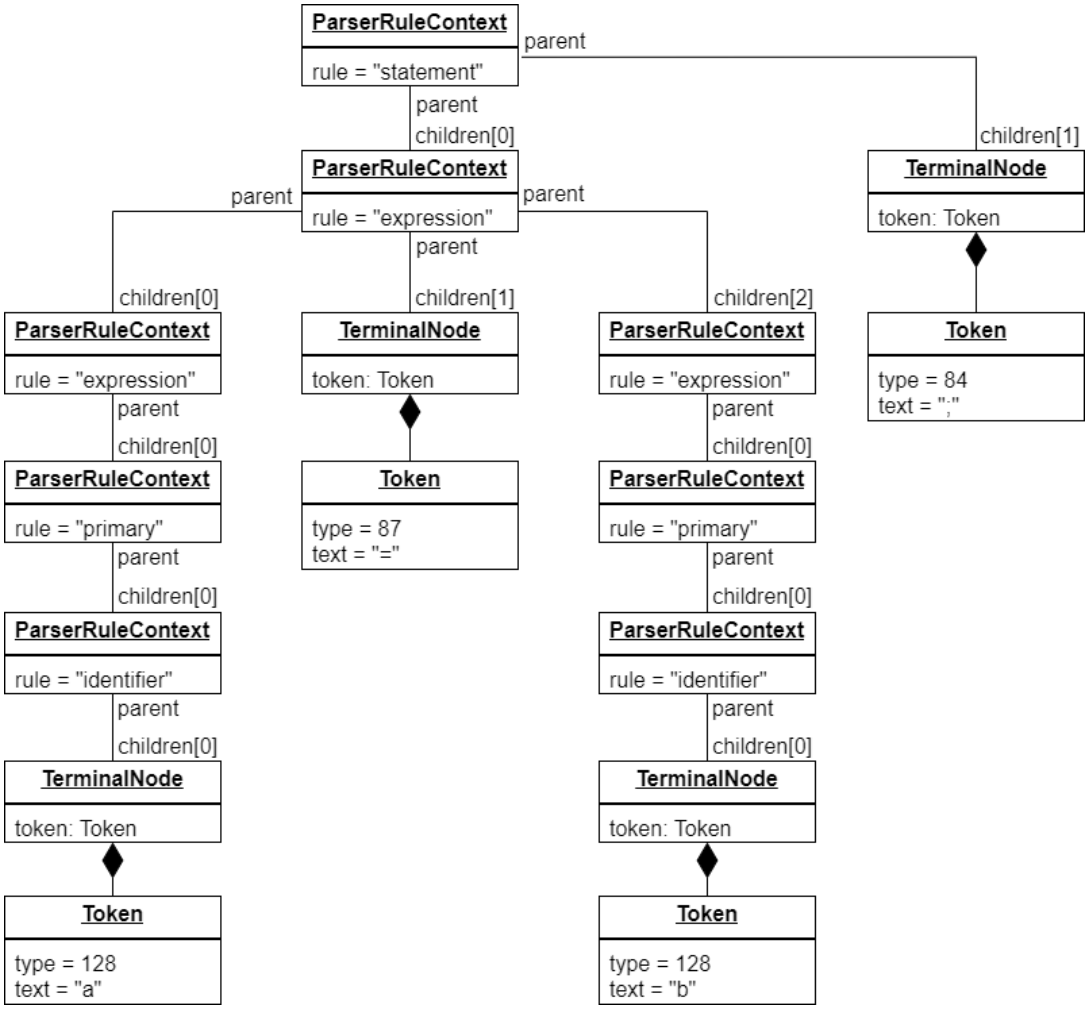


Рис. 15. Синтаксичне дерево, отримане за допомогою синтаксичного аналізатора бібліотеки ANTLR

```
[:statement
 [:expression
  [:expression [:primary [:identifier {:type 128, :text "a"}]]]
  {:type 87, :text "="}
  [:expression [:primary [:identifier {:type 128, :text "b"}]]]]
 {:type 84, :text ";"}]
```

Рис. 16. Синтаксичне дерево з рис. 15 після перетворення функцією try-convert-tree у структуру даних типу «вектор»

Висновки.

В пропонованій статті представлено засіб синтаксичного розбору програмного коду, який базується на використанні популярного генератора синтаксичних аналізаторів ANTLR, а також сучасної функційної мови програмування Clojure, конструкції якої дозволяють виконувати роботу з деревоподібними структурами даних простіше, ніж конструкції нефункційних мов програмування.

© Касянчук Д.П., Марченко О.І.

На відміну від існуючих бібліотек мови Clojure, розроблений засіб не використовує генератор ANTLR у режимі інтерпретатора, оскільки цей режим є ефективним тільки у обмеженій кількості випадків. Завдяки тому, що програмний код, написаний мовою Clojure, так само як і код, написаний мовою Java, виконується у віртуальній машині JVM (Java Virtual Machine), у розробленому засобі було використано класи бібліотек генератора ANTLR, що написані мовою Java. Це дозволило вирішити задачу реалізації доступу до генератора синтаксичних аналізаторів ANTLR з програмного коду, написаного мовою Clojure.

У якості напрямків подальших досліджень можна запропонувати такі:

1) виконання аналізу методів зіставлення з шаблоном з метою додавання такої можливості до запропонованого засобу;

2) розробка методу конструювання нових дерев синтаксичного розбору на основі вже існуючих дерев.

Список бібліографічного опису

1. Генератор синтаксичних аналізаторів ANTLR. Режим доступу: https://www.antlr.org. Дата звернення: 14.09.2024.
2. Lex & Yacc. Режим доступу: https://epaperpress.com/lexandyacc/. Дата звернення: 14.09.2024.
3. M. Mernik, M. Lenic, E. Avdicausevic, V. Zumer, «Compiler/interpreter generator system LISA», the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 2000. DOI: https://doi.org/10.1109/hicss.2000.927021. Дата звернення: 14.09.2024.
4. D. da Cruz, M.J.V. Pereira, M. Béron, R. Fonseca, P.R. Henriques, «Comparing Generators for Language-based Tools», the 1st Conf. on Compiler Related Technologies and Applications, CoRTA'07, Portugal, 2007. Режим доступу: http://surl.li/hnkvgz. Дата звернення: 14.09.2024.
5. F. Ortin, J. Quiroga, O. Rodriguez-Prieto, M. Garcia, «An empirical evaluation of Lex/Yacc and ANTLR parser generation tools», Plos one, 17(3), e0264326, 2022. DOI: https://doi.org/10.1371/journal.pone.0264326. Дата звернення: 14.09.2024.
6. Мова програмування Clojure. Режим доступу: https://clojure.org/. Дата звернення: 14.09.2024.
7. A. Sarimbekov, A. Podzimek, L. Bulej, Y. Zheng, N. Ricci, W. Binder, «Characteristics of Dynamic JVM Languages», the 7th ACM Workshop on Virtual Machines and Intermediate Languages, 2013. DOI: https://doi.org/10.1145/2542142.2542144. Дата звернення: 14.09.2024.
8. K. Kingsbury, «Бібліотека мови Clojure clj-antlr», Github. Режим доступу: https://github.com/aphyr/clj-antlr. Дата звернення: 14.09.2024.
9. M. Engelberg, «Бібліотека мови Clojure instaparse», Github. Режим доступу: https://github.com/Engelberg/instaparse. Дата звернення: 14.09.2024.
10. Antlr Project, «Набір граматик для генератора синтаксичних аналізаторів ANTLR v4», Github. Режим доступу: https://github.com/antlr/grammars-v4. Дата звернення: 14.09.2024.
11. P. Hagelberg, «Засіб Leiningen». Режим доступу: https://leiningen.org/. Дата звернення: 14.09.2024р.
12. G. McCluskey, «Using Java Reflection», Технічна стаття, Oracle, 1998. Режим доступу: http://surl.li/bwgrjz. Дата звернення: 14.09.2024.

References

1. ANTLR parser generator. Access mode: https://www.antlr.org. Accessed on: 14.09.2024.
2. Lex & Yacc. Access mode: https://epaperpress.com/lexandyacc/. Accessed on: 14.09.2024.
3. M. Mernik, M. Lenic, E. Avdicausevic, V. Zumer, «Compiler/interpreter generator system LISA», the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 2000. DOI: https://doi.org/10.1109/hicss.2000.927021. Accessed on: 14.09.2024.
4. D. da Cruz, M.J.V. Pereira, M. Béron, R. Fonseca, P.R. Henriques, «Comparing Generators for Language-based Tools», the 1st Conf. on Compiler Related Technologies and Applications, CoRTA'07, Portugal, 2007. Access mode: http://surl.li/hnkvgz. Accessed on: 14.09.2024.
5. F. Ortin, J. Quiroga, O. Rodriguez-Prieto, M. Garcia, «An empirical evaluation of Lex/Yacc and ANTLR parser generation tools», Plos one, 17(3), e0264326, 2022. DOI: https://doi.org/10.1371/journal.pone.0264326. Accessed on: 14.09.2024.
6. The Clojure programming language. Access mode: https://clojure.org/. Accessed on: 14.09.2024.
7. A. Sarimbekov, A. Podzimek, L. Bulej, Y. Zheng, N. Ricci, W. Binder, «Characteristics of Dynamic JVM Languages», the 7th ACM Workshop on Virtual Machines and Intermediate Languages, 2013. DOI: https://doi.org/10.1145/2542142.2542144. Accessed on: 14.09.2024.
8. K. Kingsbury, «The Clojure language library clj-antlr», Github. Access mode: https://github.com/aphyr/clj-antlr. Accessed on: 14.09.2024.
9. M. Engelberg, «The Clojure language library instaparse», Github. Access mode: https://github.com/Engelberg/instaparse. Accessed on: 14.09.2024.
10. Antlr Project, «A set of grammars for the ANTLR v4 parser generator», Github. Access mode: https://github.com/antlr/grammars-v4. Accessed on: 14.09.2024.
11. P. Hagelberg, «The Leiningen tool». Access mode: https://leiningen.org/. Accessed on: 14.09.2024р.
12. G. McCluskey, «Using Java Reflection», Technical Article, Oracle, 1998. Access mode: http://surl.li/bwgrjz. Accessed on: 14.09.2024.

© Касянчук Д.П., Марченко О.І.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-23
УДК 004.94
Кащенко Дмитро Олександрович, аспірант
https://orcid.org/0000-0002-3365-7186
Черкаський Державний Технологічний Університет м. Черкаси, Україна.

ТЕОРІЯ ІНФОРМАЦІЇ – СУБ'ЄКТИВНІ СИТЕМИ

Кащенко Д.О. Теорія інформації – суб'єктивні системи. Представлено аналіз теорії інформації (Claude E. Shannon, Warren Weaver) та запропонована концепція суб'єктивних систем, як варіанту автоматизації складних інформаційних динамічних систем. В статті розглядаються основні погляди на поняття «інформація» та методи визначення поняття «інформація», та залежність цих поглядів та методів від світогляду, досвіду та сфери діяльності дослідників. Пропонується інший погляд на поняття «інформація» з урахуванням ефекту спостерігача. Представлений варіант погляду на поняття «інформація» ґрунтується на біонічному підході виникнення самої інформації. І використовується не тільки для тлумачення процесів в каналах зв'язку, але і робить спробу відповісти на питання – "Навіщо потрібне саме існування інформації". І як варіант відповіді пропонується, що сама інформація, як роз'яснення, може виникнути та існувати тільки для конкретного класу систем. Подібні системи мають визначену динамічну складність і використовують інформацію для балансування себе та стабільного існування себе в навколишньому середовищі. Використовуючи це середовище для своїх потреб, як енергетичних так і функціональних. В даній моделі складних динамічних систем пропонуються використовувати не тільки методи кібернетики для самоорганізації, але і більш універсальні інші біонічні методи. Також пропонується зробити акцент, не на формуванні штучного середовища для системи. А надати складній динамічній системі повний набір механізмів для взаємодії з реальним середовищем з можливістю розвитку цих механізмів в неперервному циклі. І як висновок з цих припущень випливає залежність кількості інформації від складності подібних систем.
Ключові слова: Теорія інформації, суб'єктивні системи, моделі інформаційних систем, складні динамічні інформаційні системи, структура складних інформаційних систем

Kashchenko D. Information Theory – Subjective Systems. This paper presents an analysis of information theory (Claude E. Shannon, Warren Weaver) and proposes the concept of subjective systems as a variant for automating complex informational dynamic systems. The article discusses the main views on the concept of "information" and the methods for defining the concept of "information," as well as how these views and methods depend on the worldview, experience, and field of activity of the researchers. An alternative view on the concept of "information" is proposed, taking into account the observer effect. This perspective on the concept of "information" is based on the bionic approach to the emergence of information itself. It is used not only to interpret processes in communication channels but also attempts to answer the question – "Why does information itself need to exist?". As a possible answer, it is suggested that information, as an explanation, can arise and exist only for a specific class of systems. Such systems have a defined dynamic complexity and use information to balance themselves and maintain stable existence in their environment, utilizing this environment for their energy and functional needs. In this model of complex dynamic systems, it is proposed to use not only cybernetic methods for self-organization but also more universal bionic methods. Additionally, the focus should not be on creating an artificial environment for the system but rather on providing the complex dynamic system with a complete set of mechanisms for interacting with the real environment, allowing these mechanisms to develop in a continuous cycle. And as a conclusion from these assumptions follows the dependence of the amount of information on the complexity of such systems.
Keywords: Information Theory, subjective systems, models of information systems, complex dynamic information systems, structure of complex information systems

Як відомо теорії інформації, як і самому поняттю інформації присвячено багато статей і робіт, в яких викладені ті або інші погляди на ці поняття та їх значення в тих або інших галузях. Тому перш ніж розпочати опис даної концепції, давайте розглянемо методи визначення та філософські аспекти, які вже існують. А також зробимо окремі припущення про це, не заперечуючи інші теорії та погляди, а лише акцентуючи увагу на необхідності подібного погляду в даній моделі, яка допоможе вивести поняття автоматизації на якісно новий рівень.

Інформація (від лат. informatio – роз'яснення) – відомості, які передають усним, писемним та іншими шляхами за допомогою умовних сигналів і технічних засобів. Інформацію розглядають у контексті її змісту, структур. організації, динаміки (процесів створення, передавання, сприйняття, використання, зберігання тощо). [1]. Від серед. 20 ст. поняття «Інформація» стало одним із основних серед заг.-наук категорій, що пов'язано з розвитком інформатики, кібернетики, *теорії інформації*, інформаційних ресурсів та інше, а також із розповсюдженням комп'ютер. техніки. Обов'язкова передумови функціонування інформаційного простору – наявність джерела, користувачів та засобів передачі інформації (каналів зв'язку) (Рис.1). Носіями інформація є сигнали – матеріальні процеси, які користувачі інформації сприймають як знаки із певною семантикою. При цьому важливо не лише зафіксувати ці знаки, а й відповідно їх інтерпретувати, оскільки для різних користувачів одна й та ж інформація може мати різне значення (інколи навіть протилежне). Значна

© Кащенко Д.О.

залежність впливу інформації на користувача від її сприйняття дозволила основоположнику кібернетики та теорії штучного інтелекту Н. Вінеру прирівняти інформацію до таких понять, як «матерія» та «енергія». Розрізняють три форми інформаційних процесів: біологічна інформація всередині живих організмів і між ними, соціальна інформація у людському суспільстві, машинна інформація (всередині та між машинами). Здатність живих організмів і їхніх спільнот до самоорганізації, постійні пристосування до змін довкілля зумовили виокремлення вагомої ролі процесів інформаційної взаємодії з оточуючим світом і формування інформаційних моделей зовнішнього світу в живих організмах (Ґ. Бейтсон).[1]

Передачу ознак від клітин до клітини і від організму до організму також стали розглядати як передачу інформації. Що привело до створення біокібернетики, наукового напряму, пов'язаного з проникненням ідей, методів і технічних засобів кібернетики в біології. Зародження і розвиток біокібернетики пов'язані з еволюцією представлення зворотного зв'язку в живій системі і спробами моделювання особливостей її будови і функціонування (П. К. Анохін, Н. А. Бернштейн і т.д.). Ефективність математичного та системного підходу з дослідження живого показали багато робіт в області загальної біології (Дж. Холдейн, Е. С. Бауер, Р. Фішер, І. І. Шмальгаузен і др.). Процес «кібернетизації» біології здійснюється як у теоретичному, так і в прикладному застосуванні. Основна теоретична задача біокібернетики — вивчення загальних закономірностей управління, а також зберігання, переробки та передачі інформації в живих системах. Центральне положення поняття інформації в кібернетиці пояснюється тим, що кібернетика (з одного боку обмежує, а з іншого боку уточнює інтуїтивне представлення інформації) вивчає машини і живі організми з точки зору їх здатності сприймати певну інформацію, зберігаючи її в «пам'яті», передавати по «каналам зв'язку» і переробляти її в «сигнали», спрямовуючи їх діяльність у відповідний бік.

Кожен організм — це система, здатна до саморозвитку й управління як внутрішніми взаємозв'язками між органами й функціями, так й співвідношеннями з факторами середовища. Прагнучі зрозуміти природу живого, багато дослідників часто намагалися відшукати в організмі те, що можна було досліджувати окремо. Мета біокібернетики — вивчення організму з урахуванням основних взаємозв'язків починаючи з клітинного, тканинного, органічного рівня і закінчуючи організмом в цілому. Жива система характеризується не тільки обміном речовин і енергії, але і обміном інформації. Біокібернетика розглядає складні біологічні системи у взаємодії з середовищем саме з точки зору теорії інформації. Одним із найважливіших методів біокібернетики є моделювання структур і закономірностей поведінки живої системи; воно включає конструювання штучних систем, що відтворюючих певні сторони діяльності організмів, їх внутрішні зв'язки та відносини.

Особливість організму як складної динамічної системи — єдність централізованого та автономного управління. Саморегуляція, характерна для всіх рівнів управління живою системою, забезпечується автономними механізмами, поки не виникають такі стани, які вимагають втручання центральних механізмів керування. По-друге, була запропонована кількісна міра інформації **[5][6],** що привело до створення інформаційної теорії. Інформаційна теорія, математичні дисципліни, що досліджують процеси зберігання, перетворення та передачі інформації. В основі теорії інформації лежить визначений спосіб вимірювання кількості інформації, що міститься в яких-небудь даних («повідомленнях»). Теорія інформації виходить з уявлення про те, що повідомлення, призначені для збереження в пам'яті пристроїв, або для передачі по каналу зв'язку (Рис.1) і невідомі заздалегідь з повною визначеністю. Заздалегідь відомо лише безліч варіантів, з яких можуть бути обрані ці повідомлення, і в кращому випадку — те, як часто вибирається це або інше з цих повідомлень (тобто ймовірність повідомлень). В теорії інформації показано, що «невизначеність», з якою стикаються в подібній обстановці, допускає кількісне вираз і що саме цей вираз (а не конкретна природа самих повідомлень) визначає можливість їх зберігання і передачі. У якості такої «міри невизначеності» в інформації приймається число подвійних знаків (П. 0 або 1), необхідних для фіксації (запису) довільного повідомлення даного джерела. Іншими словами кількість інформації - це міра зменшення невизначеності якоїсь ситуації. Інформаційна теорія отримує оптимальні, в сенсі швидкості та надійності, способи передачі інформації, встановлюючи теоретичні межі досягнутої якості. Як видно з попереднього визначення, інформація носить суттєво статистичний характер, і тому значуща частина її математичних методів береться з теорії ймовірностей.

© Кащенко Д.О.

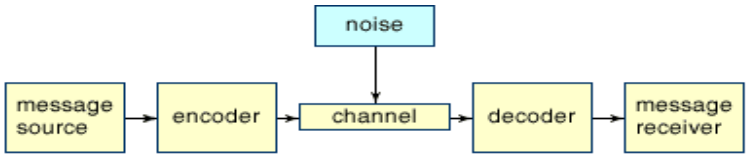


Рис. 1. Принципова схема загальної системи зв'язку(каналу)

Більш загальний, чим раніше, підхід до поняття інформації, а також поява точної кількості інформації пробудили величезний інтерес до вивчення властивостей інформації. З початку 1950-х рр. намагаються використати поняття інформації для пояснення та опису найбільш різноманітних явищ і процесів.

Дослідження проблем, пов'язаних з науковим поняттям інформації, іде в трьох основних напрямках. Перше з них міститься в розробці математичного апарату, що відображає основні властивості інформації.

Друге напрямок полягає в теоретичній розробці різних аспектів інформації на базі вже наявних математичних засобів, у дослідженні різних властивостей інформації (об'єктивності, достовірності, повноти, актуальності).

Наприклад, вже з моменту створення теорії інформації виникла складна проблема вимірювання цінності, корисності інформації з точки зору її використання. У більшості робіт за теоретичною інформацією ця властивість не враховується. Однак важливість цієї властивості безсумнівна. У кількісній теорії, виданої в 1960 році А. А. Харкевичем, цінність інформації визначається як приріст ймовірності досягнення конкретної мети в результаті використання даної інформації.[7] Близькі по сенсу роботи пов'язані з спробами дати суворе математичне визначення кількості семантичної (тобто змістовної) інформації (Р. Карнап та ін.).

Третій напрямок пов'язано з використанням інформаційних методів у лінгвістиці, біології, психології, соціології, педагогіці та ін. В лінгвістиці, наприклад, проводилось вимірювання інформативної ємності мов. Після статистичної обробки великого числа текстів, а також складання довжини перекладів одного і того ж тексту на різних мовах і багато-чисельних експериментів по вгадуванню букв тексту з'ясувалося, що при рівномірному навантаженні речових одиниць інформації тексту можна було б скоротитися в 4-5 разів. Так був з цієї точки зору встановлений факт надмірності природних мов і досить точно виміряна її величина, що знаходиться в цих мовах приблизно на одному рівні. У нейрофізіології інформаційні методи допомогли краще знайти механізм дії основного закону психофізики - закону Вебера - Фехнера, який стверджує, що досягнення інтенсивність відчуття чогось лінійно залежить від логарифму інтенсивності подразника. Саме така залежність повинна мати місце в тому випадку, якщо нервові волокна, що передають сигнали від акцепторів до мозку, володіють властивостями, присутніми ідеалізованим каналом зв'язку, фігуруючим в теорії інформації. Значущу роль інформаційний підхід зіграв в генетиці і молекулярної біології, дозволивши, зокрема, глибше усвідомити значення молекулу РНК як носіїв інформації.

Таке різноманітне використання поняття інформації спонукало надати йому загальнонаукове значення. Засновниками такого загального підходу до поняття інформації були англійський нейрофізіолог У. Р. Ешбі і французький фізик Л. Бриллюэн. Вони досліджували питання спільності поняття ентропії в теорії інформації та термодинаміки, трактуючи інформація як негативну ентропію (негентропію)**.** Бриллюен і його послідовники стали видавати інформаційні процеси під кутом зору другого початку термодинаміки, розглядаючи передачу інформації якоїсь системи як удосконалення цієї системи, що веде до зменшення її ентропії. У деяких філософських роботах висувається теза про те, що інформація є однією з основних універсальних властивостей матерії. Позитивна сторона цього підходу полягає в тому, що він зв'язує інформацію про поняття з поняттям відображення.

Філософія визначення поняття інформації відрізняються надзвичайною суперечливістю. На одному боці — безмежний панінфермізм, тобто затвердження, що весь світ і всі його властивості створені з інформації. На другому — заперечення існування інформації як реальності [2]. Фактично, можна припустити, що інформація являє собою суб'єктивну реальність. Що і визначило назву концепції даних систем.

Як можна бачити вище, під час аналізу поняття інформації дослідники зіштовхуються з декількома проблемами. Одна із них це ефект спостерігача. Так кожний спостерігач, або користувач

© Кащенко Д.О.

відштовхувався від свого досвіду та світогляду, своєї мети. Що і змінює значення поняття інформації (об'єкта) відносно спостерігача, як суб'єкта. Так і зараз читаючи цю статтю, об'єкт, ми використовуємо аналіз та категоризацію відносно себе як суб'єкта. Так виникає ефект спостерігача, тому що я (або ми) як суб'єкт(и) який(і) сприймають об'єктивну дійсність, крім цього сприймаємо інші суб'єкти як об'єкти відносно себе. Так і мене (нас) інші суб'єкти сприймають як об'єкт відносно себе. Тому пропонується використовувати при аналізі подібних систем (які мають функцію роз'яснення) користуватися принципом відносності, але не тільки відносно користувача (джерела), але й сприйняття та взаємодії. З чого виходить, що система яка володіє інформацією, тобто роз'ясненням повинна мати визначену складність для формування роз'яснення відносно себе, з повним циклом взаємодії з середовищем. Як саме утворюється межа об'єкт - суб'єкт, яку саме складність повинна мати подібна система і яку структуру і є мета подальших досліджень. Але, перш ніж розпочати опис концепції суб'єктивної системи розглянемо, базові терміни та їх значення, які використовуються на поточний момент.

Суб'єкт (від лат. subjectus - лежащий внизу, що знаходиться в основі), носій предметно-практичної діяльності та пізнання, джерело активності, спрямованої на об'єкт. Насправді термін має багато визначень в різних мовах, але в даному випадку використовуємо саме таке визначення. [3]

Об'єкт (від лат. objectum – предмет) предмет, явище, процес або їх окремі сторони, існуючі в реальній дійсності, на які спрямована діяльність, думка, почуття і т.д. Об'єкт протистоїть суб'єкту в його предметно-практичній і пізнавальної діяльності. [3]

Система (від грец. systema - ціле, складене із частей; з'єднання), множина взаємопов'язаних елементів, що утворюють єдине ціле, взаємодіють із середовищем та між собою, і мають мету. В сучасному науково-технічному знанні розробка проблематики, пов'язаної з дослідженням і конструюванням систем загального різного роду, проводиться в рамках системного підходу, теорії систем, різних спеціальних теорій систем, в кібернетиці, системотехніці, системному аналізі та інше. Виділяють статичні і динамічні системи. Для статичної системи її стан з поточним часом залишається постійним (наприклад, газ в обмеженому об'ємі - в стані рівноваги). Динамічна система змінює свій стан у часі (наприклад, живий організм). За характером взаємозв'язку системи і середовища системи діють на закриті - замкнуті (з них не виділяється речовина, відбувається лише обмін енергіями) і відкриті - незамкнуті (постійно проходять введення і виведення не тільки енергії, але і речовини).

У системному підході в різних технічних науках виділяють складні системи (складовий об'єкт, частину якого можна розглядати як систему, закономірно об'єднані в єдине ціле відповідно до визначених принципів або пов'язаних між собою заданими відносинами.). Типовим прикладом може бути самонастроювальні системи, в яких пристосування до випадково змінених умов забезпечується автоматичним зміненням налаштувань параметрів, або шляхом автоматичного пошуку оптимальної налаштувань.[3]

Вивчення відносин між елементами і підсистемами, визначення ролі і місця кожної підсистеми в загальному процесі функціонування системи складають предмет структурного аналізу складних систем.

Структура, це сукупність відносно стійких зв'язків об'єкта, що забезпечує його цілісність та тотожність самому собі, тобто збереження основних властивостей при різних зовнішніх і внутрішніх змінних. Методи структурного аналізу дозволяють виділити в складних системах набори підсистем, що знаходяться в заданих відносинах, і представити складні системи як сукупність об'єктів з досліджуваними структурами. Крім того, ці методи застосовуються для оцінки так званих структурних характеристик, які в кількості відображають ті або інші окремі властивості схеми сполучення елементів складних систем. Загальну оцінку функціональних і структурних характеристик доповнюють якісними дослідженнями, проведеними за допомогою методів так званої якісної теорії складних систем. Сюди в першу чергу входять дослідження стійкості системи, в тому числі побудовані параметри стійкості областей у складних системах, виділення типових режимів функціонування системи тощо[4].

Принцип суб'єктивних систем полягає в тому, що система (суб'єкт) є окремою частиною об'єктивної дійсності, але не роздільно зв'язаний з нею, здатна(ий) сприймати і оцінювати навколишню об'єктивну дійсність на основі і відносно своєї зовнішньої і внутрішньої структури і розвиватися відносно поточного стану цих структур і навколишньої дійсності, використовуючи ці

© Кащенко Д.О.

внутрішні і зовнішні стани цих структур і оточуючу дійсність. Таким чином, суб'єктивні системи повинні володіти, як мінімум тими функціями і механізмами, які дозволяють це реалізувати (Рис.2).

Що дозволяє інтерпретувати появу інформації (роз'яснення) у наступній трактовці. Усі процеси та явища поза системою діляться на ті, які впливають на неї або не впливають на неї, що визначає появу даних для системи, якщо ці процеси впливають на неї. У свою чергу дані, які вже є частиною зовнішньої структури системи, діляться цією системою на позитивні або негативні, відносно впливу на баланс внутрішньої і зовнішньої структури системи, що визначає появу роз'яснення, тобто інформації для системи. І якщо проявляються відносно постійні позитивні процеси, то відбувається балансування без особливих змін структури системи. Якщо це негативні процеси, то відбувається зміна структури системи для стабілізації балансування її зовнішніх і внутрішніх структур щодо навколишньої об'єктивної дійсності, якщо це можливо. Тобто керування змінами в структурі пропонованої системи відбувається на одному рівні, але в різних режимах функціонування. Але також може бути залучений вищий рівень для структурних змін. Що і дозволяє досягти автономності системи зі збереженням конфігураційної гнучкості.

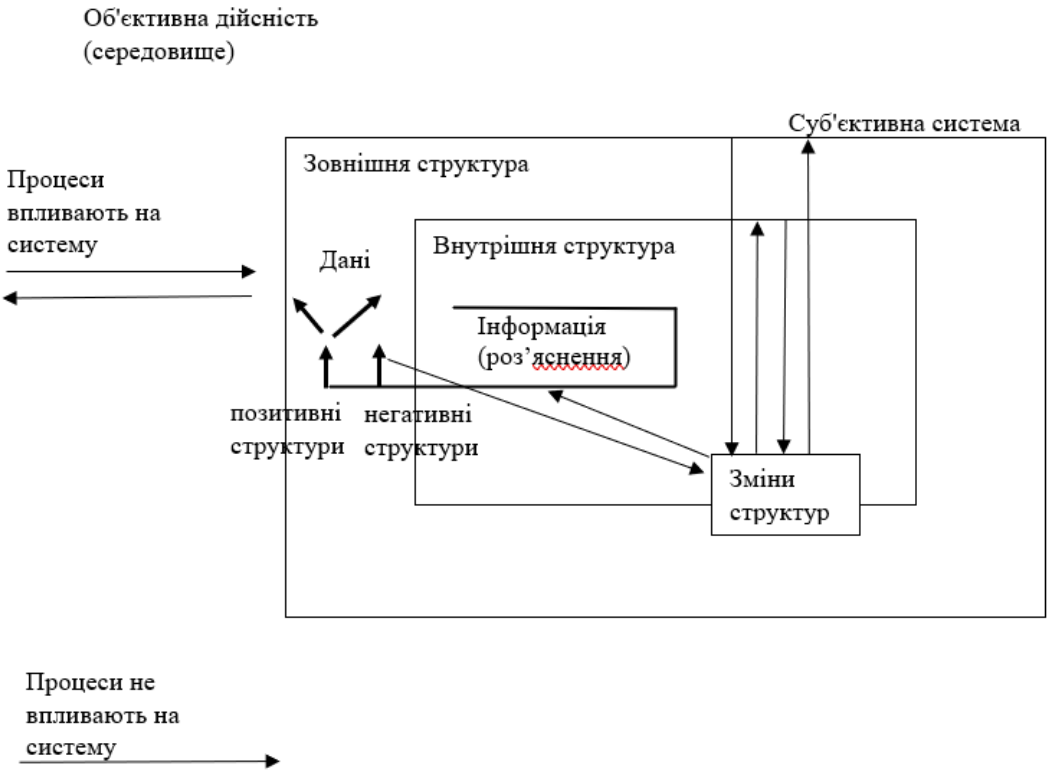


Рис.2. – Структура суб'єктивних систем.

Концепція суб'єктивних систем використовує методи біоніки і є одним із способів автоматизації складних динамічних систем. Також пропонується зробити акцент, не на формуванні штучного середовища для системи. А надати складній динамічній системі повний набір механізмів для взаємодії з реальним середовищем з можливістю розвитку цих механізмів в неперервному циклі.

Середовище в даній моделі є не тільки каналом зв'язку з шумом, але і джерелом енергії і матерії. Роз'яснення виникає під час порівняння співвідношення між процесами які сприймаються системою і вже потім в складі зовнішньої структури системи, є джерелом "даних", та процесами всіх внутрішніх структур системи та їх балансу с зовнішніми, що також є джерелом "даних" та використовуються самою системою для визначення позитивних та негативних процесів для себе. Роз'яснення цих процесів і є однією з функцій суб'єктивної системи, як єдиного організму. Інформація визначається, як здатність даної системи задавати значення як внутрішнім, так і зовнішнім процесам та структурам і розвиватися відносно цього.

© Кащенко Д.О.

Висновок: Таким чином кількість і цінність інформації знаходиться в прямій залежності від такої системи. Розуміння цієї залежності в перспективі дає можливість побудови автоматизованих складних динамічних систем з повним циклом взаємодії з середовищем.

Список бібліографічного опису

1. Інформація / В. Ф. Погорілко, Т. Я. Любива / Енциклопедія Сучасної України [Електронний ресурс] / Редкол. : І. М. Дзюба, А. І. Жуковський, М. Г. Железняк [та ін.] ; НАН України, НТШ. – К. : Інститут енциклопедичних досліджень НАН України, 2011. – Режим доступу : https://esu.com.ua/article-12485
2. Ніколаєв Є. Б. Теорія інформаційної економіки: зміст і основні суперечності: дис. на здобуття наук. ступеня канд. екон. наук : спец. 08.00.01 «Економічна теорія та історія економічної думки» / Ніколаєв Євген Борисович ; Київський нац. екон. ун-т ім. Вадима Гетьмана. – К., 2011. – 176 с.
3. Словник іншомовних слів / за ред. О. С. Мельничука. — К. : Головна редакція Української Радянської Енциклопедії, 1974.
4. Перегудов Ф. І., Тарасенко Ф. П. Введення в системний аналіз. — М.: Вища школа, 1989
5. Claude E. Shannon, Warren Weaver. The Mathematical Theory of Communication. Univ of Illinois Press, 1963
6. Теорія інформації та теорія алгоритмів. А.Н. Колмогоров. Наука, 1987. 874 с. Три підходи до визначення поняття "кількість інформації". А.Н. Колмогоров
7. Харкевич А.А. Про цінність інформації / Проблеми кібернетики. - 1960. - Вип. 4.

References

1. Information / V. F. Pohorilko, T. Ya. Lyubiva / Encyclopedia of Modern Ukraine [Electronic resource] / Editorial board: I. M. Dzyuba, A. I. Zhukovsky, M. H. Zheleznyak [and others]; NAS of Ukraine, NTSH. – K.: Institute of Encyclopedic Research of the NAS of Ukraine, 2011. – Access mode: https://esu.com.ua/article-12485
2. Nikolaiev Y. B. "Theory of Information Economy: Content and Main Contradictions." Dissertation for the degree of Candidate of Economic Sciences, specialty 08.00.01 "Economic Theory and History of Economic Thought." Kyiv National Economic University named after Vadym Hetman, 2011. 176 pages..
3. Dictionary of Foreign Words / ed. O. S. Melnychuk. — K.: Main Editorial Office of the Ukrainian Soviet Encyclopedia of the Academy of Sciences of the Ukrainian SSR, 1974.
4. Perehudov F. I., Tarasenko F. P. Introduction to System Analysis. — M.: Higher School, 1989
5. Claude E. Shannon, Warren Weaver. The Mathematical Theory of Communication. Univ of Illinois Press, 1963
6. Information Theory and Theory of Algorithms. A.N. Kolmogorov. Science, 1987. 874 pages. Three approaches to defining the concept of "amount of information." A.N. Kolmogorov
7. Kharkevich A.A. On the value of information / Problems of cybernetics. – 1960. - issue. 4.

© Кащенко Д.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-24
УДК: 004.932
Коваль Ігор Михайлович, аспірант
https://orcid.org/0009–0001–2083–1747
Суринович Олена Миколаївна, к.т.н., доцент
https://orcid.org/0000–0002–9300–0039
Луцький національний технічний університет, м. Луцьк, Україна

ПОРІВНЯЛЬНЕ ДОСЛІДЖЕННЯ МЕТОДІВ ПОПЕРЕДНЬОЇ ОБРОБКИ ТЕКСТУ В ORANGE DATA MINING ТА KNIME

Коваль І.М., Суринович О.М. Порівняльне дослідження методів попередньої обробки тексту в Orange Data Mining та KNIME. У роботі досліджується порівняння результатів попередньої обробки текстових даних у програмних системах Knime Analytics та Orange Data Mining. Представлено детальний опис методології дослідження, включаючи інструменти попередньої обробки текстових даних, налаштування та створення моделей відповідно до можливостей кожної з програм. Отримані результати аналізуються з використанням інструментів візуалізації, при цьому використовуються різні формати представлення результату. Переваги та недоліки кожного інструменту виявляються у рекомендаціях щодо застосування тієї чи іншої програмної системи у різних умовах. Результати попередньої обробки текстових даних в обох програмних системах показали, що дані були очищені від шуму, небажаних слів та синтаксичних елементів, що дозволило виділити ключові теми і тенденції із тестового матеріалу.
Ключові слова: Text Preprocess, Clustering, Orange Data Mining, Data mining, Knime Analytics, Word Cloud, попередня обробка тексту, Text to Vector

Koval I., Surynovych O. A comparative study of text preprocessing methods in Orange Data Mining and KNIME. The paper examines the comparison of the results of preprocessing of text data in the Knime Analytics and Orange Data Mining software systems. A detailed description of the research methodology is presented, including tools for preprocessing textual data, setting up and creating models according to the capabilities of each of the programs. The obtained results are analyzed using visualization tools, while different formats of the result presentation are used. The advantages and disadvantages of each tool are revealed in the recommendations for the use of this or that software system in different conditions. The results of pre-processing of text data in both software systems showed that the data were cleaned of noise, unwanted words and syntactic elements, which made it possible to highlight key themes and trends from the test material.
Keywords: Text Preprocess, Clustering, Orange Data Mining, Data mining, Knime Analytics, Word Cloud, text pre–processing, Text to Vector

Постановка проблеми. У сучасну епоху цифрових технологій аналітика даних стала одним із основних напрямків інформаційних технологій. Текстові дані становлять чи не найбільший відсоток усіх даних у світі, і кожного дня користувачі стикаються з аномально великою кількістю таких даних. Завдання кластеризації текстових даних полягає в тому, щоб впорядкувати хаотичний та непідготовлений матеріал таким чином, аби можна було отримати з нього корисну інформацію. Кластеризація та класифікація дозволяють виявити приховані патерни та структури у великих обсягах даних. Проте ефективність кластеризації значною мірою залежить від якості попередньої обробки тексту, яка може включати в себе токенізацію (Tokenization), видалення стоп-слів (Stop-Word Filter), стемінг (Stemming), лематизацію (Lemmatization) та інші методи. Попередня обробка тексту (Text Preprocessing) є обов'язковою складовою для підвищення точності кластеризації, зокрема при використанні таких програмних систем, як Knime та Orange Data Mining.

Сучасні інформаційні системи потребують автоматизованої обробки, особливо коли йдеться про великі обсяги текстових даних. Інструменти Orange та Knime надають можливість автоматизації Text Preprocessing, що, в свою чергу, зменшує час та зусилля, необхідні для роботи з текстовими даними, та підвищує продуктивність роботи. У кожному конкретному датасеті вибір оптимальних методів Text Preprocessing є критично важливим для отримання точних кластерів після застосування алгоритму. Недоліки або неправильна обробка тексту гарантовано призведуть до спотворених кластерів, які не дадуть жодної корисної інформації. Дослідження, присвячені аналізу ефективності різних методів та моделей попередньої обробки текстових даних, мають високу практичну цінність для науковців та аналітиків. Отримані результати з цієї статті можуть бути використані для розробки рекомендацій та побудови моделей, що дозволять успішно провести Text Preprocessing у запропонованих програмних системах.

Формулювання мети дослідження. Метою даного дослідження є аналіз ефективності різних методів попередньої обробки текстових даних (Text Preprocessing) та формування висновків щодо доцільності використання тієї чи іншої програмної системи у різних ситуаціях.

© Коваль І.М., Суринович О.М.

Аналіз останніх досліджень і публікацій. У сучасній науковій літературі активно досліджується питання попередньої обробки текстових даних як обов'язкового етапу перед кластеризацією. Багато дослідників підкреслюють важливість якісного Text Preprocessing для підвищення точності кластеризації та зниження рівня шуму в текстових даних. Зокрема у дослідженнях [1] акцентується увага на використанні алгоритмів лематизації для підвищення релевантності кластерів. У науковій праці [2] авторами відводиться велике дослідження впливу тих чи інших маніпуляцій з текстом на вихідні дані.

Orange Data Mining – це потужний інструмент для аналізу даних і машинного навчання, розроблений у вигляді відкритого програмного забезпечення [3]. Він надає користувачам можливість створювати візуальні робочі процеси для аналізу даних за допомогою графічного інтерфейсу користувача. Інструмент підтримує різні методи аналізу даних, включаючи кластеризацію, класифікацію, регресію та візуалізацію. Orange дозволяє інтегрувати різні методи попередньої обробки тексту, такі як токенізація, видалення стоп-слів, стемінг та лематизація, що робить його хорошим для текстового майнінгу та аналізу. Традиційно вважається, що Orange краще підходить для початківців, через зручність та інтуїтивність інтерфейсу.

KNIME (Konstanz Information Miner) – це інтегрована платформа для аналізу даних, звітності та інтеграції, яка підтримує весь процес аналізу даних через візуальне програмування [4]. KNIME дозволяє користувачам створювати робочі процеси для підготовки даних, аналізу та візуалізації без необхідності писати код. Платформа підтримує широкий спектр методів машинного навчання та статистичного аналізу, а також інтеграцію з іншими інструментами та мовами, такими як Python та R. Традиційно вважається, що Knime більше підходить для досвідчених користувачів.

Попередня обробка тексту Orange Data Mining представлена єдиним віджетом [5] який включає у себе усі можливі процедури для підвищення вихідного результату.

Схема зв'язку між віджетом Corpus (Відповідає за завантаження вхідних тексових даних) та віджетом Preprocess Text (відповідає за попередню обробку тексту) (рис. 1).

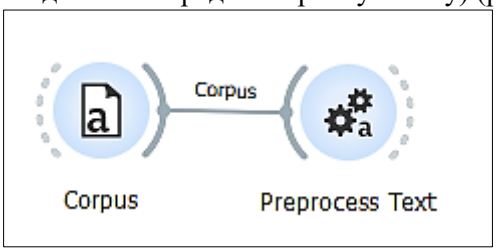


Рис 1. Віджети Orange

Функціонал (рис.2), який надає програмна система Orange Data Mining для попередньої обробки текстових даних. Представлені секції:

– Transformation (трансформація):
- Lowercase – перетворює всі символи тексту в нижній регістр;
- Remove accents – видаляє діакритичні знаки з тексту;
- Parse html – видаляє HTML–теги з тексту;
- Remove urls – видаляє URL–адреси з тексту;

– Tokenization (токенізація):
- Word Punctuation – токенізує текст, враховуючи пунктуацію;
- Whitespace – токенізує текст за пробілами;
- Sentence – токенізує текст за реченнями;
- Regexp – токенізує текст за допомогою регулярних виразів (наприклад, \w+ для слів);
- Tweet – токенізує текст, оптимізуючи для аналізу твітів;

– Filtering (фільтрація):
- Stopwords – видаляє стоп–слова. Можна вибрати мову і список стоп–слів;
- Lexicon – фільтрує текст за допомогою певного лексикону;
- Numbers – включає або виключає числа з тексту;
- Regexp – використовує регулярні вирази для фільтрації тексту;
- Document frequency – фільтрує токени за їх частотою в документі;
- Most frequent tokens – вибирає певну кількість найчастіших точені;
- POS tags – вибирає токени за їх частинами мови;

– POS Tagger (позначення частин мови):

© Коваль І.М., Суринович О.М.

- Averaged Perceptron Tagger – використовує алгоритм середнього перцептрону для позначення частин мови;
- Treebank POS Tagger (MaxEnt) – використовує Treebank POS Tagger з MaxEnt.

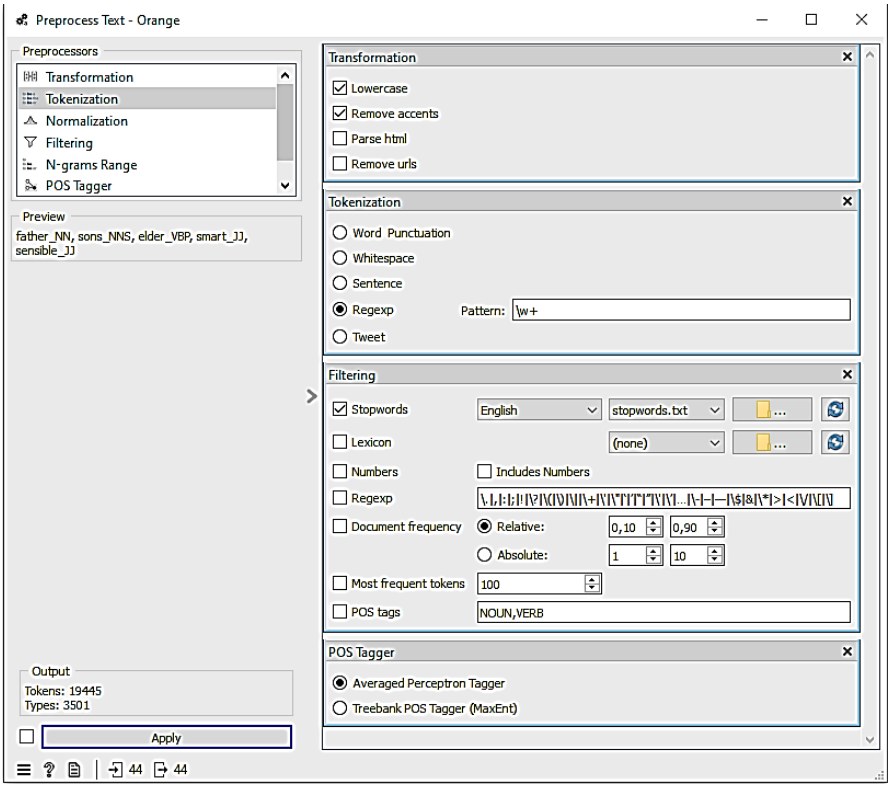


Рис 2. Інтерфейс віджету Preprocess Text

Попередня обробка тексту Knime Analytics представлена нодами [6] кожна з яких виконує відповідну функцію.

За допомогою нод, представлених на рис. 3, можна виконати основні текстові перетворення для очищення та підготовки до подальшої кластеризації. В залежності від бажаних перетворень програма пропонує відповідний функціонал:

– Case Converter – перетворює текст у верхній або нижній регістр;
– Diacritic Remover – видаляє діакритичні знаки з тексту;
– Dictionary Filter – фільтрує текст за допомогою словника;
– Dictionary Replacer – замінює слова в тексті на основі словника;
– Dictionary Replacer (File–based) – замінює слова в тексті, використовуючи словник із файлу;
– Hyphenator – розбиває слова на склади за допомогою дефісів;
– Kuhlen Stemmer – виконує стемінг (зведення слів до кореневої форми);
– Modifiable Term Filter – фільтрує змінні терміни;
– N Chars Filter – видаляє слова, довжина яких менше певної кількості символів;
– Number Filter – видаляє числа з тексту;
– Porter Stemmer – виконує стемінг за допомогою алгоритму Портера;
– Punctuation Erasure – видаляє розділові знаки з тексту;
– RegEx Filter – фільтрує текст за допомогою регулярних виразів;
– Replacer – замінює текст на основі заданих правил або регулярних виразів;
– Snowball Stemmer – виконує стемінг, використовуючи алгоритм Snowball;
– Stanford Lemmatizer – виконує лематизацію за допомогою Stanford NLP;
– Stop Word Filter – видаляє стоп–слова з тексту;
– Tag Filter – фільтрує текст за певними тегами;
– Tag Stripper – видаляє теги з тексту).

© Коваль І.М., Суринович О.М.

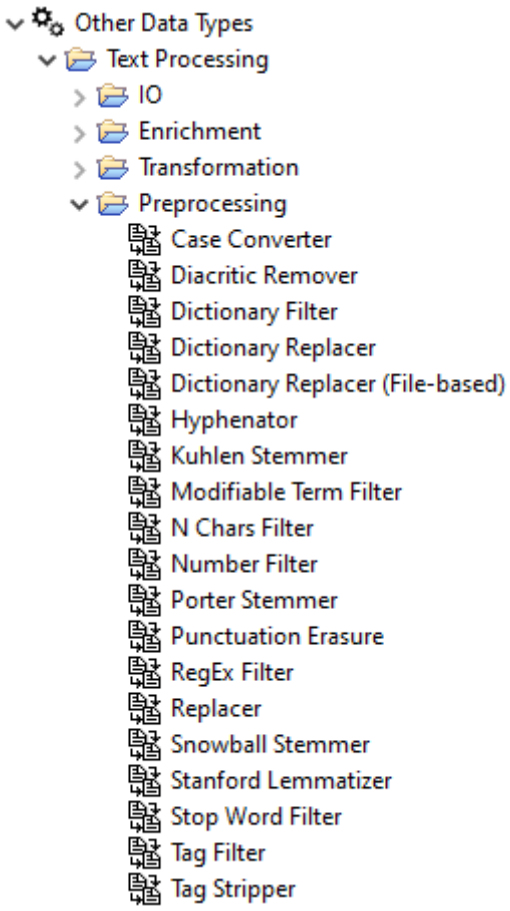


Рис. 3 Ноди гілки Preprocessing

Виклад основного матеаріалу. У ході експеременту використано 2500 заголовків новинних статтей. Для чистоти експерименту використано лише взаємовідповідний функціонал обох програм.

Опишемо створення та тестування моделі попреденьої обробки текстових даних у Orange Data Mining. Модель попередньої обробки текстових даних з використанням віджетів Text Mining відображено на рис.4.

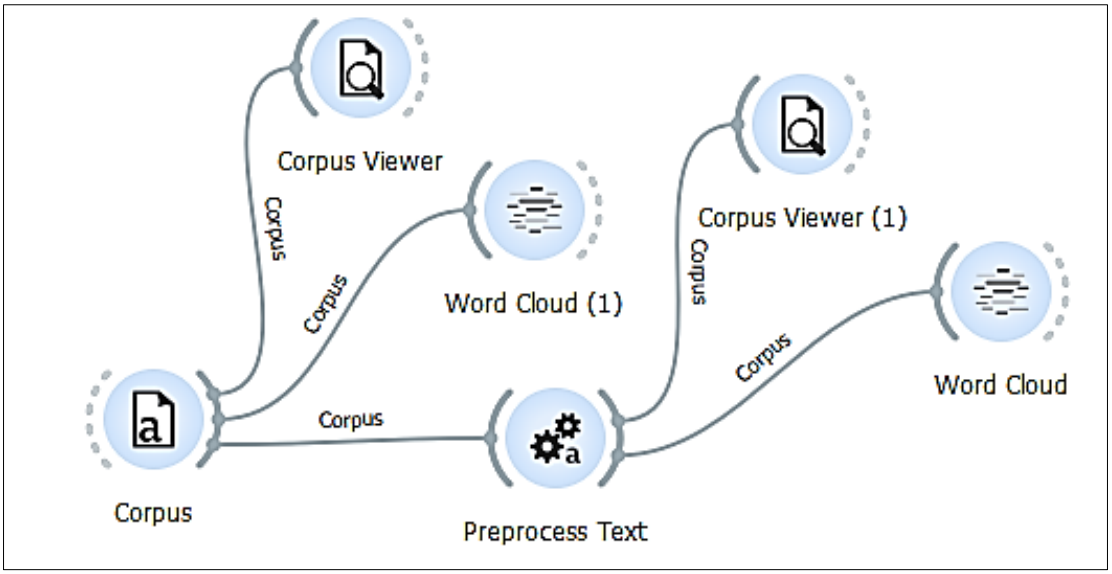


Рис. 4. Готова модель Orange Data Mining

В ході експеременту було виявлено певну проблему (рис. 5) у роботі віджету Corpus Viewer, а саме віджет Corpus Viewer (1) не відображає актуальні зміни які були застосовані інструментом Preprocess Text.

© Коваль І.М., Суринович О.М.

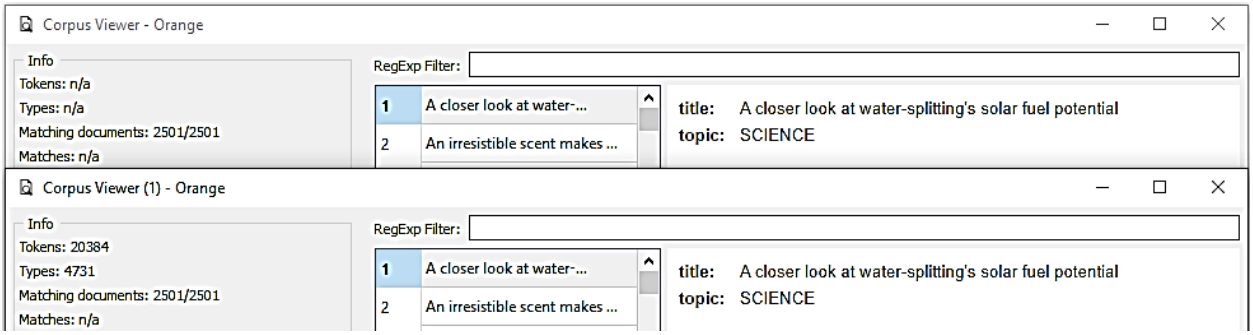


Рис. 5 Віджети Corpus Viewer

Для демонстрації результатів роботи Preprocess Text було прийняте рішення використати віджет Word Cloud, що являє собою хмару слів. Результати Word Cloud до застосування віджету Preprocess Text відображено на рис.6.

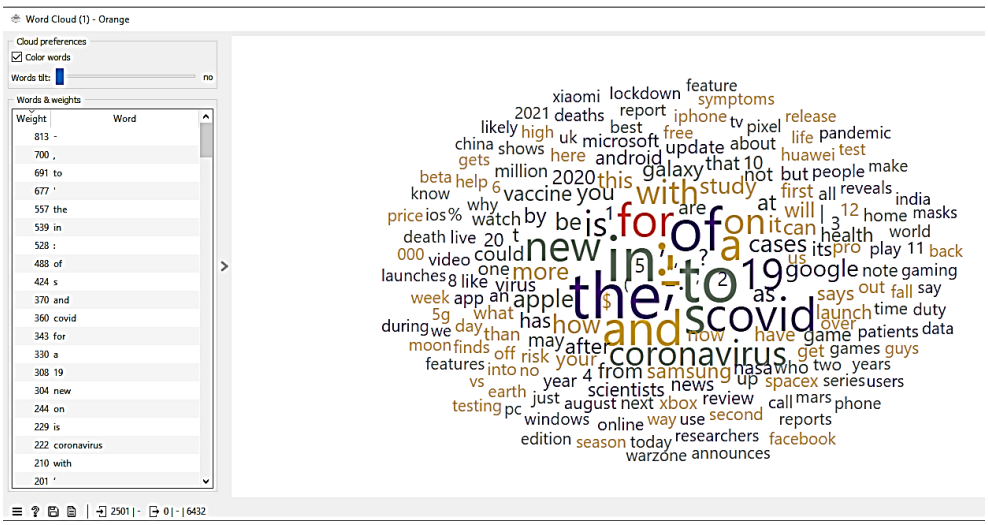


Рис. 6. Візуалізація Word Cloud

Результати Word Cloud після застосування віджету Preprocess Text відображено на рис.7.

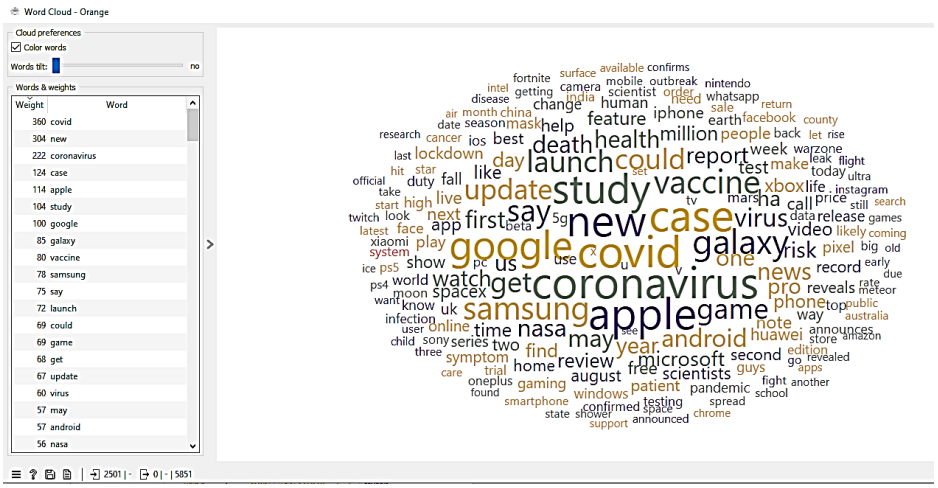


Рис. 7. Візуалізація Word Cloud після застосування віджету

Слід зазначити, що Orange Data Mining не вистачає віджетів які можна було би використовувати для перегляду проміжного результату після застосування Preprocess Text, адже Corpus Viewer не відображає зміни, а Word Cloud дає лише поверхневе уявлення та не показує результат роботи наприклад Pos Tagger. Для того, щоб побачити результат роботи Pos Tagger –

© Коваль І.М., Суринович О.М.

доводиться додатково представляти текстові дані у вигляді векторів. Принагідно варто додати, що на кінцевий результат кластеризації це не впливає.

Застосувавши інструменти Tokenization (Regexp), Normalization (WordNet Lematizer), Transformation (Lowercase, Remove accents), Filtering (Stop-word, Numbers), Pos Tagger – вдалось отримати якісний, відфільтрований результат. З даними (рис. 6) можна здійснити успішну кластеризацію текстових даних.

Опишемо створення та тестування моделі попередньої обробки текстових даних у Knime Analytics. Кінцева модель попередньої обробки текстових даних з використанням nodes (нод) у Knime (рис.8).

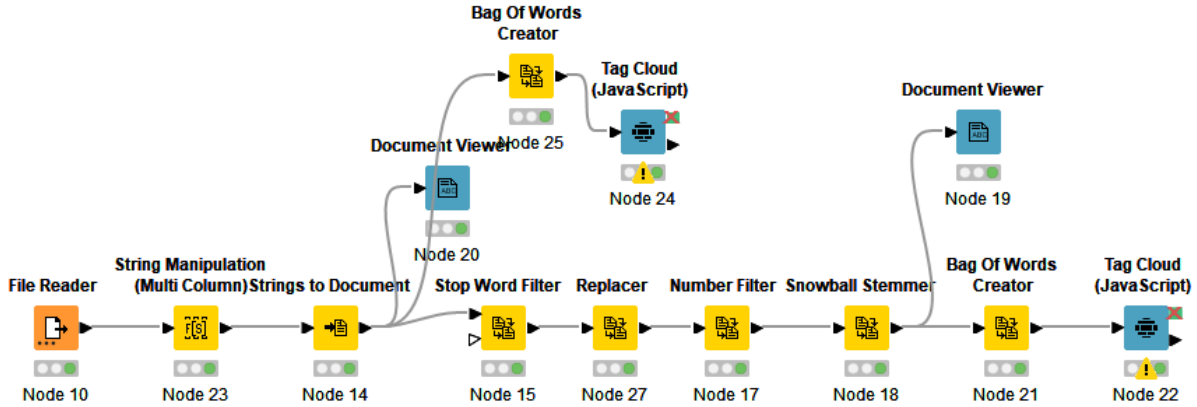


Рис. 8. Готова модель Knime Analytics

Елементарні текстові перетворення наприклад зведення до нижнього регістру, або використання регулярних виразів для перетворення даних формату string. Не зрозумілою залишається специфіка необхідності застосування нод Stop-Word, Number Filter, Snowball Stemmer оскільки для успішної реалізації усіх процесів очищення тексту вимагається перетворити вхідні текстові дані у документ застосувавши ноду Strings to Document.

На рис. 9 зображена ефективність перетворення тексту за допомогою нод з гілки Text Preprocess. За допомогою ноди Document Viewer аналізується вихідний формат строки перед кластеризацією. Як видно з рис. 9 усі інструменти застосовані успішно.

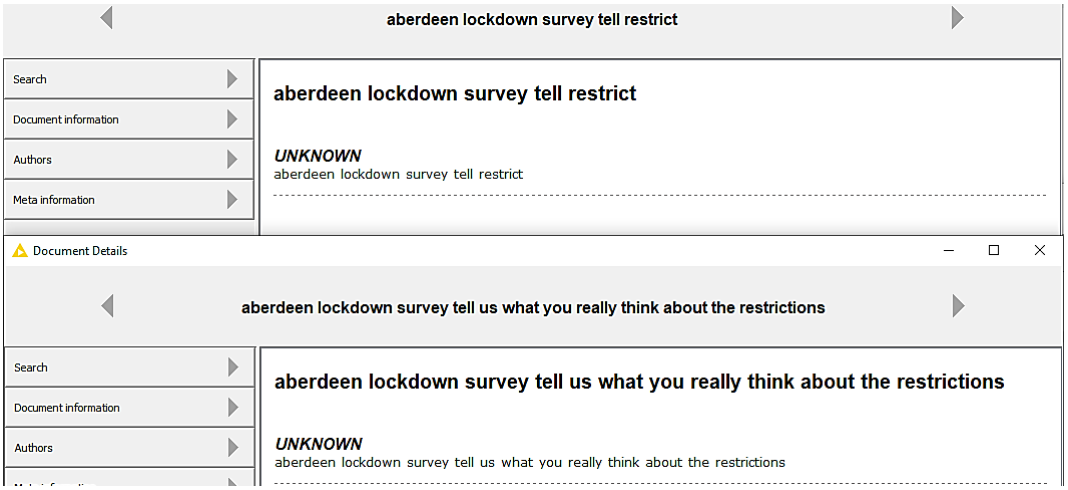


Рис. 9. Ноди Document Viewer

Для візуалізації тексту який представлений у вигляді векторів та готовий до подальшої кластеризації застосовано ноду JavaScript Tag Cloud View, що дозволяє оцінити якість початкового та кінцевого результату обробки тексту.

Хмара слів до застосування нод попередньої обробки тексту (рис. 10).

© Коваль І.М., Суринович О.М.

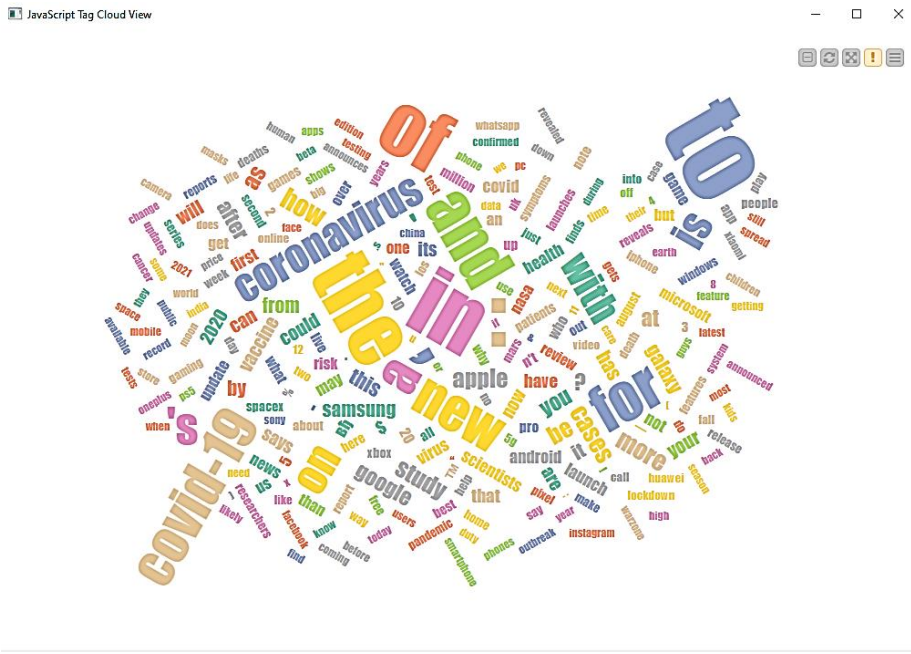


Рис. 10. Нода JavaScript Tag Cloud View

Єдиною маніпуляцією з вхідним текстом на даному етапі є застосування ноди String Manipulation (Multicolumn) всередині якої задано регулярний вираз що зводить всі складові тексту до нижнього регістру.

Хмара слів яку було отримано в результаті комплексної попередньої обробки тексту (рис. 11).

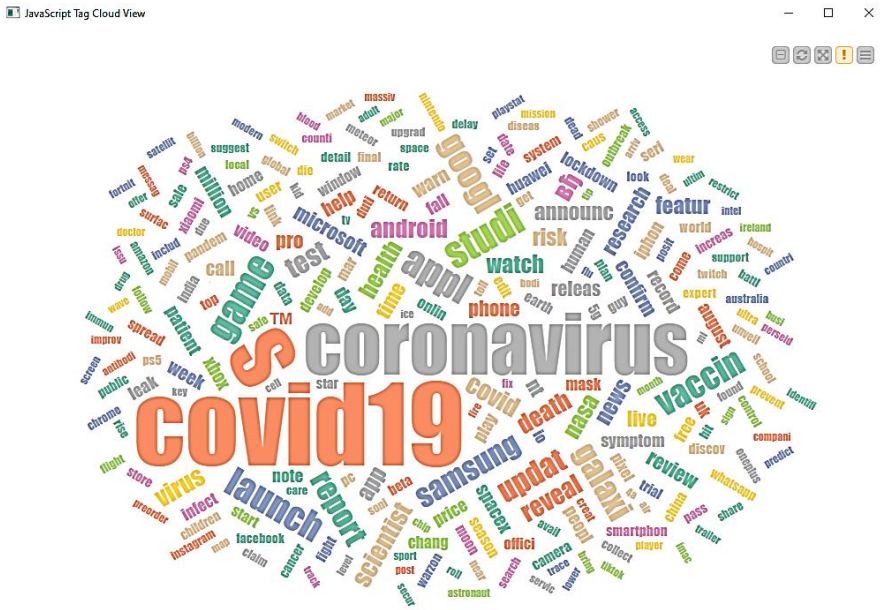


Рис 11. Нода JavaScript Tag Cloud View

Виходячи з цього, можна зробити висновок, що дані готові для подальшої кластеризації, оскільки основні теми вже виділені та очищені від шуму. Ця хмара слів наочно демонструє, що проведена попередня обробка тексту була успішною і дозволила виявити ключові тенденції та теми, присутні у вихідних даних.

У ході цього дослідження було проаналізовано та порівняно процеси попередньої обробки текстових даних, що є критичним етапом для досягнення якісних результатів кластеризації.

Висновки та перспективи подальшого дослідження. В Orange Data Mining процес попередньої обробки тексту виявився більш інтуїтивним та простим для реалізації завдяки зручному інтерфейсу та інтегрованим віджетам. Однак, обмежені можливості віджета Corpus Viewer у

© Коваль І.М., Суринович О.М.

відображенні проміжних результатів змусили використати альтернативні методи для перевірки якості очищення даних, такі як хмара слів. Це є недоліком Orange, оскільки відсутність повного перегляду проміжних результатів ускладнює аналіз окремих етапів обробки тексту.

У KNIME попередня обробка тексту потребує додаткових кроків, таких як обов'язкове перетворення текстових даних у формат документів, що додає складності у процесі. Незважаючи на це, KNIME надає більше можливостей для контролю над кожним етапом обробки тексту і дозволяє отримати більш детальний аналіз результатів. Проте, цей підхід потребує більше часу і навичок, що може бути недоліком для початківців.

Результати попередньої обробки текстових даних в обох програмних системах показали, що після застосування методів попередньої обробки тексту, дані були очищені від шуму, небажаних слів та синтаксичних елементів, що дозволило виділити ключові теми і тенденції із тестованого матеріалу.

Відповідно, можна зробити висновок, що обидві програмні системи мають свої переваги та недоліки, і вибір інструменту залежить від досвіду користувача та конкретних завдань. Orange Data Mining більше підходить для швидкої та інтуїтивної обробки даних, тоді як KNIME забезпечує більшу гнучкість та контроль на всіх етапах, але вимагає більше часу та знань. Подальші дослідження можуть бути спрямовані на застосування алгоритмів кластеризації (наприклад, K-Means, DBSCAN, Hierarchical Clustering) та формування висновків щодо ефективності тих чи інших методів попередньої обробки текстових даних у контексті точності та чистоти отриманих кластерів.

Список бібліографічного опису

1. Manning C. D., Raghavan P., & Schütze H. Introduction to Information Retrieval, 2022. URL: https://nlp.stanford.edu/IR–book/pdf/irbookonlinereading.pdf (дата звернення: 05.08.2024).
2. Charu C. Aggarwal, ChengXiang Zhai. Mining Text Data, 2012. URL: https://doi.org/10.1007/978–1–4614–3223–4 (дата звернення: 05.08.2024).
3. Orange Data Mining. URL: https://orangedatamining.com/ (дата звернення: 05.08.2024).
4. KNIME Analytics Platform.URL: https://www.knime.com/knime–analytics–platform (дата звернення: 17.08.2024).
5. Text Preprocessing Orange Blog. URL:https://orangedatamining.com/blog/text–preprocessing/ (дата звернення: 05.08.2024).
6. From data collection to text mining. URL: https://www.knime.com/blog/data–collection–to–text–mining (дата звернення: 17.08.2024).

References

1. Manning C. D., Raghavan P., & Schütze H. Introduction to Information Retrieval, 2022. URL: https://nlp.stanford.edu/IR–book/pdf/irbookonlinereading.pdf (access date: 05.08.2024).
2. Charu C. Aggarwal, ChengXiang Zhai. Mining Text Data, 2012. URL: https://doi.org/10.1007/978–1–4614–3223–4 (access date: 05.08.2024).
3. Orange Data Mining. URL: https://orangedatamining.com/ (access date: 05.08.2024).
4. KNIME Analytics Platform.URL: https://www.knime.com/knime–analytics–platform (access date: 17.08.2024).
5. Text Preprocessing Orange Blog. URL:https://orangedatamining.com/blog/text–preprocessing/ (access date: 05.08.2024).
6. From data collection to text mining. URL: https://www.knime.com/blog/data–collection–to–text–mining (access date: 17.08.2024).

© Коваль І.М., Суринович О.М.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-25
УДК 004.72
Лисенко Олександр Олександрович, магістрант
Бортник Катерина Яківна, к.т.н., доцент
https://orcid.org/0000-0001-5282-099X
Багнюк Наталія Володимирівна, к.т.н., доцент
https://orcid.org/0000-0002-7120-5455
Луцький національний технічний університет, м.Луцьк, Україна

ДОСЛІДЖЕННЯ ПОПИТУ НА МЕДИЧНІ ТОВАРИ ЗАСОБАМИ SQL

Лисенко О.О., Бортник К.Я., Багнюк Н.В. Дослідження попиту на медичні товари засобами SQL. У статті представлено детальний аналіз процесу розробки та впровадження аналітичної системи для дослідження попиту на медичні товари за допомогою SQL. Основна увага зосереджена на створенні SQL-запитів для збору, обробки та візуалізації великих обсягів даних щодо продажів медичних товарів, що включає такі аспекти, як категорії товарів, часові періоди та географічна сегментація. Для забезпечення ефективного аналізу було реалізовано інтерфейс, який дозволяє зручно налаштовувати запити та проводити сегментацію даних для більш глибокого аналізу і порівняння різних параметрів. Результати тестування системи на реальних даних підтвердили її ефективність у виявленні ключових тенденцій попиту, сезонних коливань і визначенні найбільш популярних продуктів у різних регіонах та часових періодах. Виявлено, що такі коливання суттєво впливають на продажі медичних товарів, що робить можливим оптимізацію планування закупівель та управління запасами, підвищуючи ефективність операцій. Дослідження також підкреслює важливість автоматизації процесу аналізу даних у сфері медицини, що дозволяє оперативно адаптуватися до змін попиту і реагувати на них вчасно. Окрім того, у статті проведено аналіз можливостей подальшого розширення функціоналу системи, зокрема інтеграції з іншими платформами бізнес-аналітики та впровадження методів прогнозування попиту з використанням машинного навчання для підвищення точності передбачень та більш точного планування. Розроблена система дозволяє здійснювати детальний аналіз попиту на медичні товари, що є важливим для підвищення рівня задоволення потреб споживачів та оптимізації бізнес-процесів.
Ключові слова: SQL, медичні товари, попит, сегментація даних, тенденції, сезонні коливання, управління запасами, прогнозування, оптимізація

Lysenko O., Bortnyk K., Bagniuk N. Study of the demand for medical products by means of SQL. The article presents a detailed analysis of the process of development and implementation of an analytical system for researching the demand for medical products using SQL. Focus is on creating SQL queries to collect, process and visualize large volumes of data on medical product sales, including aspects such as product categories, time periods and geographic segmentation. To ensure effective analysis, an interface was implemented that allows you to conveniently configure queries and segment data for deeper analysis and comparison of various parameters. The results of testing the system on real data confirmed its effectiveness in identifying key trends in demand, seasonal fluctuations and determining the most popular products in different regions and time periods. It was found that such fluctuations significantly affect the sales of medical products, which makes it possible to optimize procurement planning and inventory management, increasing the efficiency of operations. The study also emphasizes the importance of automating the process of data analysis in the field of medicine, which allows us to quickly adapt to changes in demand and respond to them in a timely manner. In addition, the article analyzes the possibilities of further expanding the functionality of the system, in particular, integration with other business analytics platforms and the introduction of demand forecasting methods using machine learning to increase the accuracy of predictions and more accurate planning. The developed system allows for a detailed analysis of the demand for medical products, which is important for increasing the level of satisfaction of consumer needs and optimizing business processes.
Keywords: SQL, medical products, demand, data segmentation, trends, seasonal fluctuations, inventory management, forecasting, optimization

Постановка наукової проблеми. У сучасних умовах зростаючого ринку медичних товарів виникає необхідність ефективного управління попитом та запасами продукції. Проблема полягає в тому, що традиційні методи аналізу попиту не завжди дозволяють оперативно обробляти великі обсяги даних і враховувати різні фактори, які впливають на попит, такі як сезонність, географічна специфіка та демографічні особливості. Використання засобів SQL для дослідження попиту дозволяє автоматизувати процеси збору, обробки та аналізу даних про продажі медичних товарів, що дає змогу отримати точніші прогнози та рекомендації для ефективного планування. Основною метою є розробка інструментарію для аналізу попиту на медичні товари, який дозволить оперативно отримувати інформацію про найпопулярніші товари, аналізувати тенденції продажів і виявляти фактори, що впливають на зміни в попиті. Розв'язання цієї проблеми сприятиме підвищенню ефективності управління продажами, зменшенню витрат на утримання запасів та покращенню задоволення потреб споживачів.

Аналіз останніх досліджень і публікацій. Дослідження попиту на медичні товари, як правило, базується на використанні різних інструментів бізнес-аналітики та статистичних методів.

© Лисенко О.О., Бортник К.Я., Багнюк Н.В.

Багато науковців акцентують увагу на тому, що традиційні підходи, такі як ручний аналіз продажів або використання спеціалізованих програм, не дозволяють ефективно обробляти великі обсяги даних. Для вирішення цієї проблеми все частіше використовуються сучасні технології, зокрема, бази даних та мови запитів SQL.[1]

Дослідження показують, що використання SQL для аналізу попиту на товари дає можливість швидко і точно обробляти великі масиви даних, визначати тренди та прогнозувати зміни в попиті. Зокрема, SQL дозволяє здійснювати гнучкий аналіз даних про продажі медичних товарів, сегментувати їх за категоріями, часовими періодами та географічними ознаками, що є важливим фактором для управління запасами і планування закупівель.[2]

У попередніх публікаціях наголошується на важливості автоматизації процесів аналізу та прийняття рішень у сфері медицини за допомогою баз даних. Дослідження впливу сезонності, змін в економічних умовах та демографічних факторів на попит на медичні товари активно обговорюються в наукових колах. Проте, недостатньо уваги приділяється практичному використанню SQL для глибокого аналізу даних про попит, що відкриває можливості для подальших досліджень у цій галузі.[3]

Виділення невирішених раніше частин загальної проблеми: Незважаючи на активний розвиток технологій для аналізу даних, багато аспектів дослідження попиту на медичні товари залишаються невирішеними. Зокрема, існує проблема інтеграції даних з різних джерел, таких як роздрібні мережі, аптеки та онлайн-платформи. Використання SQL дозволяє обробляти ці дані, але виклики постають через неоднорідність форматів даних і необхідність їхньої попередньої обробки для забезпечення коректності аналізу.

Крім того, питання прогнозування попиту на медичні товари на основі минулих даних ще потребує доопрацювання. Хоча SQL добре підходить для аналізу історичних даних, для більш точного прогнозування необхідно інтегрувати з SQL системи машинного навчання або інші аналітичні інструменти. Це може забезпечити більш точні прогнози щодо коливань попиту залежно від сезонності, епідемій або змін у споживчій поведінці.

Іншою важливою проблемою є автоматизація аналізу та інтерактивної візуалізації даних. Більшість існуючих рішень передбачають ручне написання запитів, що потребує певних технічних навичок. Це обмежує широке застосування таких інструментів на рівні користувачів, які не є спеціалістами з баз даних.

Таким чином, важливим завданням є розробка інструментів, які дозволять автоматизувати процес аналізу попиту на медичні товари, інтегрувати дані з різних джерел та забезпечити прогнозування на основі сучасних методів обробки великих даних.

Мета дослідження: Основною метою дослідження є розробка та впровадження ефективної системи для аналізу попиту на медичні товари за допомогою SQL. Це включає в себе створення механізмів збору, обробки та аналізу великих обсягів даних з метою виявлення тенденцій попиту, сезонних коливань та найбільш популярних категорій товарів. Дослідження також спрямоване на розробку SQL-запитів для автоматизованого отримання та візуалізації даних, що дозволить оптимізувати процес управління запасами та прогнозування попиту на медичні товари.

Основна частина дослідження: У ході дослідження було розроблено систему для аналізу попиту на медичні товари за допомогою SQL. Система базується на використанні реляційної бази даних, що містить дані про продажі медичних товарів, включаючи такі параметри, як назва товару, категорія, дата продажу, регіон та кількість проданих одиниць.

1. Створення та налаштування бази даних

Перший етап включав підготовку та структурування даних. Було створено кілька таблиць:

- Products (товари: ID товару, назва, категорія);
- Sales (продажі: ID продажу, дата, кількість, регіон);
- Customers (клієнти: ID клієнта, регіон, тип клієнта).

Дані були очищені, нормалізовані та завантажені в базу даних. Це забезпечило можливість швидкого виконання запитів для отримання необхідної аналітичної інформації.

2. SQL-запити для аналізу попиту

Для дослідження попиту було використано кілька типів SQL-запитів:

Запит дозволяє отримати загальний обсяг проданих товарів за весь період (рисунок 1).

© Лисенко О.О., Бортник К.Я., Багнюк Н.В.

```
SELECT SUM(quantity) AS TotalSales
FROM Sales;
```

Рис. 1. Запит на загальний обсяг продажів

Наступний запит дозволяє визначити, які категорії медичних товарів користуються найбільшим попитом. Це видно на рисунку 2.

```
SELECT p.category, SUM(s.quantity) AS TotalSales
FROM Sales s
JOIN Products p ON s.product_id = p.id
GROUP BY p.category;
```

Рис. 2. Аналіз попиту за категоріями товарів

Цей запит дає змогу проаналізувати попит по місяцях і виявити сезонні піки продажів. Це можна побачити на рисунку 3.

```
SELECT EXTRACT(MONTH FROM s.sale_date) AS Month, SUM(s.quantity) AS TotalSales
FROM Sales s
GROUP BY EXTRACT(MONTH FROM s.sale_date);
```

Рис.3. Сезонні коливання попиту

А далі можна зробити можливість бачити попит за регіонами, ця функція наведена на рисунку 4.

```
SELECT s.region, SUM(s.quantity) AS TotalSales
FROM Sales s
GROUP BY s.region;
```

Рис. 4. Аналіз за регіонами

Інтерактивна візуалізація даних. Зібрані та оброблені дані були візуалізовані за допомогою зовнішніх інструментів для побудови графіків та діаграм, таких як Power BI на рисуноку 5..

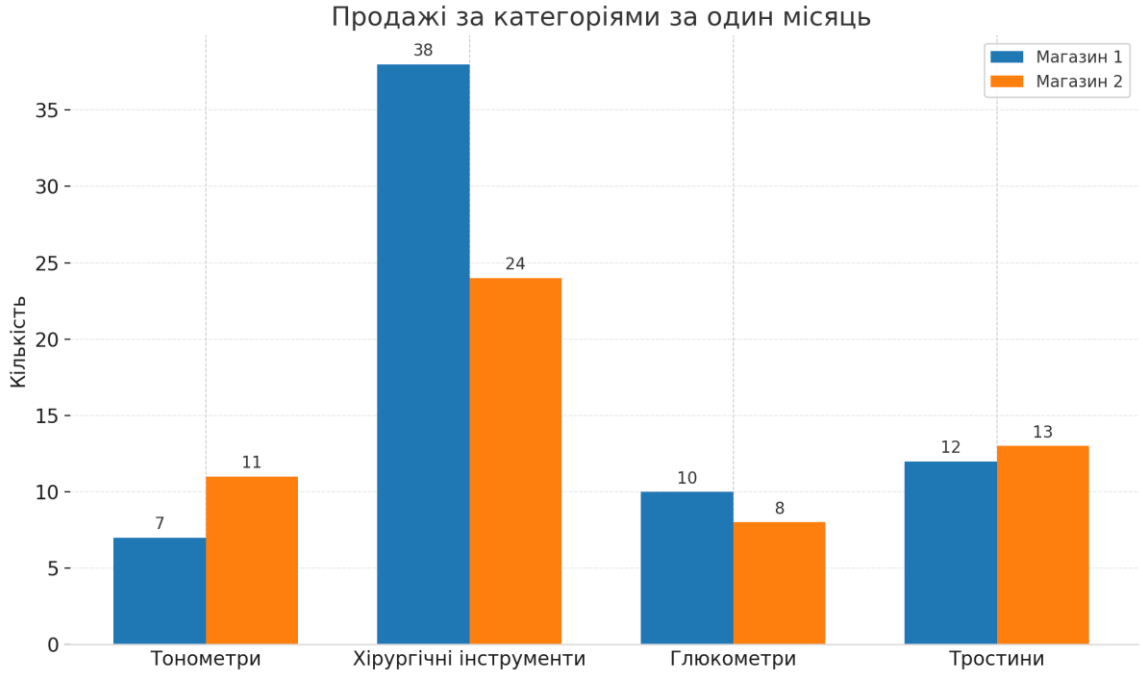


Рис.5. Графік продажів

© Лисенко О.О., Бортник К.Я., Багнюк Н.В.

Це дозволило краще зрозуміти тренди та зробити висновки на основі даних

Результати аналізу показали, що найбільший попит на медичні товари спостерігається в осінній період, коли зростає потреба в товарах для медичних закладах, або медичного коледжу. Також було виявлено, що певні магазини мають стабільно високий попит на специфічні категорії товарів, такі як хірургічні інструменти.

Висновки та перспективи подальшого дослідження. В ході дослідження було успішно розроблено та протестовано аналітичну систему для дослідження попиту на медичні товари засобами SQL. Система показала свою ефективність в аналізі великих обсягів даних, дозволяючи виявляти ключові тенденції попиту, сезонні коливання та найбільш популярні категорії медичних товарів. Використання SQL для обробки та сегментації даних за категоріями, регіонами та періодами забезпечило точність та оперативність отримання інформації, що є важливим для управління запасами та планування продажів.

Основні висновки дослідження:

1. Використання SQL для дослідження попиту значно спрощує обробку великих даних і дозволяє виявляти тренди, що допомагає оптимізувати процес управління запасами.
2. Виявлено, що сезонні коливання попиту, зокрема осінні піки, впливають на потребу в різних категоріях товарів, таких як хірургічні інструменти.
3. Інтерактивна візуалізація даних з використанням зовнішніх інструментів, таких як Power BI, значно поліпшила розуміння отриманих результатів і сприяла прийняттю обґрунтованих рішень щодо закупівель та запасів.

Перспективи подальшого дослідження:

1. Прогнозування попиту: Впровадження методів машинного навчання у поєднанні з SQL для прогнозування попиту дозволить покращити точність передбачень, враховуючи фактори сезонності, економічні зміни та зміни у споживчих вподобаннях.
2. Автоматизація процесів: Подальша автоматизація створення та виконання SQL-запитів може знизити потребу в технічних навичках користувачів, роблячи систему доступнішою для широкого кола спеціалістів.

Ці напрямки розробки сприятимуть подальшому розвитку системи та її застосуванню для більш глибокого аналізу попиту на медичні товари, підвищуючи ефективність управління бізнес-процесами.

Список бібліографічного опису

1. Gupta, A. Optimization of Inventory Management for Medical Supplies Using SQL Queries. In: Proceedings of the 10th International Conference on Healthcare Informatics, 2021, pp. 45-52 с.
2. Ponniah, P. Data Warehousing Fundamentals for IT Professionals. 2nd Edition. New York: Wiley, 2018. 323 с.
3. Korhonen, J., & Honkela, A. Seasonal Demand Prediction Using SQL and Machine Learning Techniques. Journal of Business Analytics, 2020, 17(4), 125-139 с.

References

1. Gupta, A. Optimization of Inventory Management for Medical Supplies Using SQL Queries. In: Proceedings of the 10th International Conference on Healthcare Informatics, 2021, pp. 45-52 p.
2. Ponniah, P. Data Warehousing Fundamentals for IT Professionals. 2nd Edition. New York: Wiley, 2018. – 323 p.
3. Korhonen, J., & Honkela, A. Seasonal Demand Prediction Using SQL and Machine Learning Techniques. Journal of Business Analytics, 2020, 17(4), 125-139 p.

© Лисенко О.О., Бортник К.Я., Багнюк Н.В.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-26
УДК 004.94:658.5
Марценюк Василь Петрович[1], д.т.н., професор
https://orcid.org/000-0001-5622-1038
Кіт Наталія Василівна[2], аспірант
https://orcid.org/0000-0002-9096-4393
[1]Університет в Бельсько Бяла, м. Бельсько Бяла, Польща
[2]Тернопільський національний технічний університет імені Івана Пулюя, м. Тернопіль, Україна

АНАЛІТИЧНИЙ ОГЛЯД ПІДХОДІВ ДО ПРОЕКТУВАННЯ ВИРОБНИЧИХ МЕРЕЖ

Марценюк В.П., Кіт Н.В. Аналітичний огляд підходів до проектування виробничих мереж. В даній статті проведено аналітичний огляд публікацій по проектуванню виробничої мережі, зокрема його актуальності на сучасних виробництвах, а а також проблемами, які виникають при транспортуванні сировини між логістично-виробничими вузлами. Проведено аналіз робіт по роках в наукометричній базі Scopus. Представлено авторів, які опублікували найбільшу кількість робіт, включаючи країни та навчальні заклади. А також, які саме види публікацій найбільше публікували та у якій галузі науки був до них найбільший інтерес. Представлено спонсорські фонди, які зробили найбільший вклад у видавництво даних наукових робіт. Під час проектування виробничої мережі часто виникають проблеми логістики, зокрема такі моменти, як визначення найбільш вигідних місць для розміщення виробництва об'єктів, складів і розподільних вузлів для зниження транспортних витрат, зменшити терміни доставки, а також ефективне задоволення потреб і запитів споживачів. Для вирішення проблем, які виникають під час транспортування сировини між логістичними ділянками розроблено решітчасту модель на прямокутній решітці з використанням решітчастих диференціальних рівнянь із затримкою.
Ключові слова: проектування виробничої мережі, Індустрія 4.0, кіберфізична система, решітчаста модель, Scopus

Martsenyuk V., Kit N. Analytical review of approaches manufacturing networks design. This article provides an analytical review of publications on the manufacturing network design, particularly its relevance in modern production, as well as the problems that arise during the transportation of raw materials between logistics and production sites. An analysis of works by year in the scientometric database Scopus was carried out. The authors who have published the largest number of works are represented, including countries and educational institutions. Also, what types of publications were published the most and in which field of science there was the greatest interest in them. The sponsoring funds that made the biggest contribution to the publication of these scientific works are presented. During the manufacturing network design, logistics problems often arise such points as determining the most profitable places for the location of production facilities, warehouses and distribution centers to reduce transportation costs, reduce delivery times, as well as effectively meet the needs and requests of consumers. To solve the problems that arise during the transportation of raw materials between logistics areas, a lattice model on a rectangular grid using lattice differential equations with a delay has been developed.
Keywords: manufacturing network design, Industry 4.0, cyber-physical system, lattice model, Scopus

Постановка проблеми. З розвитком технологій та впровадженням в них Індустрії 4.0 зросла потреба у проектуванні виробничих мереж та автоматизації виробничих процесів. І хоч ця тема є на даний момент дуже актуальною, особливо протягом осатннього десятиліття, бо все більше виробництв зацікавлені у автоматизації, але залишається мало дослідженою. Найменш дослідженою є тема проблем, які виникають при транспортуванні сировини між логістично-виробничими ділянками. Для того, щоб аналіз був якісним та проаналізовані дослідження були новаторськими та якісними, відповідно це не можливо зробити без попередньо вивченого та проаналізованого матеріалу на дану тематику. І саме повноцінним вирішенням данної проблеми буде через глибокий аналітичний огляд літературних джерел. В даний момент це питання можна оптимізувати за рахунок використання інформаційних пошукових систем, такий як Scopus. Також для вирішення проблем, які виникають під час транспортування сировини між логістичними ділянками буде дуже доречно розробити решітчасту модель на прямокутній решітці з використанням решітчастих диференціальних рівнянь із затримкою.

Аналіз відомих результатів досліджень. Проектування виробничої мережі має вирішальне значення, оскільки визначає оптимальне розміщення виробничих і логістичних потужностей, впливаючи на ефективність витрат, рівень обслуговування клієнтів і відіграє ключову роль у забезпеченні організацій ефективної конкуренції у все більш складній та взаємопов'язаній глобальній економіці [1]. У цьому процесі приймаються важливі рішення, пов'язані з розташуванням, пропускною спроможністю, логістикою та розгортанням технологій у мережі, враховуючи такі фактори, як вартість, якість, швидкість і гнучкість [2]. Це дозволяє компаніям

© Марценюк В.П., Кіт Н.В.

стратегічно позиціонувати свою діяльність, щоб задовольнити попит, мінімізувати витрати та швидко адаптуватися до мінливих умов ринку [3].

Кіберфізичні системи (КФС) вважаються наступним кроком у еволюції постійного вдосконалення та інтеграції функцій у рамках Industry 4.0. Напівавтономні та автономні рішення є життєво важливими компонентами Industry 4.0, де IoT, штучний інтелект (ШІ), нові інтегровані системи, розширена аналітика та інші технології відіграють важливу роль [4-6].

Огляд проектування виробничих систем у взаємодії Індустрії 4.0 та Індустрії 5.0., а також виробниче мислення та методи моделювання розглянуто у статті [7], розкриття потужності промислового штучного інтелекту для Індустрії 5.0: ідеї, шляхи та виклики в [8]. Розглянуто людино-орієнтовані системи в інтелектуальному виробництві[9], а також управління операціями та інформаційні системи для інтелектуального виробництва[10]. Тему штучного інтелекту для людино-кіберфізичних виробничих систем розкрито в [11]. Моделювання, проектування та симуляція як послуга на основі розширеної реальності (XR) в Індустрії 4.0 розглянуто в [12] та штучний інтелект для управління виробництвом і контролю на шляху до масової персоналізації глобальних мереж у [13].

У роботах [14] розглянуті підходи до проведення аналітичного пошуку літературних джерел по різних наукових тематиках в наукометричних базах Scopus та WoS.

Мета роботи: Провести аналітичний огляд публікацій по проектуванню виробничої мережі за останнє дисятиліття, зокрема його актуальності на сучасних виробництвах, а також проблемами, які виникають при транспортуванні сировини між логістично-виробничими ділянками. Провести аналіз робіт по роках в наукометричній базі Scopus, який буде включати авторів, країни, навчальні заклади, а також споснорські фонди, які зробили найбільший вклад у даному напрямку. Розглянути, які саме види публікацій найбільше видавалися та у якій галузі науки був до них найбільший інтерес. Для вирішення проблем, які виникають під час транспортування сировини між логістичними ділянками запропонувати решітчасту модель на прямокутній решітці з використанням решітчастих диференціальних рівнянь із затримкою

Аналітичний огляд. Використання комп'ютерної програми Scopus дає можливість оптимізувати релевантний пошук наукової літератури з проблеми, яка є первинною. Для того, щоб оцінити актуальність досліджень про проектування виробничої мережі в наукометричній базі Scopus було сформульовано аналітичний запит TITLE-ABS-KEY ("manufacturing") AND TITLE-ABS-KEY ("network") AND TITLE-ABS-KEY ("design")

За результатами пошуку в наукометричній базі Scopus спостерігаємо зростання кількості публікацій протягом останніх десяти років. Це також додатково підтверджує високий науковий інтерес до досліджуваної тематики, а саме проектування виробничої мережі та автоматизації виробничих процесів.

На пошуковий запит по даній темі в наукометричній базі Scopus знайдено 15980 наукових праць з 1942 до 2025 року, з них 9112 з 2014 по 2024 роки, тобто, за останнє десятиріччя. Найбільша кількість літературних джерел по досліджуваній тематиці припадає на останні 3 роки. Зокрема в 2022 році було опубліковано 1230 публікацій, 2023 – 1333, 2024 – 1100 , що підтверджує актуальність дослідження даної проблеми та невпинне зростання інтересу до неї в усьому світі (рис. 1).

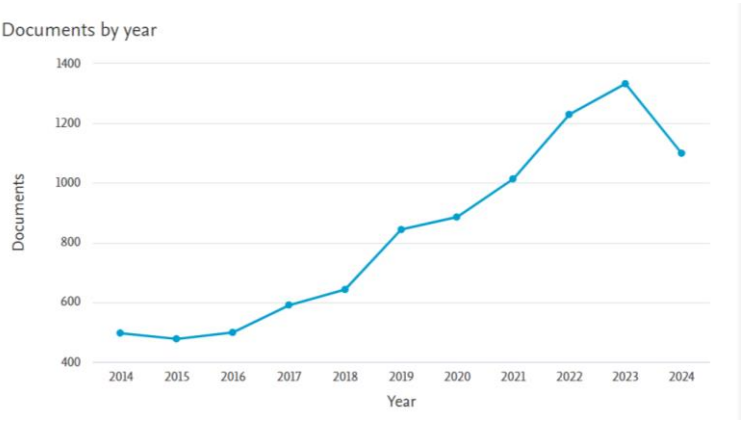


Рис. 1. Результати пошукового запиту в наукометричній базі Scopus (загальна кількість праць за роками)

© Марценюк В.П., Кіт Н.В.

Найбільше публікацій було видано у таких виданнях, як Proceedings Of SPIE The International Society For Optical Engineering(142), Proceedings Electronic Components And Technology Conference (105), Procedia CIRP (98 International Journal Of Advanced Manufacturing Technology (93), IEEE Access(91). Що свідчить про те, що багато міжнародних видань публікують статті на дану тематику та зацікавлені у її розвитку.

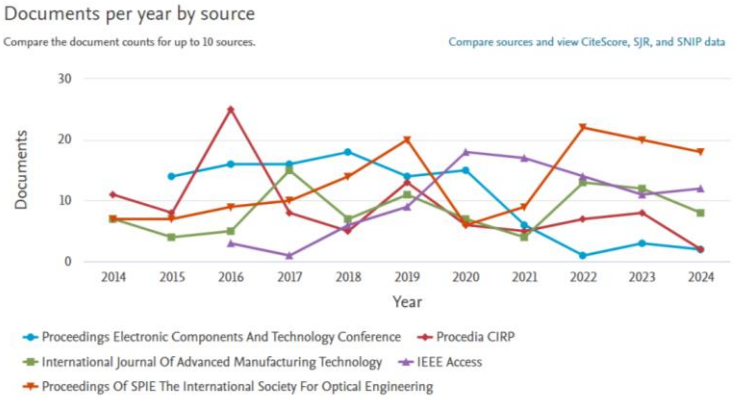


Рис. 2. Результати пошукового запиту в наукометричній базі Scopus (публікації за рік по виданнях)

Щодо основних авторів, то лідерство мають такі науковці, як Mourtzis, D. – 33, Schuh, G. – 16, Chahal, P. та Lanza, G. по 13 публікацій, Doukas, M. – 12, Chakrabarty, K. та Yu, B. по 11 публікацій, Franke, J. та Simpson, T.W. по 10 публікацій та Allen, J.K.., кількість праць якого найбільша – 9 (рис. 3).

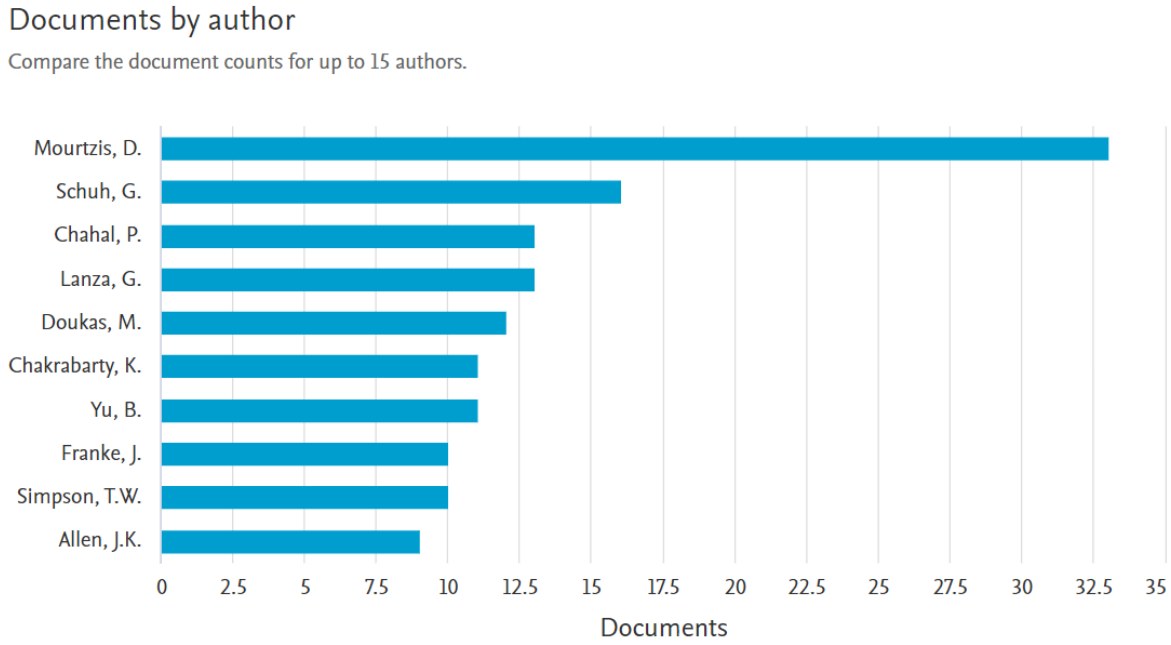


Рис. 3. Результати пошукового запиту в наукометричній базі Scopus (автори)

На рис. 4 наведено результати пошукового запиту в наукометричній базі Scopus серед наукових закладів.

© Марценюк В.П., Кіт Н.В.

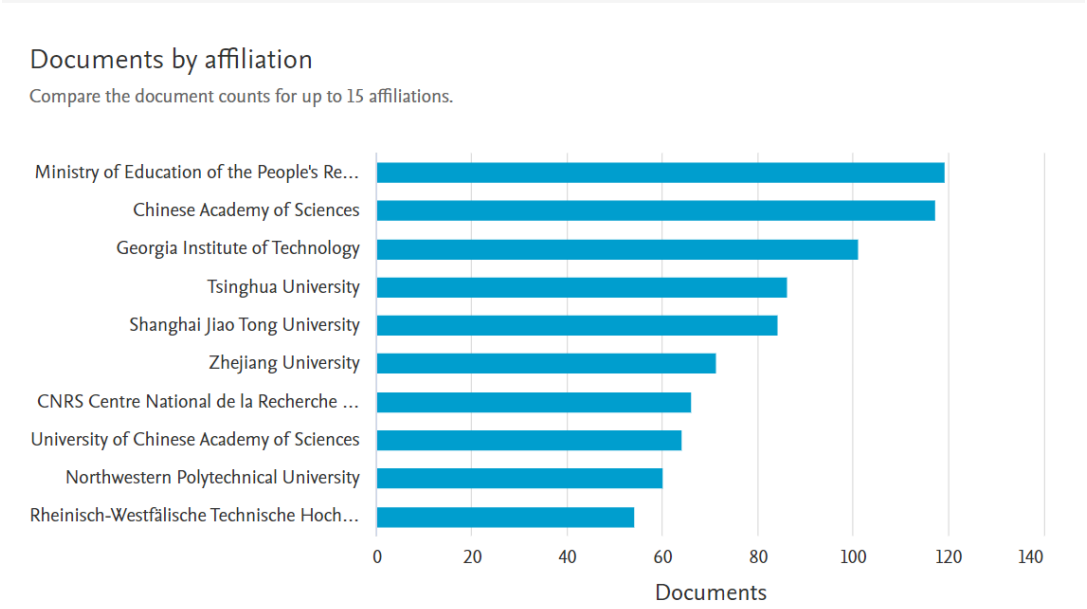


Рис. 4. Результати пошукового запиту в наукометричній базі Scopus (наукові заклади)

Серед наукових закладів найбільша кількість у Ministry of Education of the People's Republic of China-119, Chinese Academy of Sciences – 117, Georgia Institute of Technology – 101, Tsinghua University – 86, Shanghai Jiao Tong University – 84.

Аналіз серед країн показав, що кількість публікацій найбільше у Китаї – 1984, США – 1779, Індія – 817, Німеччині – 615, Об'єднаному Королівстві – 554 та Австралія – 357. Що свідчить про те, що велика кількість наукових досліджень публікується у різних країнах світу та вони зацікавлені у проектуванні виробничих мереж, що допомагає автоматизовувати їхні виробництва (рис. 5).

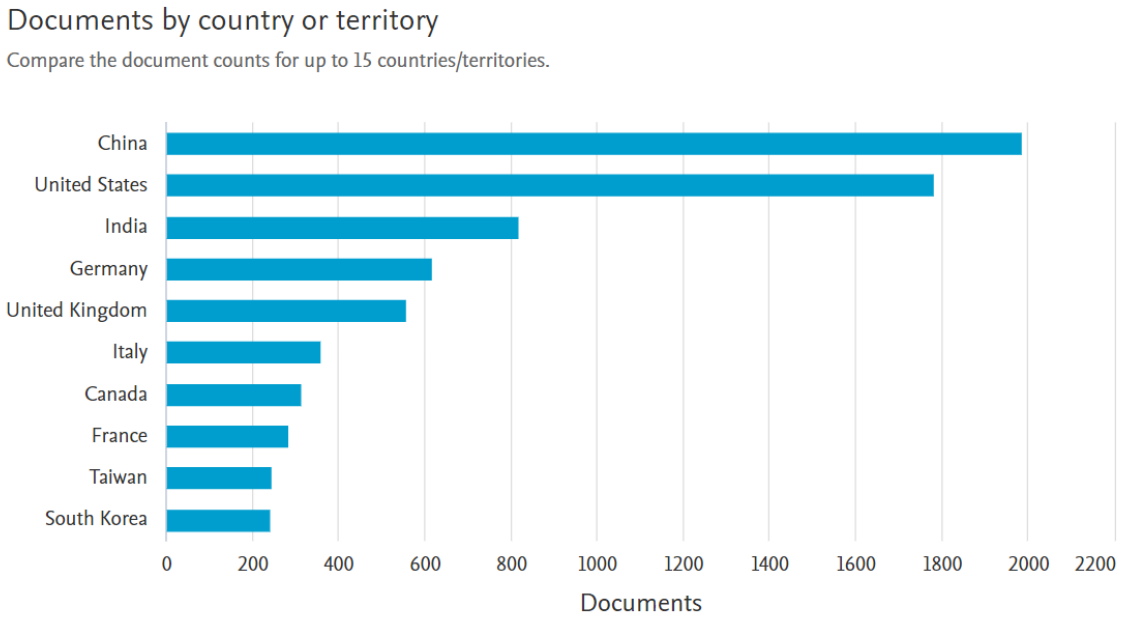


Рис. 5. Результати пошукового запиту в наукометричній базі Scopus (країни)

Серед наукових праць переважали дослідницькі статті в журналах (53,4%), тези конференцій (31,5%), огляди конференцій (6,4%), оглядові статті (3,9%), та розділи в монографіях (3,8%) (рис. 6).

© Марценюк В.П., Кіт Н.В.

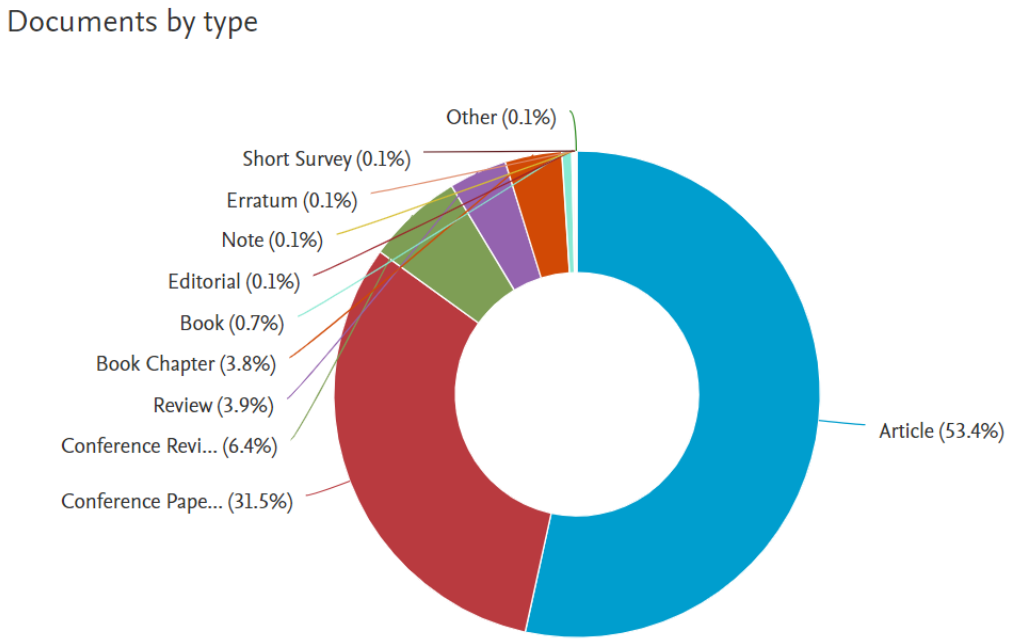


Рис. 6. Результати пошукового запиту в наукометричній базі Scopus(види публікацій)

Розглядаючи предметні галузі можна сказати, що найбільша кількість наукових публікацій належить до інженерії (29,8%), комп'ютерних наук (20,2%), матеріалознавство (8,9%), математики (6,9%), фізики та астрономії(6,6%) та бізнесу, менеджмент і бухгалтерському обліку (5,1%) (рис.7).

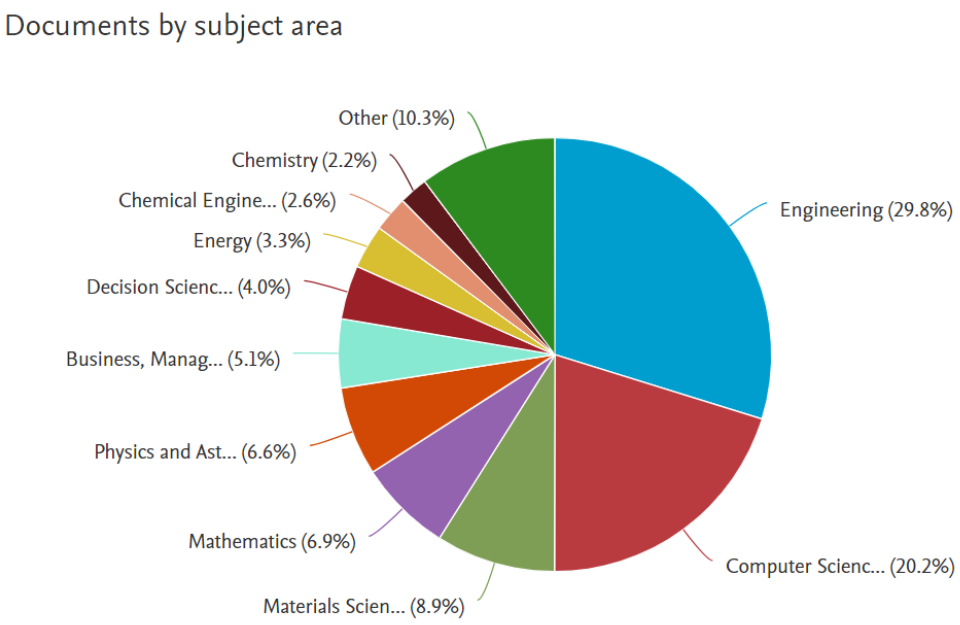


Рис. 7. Результати пошукового запиту в наукометричній базі Scopus(предметні галузі)

На рис. 8 наведено результати пошукового запиту в наукометричній базі Scopus по спонсорських фондах.

© Марценюк В.П., Кіт Н.В.

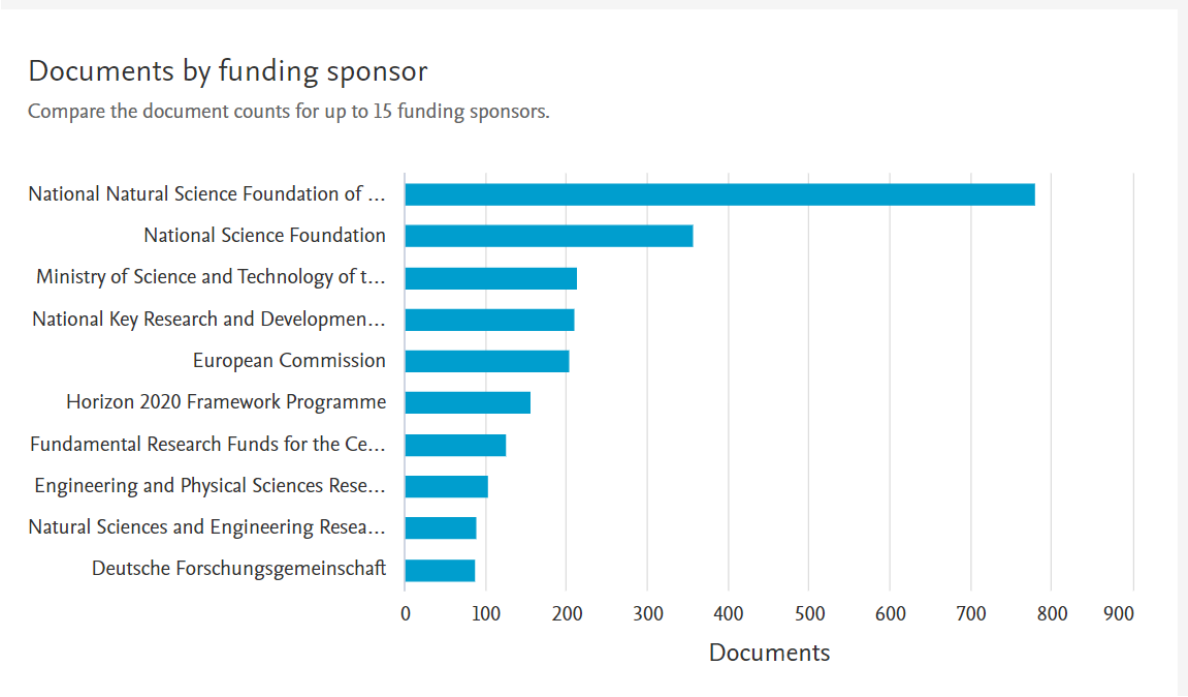


Рис. 8. Результати пошукового запиту в наукометричній базі Scopus(спонсорські фонди)

Спонсорські фонди, які зробили найбільший вклад у видавництво даних наукових робіт є National Natural Science Foundation of China – 779, National Science Foundation – 356, Ministry of Science and Technology of the People's Republic of China-213, European Commission – 203, Horizon 2020 Framework Programme – 155.

Проектування решітчастої моделі. Наведена нижче модель розроблена для проблем проектування виробничих мереж, що виникають під час транспортування сировини між логістичними вузлами на прямокутній решітці з використанням решітчастих диференціальних рівнянь із затримкою.

Термінологія моделі походить з [15]. Модель базується на ряді припущень. Припустимо, що виробнича мережа включає в себе логістично-виробничі вузли, які розташовані у вузлах прямокутної решітки (i, j), $(i, j) = \overline{1, N}$. Нехай для t, в даний момент часу, $V_{i,j}(t)$ є ресурсами, які використовуються для виробництва продукту і в даний час розташовані на вузлі, в свою чергу, $F_{i,j}(t)$ є готовим продуктом, який виробляється і зберігається на (i, j).

Модель враховує наступні параметри процесів виробництва і транспортування для довільного логістично-виробничого вузла (i, j):

1. Ресурси з'являються (можуть бути «видобуті») всередині виробничого вузла з імовірністю $\beta > 0$.
2. На виготовлення одиниці продукції потрібно $\gamma > 0$ одиниць ресурсу.
3. Використання ресурсу обмежено за допомогою коефіцієнта $\delta_v > 0$, який дозволяє нам прагнути до рівня пропускної здатності для $V_{i,j}(t)$.
4. Припустимо, що передача ресурсів може бути можлива з чотирьох сусідніх вузлів $(i-1, j)$, $(i+1, j)$, $(i, j-1)$, $(i, j+1)$ (Рис. 1) з вершинами $D_{i,j}^{i-1,j}\Delta^{-2}$, $D_{i,j}^{i+1,j}\Delta^{-2}$, $D_{i,j}^{i,j-1}\Delta^{-2}$, $D_{i,j}^{i,j+1}\Delta^{-2}$ де $D_{i,j}^{k,m} > 0$, $i, j, k, m = \overline{1, n}$ і $\Delta > 0$ - це відстань між вузлами.
5. Виробництво може бути забраковано з імовірністю $\mu_f > 0$.
6. В результаті затримок і неврахованих наслідків ми спостерігаємо збільшення вартості ресурсу, необхідного для виробництва продукції, до рівня ймовірності $\eta\gamma$.
7. Виробництво прагне до певної пропускної здатності з імовірністю $\delta_f > 0$.
8. Нехай $\tau > 0$ — час, необхідний для виготовлення одиниці продукції.

© Марценюк В.П., Кіт Н.В.

З наведених припущень ми розглядаємо приріст вартості ресурсів у вузлі (i, j) протягом часу $\Delta t : \Delta V_{i,j}(t) = V_{i,j}(t+\Delta t) - V_{i,j}(t)$ з урахуванням наступних припущень:

а) збільшення на величину $\beta V_{i,j}(t)\Delta t$, що викликано «видобутком» нового ресурсу;

б) зменшення на величину $-\gamma F_{i,j}(t-\tau)V_{i,j}(t)\Delta t$, що пояснюється ресурсами, необхідними для виробництва продукції у вузлі (i, j) в момент $(t-\tau)$;

в) збільшення вартості $-\Delta\delta_V V_{i,j}(t-\tau)V_{i,j}(t)\Delta t$ внаслідок пропускної здатності ресурсів.

Кількість продуктів у вузлі (i, j) також залежить від розподілу сировини між чотирма сусідніми вузлами, що враховується при обчисленні просторового оператора $\hat{S}\{V_{i,j}(t)\}\Delta t$ у вигляді (4).

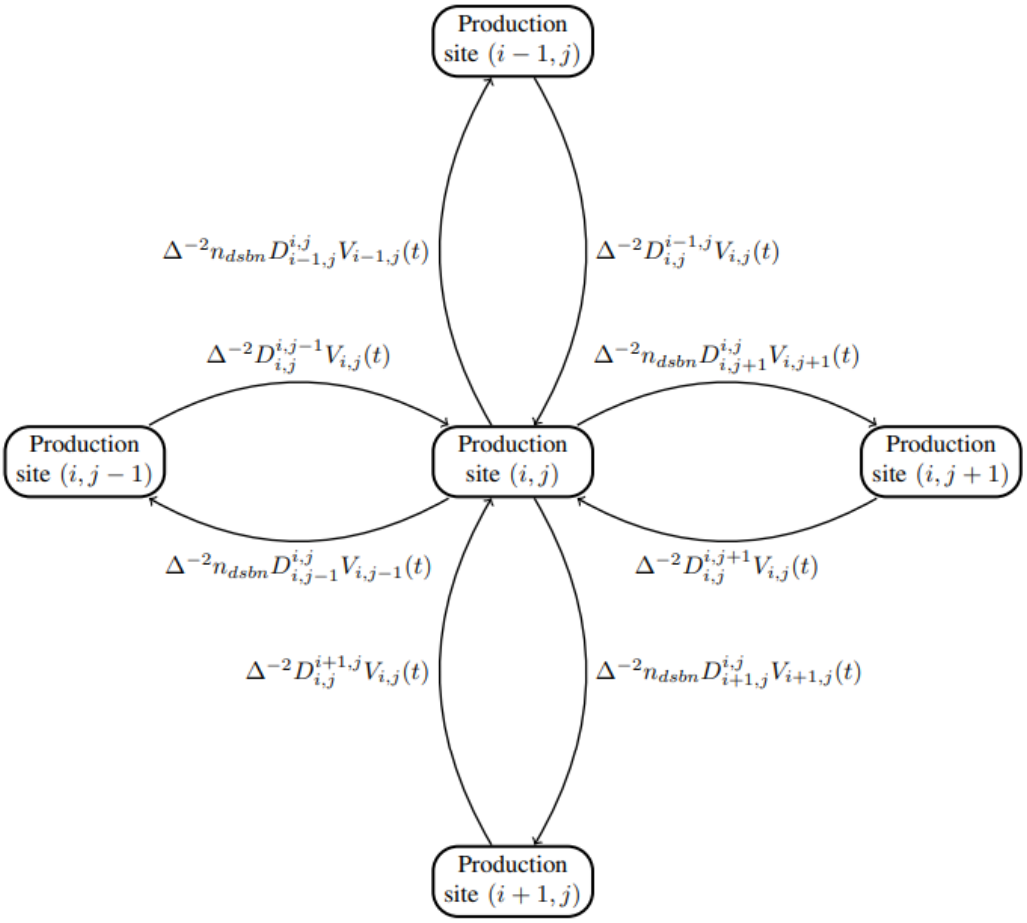
Виходячи з наведених вище припущень, збільшення кількості ресурсів $\Delta V_{i,j}(t)$ у вузлі (i, j) за певний проміжок часу можна записати у вигляді $\Delta V_{i,j}(t)\Delta t$.

Розділивши ліву і праву частини всього рівняння на Δt і додавши $\Delta t \to 0$, отримаємо рівняння для визначення кількості сировинних ресурсів $V_{i,j}$, $t > 0$.

$$\frac{dV_{i,j}}{dt} = \beta V_{i,j}(t) - \gamma F_{i,j}(t-\tau)V_{i,j}(t) - \delta_V V_{i,j}(t-\tau)V_{i,j}(t) + \hat{S}\{V_{i,j}(t)\} \qquad (1)$$

Зростання продукції у вузлі (i, j) за період часу Δt становить

$$\Delta F_{i,j}(t) = F_{i,j}(t+\Delta t) - F_{i,j}(t)$$



© Марценюк В.П., Кіт Н.В.

Рис. 9. Квадратна решітка, яка представляє собою чотири сусідні логістично-виробничі вузли

Згідно з припущеннями, зазначеними вище, наступні аналоги впливають на зміну продукту на виробничому вузлі (i, j) за період часу Δt:

а) зменшується на значення $-\mu_f F_{i,j}(t)\Delta t$ за рахунок кількості бракованих виробів;

б) збільшення на величину $\eta\gamma V_{i,j}(t-\tau)F_{i,j}(t)\Delta t$, яка визначається ресурсами, необхідними для виробництва одиниці продукції в момент часу $(t-\tau)$ (τ - значення затримки (час, витрачений на виробництво одиниці продукції));

в) зменшення на величину $-\delta_f F_{i,j}^2(t)\Delta t$, що викликано зниженням швидкості виготовлення нових виробів δ_f у зв'язку з наближенням їх до межі насичення.

З наведених причин зростання продуктів $\Delta F_{i,j}(t)$ у вузлі (i, j) за період часу Δt можна записати як $\Delta F_{i,j}(t) = -\mu_f F_{i,j}(t)\Delta t + \eta\gamma V_{i,j}(t-\tau)F_{i,j}(t)\Delta t - \delta_f F_{i,j}^2(t)\Delta t, t > 0$.

Розділивши ліву і праву частини останнього рівняння на Δt з напрямком $\Delta t \to 0$, отримаємо рівняння для визначення кількості продукції:

$$\frac{dF_{i,j}}{dt} = (-\mu_f + \eta\gamma V_{i,j}(t-\tau) - \delta_f F_{i,j}(t))F_{i,j}(t), t > 0 \qquad (2)$$

Оператор (4) включає в себе константу n_{dsbn}, яка описує можливі дисбаланси між вхідними та вихідними потоками сировини.

Кожен вузол піддається впливу ресурсів, вироблених на чотирьох сусідніх вузлах – по два вузла в кожному напрямку, розділених однаковою відстанню Δ.

$$\hat{S}\{V_{i,j}\} = \begin{cases} \Delta^{-2}\left[D_{1,2}^{1,1}V_{1,2} + D_{2,1}^{1,1}V_{2,1} - n_{dsbn}D_{1,1}^{1,2}V_{1,1} - n_{dsbn}D_{1,1}^{2,1}V_{1,1}\right], & i, j = 1 \\ \Delta^{-2}\left[D_{2,j}^{1,j}V_{2,j} + D_{1,j-1}^{1,j}V_{1,j-1} + D_{1,j+1}^{1,j}V_{1,j+1} - n_{dsbn}D_{1,j}^{2,j}V_{1,j} - n_{dsbn}D_{1,j}^{1,j-1}V_{1,j} - n_{dsbn}D_{1,j}^{1,j+1}V_{1,j}\right], & i = 1,\ j \in \overline{2, N-1} \\ \Delta^{-2}\left[D_{1,N-1}^{1,N}V_{1,N-1} + D_{2,N}^{1,N}V_{2,N} - n_{dsbn}D_{1,N}^{1,N-1}V_{1,N} - n_{dsbn}D_{1,N}^{2,N}V_{1,N}\right], & i = 1,\ j = N \\ \Delta^{-2}\left[D_{i-1,N}^{i,N}V_{i-1,N} + D_{i+1,N}^{i,N}V_{i+1,N} + D_{i,N-1}^{i,N}V_{i,N-1} - n_{dsbn}D_{i,N}^{i-1,N}V_{i,N} - n_{dsbn}D_{i,N}^{i+1,N}V_{i,N} - n_{dsbn}D_{i,N}^{i,N-1}V_{i,N}\right], & i \in \overline{2, N-1},\ j = N \\ \Delta^{-2}\left[D_{N-1,N}^{N,N}V_{N-1,N} + D_{N,N-1}^{N,N}V_{N,N-1} - n_{dsbn}D_{N,N}^{N-1,N}V_{N,N} - n_{dsbn}D_{N,N}^{N,N-1}V_{N,N}\right], & i = N,\ j = N \\ \Delta^{-2}\left[D_{N-1,j}^{N,j}V_{N-1,j} + D_{N,j-1}^{N,j}V_{N,j-1} + D_{N,j+1}^{N,j}V_{N,j+1} - n_{dsbn}D_{N,j}^{N-1,j}V_{N,j} - n_{dsbn}D_{N,j}^{N,j-1}V_{N,j} - n_{dsbn}D_{N,j}^{N,j+1}V_{N,j}\right], & i = N,\ j \in \overline{2, N-1} \\ \Delta^{-2}\left[D_{N-1,1}^{N,1}V_{N-1,1} + D_{N,2}^{N,1}V_{N,2} - n_{dsbn}D_{N,1}^{N-1,1}V_{N,1} - n_{dsbn}D_{N,1}^{N,2}V_{N,1}\right], & i = N,\ j = 1 \\ \Delta^{-2}\left[D_{i-1,1}^{i,1}V_{i-1,1} + D_{i+1,1}^{i,1}V_{i+1,1} + D_{i,2}^{i,1}V_{i,2} - n_{dsbn}D_{i,1}^{i-1,1}V_{i,1} - n_{dsbn}D_{i,1}^{i+1,1}V_{i,1} - n_{dsbn}D_{i,1}^{i,2}V_{i,1}\right], & i \in \overline{2, N-1},\quad j = 1 \\ \Delta^{-2}\left[D_{i-1,j}^{i,j}V_{i-1,j} + D_{i+1,j}^{i,j}V_{i+1,j} + D_{i,j-1}^{i,j}V_{i,j-1} + D_{i,j+1}^{i,j}V_{i,j+1} - n_{dsbn}D_{i,j}^{i-1,j}V_{i,j} - n_{dsbn}D_{i,j}^{i+1,j}V_{i,j} - n_{dsbn}D_{i,j}^{i,j-1}V_{i,j} - n_{dsbn}D_{i,j}^{i,j+1}V_{i,j}\right], & i, j \in \overline{2, N-1} \end{cases}$$

(4)

Використовується гранична умова $V_{i,j} = 0$ для індексів масиву $i, j = 0, N+1$.

Висновки: В даній статті проведено аналітичний огляд публікацій по проектуванню виробничої мережі, зокрема його актуальності на сучасних виробництвах. Проведено аналіз робіт по роках в наукометричній базі Scopus. Представлено авторів, які опублікували найбільшу кількість робіт, включаючи країни та навчальні заклади. А також, які саме публікації найбільше публікували та у якій галузі науки був до них найбільший інтерес. Представлено спонсорські фонди, які зробили найбільший вклад у видавництво даних наукових робіт. Згідно даних отриманих під час аналітичного огляду літературних джерел за допомогою програми у наукометричній базі Scopus можна зробити висновок, що є велика зацікавленість проблемою проектування виробничих мереж та тема є актуальною, і хоча протягом останнього десятиліття інтерес до неї лише зростає вона є мало дослідженою. А також, що велика кількість статей публікується у різних країнах світу, що

© Марценюк В.П., Кіт Н.В.

свідчить про те, що вони зацікавлені у проектуванні виробничих мереж та автоматизації виробництв.

Також для вирішення проблем, які виникають під час транспортування сировини між логістичними ділянками була розроблена та запропонована решітчаста модель на прямокутній решітці з використанням решітчастих диференціальних рівнянь із затримкою, яка базується на ряді припущень та дозволяє визначити час, витрачений на виробництво одиниці продукції та уникнути проблем, які часто виникають при транспортуванні.

В подальших дослідженнях планується продовжувати досліджувати дану тематику, бо хоч вона і актуальна і до неї є інтерес, але вона ще мало досліджена наковцями, а також опублікувати результати чисельного моделювання решітчастої моделі на прямокутній решітці з використанням решітчастих диференціальних рівнянь із затримкою.

Список використаних джерел

1. Jelena Milisavljevic-Syed, Janet K. Allen, Sesh Commuri, Farrokh Mistree, Design of networked manufacturing systems for Industry 4.0, Procedia CIRP, Volume 81, 2019, Pages 1016-1021, https://doi.org/10.1016/j.procir.2019.03.244.
2. Mourtzis D., Doukas M., Psarommatis F., Manufacturing Network Design for Mass Customisation using a Genetic Algorithm and an Intelligent Search Method, Procedia CIRP, vol. 7, 2013, pp. 37–42 https://doi.org/10.1016/j.procir.2013.05.007
3. D. Mourtzis, E. Vlachou, N. Boli, L. Gravias, C. Giannoulis, Manufacturing Networks Design through Smart Decision Making towards Frugal Innovation, Procedia CIRP, Volume 50, 2016, Pages 354-359, ISSN 2212-8271, https://doi.org/10.1016/j.procir.2016.04.166.
4. Kim, J. H. (2017). A Review of Cyber-Physical System Research Relevant to the Emerging IT Trends: Industry 4.0, IoT, Big Data, and Cloud Computing. In Journal of Industrial Integration and Management (Vol. 02, Issue 03, p. 1750011). World Scientific Pub Co Pte Lt. https://doi.org/10.1142/s2424862217500117
5. Mosterman, P.J., Zander, J. Industry 4.0 as a Cyber-Physical System study. Softw Syst Model 15, 17–29 (2016). https://doi.org/10.1007/s10270-015-0493-x
6. Bagheri, B., Yang, S., Kao, H.-A., & Lee, J. (2015). Cyber-physical Systems Architecture for Self-Aware Machines in Industry 4.0 Environment. In IFAC-PapersOnLine (Vol. 48, Issue 3, pp. 1622–1627). Elsevier BV. https://doi.org/10.1016/j.ifacol.2015.06.318
7. Leng, J., Guo, J., Xie, J., Zhou, X., Liu, A., Gu, X., Mourtzis, D., Qi, Q., Liu, Q., Shen, W., & Wang, L. (2024). Review of manufacturing system design in the interplay of Industry 4.0 and Industry 5.0 (Part I): Design thinking and modeling methods. In Journal of Manufacturing Systems (Vol. 76, pp. 158–187). Elsevier BV. https://doi.org/10.1016/j.jmsy.2024.07.012
8. Leng, J., Zhu, X., Huang, Z., Li, X., Zheng, P., Zhou, X., Mourtzis, D., Wang, B., Qi, Q., Shao, H., Wan, J., Chen, X., Wang, L., & Liu, Q. (2024). Unlocking the power of industrial artificial intelligence towards Industry 5.0: Insights, pathways, and challenges. In Journal of Manufacturing Systems (Vol. 73, pp. 349–363). Elsevier BV. https://doi.org/10.1016/j.jmsy.2024.02.010
9. Liu, S., Zhang, J., Yi, S., Gao, R., Mourtzis, D., & Wang, L. (2024). Human-centric systems in smart manufacturing. In Manufacturing from Industry 4.0 to Industry 5.0 (pp. 181–205). Elsevier. https://doi.org/10.1016/b978-0-443-13924-6.00006-5
10. Mourtzis, D., & Sgarbossa, F. (2024). Operations management and information systems for smart manufacturing. In Manufacturing from Industry 4.0 to Industry 5.0 (pp. 267–288). Elsevier. https://doi.org/10.1016/b978-0-443-13924-6.00009-0
11. Mourtzis, D., & Angelopoulos, J. (2024). Artificial intelligence for human–cyber-physical production systems. In Manufacturing from Industry 4.0 to Industry 5.0 (pp. 343–378). Elsevier. https://doi.org/10.1016/b978-0-443-13924-6.00012-0
12. Mourtzis, D., Ong, S. K., Wang, X. V., Panopoulos, N., Stark, R., & Wang, L. (2024). Modelling, Design and Simulation as-a-Service Based on Extended Reality (XR) in Industry 4.0. In Lecture Notes in Mechanical Engineering (pp. 99–143). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54034-9_4
13. Mourtzis, D., Panopoulos, N., Stavropoulos, P., & Papakostas, N. (2024). Artificial Intelligence for Production Management and Control Towards Mass Personalization of Global Networks. In Lecture Notes in Mechanical Engineering (pp. 267–312). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54034-9_8
14. Abubakar, S. A., Sverstyuk, A. (2024). Analytical review of publications on machine learning methods in oncology and approach to evaluating their quality. In Computer systems and information technologies (Issue 1, pp. 6–16). Khmelnytskyi National University. https://doi.org/10.31891/csit-2024-1-1
15. Martsenyuk V.P., Sverstiuk A.S., Klos–Witkowska A., Kozodii N.V., Bagriy–Zayats O.A., Zubenko I.R. Numerical Analysis of Results Simulation of Cyber–physical Biosensor Systems. In 1st International Workshop Information–Communication Techologies&Embedded Systems. 14–15 November, Mykolaiv, 2019, Vol. 1. p. 149–164.

References

16. Jelena Milisavljevic-Syed, Janet K. Allen, Sesh Commuri, Farrokh Mistree, Design of networked manufacturing systems for Industry 4.0, Procedia CIRP, Volume 81, 2019, Pages 1016-1021, https://doi.org/10.1016/j.procir.2019.03.244.

© Марценюк В.П., Кіт Н.В.

17. Mourtzis D., Doukas M., Psarommatis F., Manufacturing Network Design for Mass Customisation using a Genetic Algorithm and an Intelligent Search Method, Procedia CIRP, vol. 7, 2013, pp. 37–42 https://doi.org/10.1016/j.procir.2013.05.007
18. D. Mourtzis, E. Vlachou, N. Boli, L. Gravias, C. Giannoulis, Manufacturing Networks Design through Smart Decision Making towards Frugal Innovation, Procedia CIRP, Volume 50, 2016, Pages 354-359, ISSN 2212-8271, https://doi.org/10.1016/j.procir.2016.04.166.
19. Kim, J. H. (2017). A Review of Cyber-Physical System Research Relevant to the Emerging IT Trends: Industry 4.0, IoT, Big Data, and Cloud Computing. In Journal of Industrial Integration and Management (Vol. 02, Issue 03, p. 1750011). World Scientific Pub Co Pte Lt. https://doi.org/10.1142/s2424862217500117
20. Mosterman, P.J., Zander, J. Industry 4.0 as a Cyber-Physical System study. Softw Syst Model 15, 17–29 (2016). https://doi.org/10.1007/s10270-015-0493-x
21. Bagheri, B., Yang, S., Kao, H.-A., & Lee, J. (2015). Cyber-physical Systems Architecture for Self-Aware Machines in Industry 4.0 Environment. In IFAC-PapersOnLine (Vol. 48, Issue 3, pp. 1622–1627). Elsevier BV. https://doi.org/10.1016/j.ifacol.2015.06.318
22. Leng, J., Guo, J., Xie, J., Zhou, X., Liu, A., Gu, X., Mourtzis, D., Qi, Q., Liu, Q., Shen, W., & Wang, L. (2024). Review of manufacturing system design in the interplay of Industry 4.0 and Industry 5.0 (Part I): Design thinking and modeling methods. In Journal of Manufacturing Systems (Vol. 76, pp. 158–187). Elsevier BV. https://doi.org/10.1016/j.jmsy.2024.07.012
23. Leng, J., Zhu, X., Huang, Z., Li, X., Zheng, P., Zhou, X., Mourtzis, D., Wang, B., Qi, Q., Shao, H., Wan, J., Chen, X., Wang, L., & Liu, Q. (2024). Unlocking the power of industrial artificial intelligence towards Industry 5.0: Insights, pathways, and challenges. In Journal of Manufacturing Systems (Vol. 73, pp. 349–363). Elsevier BV. https://doi.org/10.1016/j.jmsy.2024.02.010
24. Liu, S., Zhang, J., Yi, S., Gao, R., Mourtzis, D., & Wang, L. (2024). Human-centric systems in smart manufacturing. In Manufacturing from Industry 4.0 to Industry 5.0 (pp. 181–205). Elsevier. https://doi.org/10.1016/b978-0-443-13924-6.00006-5
25. Mourtzis, D., & Sgarbossa, F. (2024). Operations management and information systems for smart manufacturing. In Manufacturing from Industry 4.0 to Industry 5.0 (pp. 267–288). Elsevier. https://doi.org/10.1016/b978-0-443-13924-6.00009-0
26. Mourtzis, D., & Angelopoulos, J. (2024). Artificial intelligence for human–cyber-physical production systems. In Manufacturing from Industry 4.0 to Industry 5.0 (pp. 343–378). Elsevier. https://doi.org/10.1016/b978-0-443-13924-6.00012-0
27. Mourtzis, D., Ong, S. K., Wang, X. V., Panopoulos, N., Stark, R., & Wang, L. (2024). Modelling, Design and Simulation as-a-Service Based on Extended Reality (XR) in Industry 4.0. In Lecture Notes in Mechanical Engineering (pp. 99–143). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54034-9_4
28. Mourtzis, D., Panopoulos, N., Stavropoulos, P., & Papakostas, N. (2024). Artificial Intelligence for Production Management and Control Towards Mass Personalization of Global Networks. In Lecture Notes in Mechanical Engineering (pp. 267–312). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-54034-9_8
29. Abubakar, S. A., Sverstyuk, A. (2024). Analytical review of publications on machine learning methods in oncology and approach to evaluating their quality. In Computer systems and information technologies (Issue 1, pp. 6–16). Khmelnytskyi National University. https://doi.org/10.31891/csit-2024-1-1
30. Martsenyuk V.P., Sverstiuk A.S., Klos–Witkowska A., Kozodii N.V., Bagriy–Zayats O.A., Zubenko I.R. Numerical Analysis of Results Simulation of Cyber–physical Biosensor Systems. In 1st International Workshop Information–Communication Techologies&Embedded Systems. 14–15 November, Mykolaiv, 2019, Vol. 1. p. 149–164.

© Марценюк В.П., Кіт Н.В.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-27
УДК: 004.02
Марченко Олексій Олександрович, асистент
https://orcid.org/0000-0002-5080-4811
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

ПОДАЛЬШИЙ РОЗВИТОК ПОШУКУ ПО ДЕРЕВУ МЕТОДОМ МОНТЕ-КАРЛО ЗА ДОПОМОГОЮ КОНТРОЛЮ ФОРМИ ДЕРЕВА ПОШУКУ

Марченко О.О. Подальший розвиток пошуку по дереву методом Монте-Карло за допомогою контролю форми дерева пошуку. Представлення інформації певної задачі у вигляді дерева і пошук найкращих рішень в рамках цієї задачі з побудовою дерева рішень методом Монте-Карло, який отримав назву MCTS (Monte Carlo Tree Search) є одним з найкращих підходів для виконання пошуку для розв'язку задач, в яких знаходження найкращих рішень є дуже складним і часовитратним процесом. Стаття присвячена запропонованому раніше варіанту подальшого розвитку пошуку по дереву методом MCTS за допомогою контролю форми дерева пошуку. Цей варіант був названий MCTS-TSC (Monte-Carlo Tree Search with Tree Shape Control) і базується на критеріях виду DW (Depth-Width) з метою контролювання форми дерева пошуку під час його побудови. В статті пропонується два нових критерії виду DW: DWC і WDC, принцип впливу яких на форму дерева пошуку, що будується, відрізняється від раніше запропонованих критеріїв. Апробація використання нових запропонованих критеріїв контролю форми дерева DWC і WDC показала підвищення ефективності пошуку кращих подальших ходів гравця, який використовував варіант пошуку MCTS-TSC з контролем форми дерева, порівняно до гравця, що використовував стандартний варіант пошуку MCTS-UCT. Гравець MCTS-TSC при використанні критерію DWC виграв у гравця MCTS-UCT на 55% партій більше, а при використанні критерію WDC – на 42% більше, що суттєво краще, ніж у запропонованих раніше критеріїв. У якості висновків зазначено, що оскільки формули обчислення критеріїв виду DW є нескладними, то їх реалізація для контролю форми дерева не буде суттєво впливати на час пошуку варіантом MCTS-TSC. Крім того, відзначено, що запропоновані нові критерії виду DW можуть бути також використані для більш ефективного розпаралелення процесу пошуку MCTS і більш ефективного використання наявних апаратних ресурсів. Запропоновані також можливі напрямки подальшого розвитку пошуку методом Монте-Карло з контролем форми дерева пошуку MCTS-TSC.

Ключові слова: пошук в дереві методом Монте-Карло, MCTS, MCTS-UCT, контроль форми дерева пошуку, критерії виду «глибина-ширина», MCTS-TSC.

Marchenko O. Further development of the Monte Carlo tree search method by controlling the search tree shape. Representing the information of a certain problem in the form of a tree and searching for the best solutions within this problem with the construction of a decision tree by the Monte Carlo method, which was named MCTS (Monte Carlo Tree Search), is one of the best approaches for performing a search for solving problems in which finding the best solutions is a very difficult and time-consuming process. The article is devoted to the previously proposed variant of further development of the tree search by the MCTS method by means of controlling the shape of the search tree. This variant was named MCTS-TSC (Monte-Carlo Tree Search with Tree Shape Control) and is based on the DW (Depth-Width) kind criteria in order to control the shape of the search tree during its construction. The article proposes two new DW kind criteria: DWC and WDC, the principle of which affects the shape of the search tree in a different way from the previously proposed criteria. Approbation of the use of the new proposed tree shape control criteria DWC and WDC showed an increase in the efficiency of searching for the best subsequent moves of the player who used the MCTS-TSC search variant with tree shape control compared to the player who used the standard MCTS-UCT search variant. The MCTS-TSC player won 55% more games than the MCTS-UCT player when using the DWC criterion, and 42% more when using the WDC criterion, which is significantly better than the previously proposed criteria showed. As conclusions, it is stated that since the formulas for calculating the criteria of the DW kind are simple, their implementation for controlling the shape of the tree will not significantly affect the search time of the MCTS-TSC variant. In addition, it was noted that the proposed new DW kind criteria can also be used for more efficient parallelization of the MCTS search process and more efficient use of available hardware resources. Possible directions for the further development of the MCTS-TSC search using the Monte Carlo method with the control of the of the search tree shape are also proposed.

Keywords: Monte-Carlo tree search method, MCTS, MCTS-UCT, search tree shape control, criteria of the "Depth-Width" kind, MCTS-TSC.

Постановка наукової проблеми

Представлення інформації певної задачі у вигляді дерева і пошук найкращих рішень в рамках цієї задачі з побудовою дерева рішень методом Монте-Карло (Monte Carlo Tree Search – MCTS) [1] зарекомендувало себе як один з найкращих підходів для розв'язку задач із галузі штучного інтелекту. Такі задачі є дуже вимогливими до швидкості знаходження якомога кращих поточних рішень, що виникають при розв'язку цих задач, оскільки інформація, яку потрібно проаналізувати, має критично великий обсяг. За час після того, як метод MCTS був запропонований вперше, було розроблено багато варіантів подальшого розвитку цього методу у різних напрямках: модифікації його основної формули UCB1 з орієнтацією на певні класи задач, пропонування нових варіантів виконання основних чотирьох етапів цього пошуку, зберігання додаткової інформації в процесі

© Марченко О.О.

пошуку для підвищення його ефективності, комбінування MCTS з методами машинного навчання, розпаралелення процесу роботи основних етапів роботи MCTS тощо. Але теоретична потужність методу MCTS, а також все зростаючі складність задач штучного інтелекту та обсяги інформації цих задач все ще залишають великий простір для подальшого розвитку методу MCTS у всіх зазначених напрямках. Одному з таких напрямків подальшого розвитку методу MCTS, а саме підвищенню ефективності цього методу за допомогою контролю форми дерева пошуку, присвячена ця стаття.

Аналіз досліджень

Пошуку по дереву методом MCTS було присвячено вже багато публікацій, виконаних як різними дослідниками, так і автором разом зі співавторами. Одними з кращих публікацій оглядового характеру щодо підходів та способів вдосконалення MCTS є роботи [1-8]. Дослідження взаємозв'язку форми дерева пошуку MCTS і ефективності знаходження цим пошуком найкращих подальших дій, крім авторських робіт, виконувалися також в [6-8]. Зокрема, в дослідженні [6] були використані такі параметри дерева пошуку як глибина і поточний степінь розгалуження вершин дерева, а в дослідженні [7] було відзначено, що (в перекладі з англійської) «… в залежності від кількості вершин в деревах, їх глибина і ширина можуть варіюватися, дозволяючи нам відповісти на питання, чи надає MCTS перевагу якимсь одним перед іншими. … ми з'ясували, що від самої гри залежить, чи надає MCTS перевагу глибоким деревам, чи великому степеню розгалуження на вершинах (тобто широким деревам), чи певному балансу між ними. Виявилося, що невеликі нюанси в правилах гри і системі оцінювання результату можуть змінювати форму дерев, яким надає перевагу MCTS».

Автор разом зі співавторами вже виконували раніше дослідження цього напрямку, зокрема у [9] був запропонований варіант подальшого розвитку методу MCTS, що базується на цій ідеї і який був названий пошуком по дереву методом Монте-Карло з контролем форми дерева (Monte-Carlo Tree Search with Tree Shape Control – MCTS-TSC), а також були запропоновані деякі критерії виду «глибина-ширина» для контролю форми дерева пошуку під час його побудови. Але наразі цей напрям досліджень отримав новий розвиток, який і описується в пропонованій статті.

Метою даної роботи є розробка нового варіанту подальшого розвитку пошуку по дереву методом з контролем форми дерева пошуку MCTS-TSC на основі нових критеріїв контролю форми дерева пошуку виду «глибина-ширина», принцип дії яких відрізняється від раніше запропонованих критеріїв.

Подальший розвиток пошуку по дереву методом Монте-Карло за допомогою контролю форми дерева пошуку (MCTS-TSC)

Головна ідея подальшого розвитку методу MCTS, що пропонується, а також критеріїв виду «глибина-ширина», була викладена раніше в [9]. Розглянемо коротко цю ідею. Якщо ми знаємо, яка типова форма дерева пошуку MCTS відповідає найбільш ефективному пошуку цим методом для конкретної гри чи іншої прикладної задачі, і ця форма є, наприклад, ширшою, ніж форма дерева, яке побудоване процесом пошуку MCTS в певний момент, то ми можемо спробувати перелаштувати параметри формули UCB1 MCTS таким чином, щоб подальшу побудову дерева спрямувати в сторону отримання більш глибокої форми дерева, і навпаки. В результаті форма дерева MCTS надалі стане більш характерною для цієї гри (задачі) і ефективність пошуку MCTS може покращитися. Оскільки ця ідея ґрунтується на контролі форми дерева під час процесу пошуку, то саме тому такий варіант подальшого розвитку MCTS був названий MCTS-TSC (MCTS with Tree Shape Control).

Форму дерева пошуку MCTS визначимо у вигляді прямокутника зі сторонами Depth та Width (рис. 1), у який це дерево може бути вписаним: Depth – глибина дерева (максимальна кількість рівнів дерева – позначено синім кольором), Width – ширина дерева (максимальна сумарна кількість вершин на всіх рівнях по горизонталі – позначено червоним кольором).

Для реалізації подальшого розвитку методу MCTS за допомогою контролю форми дерева пошуку, ідея якого описана вище, пропонується використовувати спеціальні критерії виду «глибина-ширина» DW (Depth-Width), які є формулами для оцінювання поточної форми дерева пошуку.

Критерії виду DW на кожній ітерації етапу вибору MCTS обчислюються для піддерев всіх вершин, по яких відбувається спуск на етапі вибору, від кореня до вершини, на якій зупиниться процес етапу вибору. Таке обчислення виконується на основі двох властивостей форми піддерев цих вершин – поточної глибини піддерева вершини ND (NodeDepth) та поточної ширини піддерева вершини NW (NodeWidth).

© Марченко О.О.

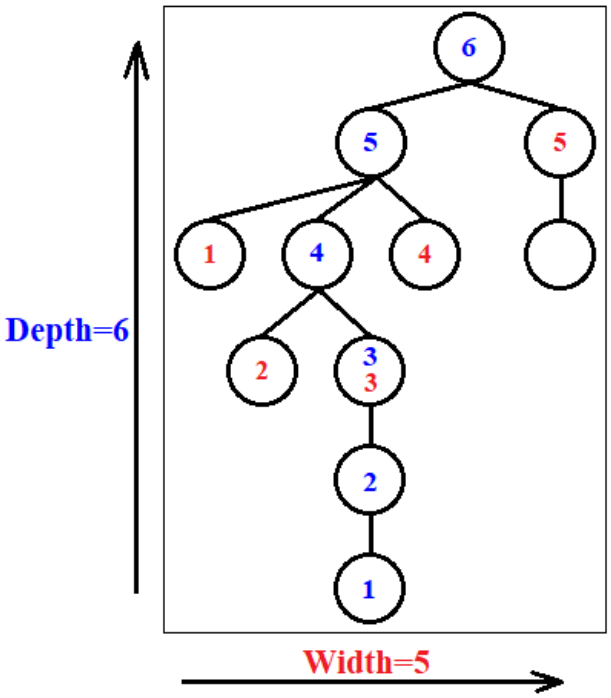


Рис. 1. Форма дерева пошуку MCTS

Процес обчислення значень властивостей ND та NW для вершин дерева під час його побудови описано в [9]. В цій же роботі для контролю форми та керування формою піддерев вершин (спрямовування подальшої побудови дерева у потрібному напрямку) на основі обчислених властивостей (ND та NW) були запропоновані деякі критерії виду DW виду «глибина-ширина», використання яких дало позитивний ефект і підвищило ефективність пошуку MCTS-TSC.

В більш пізніх дослідженнях виникла ідея подальшого розвитку контролю форми дерева пошуку MCTS на основі нових критеріїв виду DW з іншим принципом впливу на форму дерева, що будується. Ця стаття присвячена розгляду саме цього подальшого розвитку методу MCTS-TSC.

Відповідно до описаної вище ідеї контролю та керування формою піддерев пропонується підхід, який, на відміну від попередніх критеріїв, полягає у додаванні обчисленого значення критерію виду DW безпосередньо до параметра C_{UCB1} у формулі (1): UCB1_TSC (UCB1 with Tree Shape Control), а не до усього значення UCB1_TSC, як було у попередніх критеріях [9]. Тобто, нові критерії використовують загальну властивість параметра впливати на форму дерева, що будується [8] (відзначено в аналізі статті), і скеровують вплив цього параметра або в сторону поглиблення дерева, або в сторону розширення дерева.

$$UCB1_TSC = \frac{NodeWins}{NodePlayouts} + 2 * (C_{UCB1} + \boldsymbol{DW}) * \sqrt{2 * \frac{\ln(ParentPlayouts)}{NodePlayouts}} \quad (1)$$

Пропонується два нових простих критерії виду DW, які обчислюються за формулами (2) та (3) і значення яких додаються до параметра C_{UCB1} формули (1).

$$DWC = C_R * \left(\frac{ND}{NW} - C_{DW}\right) \quad (2)$$

$$WDC = -C_R * \left(\frac{NW}{ND} - C_{WD}\right) = C_R * \left(C_{WD} - \frac{NW}{ND}\right) \quad (3)$$

де

DWC (Depth/Width Criterion) – критерій «глибина/ширина»;

C_{DW} – параметр (коефіцієнт) виду «глибина/ширина» бажаної форми дерева;

WDC (Width/Depth Criterion) – критерій «ширина/глибина»;

C_{WD} – параметр (коефіцієнт) виду «ширина/глибина» бажаної форми дерева;

C_R – параметр (коефіцієнт) регуляції (зменшення/збільшення) впливу критерію DWC чи WDC.

Параметр C_{DW} визначає пропорцію «глибина/ширина» бажаної форми дерева, до якої потрібно скеровувати побудову дерева під час пошуку і обчислюється за формулою (4):

© Марченко О.О.

$$C_{DW} = \frac{DFD}{DFW} \quad (4)$$

А параметр C_{WD}, відповідно, визначає пропорцію «ширина/глибина» бажаної форми дерева, до якої потрібно скеровувати побудову дерева під час пошуку і обчислюється за формулою (5):

$$C_{WD} = \frac{DFW}{DFD} \quad (5)$$

де
DFD (Desired Form Depth) – глибина бажаної форми дерева;
DFW (Desired Form Width) – ширина бажаної форми дерева.

Критерії DWC та WDC є подібними один до одного. Різниця між ними полягає в тому, яка із властивостей, глибина чи ширина, знаходиться у чисельнику, а яка – у знаменнику. Відповідно, значення критерію DWC залежить від глибини дерева прямо пропорційно, а від ширини – оберненопропорційно, а значення критерію WDC – навпаки. В цих критеріях також навпаки розташовані доданки. У критерію DWC задана параметром C_{DW} бажана пропорція форми дерева віднімається від пропорції значень глибини та ширини піддерева поточної вершини, а у критерію WDC – навпаки, пропорція значень поточної вершини віднімається від заданої параметром C_{WD} бажаної пропорції форми дерева. Це забезпечує однаковий знак значень критеріїв DWC та WDC для однакових форм піддерев поточної вершини дерева відносно заданої бажаної форми дерева, а різниця між значеннями цих критеріїв залишиться тільки у прямій чи оберненій залежності від глибини чи ширини дерева.

При контролі форми дерева за допомогою формули UCB1_TSC (формула (1)) важливо, щоб знак критеріїв виду DW при відступах поточної пропорції глибини та ширини від заданої був саме таким, а не навпаки. Це визначається зазначеною вище роллю параметра C_{UCB1}: якщо параметру C_{UCB1} встановлюють менші значення, то дерево буде будуватися більш глибоким і вузьким, а якщо цьому параметру встановлювати більші значення, то дерево буде будуватися ширшим. Тому, якщо форма поточного дерева ширша, ніж задана пропорція форми дерева, то значення критерію DW повинно бути від'ємним, щоб скерувати побудову дерева до більш глибокої форми, ніж це визначається параметром C_{UCB1}, а якщо форма поточного дерева задовга, то навпаки значення DW повинно бути додатним.

На рис. 2 показані дві можливі ситуації поточної форми дерева пошуку (червоний прямокутник) відносно заданої форми дерева (чорний прямокутник), до якої поточну форму бажано скерувати, та відповідні знаки значень критеріїв DWC та WDC, які ці критерії будуть мати у таких ситуаціях.

Рис. 2. Ситуації поточної форми дерева пошуку відносно заданої форми дерева

Ця властивість критеріїв DW показана також у таблиці 1 на прикладі критерію DWC при заданій пропорції форми дерева C_{DW} = ND/NW = 2. Зауважимо, що ця пропорція взята винятково для демонстрації принципу роботи критеріїв, а для конкретних реальних прикладних задач та ігор може сильно відрізнятися як в одну сторону, так і в другу.

© Марченко О.О.

Таблиця 1. Значення критерію DWC при заданій пропорції форми дерева $C_{DW}=2$

DWC $C_{DW}=2$		**NW**					
		1	**2**	**3**	**4**	**5**	**6**
ND	**1**	-1,0000	-1,5000	-1,6667	-1,7500	-1,8000	-1,8333
	2	**0,0000**	-1,0000	-1,3333	-1,5000	-1,6000	-1,6667
	3	1,0000	-0,5000	-1,0000	-1,2500	-1,4000	-1,5000
	4	2,0000	**0,0000**	-0,6667	-1,0000	-1,2000	-1,3333
	5	3,0000	0,5000	-0,3333	-0,7500	-1,0000	-1,1667
	6	4,0000	1,0000	**0,0000**	-0,5000	-0,8000	-1,0000
	7	5,0000	1,5000	0,3333	-0,2500	-0,6000	-0,8333
	8	6,0000	2,0000	0,6667	**0,0000**	-0,4000	-0,6667
	9	7,0000	2,5000	1,0000	0,2500	-0,2000	-0,5000
	10	8,0000	3,0000	1,3333	0,5000	**0,0000**	-0,3333
	11	9,0000	3,5000	1,6667	0,7500	0,2000	-0,1667
	12	10,0000	4,0000	2,0000	1,0000	0,4000	**0,0000**

В таблиці червоним кольором виділено, що при збігу пропорцій поточної форми дерева із пропорцією заданої форми (2/1, 4/2, 6/3 і т.д.) значення критерію DWC дорівнює нулю, тобто корекція параметра C_{UCB1} буде відсутня. Також неважко помітити, що при ND>NW, значення критерію DWC є додатними, тобто параметр C_{UCB1} буде скоригований в сторону розширення дерева, а при ND<NW, значення критерію DWC є від'ємними, тобто параметр C_{UCB1} буде скоригований в сторону поглиблення дерева. Аналогічно буде і для критерію WDC.

Апробація

Для апробації запропонованих нових критеріїв була виконана множина партій гри Connect Four [10] між гравцями, один з яких грав, використовуючи варіант пошуку по дереву з контролем форми дерева MCTS-TSC, а другий гравець грав, використовуючи стандартний варіант MCTS-UCT. Було виконано сумарно по 120 партій для критерію DWC та критерію WDC при попередньо визначеному найкращому для гри Connect Four значенні параметра $C_{UCB1}=0.55$, різних значеннях параметрів C_{DW}, C_{WD}, C_R критеріїв DWC і WDC відповідно, та різному обчислювальному бюджеті на знаходження наступного ходу. Було проведено по 20 партій при кожному з шести обчислювальних бюджетів (кількості ітерацій пошуку): 50000, 1000000, 150000, 200000, 250000, 300000. В десяти партіях з двадцяти першим ходив гравець MCTS-TSC, а ще в десяти партіях першим ходив гравець MCTS-UCT. Сумарні результати проведення цих партій, що наведені на рис. 3, підтверджують ефективність застосування нових критеріїв DWC та WDC.

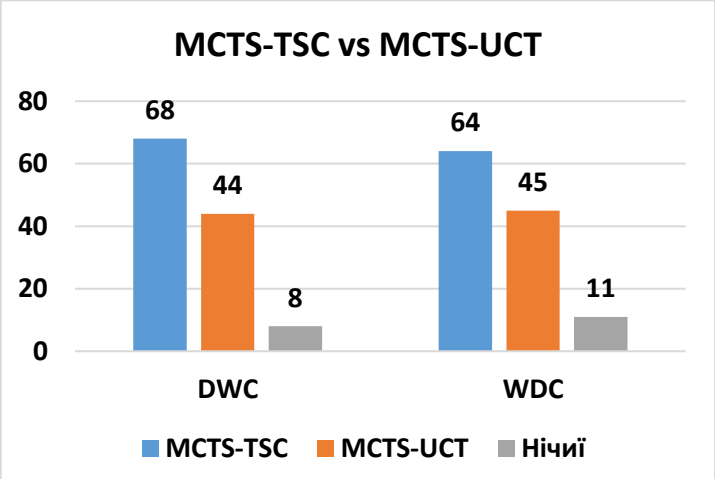


Рис. 3. Результати експериментальної перевірки ефективності критеріїв DWC та WDC

© Марченко О.О.

Гравець MCTS-TSC при використанні критерію DWC виграв у гравця MCTS-UCT на 55% партій більше, а при використанні критерію WDC – на 42% більше, що суттєво краще, ніж у запропонованих раніше критеріїв.

Автором планується, що розширеній експериментальній перевірці ефективності цих критеріїв та порівняльній характеристиці отриманих результатів буде присвячена окрема стаття.

Висновки

В статті запропоновані нові критерії виду DW («глибина-ширина») для контролю форми дерева пошуку методу MCTS. Ці критерії дозволяють контролювати форму дерева гри під час процесу пошуку, тобто оцінювати поточну форму і спрямовувати подальшу побудову дерева у напрямку до бажаної для певної задачі форми.

Якщо завчасно відомо, яка форма дерева пошуку (вузька, «скелетоподібна», широка, неглибока, глибока, дуже глибока тощо) є типовою для ефективного пошуку кращих ходів для конкретної гри чи прикладної задачі, то, враховуючи таку інформацію, з'являються нові можливості для динамічного контролю форми дерева пошуку, що дозволяє скерувати подальший процес побудови дерева в потрібному напрямку і, в результаті, підвищити ефективність пошуку кращих наступних дій (ходів) методом MCTS.

Оскільки формули обчислення критеріїв виду DW є нескладними, то їх реалізація для контролю форми дерева не буде суттєво впливати на час пошуку варіантом MCTS-TSC.

Запропоновані критерії виду DW можуть бути також використані для більш ефективного розпаралелення процесу пошуку MCTS і більш ефективного використання наявних апаратних ресурсів.

Напрямки подальших досліджень

Окрім запропонованих в цій статті нових критеріїв виду DW, можливі й інші підходи до подальшого розвитку пошуку MCTS за допомогою контролю форми дерева пошуку. Наприклад, можна запропонувати підхід, який полягає у використанні значень критеріїв DW на етапі вибору для прийняття рішення, чи зупиняти вибір вершини і починати виконувати етапи розширення та моделювання, чи продовжувати процес етапу вибору вниз по дереву. Ще одним з варіантів застосування контролю форми дерева TSC, що базується на критеріях виду DW, може бути також його сумісне використання з іншими варіантами подальшого розвитку методу пошуку MCTS.

Крім того, параметри C_{DW}, C_{WD} та C_R, що контролюють вплив форми дерева, можуть бути встановлені як однакові константи для всіх вершин, або як різні константи для вершин, що знаходяться на різних рівнях глибини (прогресивні значення), або як змінні, що залежать від деяких евристик та знань предметної області, або натреновані за допомогою деякого способу машинного навчання. Іншим варіантом керування параметрами C_{DW}, C_{WD} та C_R може бути встановлення їм більших значень для вершин, що мають багато ще невідвіданих та/або мало відвіданих вершин-нащадків, для того, щоб отримати більш широку форму дерева пошуку з кращим виконанням дослідження.

Варто також зазначити, що використання контролю форми дерева MCTS-TSC для налаштування формули UCB1 не виключає одночасного використання інших способів і, швидше за все, найкраще застосування MCTS-TSC буде якраз сумісно з іншими налаштовуючими, відсікаючими та навчаючими способами, включно зі способами, що використовують деякі критерії, орієнтовані на форму дерева.

Всі ці варіанти подальшого розвитку методу MCTS можуть бути предметом подальших досліджень.

Список бібліографічного опису

1. Cameron Browne, Edward Powley, Daniel Whitehouse, Simon M. Lucas, Peter I. Cowling, et al., "A Survey of Monte Carlo Tree Search Methods", *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, pp. 1–43, March 2012. doi: 10.1109/TCIAIG.2012.2186810.
2. Marco Kemmerling, Daniel Lütticke, and Robert H. Schmitt, "Beyond games: a systematic review of neural Monte Carlo tree search Applications", *Applied Intelligence*, vol.54, 2024, pp. 1020–1046. doi: 10.1007/s10489-023-05240-w.
3. Jorik Jooken, Pieter Leyman, Tony Wauters, Patrick De Causmaecker, "Exploring search space trees using an adapted version of Monte Carlo tree search for combinatorial optimization problems", *Computers & Operations Research*, ISSN: 0305-0548, vol. 150, pp. 1–46. doi: 10.1016/j.cor.2022.106070.
4. T. Ou, Y. Lu, X. Wu and J. Cao, "Monte Carlo Tree Search: A Survey of Theories and Applications," *2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, Xi'an, China,

© Марченко О.О.

2022, pp. 388-396, doi: 10.1109/ICBAIE56435.2022.9985899.
5. G. M. J.-B. Chaslot, M. H. M. Winands, and H.J. van den Herik, "Parallel Monte-Carlo Tree Search", *Computers and Games. CG 2008. Lecture Notes in Computer Science*, vol. 5131, pp.60–71, 2008, Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-87608-3_6.
6. Gareth M. J. Williams, "Determining Game Quality Through UCT Tree Shape Analysis", *MSc. Thesis. Imperial College London,* 2010. [Online]. Available: https://www.doc.ic.ac.uk/teaching/distinguished-projects/2010/g.williams.pdf. Accessed on: Sep. 05, 2024.
7. Hilmar Finnsson and Yngvi Björnsson, "Game-Tree Properties and MCTS Performance", *GIGA 2011: Proceedings of the 2nd International General Game Playing Workshop,* Jan. 2011, pp. 23–30. Available: http://surl.li/hlmyrs. Accessed on: Sep. 05, 2024
8. Joel Veness, Kee Siong Ng, Marcus Hutter, William Uther, and David Silver, "A Monte-Carlo AIXI Approximation", *Journal of Artificial Intelligence Research*, vol. 40, no. 1, pp. 95–142, January 2011. Available: https://dl.acm.org/doi/abs/10.5555/2016945.2016949. Accessed on: Sep. 05, 2024.
9. Oleksandr I. Marchenko and Oleksii O. Marchenko, "Monte-Carlo tree search with tree shape control", *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Kyiv, Ukraine, 2017, pp. 812-817, doi: 10.1109/UKRCON.2017.8100358.
10. Connect Four, *Wikipedia, The Free Encyclopedia*. Available: http://surl.li/zhdnfm. Accessed on: Sep. 05, 2024.

References

1. Cameron Browne, Edward Powley, Daniel Whitehouse, Simon M. Lucas, Peter I. Cowling, et al., "A Survey of Monte Carlo Tree Search Methods", *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, no. 1, pp. 1–43, March 2012. doi: 10.1109/TCIAIG.2012.2186810.
2. Marco Kemmerling, Daniel Lütticke, and Robert H. Schmitt, "Beyond games: a systematic review of neural Monte Carlo tree search Applications", *Applied Intelligence*, vol.54, 2024, pp. 1020–1046. doi: 10.1007/s10489-023-05240-w.
3. Jorik Jooken, Pieter Leyman, Tony Wauters, Patrick De Causmaecker, "Exploring search space trees using an adapted version of Monte Carlo tree search for combinatorial optimization problems", *Computers & Operations Research*, ISSN: 0305-0548, vol. 150, pp. 1–46. doi: 10.1016/j.cor.2022.106070.
4. T. Ou, Y. Lu, X. Wu and J. Cao, "Monte Carlo Tree Search: A Survey of Theories and Applications," *2022 3rd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, Xi'an, China, 2022, pp. 388-396, doi: 10.1109/ICBAIE56435.2022.9985899.
5. G. M. J.-B. Chaslot, M. H. M. Winands, and H.J. van den Herik, "Parallel Monte-Carlo Tree Search", *Computers and Games. CG 2008. Lecture Notes in Computer Science*, vol. 5131, pp.60–71, 2008, Springer, Berlin, Heidelberg. doi: 10.1007/978-3-540-87608-3_6.
6. Gareth M. J. Williams, "Determining Game Quality Through UCT Tree Shape Analysis", *MSc. Thesis. Imperial College London,* 2010. [Online]. Available: https://www.doc.ic.ac.uk/teaching/distinguished-projects/2010/g.williams.pdf. Accessed on: Sep. 05, 2024.
7. Hilmar Finnsson and Yngvi Björnsson, "Game-Tree Properties and MCTS Performance", *GIGA 2011: Proceedings of the 2nd International General Game Playing Workshop,* Jan. 2011, pp. 23–30. Available: http://surl.li/hlmyrs. Accessed on: Sep. 05, 2024
8. Joel Veness, Kee Siong Ng, Marcus Hutter, William Uther, and David Silver, "A Monte-Carlo AIXI Approximation", *Journal of Artificial Intelligence Research*, vol. 40, no. 1, pp. 95–142, January 2011. Available: https://dl.acm.org/doi/abs/10.5555/2016945.2016949. Accessed on: Aug 14, 2024.
9. Oleksandr I. Marchenko and Oleksii O. Marchenko, "Monte-Carlo tree search with tree shape control", *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Kyiv, Ukraine, 2017, pp. 812-817, doi: 10.1109/UKRCON.2017.8100358.
10. Connect Four, *Wikipedia, The Free Encyclopedia*. Available: http://surl.li/zhdnfm. Accessed on: Sep. 05, 2024.

© Марченко О.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-28
УДК 004.42
Мороз Борис Іванович, д.т.н., професор,
https://orcid.org/0000-0002-5625-0864
Круглик Андрій Сергійович, аспірант
Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна.

ПРОБЛЕМИ ЗАСТОСУВАННЯ ГРУП БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ. КОНЦЕПТУАЛЬНІ ЗАСАДИ ЇХ ВИРІШЕННЯ

Мороз Б.І., Круглик А.С. Проблеми застосування груп безпілотних літальних апаратів. Концептуальні засади їх вирішення. У межах статті розглянуто можливості застосування груп безпілотних літальних апаратів військового призначення та проблеми, пов'язані з керуванням такими групами. Проаналізовано тенденції до застосування груп безпілотних літальних апаратів, визначено основні напрямки. Проаналізовано стан сучасних наробок у сфері застосування груп БПЛА для виконання завдання. Описано задачі, що висуваються до програмного забезпечення для керування групою літальних апаратів. Розглянуті неточності при використанні класичного підходу до керування групою ведучий – ведені. Описані шляхи вдосконалення і вирішення задач керування групою лідером групи. Розглянуті концептуальні шляхи вирішення проблем застосування груп БПЛА для успішного виконання завдання групою. Досліджено питання використання клієнт-серверної архітектури у комунікації групи БПЛА. Запропоновані шляхи вдосконалення такої архітектури саме для використання у цілях військового призначення, де комунікація між учасниками групи може бути порушена декількома факторами. Запропоновано дублювання ролей у групі, тобто модулів програмного забезпечення, для підтримки сталості групи у випадку виникнення непередбачуваної ситуації. Наведено приклади використання алгоритмів узгодження зміни лідера групи для продовження роботи групи у випадку втрати попереднього лідера. Описано проблеми використання групи у випадку неможливості комунікації всередині групи та перспективні шляхи вирішення таких проблем.
Ключові слова: рій, клієнт-серверна архітектура рою, керування групою, БПЛА, інтенсивність повідомлень, цінність повідомлення

Moroz B., Kruhlyk A. The possible issues when using groups of unmanned aircraft. Conceptual basis of their solution. The article describes the possibilities of using groups of unmanned aerial vehicles for military purposes. Also, the issues associated with the management of such groups are considered. The trends in the use of groups of unmanned aerial vehicles were analyzed, and the main directions were determined. The state of modern developments in the field of application of UAV groups for task performance is analyzed. The issues that related to the software for controlling a group of aircraft are described. Considered inaccuracies when using the classic approach to managing the leader-followed group. Conceptual ways of solving the issues of using groups of UAVs for the successful performance of the task by the group are considered. The question of using a client-server architecture in the communication of a group of UAVs has been investigated. Ways to improve such an architecture are proposed specifically for use in military purposes, where communication between group members may be disrupted by different factors. Duplication of group roles, i.e. software modules, is proposed to maintain group stability in the event of an unexpected situation. Examples of using algorithms for coordinating the change of the group leader for the continuation of the group's work in case of the loss of the previous leader are proposed. The issues of using a group in case of impossibility of communication within the group and promising ways of solving such issues are described.
Keywords: UAV swarm, UAV swarm client-server architecture, group control, UAV, message intensity, message value

Постановка проблеми. Останніми роками помітно виріс інтерес до використання безпілотних літальних апаратів у повсякденному житті. Звичайно, в першу чергу це стосується військової сфери, оскільки мабуть очевидно, що в умовах військової агресії з боку Російської Федерації проти України, абсолютне значення кількісної характеристики особового складу перевищує українську. Тому повинно гостро стояти питання впровадження безпілотних систем, особливо літальних, для мінімізації втрат особового складу. Окрім того, впровадження безпілотних систем дозволяє виконувати завдання, які неможливо або занадто ризиковано виконувати за допомогою безпосередньої участі людини. Саме тому останнім часом з'являється усе більше публікацій, де автори обговорюють концепції використання безпілотних (літальних) апаратів для розв'язання військових задач. Більш того, з'являються роботи, де науковці пропонують використовувати так звані рої або зграї безпілотних літальних апаратів для розв'язання поставленої задачі. Сукупність БПЛА , які одночасно використовуються для виконання завдання чи групи завдань, називається роєм. Такі групи БПЛА працюють в чіткій координації між собою, а також між центром управління, що повинно давати можливість забезпечити виконання більш складного чи комплексного завдання. Впровадження системи керування роєм безпілотних літальних апаратів є задачею не тривіальною.

© Мороз Б.І., Круглик А.С.

Впровадження використання рою безпілотних літальних апаратів для виконання завдання військового призначення підвищує шанс на успішність виконання такого завдання чи завдань. Разом з тим виникає цілий ряд задач, які потребують вирішення для успішного керування роєм і основне, для успішного виконання покладених на такий рій завдань.

Аналіз останніх досліджень і публікацій. Використання групи літальних апаратів в Україні поки що не можна назвати повністю впровадженим. Це пов'язано з недостатнім технічним розвитком таких апаратів. Але усі наробки останніх років вказують на можливість і необхідність удосконалення технічних можливостей літальних апаратів для їх використання в зграї. Це обумовлено, з поміж іншого, удосконаленням та розвитком систем противника. Що в свою чергу унеможливлює або робить складнішим виконання завдання військового призначення з використанням одного дрону. Використання групи дронів є перспективним напрямком. Цікавою на тему використання рою дронів автори вважають роботу [2], де автори класифікували БПЛА за різними критеріями, а також розглянули декілька проектів, пов'язаних із розробкою та використання рою БПЛА. Окрім того, автори роботи наводять три можливі стратегії управління роєм. У роботі [3] автор розглядає концепцію використання безпілотних літальних систем для охорони державного кордону. Також хотілося б виокремити роботу [5], де автори проаналізували перспективність групового застосування БПЛА військового призначення для успішного виконання покладеного на таку групу завдання. У роботі [6], яке автори вважають авторитетним у галузі розробки і впровадження безпілотних літальних апаратів у військовій сфері, було описано основні тенденції розвитку та застосування груп БПЛА. У роботі було описано роботу ройового інтелекту, що може внести вагомий вклад у розвиток та використання ройових груп військового призначення в цілому. У роботі автори також вказали на способи протидії групі БПЛА наявними засобами.

Постановка завдання. Метою роботи є дослідження проблем застосування груп БПЛА для виконання завдань військового призначення. Способи їх вирішення.

Викладення основного матеріалу дослідження. Застосування групи літальних апаратів ставить завдання до переосмислення підходу до управління такою системою. Існує дуже багато проблем, які потребують вирішення при розробці програмного забезпечення (ПЗ) для керування такою системою. У попередній роботі [4] автори показали загальну концепцію, як можливо представити систему управління безпілотними (літальними) апаратами. В даній роботі пропонується більш детально розглянути проблеми, які повинні враховуватись при розробці ПЗ для управління конкретною групою БПЛА для вирішення поставленого завдання військового призначення.

Кожна з трьох запропонованих стратегій управління [2] (централізована, децентралізована, змішана) передбачає, що у рою буде лідер а також підпорядковані йому. Лідер – керуючий елемент групи, без якого група перестає існувати як структурована одиниця. Іншими словами, ведучий та ведені. Таку систему в класичному вигляді можна представити як клієнт-серверна архітектура, де сервер це лідер, який роздає завдання, а клієнти, тобто ведені, виконують завдання і інформують літера про статус виконання. Сервер обробляє потоки повідомлень від усіх клієнтів, класифікує їх за різними характеристиками, такими як цінність повідомлення, час його старіння та ін. Аналізує повідомлення за заданими алгоритмами і на основі аналізу приймає відповідні рішення. Така архітектура є досить сталою в комп'ютерних мережах, і для вирішення загальних задач є такою, що задовольняє вимогам. Але автори вважають, що таку систему не раціонально використовувати для керування групою БПЛА військового призначення. Це пов'язано з тим, що БПЛА є мішенню для ворожих засобів ППО. У випадку, коли буде вражено лідера рою першим, уся група стає некерованою або діє по заздалегідь передбаченому алгоритму. Але, у випадку повернення до базового алгоритму керування, ми втрачаємо групу. Це будуть окремі одиниці безпілотних літальних апаратів і про виконання успішного завдання групою мова вже не йде. Тому до розробки ПЗ для керування групою повинні пред'являтись інші вимоги, ніж до класичного ПЗ з клієнт-серверною архітектурою.

Як правило, кожен рій складається із декількох груп або типів БПЛА, які відрізняються за тактико-технічними характеристиками (ТТХ) і призначенням. Логічно, що в групу входять як ударні дрони, дрони – розвідники, дрони зв'язку і ретрансляції, можливо дрони логістичного забезпечення та ін. Зрозуміло, що функція ударних дронів лише одна – вразити задану мішень. Але функції інших дронів повинні бути частково взаємодоповнюючими і взаємозамінюючими настільки, наскільки це дозволяє їх ТТХ. Пропонується розглянути схему управління, при якій в кожен момент часу може

© Мороз Б.І., Круглик А.С.

бути тільки один сервер (лідер), але інші лідери також присутні в групі у так званому «сплячому» стані (див. рис. 1). Така система дозволить залишитись роєм, навіть при втраті лідера. Для реалізації такої системи управління групою потрібно створити такі алгоритми взаємодії, що дозволять раціонально визначати нового лідера, при втраті попереднього. Визначення нового лідера є тривіальною задачею і не потребує додаткових пояснень, в якості алгоритму можна використати класичне «кидання кубика». Звичайно, «кидання кубика» повинно відбуватись іще до польоту, і в призначенні «сплячих» лідерів повинні приймати участь лише ті БПЛА, що за своїми ТТХ здатні виконувати роль лідера. Таким чином, іще до польоту можливо частково визначити кількість потенційних лідерів і інших БПЛА для коригування виконання поставленої задачі у нештатних випадках, таких як втрата безпілотного літального апарату через несправність або через роботу ворожого ППО.

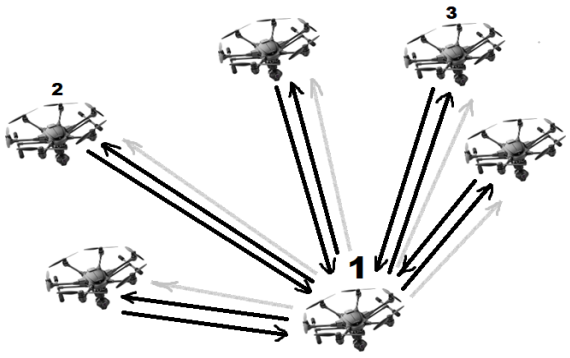


Рис. 1. Схематичне зображення можливої взаємодії між лідером і учасниками групи. Де цифрами позначені діючий лідер і сплячі.

При описаній вище концепції ПЗ зі «сплячими» лідерами з'являється додаткова вимога, яку повинні також враховувати при розробці ПЗ, а саме той фактор, що група повинна чітко розуміти, хто в кожний момент часу є лідером. І лідер повинен бути тільки один, для прийняття узгодженого рішення на основі алгоритмів ПЗ лідера. Звідси постає питання забезпечення додаткової комунікації і реплікації вихідних потоків повідомлень між членами групи. Під час роботи кожен ведений взаємодіє із ведучим по каналу зв'язку. Ведучий обробляє інформацію і приймає відповідні рішення. Кожен із учасників працює ізольовано один від одного, забезпечуючи виконання частини покладеного на нього завдання. Але ведучий, окрім взаємодії між кожним конкретним літальним апаратом (ЛА), повинен з визначеною періодичністю посилати повідомлення всім учасникам групи. У класичному вигляді це буде повідомлення зі службовою інформацією, яка, з-поміж іншого, даватиме змогу членам групи «зрозуміти» про наявність лідера в групі. У той же час, службові повідомлення дозволять «сплячому» лідеру з найвищим пріоритетом розуміти, коли виникає необхідність узяти на себе роль повноцінного лідера. У випадку втрати лідера, «сплячий» лідер перехоплює ініціативу на себе, запускає додаткові модулі ПЗ лідера і робить розсилку службових повідомлень вже як повноцінний лідер. Уся група перемикається на взаємодію із новим лідером і останній вже на основі існуючих можливостей корегує польотну місію.

У такому комплексному підході до вирішення питання перемикання між діючими лідерами на основі можливостей групи, виникає питання раціональної організації обробки повідомлень у такій групі між ведучим та веденими. Це питання виникає тому, що група БПЛА військового призначення знаходиться у дуже динамічному середовищі. Мова йде не про переміщення у простоті, а саме про можливість раптової втрати будь якого члена групи, яку неможливо зі 100% імовірністю передбачити. Так само, як і не можливо передбачити втрату конкретних ЛА, які буде уражено засобами протидії противника. Саме тому, окрім описаного вище способу взаємодії в групі, потрібно забезпечити раціональну організацію обробки повідомлень між ведучим та веденим. Така взаємодія, на думку авторів, повинна носити атомарний характер. Тобто, ведений відправляє повідомлення ведучому, ведучий отримує повідомлення та інформує веденого про статус повідомлення як отриманого. У разі неотримання відповіді від ведучого, ведений повторює відправку повідомлення ведучому. Іншими словами, повідомлення вважається доставленим до сервера, коли клієнт отримав відповідь від сервера про доставку. У цей же час БПЛА групи продовжує виконувати покладене на нього завдання, та з визначеною періодичністю повторювати відправку повідомлення лідеру. Це

© Мороз Б.І., Круглик А.С.

може бути завдання спостереження, картографування, пошук військових об'єктів або будь яке інше, що можливо виконати засобами БПЛА. Така взаємодія забезпечить однозначність, що у випадку із втратою лідера дозволить веденому в встановленому порядку переключитись на взаємодію з новим лідером. Усі накопичені повідомлення будуть відправлені новому лідеру а той у свою чергу класифікує їх за характеристиками цінності і старіння та відреагує на них відповідним чином.

Існує інша, не менш важлива проблема, яку потрібно враховувати при управлінні групою БПЛА військового призначення. Як було сказано вище, для існування групи повинна відбуватись комунікація між учасниками групи. Без комунікації кожен учасник стає автономним і про виконання групового завдання мова не йде. Але як забезпечити комунікацію групи в умовах активного подавлення зв'язку засобами радіоелектронної боротьби (РЕБ) противника. В представленому варіанті комунікації між літером і учасниками може виникнути ситуація, коли через ворожу РЕБ буде втрачено зв'язок одразу з усіма учасниками такого рою. У роботі [6] автори навели розрахунки, наскільки впливають засоби РЕБ на керування групою. Також автори роботи знайшли часткове рішення проблеми комунікації в середині рою, а саме зменшення дистанції між її учасниками. Зменшення дистанції між учасниками дозволить продовжити виконувати завдання, можливо не повністю, але найбільш цінні пункти завдання все ж таки зможуть бути виконаними групою. У випадку комунікації пункту управління з учасником чи учасниками групи, такий член групи повинен бути забезпечений супутниковим зв'язком для сприймання можливості взаємодії в умовах роботи РЕБ. Однак, досвід війни в України показав, що ворожі РЕБ також удосконалюються і стають потужнішими для активного протистояння групам ЛА. Вже з'являються стаціонарні пункти РЕБ, здатні подавляти велику кількість частот, на яких працюють БПЛА. У статті [7] було вказано, що ворог почав блокувати супутниковий зв'язок на активних ділянках фронту.

Саме тому, передача і обробка інформаційних потоків між учасниками рою в умовах активного подавлення засобами ворожих РЕБ потребує архітектурного підходу до вирішення проблеми ще на етапі проектування ПЗ. Таке ПЗ повинно бути модульним, тобто легко розширюваним та адаптованим до динамічних змін. Модульність, зокрема, повинна передбачати розробку драйверів і заглушок, що забезпечить виконання завдання в умовах швидкої заміни ПЗ на інше, з урахуванням загальної польотної місії. Наприклад, вже зараз в ручному режимі оператор БПЛА може зробити переключення частот, що використовуються для зв'язку, для забезпечення керування таким ЛА. Такий самий принцип можливо використати і в групах. Але, як було зазначено раніше, група не є константною величиною і може мати у своєму складі найрізноманітніші дрони, залежно від військової місії. Окрім того, такі БПЛА можуть бути різних виробників із різних країн, що в свою чергу призводить до відмінностей навіть у технічних складових. Як наслідок, члени групи можуть підтримувати різний діапазон частот, на яких вони можуть комунікувати і бути керованими. В даному випадку мова йде безпосередньо про технічну складову такого дрону. Тому драйвери керування модулем зв'язку повинні розроблюватись з урахуванням абстрактності. Приховувати деталі взаємодії між апаратним забезпеченням конкретного модуля зв'язку конкретного ЛА, та в той же час надавати прості механізми взаємодії із іншими модулями. Це дозволить на етапі передполітної підготовки узгодити використання лише тих частот, що підтримуються усіма членами групи. В цьому випадку, учасники групи зможуть робити переключення між доступними групі частотами комунікації і таким чином частково протидіяти засобам РЕБ. На рис. 2 показано приклад, як на етапі передполітної підготовки можливо узгодити діапазон частот, на яких усі дрони, незалежно від виробників, зможуть працювати у складі групи і комунікувати між собою, забезпечуючи автоматичне переключення лише між узгодженими частотами.

У випадку активної протидії засобів РЕБ, зв'язок між членами групи може бути втрачено. Цю проблему повинні завжди враховувати і допускати, що кожен ЛА групи повинен вміти діяти як самостійна одиниця. В такому випадку, про успішне виконання всіх завдань, покладених на групу, мова вже не йде. На такий випадок потрібно мати певну автономність а також синхронність дій для кожного з учасників окремо. Розглянемо приклад, коли група займається пошуком і знешкодженням ворожих об'єктів, в тилу, де активно застосовуються засоби РЕБ. Група складається із декількох менших підгруп безпілотників, серед яких одна підгрупа дронів спостереження, друга підгрупа дронів ураження, третя підгрупа дрони донаведення і аналізу результатів, четверта підгрупа БПЛА дальнього зв'язку для передавання інформації центру управління (ЦУ).

© Мороз Б.І., Круглик А.С.

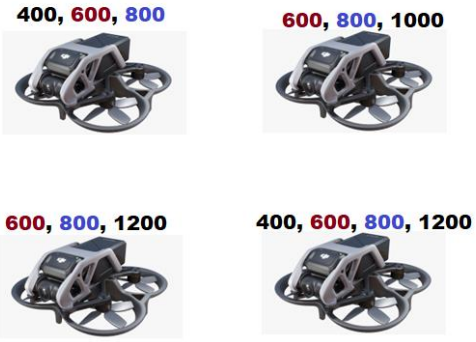


Рис.2. Налаштування частот комунікації, доступних для усіх учасників групи.

Підгрупа спостереження має потужні камери для чіткого фотографування та відеофільмування місцевості. Результати фотографування обробляє самостійно і результати обробки передає лідеру. Друга підгрупа це безпосередньо дрони – камікадзе, задача яких уразити мішень за командою лідера. Третя підгрупа, можливо має засоби дальноміру та засобів відеозапису для аналізу виконання завдання. Четверта підгрупа, яка може складатись навіть з одного ЛА, який буде забезпечувати зв'язок між ЦУ та групою. На певному проміжку політної дистанції група повністю втрачає зв'язок. У цьому випадку перша підгрупа залишається активною і продовжує виконувати поставлене на неї завдання фотографування чи відеофільмування території. Решта БПЛА втрачають свою значимість і повинні повертатись назад, на ту дистанцію, де вони є недосяжними для засобів РЕБ ворога. Перша підгрупа працює вже як окремі БПЛА в автономному режимі. В залежності від поставлених задач і можливостей БПЛА до аналізу, вони можуть самостійно виявляти ворожі цілі засобами машинного зору. Після фотографування підгрупа повертається до основної групи, де надає інформацію лідеру стосовно цілей. Лідер аналізує вхідні потоки, визначає цінність кожного з повідомлень і на основі заданих алгоритмів або зв'язується з ЦУ для прийняття рішення, або приймає рішення самостійно, якщо для цього є можливості. Звичайно, немає сенсу відправляти дрони – камікадзе на ураження колони що рухається, тому що цей процес відбувається не в реальному часі і колона вже змінила своє розташування. Або навпаки, якщо було знайдено підгрупою склади з боєприпасами, то шанс що вони зникли зі знайдених координат дуже невеликий. Тому автори пропонують розглядати ПЗ до керування групою ЛА комплексно, тобто не тільки як група, але й мати обов'язкову можливість працювати кожному БПЛА в автономному режимі. Це може забезпечити часткове виконання завдання в умовах активного подавлення зв'язку засобами РЕБ а також забезпечить «виживання» кожного окремого ЛА для подальшого відновлення складу групи, перегрупування і можливих подальших використань.

Висновки. У межах даної роботи було розглянуто проблеми використання груп безпілотних літальних апаратів для виконання завдання військового призначення. Розглянуто потенційні виклики, що виникають при розробці ПЗ для керування груп БПЛА. Описані абстрактні підходи до розробки ПЗ для комунікації в групі. Сформовано бачення стосовно застосування групи а також важливого збереження автономності кожного окремого ЛА як одиниці у випадках, коли неможливо керувати групою в динамічному середовищі. Запропоновано концептуальний підхід до розробки ПЗ для керування групою БПЛА з урахуванням проблем, що розглянуто.

Перспективи подальшого дослідження. У роботі розглянуто клієнт-серверну архітектуру із декількома можливими серверами, між якими можливо переключення для забезпечення існування лідеру групи. Клієнт-серверна архітектура є класичним варіантом взаємодії. У подальшому необхідно проаналізувати інші підходи до комунікації в групі, такі як без серверна архітектура та можливість застосування такої архітектури для використання при керуванні групою БПЛА. Окрім того, в межах даної роботи не розглянуто проблеми шифрування повідомлень, що є обов'язковим в умовах ведення бойових дій.

Список бібліографічного опису

1. Мороз Б.І., Круглик А.С., Мороз Д.М., Мартиненко А.А. Математична модель і загальний алгоритм вирішення задачі обробки повідомлень з урахуванням їх цінності і старіння в системах літальних апаратів. *Системні технології. Регіональний міжвузівський збірник наукових праць*. 2024. № 5(154). С. 3 – 18. DOI: 10.34185/1562-9945-5-154-2024-01

© Мороз Б.І., Круглик А.С.

2. Бінько І. В., Шевель В. В., Крицький Д. В. Комплексний підхід до управління формуванням групи роботів. *Сучасний стан наукових досліджень та технологій в промисловості.* 2024. №2(28). С. 17 – 32. DOI: https://doi.org/10.30837/2522-9818.2024.2.017
3. V. Nikiforenko. Conceptual Framework for the Use of Unmanned Aerial Systems in the Protection of the State Border. *Science and Innovation*, №20(2). С. 69–78. DOI: https://doi.org/10.15407/scine20.02.069
4. Мороз, Б., & Круглик, А. (2024). Концептуальна модель обробки інформаційних потоків в системі безпілотних літальних апаратів. КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ: ОСВІТА, НАУКА, ВИРОБНИЦТВО, (55), 155-160. DOI: https://doi.org/10.36910/6775-2524-0560-2024-55-19
5. Артюшин Л. М., Кононов О. А., Невзгляденко Ю. О. Аналіз перспектив реалізації групового застосування безпілотних літальних апаратів військового призначення. *Збірник наукових праць Державного науково-дослідного інституту авіації.* 2023. № 19(26). С. 42 – 48. DOI: 10.54858/dndia.2023-19-5
6. Лупандін В. А. Основні тенденції створення та застосування груп безпілотних літальних апаратів. *Наука і техніка Повітряних Сил Збройних Сил України.* 2019. №2. -С. 88-96. DOI: 10.30748/nitps.2019.35.11
7. Росіяни вперше стали глушити українські Старлінки у Бахмуті - військовий експерт розповів, як з цим боролися / New Voice : веб-сайт. URL: https://nv.ua/ukr/ukraine/events/yak-okupanti-glushili-starlinki-u-bahmuti-ta-yak-z-cim-borolis-ukrajinski-viyskovi-ekspert-50423570.html. (дата звернення 02.09. 2024)
8. Що таке РЕБ: принцип роботи на прикладі стаціонарних систем, окопних рішень та антидронових рушниць. Веб-сайт. URL: https://lander.kiev.ua/shcho-take-reb-pryntsyp-roboty-na-prykladi-statsionarnykh-system-okopnykh-rishen-ta-antydronovykh-rushnyts/. (дата звернення 12.09.2024)
9. O. Hutsa, D. Yelchaninov, D. Yanushkevych, I. Tolkunov, L. Ivanov, R. Petrova, A. Morozova. Conceptualization of Intelligent Control System for Humanitarian Demining Robotic Complexes Based on Verbal Methods. *Science and Innovation*. 2024. 20(3), 82–95. https://doi.org/10.15407/scine20.03.082
10. Павленко П. Модель управління групою безпілотних маневрених літальних апаратів із урахуванням їх безпеки польотів / P. Pavlenko, Ye. Samborskyi, H. Krykhovetskyi, I. Samborskyi // *Системи управління, навігації та зв'язку. Збірник наукових праць.* – Полтава: ПНТУ, 2023. – Т. 3 (73). – С. 58-63. DOI: https://doi.org/10.26906/SUNZ.2023.3.058
11. Гуляницький, Л. Ф. Оптимізація рішень при плануванні місій безпілотних літальних апаратів: Стенограма виступу на сесії Загальних зборів НАН України 27 квітня 2023 р. Visnik Nacional Noi Academii Nauk Ukrai Ni. 2023. № 5, 75–78. DOI: https://doi.org/10.15407/visn2023.05.075
12. Гуляницький Л. Ф., Рибальченко О. В. Оптимізація маршрутів при плануванні місій гібридних транспортних систем «дрон + транспортний засіб». *Кібернетика та комп'ютерні технології.* 2023. № 3. С. 44 – 58. DOI: *https://doi.org/10.34229/2707-451X.23.3.4*

References

1. B. Moroz, A. Kruhlyk. Mathematical model and general algorithm for solving the problem of processing messages taking into account their value and aging in aircraft systems. *System technologies*. 2024. № 5(154). P. 3 – 18. DOI 10.34185/1562-9945-5-154-2024-01
2. I. Binko, V. Shevel, D. Krytskyi. (2024). A comprehensive approach to managing robot group formation. INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES, (2(28), P. 17–32. DOI: https://doi.org/10.30837/2522-9818.2024.2.017
3. V. Nikiforenko. Conceptual Framework for the Use of Unmanned Aerial Systems in the Protection of the State Border. *Science and Innovation*, №20(2). P. 69–78. DOI: https://doi.org/10.15407/scine20.02.069
4. B. Moroz, A. Kruhlyk. Conceptual model of information flow processing in the system of unmanned aerial vehicles. COMPUTER-INTEGRATED TECHNOLOGIES: EDUCATION, SCIENCE, PRODUCTION. 2024. (55), 155-160. https://doi.org/10.36910/6775-2524-0560-2024-55-19
5. L. Artushin, O. Kononov, Y. Nevzgladenko. Analysis of prospects for the implementation of group application of unmanned aircraft for military purposes. COLLECTION OF SCIENTIFIC WORKS State research Institute of Aviation. 2023. № 19(26). P. 42 – 48. DOI: 10.54858/dndia.2023-19-5
6. V. Lupandin. Major trends of the development and application of a unmanned aerial vehicle groups. *Science and technology of the Air Force of the Armed Forces of Ukraine*. 2019. № 2. - P. 88-96. DOI: 10.30748/nitps.2019.35.11
7. For the first time, the Russians began jamming Ukrainian Starlinks in Bakhmut - a military expert told how they fought it. / New Voice : website. URL: https://nv.ua/ukr/ukraine/events/yak-okupanti-glushili-starlinki-u-bahmuti-ta-yak-z-cim-borolis-ukrajinski-viyskovi-ekspert-50423570.html. (date of app. 02.09. 2024)
8. What is EW: the principle of operation on the example of stationary systems, trench solutions and anti-drone guns. Website. URL: https://lander.kiev.ua/shcho-take-reb-pryntsyp-roboty-na-prykladi-statsionarnykh-system-okopnykh-rishen-ta-antydronovykh-rushnyts/. (date of app. 12.09.2024)
9. O. Hutsa, D. Yelchaninov, D. Yanushkevych, I. Tolkunov, L. Ivanov, R. Petrova, A. Morozova. Conceptualization of Intelligent Control System for Humanitarian Demining Robotic Complexes Based on Verbal Methods. Science and Innovation. 2024. 20(3), 82–95. https://doi.org/10.15407/scine20.03.082
10. Pavlenko P. Control model of a group of maneuverable unmanned aerial vehicles taking into account their flight safety / P. Pavlenko, Ye. Samborskyi, H. Krykhovetskyi, I. Samborskyi. *Control, Navigation and Communication Systems. Academic Journal*. 2023. №. 3 (73). P. 58-63. DOI: doi:https://doi.org/10.26906/SUNZ.2023.3.058.
11. L. Hulianytskyi. Optimization of decisions in mission planning for unmanned aerial vehicles. Transcript of speech at the session of the General Meeting of the NAS of Ukraine, 2023. № 5, 75–78. DOI: https://doi.org/10.15407/visn2023.05.075
12. L. Hulianytskyi, O. Rybalchenko. Route Optimization in Mission Planning for Hybrid DRONE+VEHICLE Transport Systems. *Cybernetics and computer technologies.* 2023. № 3. P. 44 – 58. DOI:10.34229/2707-451X.23.3.4

© Мороз Б.І., Круглик А.С.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-29
УДК 004.8:004.93'1
Панаскін Денис Валентинович, магістр, головний технічний директор
https://orcid.org/0009-0002-6900-5902
IT компанія Varteq, Чикаго, Іллінойс, США

ОЦІНКА СТІЙКОСТІ НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ ДО ШУМУ ТА АРТЕФАКТІВ У СКЛАДНИХ УМОВАХ ЕКСПЛУАТАЦІЇ

Панаскін Д.В. Оцінка стійкості нейромережевих моделей до шуму та артефактів у складних умовах експлуатації. Нейронні мережі широко використовуються для вирішення різноманітних складних завдань: від розпізнавання зображень до обробки природної мови. Багато сучасних наукових робіт спрямовано на підвищення ефективності використання нейронних мереж у різноманітних додатках та роботі програмного забезпечення. У ході цієї роботи було проведено аналіз питання стійкості нейромережевих моделей до шуму та артефактів у складних умовах експлуатації та надано її оцінку. Виявлено, що в будь-якій системі існують шуми, створені внутрішніми фізичними властивостями як пристроїв пам'яті, так і периферійних схем. Вплив цих шумів ще більше зростає у складних умовах експлуатації, що їх викликають додаткові артефакти. Згадано, що шумозаглушення зображень є однією із фундаментальних проблем у сфері обробки зображень, оскільки є обов'язковим етапом у багатьох додатках, пов'язаних із комп'ютерним зором. Особливо це актуально в медичній та військовій сферах, де алгоритми зменшення шуму використовують для отримання якісних рентгенівських знімків у системах комп'ютерної томографії та супутниковій зйомці. Також важливим є питання збереження якості зображень, що їх надають дрони-розвідники, оскільки деякі засоби радіоелектронної боротьби здатні псувати якість зображень чи робити його абсолютно нерозбірливим. Для захисту чіткості зображень зараз активно використовують нейромережеві моделі, оскільки вони демонструють достатньо непогані результати та здатні до наступної адаптації та тренування. Дослідження підкреслює потенціал нейромереж для синтезу та доповнення конфіденційних медичних даних у контексті виявленні клінічних респіраторних захворювань. Отримані результати цього дослідження можуть бути залучені надалі до подібних задач у системах комп'ютерного зору, наприклад, під час аналізу зображень у системах відеоспостереження.
Ключові слова: шум, артефакт, нейромережа, модель, зображення, мережева архітектура, згорткова нейронна мережа, аускультація, автоматизація

Panaskin D. Assessment of the resistance of neural network models to noise and artifacts in complex operating conditions. Assessment of robustness of neural network models to noise and artifacts in complex operating conditions. Neural networks are widely used to solve a variety of complex tasks: from image recognition to natural language processing. Many modern scientific works are aimed at increasing the efficiency of the use of neural networks in various applications and software operation. In the course of this work, an analysis of the issue of robustness of neural network models to noise and artifacts in difficult operating conditions was carried out and its assessment was provided. It was found that in any system there are noises created by the internal physical properties of both memory devices and peripheral circuits. The influence of these noises increases even more in difficult operating conditions, which cause additional artifacts. It is mentioned that image denoising is one of the fundamental problems in the field of image processing, as it is a mandatory step in many applications related to computer vision. This is especially relevant in the medical and military fields, where noise reduction algorithms are used to obtain high-quality X-ray images in computer tomography systems and in satellite imaging. The issue of preserving the quality of the image provided by reconnaissance drones is also important, since some means of radio-electronic warfare can spoil the quality of the image or make it completely illegible. To protect clarity, neural network models are now actively used, as they show quite good results and are capable of subsequent adaptation and training. The study highlights the potential of neural networks to synthesize and supplement sensitive medical data in the context of clinical respiratory disease detection. The obtained results of the current study can be involved in the future for similar tasks in computer vision systems, for example, during the analysis of images in video surveillance systems.
Keywords: noise, artifact, neural network, model, image, network architecture, convolutional neural network, auscultation, automation

Постановка проблеми. Проблема зниження впливу шуму під час оброблення зображень для створення якісного фото- і відеоконтенту сьогодні є однією із найважливіших, особливо в контексті ефективності роботи систем комп'ютерного зору [1]. Надзвичайно актуальним це є для зменшення шумів на зображеннях комп'ютерної томографії з низькими дозами радіації, бо через небезпеку високої дози радіаційного опромінення для пацієнтів та лікарів [2] доводиться їх знижувати, що збільшує кількість шумів і артефактів, змушуючи використовувати пост-обробку рентгенівських знімків [3].

Шум здатний накопичуватися в зображеннях під час їхньої обробки чи переходу в цифровий вигляд. Усе це спричиняє зниження чіткості зображення та виникнення артефактів, які заважають зрозуміти масштаб, розміри і границі об'єктів на фото чи відео [4]. У медицині низька якість знімків, виконаних шляхом використання комп'ютерного томографа, здатна спричинити постановку лікарем неправильного діагнозу, що є небезпечним для пацієнтів [5]. Для розвідки місцевості за допомогою супутників однією із головних проблем є високочастотний спекл-шум, що змушує проводити попередню фільтрацію отриманих зображень ще до їхньої обробки [6].

@ Панаскін Д. В.

Таким чином, під час обробки зображень необхідно знайти компроміс між точним зменшенням шуму і збереженням деталей зображення, що вирішується використанням нейронних мереж. Хоч раніше нейромережі вже і використовувалися в медицині для аускультації легень, але не було виявлено ефективності конкретних нейромереж саме для цієї методики в умовах стороннього шуму [7]. Оскільки різні нейромережеві моделі мають відмінні показники в контексті зменшення шуму, то необхідно провести огляд найпоширеніших із них та оцінити їх з точки зору ефективності боротьби з шумом й артефактами в складних чи специфічних умовах їхньої експлуатації.

Формулювання мети і завдань дослідження. Метою статті є огляд нейромережевих моделей та подальша оцінка їхньої стійкості до шуму та артефактів у складних умовах експлуатації.

Для досягнення поставленої мети було вирішено ряд таких завдання цього дослідження:

1) пояснено особливості різних моделей згорткових нейронних мереж;
2) досліджено використання згорткової нейромережевої моделі в контексті зменшення шуму і боротьби з появою артефактів на зображеннях та при аускультації легень;
3) визначено основні переваги і недоліки використання згорткових нейронних мереж;
4) на основі попередніх кроків було надано оцінку стійкості розглянутих нейромережевих моделей до шуму та артефактів у складних умовах експлуатації.

Для вирішення всіх завдань дослідження і досягнення його мети було використано такі загальнонаукові методи:

– метод моніторингу – для збору, систематизації та аналізу інформації про проблеми шуму і артефактів;

– метод порівняння допоміг під час дослідження різних мережевих моделей та їх порівняння;

– метод абстрагування було використано в ході дослідження з метою визначення головних понять та категорій;

– методи аналізу та синтезу були застосовані для ідентифікації факторів і етапів розвитку, а також найвизначніших елементів досліджуваного об'єкту.

Для розв'язання загальних завдань поточного дослідження було використано декілька груп спеціальних методів, серед яких можна назвати такі: методи збору та обробки інформації (дослідження онлайн-джерел); проведення аналітичної роботи (виявлення впливу різних факторів, узагальнення) та метод обґрунтувань.

Актуальність досліджуваної теми аргументується важливістю питання контролю якості зображень та процесу аускультації, що дозволяє користувачам фото- і відеотехніки передати вигляд об'єктів точніше, а лікарям – визначати точні діагнози, на основі чого вони можуть вилікувати пацієнтів набагато швидше, якісніше та, головне, безпечніше, без залучення часто непотрібних медичних препаратів чи практик.

Аналіз останніх досліджень та публікацій. Натепер подолання впливу шуму та артефактів шляхом використання різноманітних нейромережевих моделей, як і їх похідних, цікавить багатьох українських і закордонних дослідників та практиків у галузі інформаційних технологій та інженерії програмного забезпечення. Серед вітчизняних вчених слід згадати А. Олексія, Г. Пуху, О. Жеребуха, І. Фармагу, О. Переверзєва, К. Трапезона, В. Шликова, О. Воляника, А. Білінську, Я. Біньковського, А. Головатюка, Д. Мельничука, Т. Говорущенко.

У спільній статті А. Олексія та Г. Пухи розлянуто питання створення на основі нейронної мережі фреймворку, що може працювати з акустичним шумом, у якому цільовий сигнал забруднено фоновим шумом водного середовища, що відповідає реальним умовам [8].

О. Жеребуха та І. Фармага у своїй науковій праці розглянули процес використання згорткових нейронних мереж для роботи із зображеннями, отриманими з камер відеоспостереження, для створення оптимізованого алгоритму, щоб ефективно виявляти і класифікувати об'єкти на відео в контексті зменшення шуму [9].

О. Переверзєв та К. Трапезон із метою зменшення адаптивного білого шуму в зображеннях на основі стратегії залишкового навчання розробили і побудували власну наскрізну глибоку нейронну мережу згорткового типу [10].

У своїй науковій праці В. Шликов та О. Воляник обґрунтували потребу у створенні системи для виявлення артефактів на КТ-зображеннях та розробили свою архітектуру нейромережі в програмному середовищі Python шляхом використання спеціалізованої бібліотеки Keras [11].

@ Панаскін Д. В.

Цілий колектив вчених, серед яких А. Білінська, Я. Біньковський, А. Головатюк, Д.Мельничук та Т. Говорущенко, дослідили питання можливості використання штучного інтелекту та нейронних мереж у комп'ютерному зорі для підвищення швидкості та точності виявлення порушників, що вимагає зменшення рівня шуму та відсутності артефактів [12].

Серед іноземних дослідників питання подолання впливу шуму та артефактів шляхом використання різноманітних нейромережевих моделей розглядали M. Momeny, A. Latif, M.Sarram, R. Sheikhpour, Y. Zhang, A. Ilesanmi, T. Ilesanmi, W. Zhang, L. Jin, E. Song, X. Xu, Y. Wang, X. Song, K. Chen, S. Kaji та S. Kida.

У статті M. Momeny, A. Latif, M. Sarram, R. Sheikhpour, Y. Zhang пропонується створення шумостійкої згорткової нейронної мережі (NR-CNN) для класифікації зашумлених зображень без будь-якої попередньої обробки для видалення шуму та покращення продуктивності класифікації зашумлених зображень у згорткових нейронних мережах [13].

A. Ilesanmi та T. Ilesanmi провели детальне дослідження різних методів згорткової нейронної мережі, що використовуються для усунення шумів на зображенні [14].

W. Zhang, L. Jin, E. Song та X. Xu запропонували нову структуру усунення шумів на основі глибокої згорткової нейронної мережі для придушення імпульсного шуму в кольорових зображеннях. [15].

Y. Wang, X. Song та K. Chen дослідили нову нейронну мережу з канальною та просторовою увагою (CSANN) для згладжування зображень. У CSANN було об'єднано рівень шуму із середнім та максимальним значенням кожного каналу на вході та згорткову мережу для вивчення взаємозв'язку між цими каналами [16].

S. Kaji та S. Kida провели огляд питання трансляції зображень за допомогою глибоких нейронних мереж у контексті усунення шуму й перетворення модальності та реконструкції в медичній візуалізації [17].

Проаналізувавши сучасні матеріали, які сьогодні є у відкритому доступі: наукові збірники, монографії, статті на профільних сайтах – можна зробити висновок, що тема подолання впливу шуму та артефактів шляхом використання різноманітних нейромережевих моделей є актуальною та викликає інтерес у багатьох дослідників з будь-якого куточка Землі.

Також, визнаючи вагомий внесок усіх зазначених дослідників в огляд цієї теми, слід відзначити відсутність порівняння й оцінки різноманітних нейромережевих моделей у їхніх наукових працях.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.

Для початку необхідно зазначити, що існує безліч видів забруднення, серед яких шумові складають значний відсоток, хоч їхній вплив на організм людини і не настільки шкідливий, як, наприклад, хімічних [18]. Для зображень найбільш небезпечним забрудненням є саме шумове. Розглянемо детальніше його класифікацію:

– адитивний білий шум Гауса (AWGN) – базова модель шуму, що використовується для імітації ефекту випадкових процесів, які відбуваються в природі. Він додається до будь-якого шуму, який може бути притаманний системі та має рівномірну потужність у смузі частот цієї системи й нормальний розподіл у часовій області з її середнім значенням – нуль [19];

– імпульсний шум – вид шуму, який визначається заміною частини пікселів значеннями фіксованої чи випадкової величини, що пов'язано із втратами при передачі зображень каналами зв'язку [20];

– шум квантування – це різниця між значеннями вихідного й квантованого за рівнем аудіосигналу;

– шум Пуассона (дробовий) – складається із послідовності стрибків, які поступово зменшуються з часом [21];

– спекл-шум – мультиплікативний шум, що знижує якість зображення при діагностичних дослідженнях, надаючи знімкам вигляд розсіяної хвилі, створеної великою кількістю крихітних розсіяних віддзеркалень, які проходять через внутрішні органи.

Головною проблемою при усуненні шуму на зображенні є те, як відрізнити сам шум і краї та текстуру зображення, бо всі вони мають високочастотні компоненти.

Лінійні, нелінійні та неадаптивні фільтри були першими, що використовувалися для очистки зображень. Фільтри зменшення шуму поділяються на шість категорій (лінійні, нелінійні, адаптивні, вейвлет-фільтри, фільтри рівняння в частинних похідних (PDE) і фільтри загальної варіації). Особливості їхнього використання можна побачити у табл. 1.

Таблиця 1. Особливості використання фільтрів зменшення шуму

№	Назва фільтра	Головні особливості
1	Лінійний	Відповідають вихідним пікселям із вхідними сусідніми пікселями (використовуючи процедуру множення матриці), щоб зменшити шум
2	Нелінійний	Зберігають інформацію про краї та пригнічують шум. У більшості застосувань фільтрації нелінійний фільтр використовується замість лінійного
3	Адаптивний	Використовують статистичні компоненти для додатків реального часу (прикладами є найменше середнє квадратичне і рекурсивне середньоквадратичне)
4	Вейвлет	Дають можливість, змінюючи свої налаштування, не тільки зменшувати шуми, але і «витягати» невиявлені компоненти корисного сигналу
5	Рівняння в частинних похідних	Дозволяє фільтрації вздовж орієнтації смуги для всього зображення шляхом чисельного розв'язання рівняння частинних похідних, без необхідності трудомісткого встановлення невеликого вікна фільтрації вздовж орієнтації смуги і переміщення цього вікна фільтрації над кожним пікселем на зображенні
6	Загальної варіації	Є ітераційним фільтром, який зазвичай використовується для зменшення різних типів адитивного або мультиплікативного шуму зі збереженням різких переходів

Більшість згаданих вище фільтрів показали досить позитивні результати для зменшення рівнів шуму на зображеннях, однак і вони мають деякі недоліки, серед яких погана оптимізація фази тестування, налаштування параметрів вручну та складність спеціальних моделей шумозаглушення.

Але використання цих фільтрів не є панацеєю, оскільки вимагає виконання налаштувань в «ручному» режимі, що не актуально для багатьох сфер життя, особливо тих, де потрібна швидкість та відсутність помилок. Такою сферою є медицина, особливо, в контексті аускультації внутрішніх органів, що використовується ще з 19 століття для діагностики їх порушень. Для цього процесу критичним є питання зменшення зашумлення, щоб «відрізати» не потрібні в діагностиці захворювань «зайві» шуми.

Згадуючи питання зменшення шуму при аускультації легенів, потрібно зауважити, що, для початку, потрібно класифікувати звуки легенів. Зазвичай така класифікація створюється на основі наступних кроків: вибір ознак, побудова та навчання класифікаторів та оцінка продуктивності.

Оскільки вибір ознак для класифікації звуків легенів є складним та працезатратним процесом, то його виконання в «ручному» режимі вже давно морально застаріло та потребує заміни автоматизованими системами на основі алгоритмів машинного, глибинного чи ансамблевого навчання

Машинні методи навчання хоч і можуть бути використаними для автоматизації класифікації звуку легенів, але вони потребують додаткового початкового етапу екстракції ознак, що виконується «вручну». Такий відбір може призвести до помилок та неточностей, що критично в питанні моніторингу та подальшого призначення лікування для пацієнта, тому значно цікавішими для застосування в аускультації є два інші алгоритма навчання (глибинне та ансамблеве).

Розглядаючи глибоке навчання, необхідно зауважити, що нейронні мережі, які побудовані на основі використання його алгоритмів, витіснили традиційні методи машинного навчання у побудові автоматичних систем для класифікації звуків легень. В порівнянні з фільтрами звуку, вони мають значно менше недоліків. Серед великої кількості штучних нейромереж широко використовуються згорткові нейронні мережі (CNN), що призначені для екстракції ознак та класифікації звуків легень. Використання цих мереж не обмежується лише загальним усуненням шумів у зображенні, CNN дає чудові результати також і для сліпого усунення шумів.

В виявленні хронічних, нехронічних та нормальних звуків дихання CNN здатні показати високу чутливість навіть в роботі з більш складними та комплексними наборами даних [22]. А

залишкова CNN, що була розроблена в 2015 році, ResNets взагалі показала в мультикласифікації звуків дихання результати, що були близькі до 100% точності [23].

Ансамблеві моделі навчання будуються на основі поєднання кількох алгоритмів глибокого навчання в одній архітектурі. Можливості цих систем в діагностиці вражають. Потрібно зауважити, що включення згорткових нейронних мереж до ансамблів помітно підвищує точність класифікації випадкових звуків легенів. Ансамблеві моделі мають великий потенціал для створення, на їх основі, автоматизованих систем з метою класифікації звуків легень та ідентифікації респіраторних захворювань. Наприклад, поєднання згорткової нейронної мережі (CNN) і двонаправленої нейронної мережі з довготривалою пам'яттю (BiLSTM) дозволило досягти точності у 99,62% при класифікації різноманітних респіраторних захворювань [24].

Знаючи про таку високу точність роботи CNN та моделей з її інтеграцією, слід описати основні практики, які використовуються для ефективної роботи із згортковими нейронними мережами в контексті усунення шуму. Описувати ансамблеві моделі немає сенсу, бо вони завжди можуть мати інший вигляд, бо є поєднанням декількох інших.

Оскільки, шуми є проблемою не тільки для процесу аускультації легень, але і при проведенні комп'ютерної томографії і флюрографії, то для класифікації різноманітних респіраторних захворювань необхідно розглянути і питання боротьби з шумом на зображеннях.

Отже, найкраща та найсучасніша практика обробки шумів за допомогою згорткової нейронної мережі – це розробка мережевої архітектури, яка може впоратися із шумом і «витягувати» значущу інформацію із зображень. Деякі фактори, що впливають на мережеву архітектуру, – це кількість і розмір згорткових шарів, тип і розмір фільтрів або ядер, функції активації, рівні об'єднання та методи регуляризації. Хороша мережева архітектура повинна мати достатню кількість згорткових шарів для захоплення складних візерунків і функцій на зображеннях, але не надто багато, щоб уникнути перевантаження та обчислювальної складності.

Фільтри або ядра обираються відповідно до типу та масштабу шуму, а функції активації мають бути нелінійними та плавними, щоб уникнути зникнення насиченості та градієнта. Об'єднані шари повинні зменшувати розмірність і зберігати важливу інформацію зображень, а методи регуляризації – запобігати переобладнанню та покращувати узагальнення, наприклад, випадання, нормалізацію партії та зменшення ваги.

Ще однією ефективною практикою обробки зображень із шумом за допомогою згорткової нейронної мережі є прийняття стратегії навчання, яка може оптимізувати параметри мережі та мінімізувати функцію втрат. Деякі з факторів, що впливають на стратегію навчання, – це швидкість навчання, оптимізатор, функція втрат, розмір партії та кількість епох.

Швидкість навчання має регулюватися динамічно відповідно до прогресу та продуктивності мережі, а оптимізатор має бути адаптивним і ефективним, наприклад, Adam, RMSProp або SGD із імпульсом. Функція втрат має відображати мету та характер проблеми, наприклад, крос-ентропію для класифікації або середню квадратичну помилку для регресії. Розмір партії має бути збалансований між різноманітністю та стабільністю оновлень градієнта, а кількість епох має бути достатньою для досягнення конвергенції та уникнення надмірного або недостатнього оснащення.

Також необхідно пояснити алгоритм роботи згорткової нейронної мережі із зображеннями, оскільки це дозволяє зрозуміти специфіку використання цих мереж. Виконання будь-якої операції щодо зображень за допомогою згорткової нейронної мережі складається з шести кроків (шарів) що показано в табл. 2.

Таблиця 2. Робота із зображеннями за допомогою згорткової нейронної мережі

№	Назва шару	Дії
1	Вхідний шар	Отримує кольорові зображення у форматі RGB, де кожен піксель представлений значеннями інтенсивності кольорів
2	Згортковий шар	Застосовує фільтри до отриманого раніше зображення, щоб виділити основні характеристики
3	Шар функції активації (ReLU)	Додає нелінійність, використовуючи функцію активації ReLU для виходу згорткового шару
4	Шар пулінга	Зменшує розмірність основних характеристик, вибираючи максимальні значення на кожній ділянці карти ознак

5	Повторення шарів	Об'єднує безліч згорткових та пулінгових шарів для вилучення все більш складних характеристик із вхідного зображення
6	Шар розгладжування	Перетворює вихід попередніх шарів в одномірний вектор, що представляє всі характеристики
7	Повнозв'язковий шар	Приймає вихідні дані та застосовує ваги для класифікації зображення

Після пояснення процесу роботи із зображеннями за допомогою згорткової нейронної мережі слід згадати основні моделі цих мереж. Всього існує чотири основні моделі згорткових нейронних мереж, які показано на рис. 1.

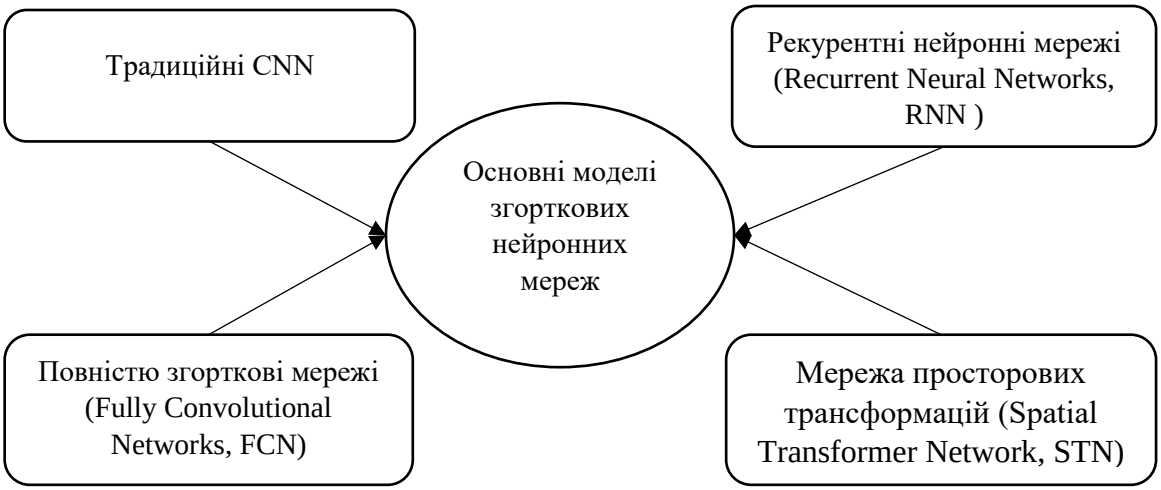


Рис. 1. Моделі згорткових нейромереж

Тепер слід пояснити особливості застосування цих моделей нейромереж у контексті роботи із зображеннями:

1. Традиційні згорткові нейронні мережі (CNN) складаються із серії згорткових та субдискретизуючих шарів, за якими знаходяться один чи декілька повнозв'язкових шарів. Кожен згортковий шар у цій мережі реалізує згортки із використанням фільтрів, що навчаються, для вилучення ознак із вхідного зображення. В основному використовується для аналізу зображень, розбиваючи зображення на дрібні частини та шукаючи шаблони, що допомагають розпізнавати об'єкти на зображенні.
2. Рекурентні нейронні мережі (Recurrent Neural Networks, RNN) здатні послідовно обробляти дані в контексті попередніх значень. На відміну від традиційних нейронних мереж, які обробляють дані у фіксованому порядку, RNN можуть обробляти вхідні дані змінної довжини та робити висновки залежно від попередніх вхідних даних. Рекурентні нейронні мережі зазвичай використовуються в обробці природної мови. При роботі з текстом вони можуть не тільки генерувати його, а й виконувати переклади. Також вони розуміють контекст, враховуючи попередні вхідні та вихідні дані, що дозволяє їм точно перекладати навіть складний текст.
3. Повністю згорткові мережі (Fully Convolutional Networks, FCN): їх часто застосовують для вирішення задач комп'ютерного зору, таких як сегментація зображень, виявлення об'єктів та класифікація зображень. Вони навчаються весь час з використанням методу зворотного розповсюдження помилки для категоризації чи сегментації зображень. Цей метод дає можливість нейромережі розрахувати градієнти функції втрат за вагами, що використовується для обрахунку того, як передбачається очікуваний результат для заданого входу за допомогою моделі машинного навчання. У повністю згорткових мережах відсутні повнозв'язкові шари і вони базуються на основі згорткових шарів, на відміну від традиційних згорткових нейромереж. Це робить їх гнучкішими та ефективнішими для проведення обчислень.
4. Мережа просторових трансформацій (Spatial Transformer Network, STN) використовується для вирішення задач комп'ютерного зору в контексті покращення можливості розпізнавати об'єкти та візерунки на зображенні незалежно від місця їхнього розташування,

@ Панаскін Д. В.

орієнтації або масштабу. Цю особливість STN називають просторовою інваріантністю. Прикладом використання чого є мережа, яка здійснює перетворення вхідного зображення ще перед етапом його обробки. Таке перетворення може включати вирівнювання об'єктів на зображенні та виправлення перспективних спотворень або інших змін для покращення роботи мережі в процесі виконання конкретних задач. Мережа просторових трансформацій допомагає працювати із зображеннями, враховуючи особливості їхнього знаходження в просторі, та поліпшує здатність розпізнавати об'єкти в складних умовах експлуатації.

Як і будь-яка нейромережа, використання згорткових нейромереж має свої переваги та недоліки (табл. 3).

Таблиця 3. Переваги та недоліки використання згорткових нейромереж

№	Переваги	Недоліки
1	Інваріантність до зсуву	Великий обсяг розмічених даних для навчання
2	Загальне використання параметрів	Процес навчання займає багато часу
3	Здатність моделювати складні структури даних	Високі вимоги до обчислювальної потужності
4	Навчання моделі відбувається упродовж усього шляху: від вхідних даних до висновку	Глибокі архітектури страждають від перенавчання
5	Можливість вивчати різні рівні параметрів вхідного зображення	Обмеження в завданнях, де треба контекстне розуміння
6	Здатність мінімізувати втрати інформації у процесі навчання	Погано застосовує набуті знання до невідомих даних

Джерело: власна розробка автора

Для тренування згорткової мережі в контексті зменшення шуму і вилучення артефактів необхідно створити тренувальний набір із великою кількістю зображень (із артефактами та шумом і без). Далі потрібно створити базову згортку із певною кількістю шарів (чим більше шарів, тим вища точність роботи системи). Потім треба обрати кількість епох, через які пройде мережа (залежно від потреб). Подальші показники точності мережевих моделей будуються на основі відсоткових значень точності на тренувальному, валідаційному та тестовому наборах.

Порівнюючи й обираючи найкращу із моделей згорткових нейромереж для роботи із зображеннями в контексті зменшення шуму і вилучення артефактів згідно з їхніми особливостями, запропоновано зупинитися на моделі повністю згорткових мереж (STN), оскільки вона здатна до просторової інваріантності, що є визначальним у цьому питанні, бо застосовує перетворення вхідного зображення перед його обробкою.

В контексті ансамблевих моделей цікавим виглядає, для класифікації звуків легень, поєднання традиційної згорткової мережі (CNN) та двонаправленої нейронної мережі з довготривалою пам'яттю (BiLSTM), де середнє гармонійне значення між точністю і повнотою склало 0,9845, де наближення значення до 1 представляє ідеальну продуктивність.

Підсумовуючи, слід зазначити, що зменшення шуму і вилучення артефактів є важливим питанням для галузі медичної візуалізації, дистанційного зондування, військового спостереження, біометрії та криміналістики. Уже існують ефективні засоби вирішення цього питання, але складні умови експлуатації створюють ризики додаткових шумів та артефактів, що вимагає розробки новіших мережевих архітектур.

Висновки та перспективи подальшого дослідження.

Отже, в сучасних умовах інтенсивного розвитку інформаційних технологій неможливо уявити життя людини без використання та обміну зображеннями різних об'єктів. У ході роботи було проведено огляд нейромережевих моделей та подальшу оцінку їхньої стійкості до шуму та артефактів у складних умовах експлуатації. Виявлено основні компоненти шумового забруднення зображень. Згадано особливості використання фільтрів зменшення шуму, використання згорткових нейромереж для зменшення шуму та вилучення артефактів. Описано основні моделі згорткових нейромереж та їхні особливості. Перераховано переваги та недоліки використання згорткових

нейромереж та їхнього процесу навчання. Також визначено, що для очищення від шуму зображень актуальною є модель повністю згорткових мереж (STN), через її здатність до просторової інваріантності, а для атоматизації аускультації легенів оптимальним є використання ансамблевої моделі у вигляді поєднання CNN та BiLSTM.

Отримані результати поточного дослідження можуть бути залучені надалі до подібних задач у системах комп'ютерного зору, наприклад, під час аналізу зображень у системах відеоспостереження та у розробці автоматизованих систем діагностики загального стану внутрішніх органів та здоров'я людей та тварин. Особливо цікавим виглядає залучення ансамблевих моделей у ветеринарії, оскільки тварини не здатні самостійно пояснити свої симптоми, на відміну від людей.

Список бібліографічного опису

1. Li Z. Image Denoising based on Deep Learning. *Highlights in Science, Engineering and Technology*. 2023. Вип. 39. С. 1245-1251. URL: https://doi.org/10.54097/hset. v39i.6749 (дата звернення: 20.08.2024).
2. Abuelhia E., Alghamdi A. Evaluation of arising exposure of ionizing radiation from computed tomography and the associated health concerns. *Journal of Radiation Research and Applied Sciences*. 2020. Вип. 13. №. 1. С. 295-300. URL: https://www.sciencedirect.com/science/article/pii/S1687850721002806 (дата звернення: 20.08.2024).
3. Zheng W. et al. Low-dose CT image post-processing based on learn-type sparse transform. *Sensors.* 2022. Вип. 22. №. 8. С. 1-17. URL: https://www.mdpi.com/1424-8220/22/8/2883 (дата звернення: 20.08.2024).
4. Abhadiomhen S. E. et al. Image edge preservation via low-rank residuals for robust subspace learning. *Multimedia Tools and Applications*. 2024. Вип. 83. №. 18. С. 53715-53741. URL: https://link.springer.com/article/10.1007/s11042-023-17423-1 (дата звернення: 20.08.2024).
5. Yanase J., Triantaphyllou E. A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications.* 2019. Вип. 138. С. 1-46. URL: https://doi.org/10.1016/j.eswa.2019.112821 (дата звернення: 20.08.2024).
6. Костенко П. Ю. та ін. Метод фільтрації мультиплікативного шуму на зображеннях з використанням технології сурогатних даних. *Системи обробки інформації*. 2021. Вип. 4. №. 167. С. 45-57. URL: https://journal-hnups.com.ua/index.php/soi/article/view/786 (дата звернення: 20.08.2024).
7. Панаскін Д., Білоконь Є., Бабко Д. Машинне навчання в діагностиці захворювань легеневої системи. *Технічні науки та технології*. Вип. 2. №. 28. С. 76–87. URL: https://doi.org/10.25140/2411-5363-2022-2(28)-76-87 (дата звернення: 20.08.2024).
8. Олексій А., Пуха Г. Створення датасету акустичних сигналів водного середовища для тренування нейромережі для придушення шумів. *Інформаційні технології та суспільство*. 2024. Вип. 2. №. 13. С. 56-60. URL: https://doi.org/10.32689/maup.it.2024.2.8 (дата звернення: 20.08.2024).
9. Жеребух О., Фармага І. Використання нейронних мереж для визначення об'єктів на зображенні. *Computer design systems. theory and practice.* 2024. Вип. 6. №. 1. С. 232-240. URL https://science.lpnu.ua/sites/default/files/journal-paper/2024/apr/34361/42.pdf (дата звернення: 20.08.2024).
10. Переверзєв О., Трапезон К. Особливості використання нейронних мереж у задачах зменшення шуму цифрових зображень. *Вісник КрНУ імені Михайла Остроградського.* 2023. Вип. 3. №. 140. С. 44-51. URL: https://visnikkrnu.kdu.edu.ua/statti/2023_3_2023_3_44.pdf (дата звернення: 20.08.2024).
11. Шликов В., Воляник О. Нейронна мережа для виявлення артефактів на КТ-зображеннях. *Біомедична інженерія і технологія.* 2020. №. 3. С. 17-24. URL: http://biomedtech.kpi.ua/article/view/195555 (дата звернення: 20.08.2024).
12. Білінська А. та ін. Автоматичне виявлення автомобільних порушників за допомогою комп'ютерного зору в рамках кіберфізичної системи запобігання аварійним ситуаціям. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2024. №. 1. С. 176-185. URL: https://vottp.khmnu.edu.ua/index.php/vottp/article/view/261 (дата звернення: 20.08.2024).
13. Momeny M. et al. A noise robust convolutional neural network for image classification. *Results in Engineering*. 2021. Вип. 10. С. 1-12. URL: https://helda.helsinki.fi/server/api/core/bitstreams/d04c9a76-c233-4460-955c-313cb029b9fc/content (дата звернення: 20.08.2024).
14. Ilesanmi A. E., Ilesanmi T. O. Methods for image denoising using convolutional neural network: a review. *Complex & Intelligent Systems*. 2021. Вип. 7. №. 5. С. 2179-2198. URL: https://link.springer.com/article/10.1007/S40747-021-00428-4 (дата звернення: 20.08.2024).
15. Zhang W. et al. Removal of impulse noise in color images based on convolutional neural network. *Applied Soft Computing*. 2019. Вип. 82. С. 1-11. URL: https://dl.acm.org/doi/abs/10.1016/j.asoc.2019.105558 (дата звернення: 20.08.2024).
16. Wang Y., Song X., Chen K. Channel and space attention neural network for image denoising. *IEEE Signal Processing Letters.* 2021. Вип. 28. С. 424-428. URL: https://ieeexplore.ieee.org/abstract/document/9350169 (дата звернення: 20.08.2024).
17. Kaji S., Kida S. Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiological physics and technology*. 2019. Вип. 12. №. 3. С. 235-248. URL: https://arxiv.org/pdf/1905.08603 (дата звернення: 20.08.2024).
18. Mykhailenko V., Safranov T. The problem of evaluation of individual persistent organic pollutants emissions from road transport (illustrated by the case of Odessa industrial-and-urban agglomeration). *Environmental problems*. 2022. Вип. 7. №. 1. С. 39-46. URL: https://doi.org/10.23939/ep2022.01.039 (дата звернення: 20.08.2024).

@ Панаскін Д. В.

19. Kokil P., Pratap T. Additive white gaussian noise level estimation for natural images using linear scale-space features. *Circuits, Systems, and Signal Processing*. 2021. Вип. 40. №. 1. С. 353-374. URL: https://link.springer.com/article/10.1007/s00034-020-01475-x (дата звернення: 20.08.2024).
20. Chen J., Zhan Y., Cao H. Adaptive sequentially weighted median filter for image highly corrupted by impulse noise. *IEEE Access*. 2019. Вип. 7. С. 158545-158556. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8886371 (дата звернення: 20.08.2024).
21. Кобилін О. А., Путятіна О. Є. Знешумлення зображень, зіпсованих дробовим шумом, у реальному часі. *Системи обробки інформації*. 2024. Вип.1. №. 176. С. 46-51. URL: https://doi.org/10.30748/soi.2024.176.06. (дата звернення: 20.08.2024).
22. Garcia-Ordas M. T. et al. Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data . *Sensors*. 2020. Vol. 20. №. 4. С. 1-16. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7070339/ (дата звернення: 20.08.2024).
23. Chen H. et al. Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access*. 2019. Vol. 7. С. 32845-32852. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8663379 (дата звернення: 20.08.2024).
24. Fraiwan M. et al. Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. *Journal of Ambient Intelligence and Humanized Computing*. 2022. С. 1-13. URL: https://link.springer.com/article/10.1007/s12652-021-03184-y (дата звернення: 20.08.2024).

References

1. Li, Z. (2023). Image Denoising based on Deep Learning. *Highlights in Science, Engineering and Technology*, *39*, 1245-1251. Retrieved from https://dev.ua/news/63-naselennia-planety-maiut-internet [in English].
2. Abuelhia, E., & Alghamdi, A. (2020). Evaluation of arising exposure of ionizing radiation from computed tomography and the associated health concerns. *Journal of Radiation Research and Applied Sciences*, *13*(1), 295-300. Retrieved from https://www.sciencedirect.com/science/article/pii/S1687850721002806 [in English].
3. Zheng, W., Yang, B., Xiao, Y., Tian, J., Liu, S., & Yin, L. (2022). Low-dose CT image post-processing based on learn-type sparse transform. *Sensors*, *22*(8), 1-17. Retrieved from https://www.mdpi.com/1424-8220/22/8/2883 [in English].
4. Abhadiomhen, S. E., Shen, X. J., Song, H., & Tian, S. (2024). Image edge preservation via low-rank residuals for robust subspace learning. *Multimedia Tools and Applications*, *83*(18), 53715-53741. Retrieved from https://link.springer.com/article/10.1007/s11042-023-17423-1 [in English].
5. Yanase, J., & Triantaphyllou, E. (2019). A systematic survey of computer-aided diagnosis in medicine: Past and present developments. *Expert Systems with Applications*, *138*, 1-46. Retrieved from https://doi.org/10.1016/j.eswa.2019.112821 [in English].
6. Kostenko, P. Y., Slobodyanyuk, V. V., Karlov, D. V., Vetluhin, Y. V., & Lyebyedyev, V. O. (2021). Metod filtratsiyi multyplikatyvnoho shumu na zobrazhennyakh z vykorystannyam tekhnolohiyi surohatnykh danykh [A method of filtering multiplicative noise on images using surrogate data technology]. *Systemy obrobky informatsiyi - Information processing systems*, *4*(167), 45-57. Retrieved from https://journal-hnups.com.ua/index.php/soi/article/view/786 [in Ukrainian].
7. Panaskin, D., Bilokon, Y., & Babko, D. (2022). Mashynne navchannya v diahnostytsi zakhvoryuvan lehenevoyi systemy [Machine learning in the diagnosis of diseases of the pulmonary system]. *Tekhnichni nauky ta tekhnolohiyi - Technical Sciences and Technologies, 2*(28), 76–87. Retrieved from https://doi.org/10.25140/2411-5363-2022-2(28)-76-87 [in Ukrainian].
8. Oleksiy, A., & Pukha, H. (2024). Stvorennya datasetu akustychnykh syhnaliv vodnoho seredovyshcha dlya trenuvannya neyromerezhi dlya prydushennya shumiv [Creation of a dataset of acoustic signals of the aquatic environment for training a neural network for noise suppression]. *Informatsiyni tekhnolohiyi ta suspilstvj - Information Technology and Society, 2*(13), 56-60. Retrieved from https://doi.org/10.32689/maup.it.2024.2.8 [in Ukrainian].
9. Zherebukh, O., & Farmaha, I. (2024). Vykorystannya neyronnykh merezh dlya vyznachennya obyektiv na zobrazhenni [Using neural networks to detect objects in an image]. *Computer design systems. theory and practice*, *6*(1), 232-240. Retrieved from https://science.lpnu.ua/sites/default/files/journal-paper/2024/apr/34361/42.pdf [in Ukrainian].
10. Pereverzyev, O., & Trapezon, K. Osoblyvosti vykorystannya neyronnykh merezh u zadachakh zmenshennya shumu tsyfrovykh zobrazhen [Features of the use of neural networks in the problems of digital image noise reduction]. *Visnyk KrNU imeni Mykhayla Ostrohradskoho - Bulletin of Mykhailo Ostrogradsky KrNU*, *3*(140), 44-51. Retrieved from https://visnikkrnu.kdu.edu.ua/statti/2023_3_2023_3_44.pdf [in Ukrainian].
11. Shlykov, V., & Volyanyk, O. (2020). Neyronna merezha dlya vyyavlennya artefaktiv na KT-zobrazhennyakh [A neural network for artifact detection in CT images]. *Biomedychna inzheneriya i tekhnolohiya - Biomedical engineering and technology*, (3), 17-24. Retrieved from http://biomedtech.kpi.ua/article/view/195555 [in Ukrainian].
12. Bilinska, A., Binkovskyy, Y., Holovatyuk, A., Melnychuk, D., & Hovorushchenko, T. (2024). Avtomatychne vyyavlennya avtomobilnykh porushnykiv za dopomohoyu kompyuternoho zoru v ramkakh kiberfizychnoyi systemy zapobihannya avariynym sytuatsiyam [Automatic detection of vehicle violators with the help of computer vision within the cyberphysical system of emergency prevention]. *Vymiryuvalna ta obchyslyuvalna tekhnika v tekhnolohichnykh protsesakh - Measuring and computing devices in technological processes*, (1), 176–185. Retrieved from https://vottp.khmnu.edu.ua/index.php/vottp/article/view/261 [in Ukrainian].
13. Momeny, M., Latif, A. M., Sarram, M. A., Sheikhpour, R., & Zhang, Y. D. (2021). A noise robust convolutional neural network for image classification. *Results in Engineering*, *10*, 1-12. Retrieved from https://helda.helsinki.fi/server/api/core/bitstreams/d04c9a76-c233-4460-955c-313cb029b9fc/content [in English].

14. Ilesanmi, A. E., & Ilesanmi, T. O. (2021). Methods for image denoising using convolutional neural network: a review. *Complex & Intelligent Systems*, *7*(5), 2179-2198. Retrieved from https://link.springer.com/article/10.1007/S40747-021-00428-4 [in English].
15. Zhang, W., Jin, L., Song, E., & Xu, X. (2019). Removal of impulse noise in color images based on convolutional neural network. *Applied Soft Computing*, *82*, 1-11. Retrieved from https://dl.acm.org/doi/abs/10.1016/j.asoc.2019.105558 [in English].
16. Wang, Y., Song, X., & Chen, K. (2021). Channel and space attention neural network for image denoising. *IEEE Signal Processing Letters*, *28*, 424-428. Retrieved from https://ieeexplore.ieee.org/abstract/document/9350169 [in English].
17. Kaji, S., & Kida, S. (2019). Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiological physics and technology*, *12*(3), 235-248. Retrieved from https://arxiv.org/pdf/1905.08603 [in English].
18. Mykhailenko, V., & Safranov, T. (2022). The problem of evaluation of individual persistent organic pollutants emissions from road transport (illustrated by the case of Odessa industrial-and-urban agglomeration). *Environmental problems*, *7*(1), 39-46. Retrieved from https://doi.org/10.23939/ep2022.01.039 [in English].
19. Kokil, P., & Pratap, T. (2021). Additive white gaussian noise level estimation for natural images using linear scale-space features. *Circuits, Systems, and Signal Processing*, *40*(1), 353-374. Retrieved from https://link.springer.com/article/10.1007/s00034-020-01475-x [in English].
20. Chen, J., Zhan, Y., & Cao, H. (2019). Adaptive sequentially weighted median filter for image highly corrupted by impulse noise. *IEEE Access*, *7*, 158545-158556. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8886371 [in English].
21. Kobylin, O. A., & Putyatina, O. Y. (2024). Zneshumlennya zobrazhen, zipsovanykh drobovym shumom, u realnomu chasi [Real-time denoising of images corrupted by shot noise]. *Systemy obrobky informatsiyi - Information processing systems*, *1*(176), 46-51. Retrieved from https://doi.org/10.30748/soi.2024.176.06 [in Ukrainian].
22. Garcia-Ordas, M. T., Benítez-Andrades, J. A., Garcia-Rodríguez, I., Benavides, C., & Alaiz-Moreton, H. (2020). Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data. *Sensors*, *20*(4), 1-16. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7070339/ [in English].
23. Chen, H., Yuan, X., Pei, Z., Li, M., & Li, J. (2019). Triple-classification of respiratory sounds using optimized s-transform and deep residual networks. *IEEE Access*, *7*, 32845-32852. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8663379 [in English].
24. Fraiwan, M., Fraiwan, L., Alkhodari, M., & Hassanin, O. (2022). Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. *Journal of Ambient Intelligence and Humanized Computing*, 1-13. Retrieved from https://link.springer.com/article/10.1007/s12652-021-03184-y [in English].

@ Панаскін Д. В.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-30
УДК: 004.438:004.4'042.42/.43:004.432.4:004.432.23:004.382.745
Пахомов Сергій Володимирович[1], архітектор рішень
https://orcid.org/0009-0007-1571-4611
Муляревич Олександр Володимирович[2], к.т.н., доцент
https://orcid.org/0000-0002-4644-7962
Боярінова Юлія Євгенівна[1], к.т.н., с.н.с., доцент
https://orcid.org/0000-0002-8974-529X
[1] Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна
[2] Національний університет «Львівська політехніка», м. Львів, Україна

РОЗРОБКА ТА ОПТИМІЗАЦІЯ МОБІЛЬНИХ ДОДАТКІВ ДЛЯ РІЗНИХ ПЛАТФОРМ ІЗ ВИКОРИСТАННЯМ МОВ ПРОГРАМУВАННЯ C, C++, C#, JAVA

Пахомов С.В., Муляревич О.В., Боярінова Ю.Є. Розробка та оптимізація мобільних додатків для різних платформ із використанням мов програмування C, C++, C#, Java. У процесі роботи було здійснено огляд методів розробки та оптимізації мобільних додатків для різних платформ на основі використання різноманітних мов програмування C, C++, C#, Java. Виявлено основні вимоги платформ та їх користувачів до мобільних додатків. Згадано фреймворки та бібліотеки, що допомагають реалізувати мультиплатформність та переваги й недоліки їх використання. Також запропоновано застосування обраних мов програмування відповідно до особливостей кожної з них, що зробить розробку мобільних додатків простішою та ефективнішою, на основі чого були надані практичні рекомендації для оптимізації мобільних застосунків. Ці рекомендації можуть бути задіяні як на рівні відділів корпорацій, так і індивідуальними розробниками чи фрілансерами, що робить їх універсальними. Отримані рекомендації можуть стати основою для зміни парадигми розробки та оптимізації мобільних додатків, що здатне змінити парадигму програмування з використанням мов C, C++, C#, Java загалом. Саме питання зміни парадигми програмування з використанням мов C, C++, C#, Java не тільки для мобільних додатків, але й для десктопних застосунків (програм, які встановлюються на локальний комп'ютер або ноутбук) визначено актуальним для наступних досліджень.
Ключові слова: смартфони**,** парадигми програмування, мобільні застосунки, операційна система, бібліотеки, фреймворки

Pakhomov S., Muliarevych O., Boiarinova Yu. Development and optimization of mobile applications for various platforms using C, C++, C#, Java programming languages. In the course of the work, a review of methods for developing and optimizing mobile applications for different platforms was conducted based on the use of various programming languages C, C++, C#, Java. The main requirements of platforms and their users to mobile applications are identified. The author mentions frameworks and libraries that help to implement multiplatform and the advantages and disadvantages of using them. It is also proposed to use the selected programming languages in accordance with the features of each of them, which will make the development of mobile applications easier and more efficient, based on which practical recommendations for optimizing mobile applications were provided. These recommendations can be used both at the level of corporate departments and by individual developers or freelancers, which makes them universal. The obtained recommendations can become the basis for changing the paradigm of mobile application development and optimization, which can change the paradigm of programming using C, C++, C#, and Java languages in general. The issue of changing the programming paradigm using C, C++, C#, and Java languages not only for mobile applications but also for desktop applications (programs that are installed on a local computer or laptop) is identified as relevant for further research.
Keywords: smartphones, programming paradigms, mobile applications, operating system, libraries, frameworks

Постановка наукової проблеми. Станом на 2022 рік 67% людей у всьому світі користувалися мобільними телефонами. Смартфони мали 80% з них, а з інтернет-користувачів – 96% [1]. Це свідчить про важливість, яку ці гаджети відіграють у суспільстві: від дзвінків та ігор до отримання кредитної інформації від банку чи прогнозу погоди. Зараз створюється все більше різноманітних мобільних застосунків, які допомагають їх користувачам виконувати широкий спектр завдань.

Щоб забезпечити успішність використання мобільного застосунку, розробники прагнуть максимізувати якість користувацького досвіду, а отже, і задоволеність користувачів [2]. Було виявлено що успіх мобільних застосунків, в контексті кількості їх завантажень, корелює з рейтингом, який отримує додаток в магазині додатків (наприклад, Google Play чи Apple Store) [3]. У 2018 році загальна кількість завантажених мобільних застосунків склала 194 мільярди, при цьому кожен користувач смартфона має багато встановлених на своєму телефоні додатків [4].

Головними проблемами в роботі мобільних застосунків зараз стали питання безпеки, складності обчислень та високого енергоспоживання. Серед усіх причин низьких оцінок додатків надмірне споживання ними енергії спричинило найбільшу кількість видалень зі смартфонів [5].

@ Пахомов С.В., Муляревич О.В., Боярінова Ю.Є.

Одним зі способів розв'язання цієї потрійної проблеми є створення гібридних додатків, які використовують дві або більше мов програмування. Такі застосунки будуть оптимізувати та використовувати кожну мову програмування. Усе це потребує зміни парадигми програмування та уніфікації, що можливе тільки при використанні бібліотек, фреймворків і розширень, які не конфліктують між собою.

Формулювання мети і завдань дослідження. Метою статті є огляд та виявлення особливостей використання методів розробки та оптимізації мобільних додатків для різних платформ на основі мов програмування C, C++, C# та Java.

Для досягнення поставленої мети були сформульовані такі завдання:

1. З'ясувати особливості мов програмування C, C++, C# та Java в контексті фреймворків, бібліотек і розширень.
2. Дослідити сумісність різних мобільних платформ (операційних систем) та додатків.
3. Визначити головні вимоги потенційних користувачів до мобільних застосунків.
4. Розробити практичні рекомендації для оптимізації мобільних застосунків.

Серед загальнонаукових методів для розв'язання завдань дослідження й досягнення його мети були використані такі:

– метод моніторингу: застосовувався для збору, систематизації та аналізу інформації про мови програмування та мобільні застосунки;

– метод порівняння: використовувався під час дослідження різних мобільних платформ та їх порівняння;

– метод абстрагування: застосовувався для виокремлення основних понять та категорій;

– методи аналізу та синтезу: використовувалися в процесі ідентифікації етапів і факторів розвитку, а також найвпливовіших елементів досліджуваного об'єкта.

Для розв'язання загальних завдань поточного дослідження було використано такі групи спеціальних методів: методи збору інформації; методи обробки інформації; методи проведення аналітичної роботи; метод обґрунтувань.

Актуальність досліджуваної теми аргументується важливістю питання контролю якості мобільних застосунків, що дозволяє користувачам робити своє життя простішим та зручнішим, а розробникам – залучати кошти і, на основі попереднього досвіду, створювати все якісніші та сучасніші мобільні застосунки, які позбавлені минулих недоліків та помилок.

Аналіз останніх досліджень та публікацій. Сьогодні тема розробки та оптимізації мобільних додатків для різних платформ із використанням мов програмування C, C++, C# та Java перебуває у фокусі наукової уваги багатьох вітчизняних і закордонних дослідників та практиків у галузі інформаційних технологій. Серед українських учених необхідно згадати О. Сікору, Т. Кобильника, Н. Єрьоміну, М. Шапошника, В. Хамбіра, Н. Ічанську, С. Улько, І. Гунька та інших.

У статті О. Сікори та Т. Кобильника розглянуто питання вибору мови програмування в шкільній освіті. Визначено, що такою може бути Java, не в останню чергу через можливість її використання для створення мобільних додатків [6].

Н. Єрьоміна та М. Шапошник вивчали методи тестування мобільних застосунків, де Java була визнана як зручна мова програмування для написання тестів (автоматизованого тестування), оскільки саме Java-додатки набагато простіше запускати на різних платформах [7]. В. Хамбір також згадував про критичну важливість автоматизації тестування мобільних застосунків для їх подальшої оптимізації [8].

Н. Ічанська та С. Улько описали основні відмінності нативної й багатоплатформної розробок мобільних додатків у контексті мов програмування, бібліотек, фреймворків та розширень [9].

У своїй науковій праці І. Гунько виклав своє бачення питання тестування програмного забезпечення через призму передових тенденцій: використання засобів штучного інтелекту та машинного навчання, практик DevOps та блокчейну [10].

З іноземних дослідників питання розробки та оптимізації мобільних додатків для різних платформ із використанням мов програмування C, C++, C# та Java варто зазначити таких учених, як M. Hort, M. Kechagia, F. Sarro, M. Harman, M. Farooq, S. Riaz, A. Alvi, A. Ali, I. Rehman, J. Ogala та D. Ojie.

У статті таких науковців, як M. Hort, M. Kechagia, F. Sarro, M. Harman представлено всебічний огляд оптимізації нефункціональної продуктивності для Android-додатків на основі показників швидкості відгуку, часу запуску, споживання пам'яті та енергії [11].

@ Пахомов С.В., Муляревич О.В., Боярінова Ю.Є.

M. Farooq, S. Riaz, A. Alvi, A. Ali та I. Rehman розглядали підходи та інструменти розробки кросплатформних мобільних додатків. Ними також були визначені основні виклики у сфері кросплатформності та шляхи їх подолання [12].

J. Ogala та D. Ojie провели дослідження мов програмування C, C++, C# та Java. У процесі їхньої роботи було узагальнено технічні особливості цих чотирьох мов програмування та здійснено їх порівняння шляхом реалізації бенчмарку для кожної з них [13].

Проаналізувавши матеріали, які на сьогодні є у відкритому доступі, можна зробити висновок, що тема розробки та оптимізації мобільних додатків для різних платформ із використанням мов програмування C, C++, C# та Java є актуальною та цікавою для багатьох дослідників. Проте здійснений аналіз дозволяє констатувати відсутність єдиного універсального підходу до вивчення питання трансформації сучасної парадигми програмування.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Для ґрунтовного висвітлення обраної проблематики дослідження необхідно визначити особливості мов програмування C, C++, C# та Java. C – це мова програмування загального призначення, що була створена в 1970-х роках Деннісом Рітчі й залишається широко використовуваною та впливовою навіть зараз. За задумом її творця, функції мови C чітко відображають можливості цільових процесорів. Вона знайшла тривале застосування в коді операційних систем (особливо в ядрах), драйверах пристроїв і стеках протоколів, але її використання в прикладному програмному забезпеченні зменшується. Також вона широко використовується в комп'ютерних архітектурах, які варіюються від найбільших суперкомп'ютерів до найменших мікроконтролерів та вбудованих систем [14].

Потрібно зазначити що C – це імперативна процедурна мова, що підтримує структуроване програмування, лексичну область видимості змінних та рекурсію зі статичною системою типів. Вона була розроблена для компіляції, щоб забезпечити низькорівневий доступ до пам'яті та мовних конструкцій, які ефективно відображаються в машинній інструкції, і все це з мінімальною підтримкою під час виконання. Попри свої низькорівневі можливості, мова була розроблена для заохочення кросплатформного програмування. Відповідно до стандартів програма C, написана з думкою про переносність, може бути скомпільована для широкого спектра комп'ютерних чи мобільних платформ та операційних систем з невеликими змінами у вихідному коді [15].

Мова програмування C призначена не лише для настільних комп'ютерів та серверів. При правильному підході вона може стати цінним інструментом для розробки мобільних додатків. Дуже важливо мати сприятливе середовище для розробки. Це гарантує, що процес програмування буде безперебійним і без непотрібних затримок. Першим кроком є вибір інтегрованого середовища розробки (IDE), придатного для програмування мовою C. Існує багато IDE, але найпопулярнішими з них для C є Eclipse CDT, Code::Blocks та CLion. Особливості їх використання представлені в табл. 1.

Таблиця 1. Особливості інтегрованого середовища розробки Eclipse CDT, Code::Blocks та CLion

№	Назва IDE	Головні особливості
1	Eclipse CDT	Підтримує статичний аналіз коду
		Має чудовий графічний інтерфейс користувача з функцією перетягування.
		Інтегрується з Git
2	Code::Blocks	Інтерфейс із вкладками, завершення коду, зручна навігація
		Підтримує різні компілятори – GCC, Clang та Visual C++
		Доступна повна підтримка точок зупинки
3	CLion	Забезпечує безпечний рефакторинг
		Дозволяє тестувати окремі одиниці вихідного коду
		Містить вбудований термінал
		Дозволяє проводити автоматичний аналіз коду

Розробка мобільних додатків часто вимагає спеціальних бібліотек для взаємодії з мобільною операційною системою. Наприклад, якщо розробка створена для Android, то може знадобитися NDK (Native Development Kit). Він дозволяє запускати код написаний на мовах програмування C та

C++ на пристроях Android. Компілятор необхідний для перетворення С-коду в машинний код. Для розробки мобільних додатків необхідно переконатися, що компілятор налаштований на відповідну мобільну архітектуру, наприклад ARM або x86.

Емулятор дозволяє протестувати створений мобільний додаток на комп'ютері без використання реального пристрою. Такі інструменти, як Android Studio, мають вбудовані емулятори для пристроїв Android. Важливо переконатися, що емулятор може безперебійно запускати програми на мові С. При розробці мобільних застосунків на мові програмування С бібліотеки та фреймворки, які обирає програміст, можуть суттєво вплинути на продуктивність, можливості та швидкість розробки його додатку, тому правильний вибір цих інструментів має вирішальне значення. Якщо для розробки на Android незамінним є NDK (Native Development Kit), то для iOS потрібно обрати бібліотеки, які взаємодіють з Objective-C, оскільки вона є підмножиною С.

Такі фреймворки, як SDL (Simple DirectMedia Layer) та Cocos2d-x, дозволяють програмісту написати додаток один раз і розгорнути його на різних платформах. Вони особливо корисні, якщо творці мобільного застосунку прагнуть орієнтуватися як на Android, так і на iOS без переписування коду. Саме ці фреймворки допомагають реалізувати мультиплатформність. Одним із найсучасніших універсальних фреймворків, що підтримує всі основні мови програмування та сумісний з Objective-C, є React Native. Як і будь-яка технологія його використання має свої переваги та недоліки (табл. 2).

Таблиця 2. Переваги та недоліки кросплатформного фреймворку React Native

№	Переваги	Недоліки
1	Багаторазове використання коду	Обмежений нативний доступ
2	Нативна продуктивність	Обмеження продуктивності
3	Гаряче перезавантаження	Відсутність кастомізації

Пояснимо переваги та недоліки фреймворку React Native більш детально:

1. Багаторазове використання коду: фреймворк React Native дозволяє розробнику написати код один раз і використовувати його на обох мобільних платформах (Android та iOS).
2. Нативна продуктивність: розробник може створювати додатки, які мають такий же рівень продуктивності, як і нативні.
3. Гаряче перезавантаження: фреймворк має функцію гарячого перезавантаження, яка дозволяє програмісту виправляти код з огляду на зміни в реальному часі.
4. Обмежений нативний доступ: розробник може стикатися з певними обмеженнями при створенні застосунків, які вимагають доступу до нативних API.
5. Обмеження продуктивності: хоч фреймворк і дозволяє створювати додатки, які мають ознаки нативних, але можуть даватися взнаки обмеження продуктивності при створенні саме складних застосунків.
6. Відсутність кастомізації: React Native використовує раніше зібрані компоненти, які не можуть бути налаштовані так, як цього потребують деякі додатки.

Для обчислювально-інтенсивних завдань можна використовувати такі бібліотеки, як OpenCL та OpenGL. Вони допомагають досягти кращого рендерингу графіки та паралельних обчислень на мобільних пристроях. Важливим також є питання створення надійної та масштабованої архітектури мобільного додатку, оскільки це має вирішальне значення для його подальшого успіху. Модульний дизайн застосунку гарантує, що кожен його компонент функціонує незалежно, що забезпечує кращу масштабованість та придатність до виправлень. Також додаток необхідно розділити на різні шари, такі як презентація, бізнес-логіка та доступ до даних, що гарантує незмінність усього додатку при зміні в одному з них. Для підвищення продуктивності роботи додатка потрібно використати принцип паралелізму, оскільки більшість сучасних смартфонів мають багатоядерні процесори [17].

Після створення застосунку необхідно його оптимізувати, зважаючи на обмеження щодо пам'яті, місткості акумулятора та обчислювальної потужності смартфонів. Жодний із сучасних додатків не може стати успішним серед користувачів, якщо буде мати критичні проблеми з надмірним споживанням енергії, оскільки це не дозволить користуватися ним регулярно. Найважливіші шляхи оптимізації застосунку, що створений із використанням мови С, показано на рис. 1.

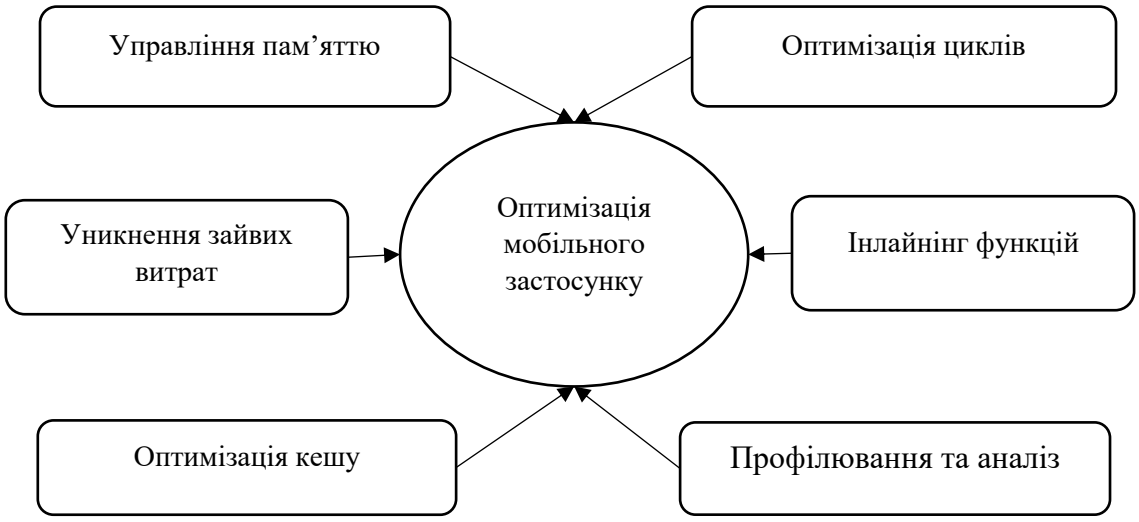


Рис. 1. Шляхи оптимізації мобільного застосунку

Сутність оптимізації мобільних застосунків, що розроблені на основі мови С, пояснена далі детальніше:

1. Управління пам'яттю: ефективне управління пам'яттю має вирішальне значення, тому потрібно уникати витоків пам'яті й переконатися в тому, що динамічно виділена пам'ять звільняється, коли вона більше не потрібна.
2. Оптимізація циклів: розгортання циклів та мінімізація накладних витрат на цикли може призвести до значного прискорення роботи.
3. Уникнення зайвих витрат: деякі операції, такі як ділення або математичні функції, можуть бути витратними, тому там, де це можливо, необхідно використовувати побітові операції або таблиці пошуку.
4. Інлайнінг функцій: вбудовування невеликих функцій може зменшити накладні витрати на виклики функцій, що призводить до швидшого виконання.
5. Оптимізація кешу: оптимізація використання кешу може значно підвищити продуктивність. Відповідно, треба переконатися, що дані, до яких часто звертаються, зберігаються безперервно, щоб скористатися перевагами локальності кешу.
6. Профілювання та аналіз: такі інструменти, як gprof або Valgrind, можуть допомогти визначити області, які потребують оптимізації [18].

Узагальнюючи питання оптимізації мобільних застосунків, які розроблені на основі мови С, потрібно зауважити, що концептуально вона визначається як поєднання розуміння апаратних обмежень та використання низькорівневих можливостей С.

На жаль, Java та C++ не мають такої універсальності в контексті створення додатків як С та C#, які можуть використовуватися для роботи з платформами Android і iOS. Java та C++ можна використовувати виключно для створення застосунків для Android, що є серйозним обмеженням для розробників. Саме тому для багатоплатформної розробки застосовують фреймворки, де основою є такі мови, як С та C#. Існує чотири основні способи створення додатків для Android та iOS:

1. Окремі нативні додатки для кожної операційної системи: при створенні нативних додатків розробники створюють додаток для певної операційної системи й покладаються на інструменти та мови програмування, розроблені спеціально для однієї платформи: Kotlin або Java для Android, Objective-C або Swift для iOS.
2. Прогресивні вебдодатки (PWA): розробники створюють PWA за допомогою вебтехнологій, таких як JavaScript, HTML, CSS і WebAssembly. Вони не потребують окремого пакування чи поширення й можуть бути опубліковані в Інтернеті та доступні через браузер на комп'ютері, смартфоні, планшеті. До того ж їх не потрібно встановлювати через Google Play або App Store.

3. Кросплатформні додатки: багатоплатформні застосунки призначені для однакової роботи на різних мобільних платформах. Кросплатформні фреймворки дозволяють писати спільний і багаторазовий код для розробки таких додатків.

4. Гібридні додатки: гібридна розробка додатків – це підхід, який поєднує нативні та вебтехнології. Він вимагає вбудовування коду, написаного мовою веброзробки, наприклад HTML, CSS або JavaScript, у нативний додаток. Це можна зробити за допомогою фреймворків, таких як Ionic Capacitor та Apache Cordova, використовуючи додаткові плагіни для доступу до нативних функцій платформ [19].

Вибір між кросплатформною та нативною розробкою залежить від цілей самого розробника застосунку. Якщо мета проєкту – довгостроковий розвиток, де необхідна безперебійна робота й швидке реагування, то бажано обирати нативну розробку. Також, хоч офіційно і вважається що кросплатформна розробка робить сам процес менш затратним, але вона призводить до збільшення необхідного часу розробки та інших неявних негативних факторів у процесі створення застосунку.

Попри те, що для створення мобільних застосунків можуть використовуватися різні мови програмування, фреймворки та бібліотеки, користувачі здебільшого мало цікавляться деталями розробки або й зовсім нічого не розуміють у цьому. Їх вимоги до роботи застосунку відрізняються залежно від особистих уподобань, але можна виокремити декілька основних із них, без задоволення яких мобільний додаток не здобуде успіху.

Вимоги користувачів часто збігаються з вимогами магазинів різних платформ, таких як AppStore та PlayMarket. Ці вимоги переважно базуються на гайдлайнах – збірниках головних правил, які обов'язкові до виконання для створення додатка. Так прийнято називати вимоги до дизайну застосунку та принципів зворотного зв'язку, якими повинні керуватися розробники при створенні додатка.

Більшість магазинів висувають досить реалістичні та інтуїтивно зрозумілі вимоги до розроблених додатків. Для розробки додатків під iOS використовуються Human Interface Guidelines, для Android – Material Design. Основні вимоги до дизайну додатків згідно з гайдлайнами зображено на рис. 2.

Рис. 2. Вимоги до дизайну додатків згідно з гайдлайнами

Серед додаткових вимог AppStore та PlayMarket до застосунку є наявність брендингу (кольори, формулювання, шрифти чи анімації розробників), оптимізації, свободи вибору й високого рівня безпеки.

На основі попередньо розглянутих даних можна надати низку рекомендацій, що допоможуть розробникам ефективно оптимізувати їхні мобільні застосунки. Оптимізація продуктивності мобільних додатків передбачає підвищення швидкості їх роботи та відгуку на дії користувачів і збільшення ефективності роботи програми. Добре оптимізований додаток не тільки покращує користувацький досвід, але й сприяє успіху бізнесу та розробників. Для покращення ефективності оптимізації мобільних додатків треба використовувати такі практики:

– проводити тестування продуктивності перед запуском застосунку (допомагає визначити час відгуку та обсяг використання пам'яті);

– видаляти невикористаний код і ресурси, стискати зображення та відео- і аудіофайли (чим менший розмір програми, тим швидше вона завантажується);

– зменшити кількість HTTP-запитів, мінімізувати споживання ресурсів і кешування даних, які часто використовуються (цей метод оптимізації допомагає підвищити швидкість програми та скоротити час її завантаження);

– використовувати мережу доставки вмісту (CDN), що поширює вміст програми серед користувачів на основі їхнього географічного розташування (покращує взаємодію з користувачем та зменшує навантаження на сервер);

– регулярно оновлювати додаток та виправляти помилки (допомагає покращити продуктивність програми);

– використовувати ефективні методи кодування, обмежити фонові процеси та зменшити використання анімації та візуальних ефектів (менше споживання заряду батареї).

Підсумовуючи, зазначимо, що оптимізація продуктивності мобільного додатка має вирішальне значення для його популярності серед користувачів та успіху бізнесу розробників. Проводячи тестування продуктивності, зменшуючи розмір програми, оптимізуючи код програми, використовуючи CDN, впроваджуючи оновлення, мінімізуючи розряд акумулятора та оптимізуючи для різних пристроїв, розробники здатні покращити продуктивність програми та вивести взаємодію з користувачем на більш високий рівень.

Висновки та перспективи подальшого дослідження. У сучасних умовах інтенсивного розвитку інформаційних технологій неможливо уявити життя людини без мобільних застосунків. У процесі роботи було здійснено огляд методів розробки та оптимізації мобільних додатків для різних платформ на основі використання різноманітних мов програмування C, C++, C#, Java, зокрема нативних та кросплатформних методів. Визначно основні вимоги платформ та їх користувачів до мобільних додатків, що збігаються з вимогами гайдлайнів магазинів. Згадано фреймворки та бібліотеки, що допомагають реалізувати мультиплатформність та переваги й недоліки їх використання. Також запропоновано застосування обраних мов програмування відповідно до особливостей кожної з них, що зробить розробку мобільних додатків простішою та ефективнішою. На цій підставі було надано практичні рекомендації для оптимізації мобільних застосунків. Ці рекомендації можуть бути задіяні як на рівні відділів корпорацій, так і індивідуальними розробниками чи фрілансерами, що робить їх універсальними. Отримані рекомендації повинні стати основою для зміни парадигми розробки та оптимізації мобільних додатків, що здатне змінити парадигму програмування з використанням мов C, C++, C#, Java загалом. Питання зміни парадигми програмування з використанням мов C, C++, C#, Java не тільки для мобільних додатків, але й для десктопних застосунків (програм, які встановлюються на локальний комп'ютер або ноутбук) визначено актуальним для наступних досліджень.

Список бібліографічного опису

1. Нагорна А. Дослідження: 63% населення планети мають доступ до інтернету, а в 67% є мобільний телефон. *dev.ua*. URL: https://dev.ua/news/63-naselennia-planety-maiut-internet (дата звернення: 22.07.2024).
2. Al-Subaihin A. A., Sarro F., Black S., Capra L., Harman M. App store effects on software engineering practices. *IEEE Transactions on Software Engineering*. 2019. Vol. 47(2). P. 300–319. DOI: 10.1109/TSE.2019.2891715
3. Sällberg H., Wang S., Numminen E. The combinatory role of online ratings and reviews in mobile app downloads: an empirical investigation of gaming and productivity apps from their initial app store launch. *Journal of Marketing Analytics*. 2023. Vol. 11(3). P. 426–442. URL: https://link.springer.com/article/10.1057/s41270-022-00171-w
4. Simon J. P. The global internet market (s): a reconstruction of the views of the industry. *Digital Policy, Regulation and Governance*. 2020. Vol. 22(2). P. 109–133. URL: https://www.emerald.com/insight/content/doi/10.1108/DPRG-10-2019-0092/full/html (дата звернення: 22.07.2024).
5. Dahdouh M., Bouchi A., Khawatmi S., Ayman M. Programmatic effect of optimized smali code on saving energy of android applications. *Int. J. Comput. Appl*. 2020. Vol. 177(42). P. 33–41. DOI:10.5120/ijca2020919928
6. Сікора О. В., Кобильник Т. П. Java як засіб навчання учнів основ програмування. *Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки*. 2023. № 34(73). С. 224–230. DOI: https://doi.org/10.32782/2663-5941/2023.5/35
7. Єрьоміна Н., Шапошник, М. Методи тестування мобільних застосунків. *Scientific Collection «InterConf»*. 2023. № 178. С. 322–325. URL: https://archive.interconf.center/index.php/conference-proceeding/article/view/4715 (дата звернення: 22.07.2024).
8. Хамбір В. Автоматизація процесів тестування мобільних застосунків. *Computer-integrated technologies: education, science, production*. 2024. № 55. С. 213–224. DOI: https://doi.org/10.36910/6775-2524-0560-2024-55-27
9. Ічанська Н., Улько С. Основні аспекти створення мобільних додатків та вибір інструментів їх розробки. *Системи управління, навігації та зв'язку. Збірник наукових праць*. 2020. Вип. 1(59). С. 74–78. DOI: https://doi.org/10.26906/SUNZ.2020.1.074
10. Гунько І. Software testing in 2023: new trends and challenges. *Herald of Kyiv Institute of Business and Technology*. 2023. Вип. 49(1-2). С. 25–36. DOI: https://doi.org/10.37203/kibit.2023.49.03
11. Hort M., Kechagia M., Sarro F., Harman M. A survey of performance optimization for mobile applications. *IEEE Transactions on Software Engineering*. 2021. Vol. 48(8). P. 2879–2904. URL: https://ieeexplore.ieee.org/abstract/document/9397392 (дата звернення: 22.07.2024).
12. Farooq M. S., Riaz S., Alvi A., Ali A., Rehman I. U. Cross-Platform Mobile Development Approaches and Frameworks. *VFAST Transactions on Software Engineering*. 2022. Vol. 10(2). P. 79–93. DOI: https://doi.org/10.21015/vtse.v10i2.978
13. Ogala J. O., Ojie D. V. Comparative analysis of C, C++, C# and Java programming languages. *GSJ*. 2020. Vol. 8(5). P. 1899–1913. URL: https://unidel.edu.ng/cms/uploads/publications/unidel_pub_1679733436.pdf (дата звернення: 22.07.2024).
14. Klabnik S., Nichols C. The Rust programming language. No Starch Press, 2023. URL: https://ejudge.lksh.ru/lang_docs/The%20Rust%20Programming%20Language.pdf (дата звернення: 22.07.2024).

15. Artigues V., Kormann K., Rampp M., Reuter K. Evaluation of performance portability frameworks for the implementation of a particle-in-cell code. *Concurrency and Computation: Practice and Experience*. 2020. Vol. 32(11). P. 1–23. DOI: https://doi.org/10.1002/cpe.5640
16. GeeksforGeeks. 7 Best IDEs For C/C++ Developers in 2024. URL: https://www.geeksforgeeks.org/best-ides-for-c-c-plus-plus-developers/ (дата звернення: 22.07.2024).
17. MarketSplash. C Programming Mobile App Development: Techniques And Tips. URL: https://marketsplash.com/c-programming-mobile-app-development/ (дата звернення: 22.07.2024).
18. Mobile Growth Association. How to Optimize Your Code in Mobile Applications – 10 Pro Tips. URL: https://mobilegrowthassociation.com/how-to-optimize-your-code-in-mobile-applications-10-pro-tips/ (дата звернення: 22.07.2024).
19. Stoyko T. How to Make an App both for Android and iOS. *Incora*. URL: https://incora.software/insights/make-app-both-iOS-Android (дата звернення: 22.07.2024).

References

1. Nahorna, A. (2022). Doslidzhennya: 63% naselennya planety mayut dostup do internetu, a v 67% ye mobilnyy telefon [Research: 63% of the world's population has access to the Internet, and 67% have a mobile phone]. Retrieved from https://dev.ua/news/63-naselennia-planety-maiut-internet [in Ukrainian].
2. Al-Subaihin, A. A., Sarro, F., Black, S., Capra, L., & Harman, M. (2019). App store effects on software engineering practices. *IEEE Transactions on Software Engineering*, 47(2), 300–319. DOI: 10.1109/TSE.2019.2891715 [in English].
3. Sällberg, H., Wang, S., & Numminen, E. (2023). The combinatory role of online ratings and reviews in mobile app downloads: an empirical investigation of gaming and productivity apps from their initial app store launch. *Journal of Marketing Analytics*, 11(3), 426–442. Retrieved from https://link.springer.com/article/10.1057/s41270-022-00171-w [in English].
4. Simon, J. . (2020). The global internet market (s): a reconstruction of the views of the industry. *Digital Policy, Regulation and Governance*, 22(2), 109–133. Retrieved from https://www.emerald.com/insight/content/doi/10.1108/DPRG-10-2019-0092/full/html [in English].
5. Dahdouh, M., Bouchi, A., Khawatmi, S., & Ayman, M. (2020). Programmatic effect of optimized smali code on saving energy of android applications. *Int. J. Comput. Appl*, 177(42), 33–41. DOI:10.5120/ijca2020919928 [in English].
6. Sikora, O. V., & Kobylnyk, T. P. (2023). Java yak zasib navchannya uchniv osnov prohramuvannya [Java as a means of teaching students the basics of programming]. *Vcheni zapysky TNU imeni V. I. Vernadskoho. Seriya: Tekhnichni nauky – Academic notes of TNU named after V. I. Vernadskyi. Series: Technical sciences*, 34(73), 224–230. https://doi.org/10.32782/2663-5941/2023.5/35 [in Ukrainian].
7. Yeromina, N., & Shaposhnyk, M. (2023). Metody testuvannya mobilnykh zastosunkiv [Methods of testing mobile applications]. *Scientific Collection «InterConf»*, 178, 322–325. Retrieved from https://archive.interconf.center/index.php/conference-proceeding/article/view/4715 [in Ukrainian].
8. Khambir, V. (2024). Avtomatyzatsiya protsesiv testuvannya mobilnykh zastosunkiv [Automation of mobile application testing processes]. *Computer-integrated technologies: education, science, production*, 55, 213–224. https://doi.org/10.36910/6775-2524-0560-2024-55-27 [in Ukrainian].
9. Ichanska, N., & Ulko, S. (2020). Osnovni aspekty stvorennya mobilnykh dodatkiv ta vybir instrumentiv yikh rozrobky [The main aspects of creating mobile applications and the choice of tools for their development]. *Systemy upravlinnya, navihatsiyi ta zvyazku. Zbirnyk naukovykh prats – Control, navigation and communication systems. Collection of scientific works*, 1(59), 74–78. https://doi.org/10.26906/SUNZ.2020.1.074 [in Ukrainian].
10. Hunko, I. (2023). Software testing in 2023: new trends and challenges. *Herald of Kyiv Institute of Business and Technology*, 49(1–2), 25–36. https://doi.org/10.37203/kibit.2023.49.03 [in Ukrainian].
11. Hort, M., Kechagia, M., Sarro, F., & Harman, M. (2021). A survey of performance optimization for mobile applications. *IEEE Transactions on Software Engineering*, 48(8), 2879–2904. Retrieved from https://ieeexplore.ieee.org/abstract/document/9397392 [in English].
12. Farooq, M. S., Riaz, S., Alvi, A., Ali, A., & Rehman, I. U. (2022). Cross-Platform Mobile Development Approaches and Frameworks. *VFAST Transactions on Software Engineering*, 10(2), 79–93. https://doi.org/10.21015/vtse.v10i2.978 [in English].
13. Ogala, J. O., & Ojie, D. V. (2020). Comparative analysis of C, C++, C# and Java programming languages. *GSJ*, 8(5), 1899–1913. Retrieved from https://unidel.edu.ng/cms/uploads/publications/unidel_pub_1679733436.pdf [in English].
14. Klabnik, S., & Nichols, C. (2023). *The Rust programming language*. No Starch Press. Retrieved from https://ejudge.lksh.ru/lang_docs/The%20Rust%20Programming%20Language.pdf [in English].
15. Artigues, V., Kormann, K., Rampp, M., & Reuter, K. (2020). Evaluation of performance portability frameworks for the implementation of a particle-in-cell code. *Concurrency and Computation: Practice and Experience*, 32(11), 1–23. https://doi.org/10.1002/cpe.5640 [in English].
16. GeeksforGeeks. (2024). *7 Best IDEs For C/C++ Developers in 2024*. Retrieved from https://www.geeksforgeeks.org/best-ides-for-c-c-plus-plus-developers/ [in English].
17. MarketSplash. (2023). *C Programming Mobile App Development: Techniques And Tips*. Retrieved from https://marketsplash.com/c-programming-mobile-app-development/ [in English].
18. Mobile Growth Association. (2023). *How to Optimize Your Code in Mobile Applications – 10 Pro Tips*. Retrieved from https://mobilegrowthassociation.com/how-to-optimize-your-code-in-mobile-applications-10-pro-tips/ [in English].
19. Stoyko, T. (2023). How to Make an App both for Android and iOS. *Incora*. Retrieved from https://incora.software/insights/make-app-both-iOS-Android [in English].

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-31
УДК 004.94:620.17
Попов Олексій Ігорович, аспірант
https://orcid.org/0009-0000-2385-5878
Запорізький національний університет, м. Запоріжжя, Україна

УПРОВАДЖЕННЯ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ПРОГНОЗУВАННЯ ВЛАСТИВОСТЕЙ МЕТАЛЕВИХ МАТЕРІАЛІВ

Попов О.І. Упровадження штучного інтелекту для прогнозування властивостей металевих матеріалів. Штучний інтелект (ШІ) змінює сферу матеріалознавства, забезпечуючи передові інструменти для прогнозування властивостей металевих матеріалів, які мають важливе значення для різних промислових застосувань. Актуальність цієї теми полягає у зростанні потреби в ефективних і точних методах оптимізації характеристик і дизайну матеріалів, що зумовлено попитом на високоефективні матеріали в таких галузях, як аерокосмічна, автомобільна та виробнича. Традиційні підходи до прогнозування властивостей матеріалів часто передбачають проведення масштабних експериментальних випробувань і комп'ютерного моделювання, що займає багато часу і є дороговартісним. На противагу цьому, методи штучного інтелекту, зокрема алгоритми машинного навчання, дозволяють підвищити точність прогнозування й скоротити час розробки завдяки використанню великих наборів даних і виявленню складних закономірностей, які можуть не визначатися традиційними методами. Мета статті – аналіз застосування штучного інтелекту, зокрема методів машинного навчання, для прогнозування властивостей металевих матеріалів. У роботі увага акцентується на інтеграції моделей машинного навчання з існуючими експериментальними і комп'ютерними методами для підвищення точності й ефективності прогнозів. У цій статті розглядається застосування штучного інтелекту для прогнозування властивостей металевих матеріалів із акцентом на інтеграції моделей машинного навчання та даних, керованих методами. Результати дослідження свідчать про значний прогрес у точності та ефективності прогнозування властивостей, показуючи, як штучний інтелект може ефективно моделювати складні взаємозв'язки між складом матеріалу, умовами обробки та отриманими в результаті властивостями. Зроблені в результаті цих досліджень висновки підкреслюють трансформаційний вплив ШІ на матеріалознавство. Методи на основі штучного інтелекту не лише спрощують процес прогнозування, але й дозволяють більш точно та індивідуально розробляти матеріали, що призводить до підвищення продуктивності та впровадження інновацій. Інтеграція ШІ в цю сферу є значним кроком вперед, порівняно з традиційними підходами, і відкриває широкі перспективи для майбутніх досліджень і розробок. Продовжуючи вдосконалювати моделі штучного інтелекту й розширювати їхнє застосування, науковці та інженери можуть ще більше розкрити потенціал металевих матеріалів, стимулюючи прогрес у різних галузях промисловості та сферах застосування.
Ключові слова: штучний інтелект, нейромережі, машинне навчання, матеріали, хімічний склад, властивості матеріалів, метали

Popov O. Implementation of Artificial Intelligence for Predicting the Properties of Metallic Materials. Artificial intelligence (AI) is transforming the field of materials science by providing advanced tools for predicting the properties of metallic materials, which are crucial for various industrial applications. The relevance of this topic stems from the increasing demand for effective and accurate methods for optimizing material characteristics and design, driven by the need for high-performance materials in industries such as aerospace, automotive, and manufacturing. Traditional approaches to predicting material properties often involve extensive experimental testing and computational modeling, which can be time-consuming and costly. In contrast, AI methods, particularly machine learning algorithms, enhance prediction accuracy and reduce development time by leveraging large datasets and uncovering complex patterns that may be invisible through traditional methods. The purpose of the article is to analyse the use of artificial intelligence, in particular machine learning methods, to predict the properties of metal materials. The article focuses on the integration of machine learning models with existing experimental and computer methods to improve the accuracy and efficiency of predictions. This paper explores the application of AI for predicting the properties of metallic materials with a focus on integrating machine learning models and data-driven methodologies. The research findings indicate significant progress in prediction accuracy and efficiency, demonstrating how AI can effectively model complex relationships between material composition, processing conditions, and resulting properties. The conclusions drawn from these studies highlight the transformative impact of AI on materials science. AI-based methods not only simplify the prediction process but also enable more precise and customized material development, ultimately leading to enhanced performance and innovation. The integration of AI into this field represents a significant advancement over traditional approaches, opening up broad prospects for future research and development. By continuing to refine AI models and expand their applications, researchers and engineers can further unlock the potential of metallic materials, driving progress across various industrial sectors and applications.
Keywords: artificial intelligence, neural networks, machine learning, materials, chemical composition, material properties, metals

Постановка проблеми. Стрімкий розвиток сучасних технологій штучного інтелекту (ШІ) істотно вплинув на різні галузі науки, у тому числі й на матеріалознавство. Останнім часом потреба в прогнозуванні властивостей металевих матеріалів із високою точністю стає дедалі актуальнішою через щораз вищий попит на сучасні матеріали зі специфічними властивостями в таких галузях, як аерокосмічна, автомобільна та електронна промисловість. Зазвичай, визначення та прогнозування

@ Попов О. І.

властивостей металевих матеріалів значною мірою залежить від експериментальних методів, які часто є трудомісткими, вартісними та вимагають значних ресурсів. Незважаючи на те, що ці методи є ефективними, вони мають певні обмеження з точки зору масштабованості та можливості досліджувати широкий спектр потенційного складу матеріалів та умов їхньої обробки.

Використання штучного інтелекту для прогнозування властивостей металевих матеріалів є перспективним рішенням цих проблем. Завдяки використанню алгоритмів машинного навчання та підходів, що базуються на даних, ШІ може аналізувати великі масиви даних, визначати складні закономірності та надавати точні прогнози властивостей матеріалів: міцності, пластичності, електропровідності та корозійної стійкості. Ці можливості мають особливу користь у розробці нових сплавів, оптимізації вже наявних матеріалів і скороченні часу та витрат, пов'язаних із експериментальними випробуваннями [1, с. 102].

Незважаючи на значні можливості, упровадження штучного інтелекту в галузі матеріалознавства все ще перебуває на початковій стадії, і перед ним стоїть низка викликів, серед яких – потреба у високоякісних, всеосяжних наборах даних для навчання ШІ-моделей, розробка алгоритмів, здатних максимально точно відображати основні фізичні явища, а також інтеграція прогнозів ШІ з традиційними експериментальними та обчислювальними методами. Водночас зростає потреба у вирішенні питань інтерпретованості та надійності ШІ-моделей при прогнозуванні властивостей матеріалів, оскільки ці прогнози мають безпосередній вплив на безпеку та продуктивність критично важливої техніки.

Актуальність цієї теми посилюється зростанням тиску на промисловість щодо розробки матеріалів, які відповідають високим експлуатаційним критеріям, мінімізуючи при цьому вплив на навколишнє середовище. Можливість точного прогнозування властивостей матеріалів за допомогою штучного інтелекту може зробити справжній переворот у проєктуванні та розробці металевих матеріалів, що призведе до ефективнішого використання ресурсів, прискорення інноваційних циклів і поліпшення експлуатаційних характеристик продукції. Тому впровадження ШІ у цій галузі є значним кроком у майбутнє матеріалознавства, де методи, що ґрунтуються на даних, доповнюють традиційні підходи для досягнення більш стійких і досконалих матеріальних рішень.

Аналіз останніх досліджень і публікацій. Застосування штучного інтелекту (ШІ) для прогнозування властивостей металевих матеріалів стало одним із головних напрямків сучасних досліджень, що зумовлено потребою в матеріалах із властивостями, пристосованими до конкретних застосувань. У цій сфері, що знаходиться на перетині матеріалознавства, інформатики та комп'ютерного моделювання, спостерігається значний прогрес, проте вона продовжує зазнавати впливу невирішених питань.

Одним із перших досягнень у цій галузі стало застосування алгоритмів машинного навчання (ML) для прогнозування механічних властивостей сплавів. Зокрема, такі вчені, як Р. Абдікєєв [2, с. 27] та Б. Колупаєв [3, с. 338], розробили фреймворки, що використовують велику кількість даних про властивості матеріалів у поєднанні з передовими методами ML для прогнозування поведінки нових матеріалів. Їхні роботи показали потенціал ШІ для пришвидшення відкриття високоефективних сплавів, але також висвітлили обмеження уже наявних моделей у потребі в обширних і якісних даних, які часто не є загальнодоступними для багатьох класів матеріалів.

Дослідження О. Крюкова [4, с. 63] та І. Мельник [5, с. 154] ще більше розвинули цю сферу, застосувавши ML для прогнозування фазових діаграм, які мають важливе значення для вивчення стабільності матеріалів за різних умов. Їхнє дослідження продемонструвало, як ШІ може потенційно замінити або доповнити традиційні методи розрахунку фазових діаграм, які є довготривалими і витратними в обчислювальному плані. Водночас вони також указали на проблему інтерпретації моделей ШІ, які часто діють як «чорні скриньки», що ускладнює дослідження причин, які лежать в основі передбачуваних властивостей матеріалу.

Останнім часом використання моделей глибокого навчання (DL) набуло популярності, і такі науковці, як О. Герцик [6, с. 132] та Т. Кваша [7, с. 158] досліджують їхнє використання для прогнозування складної механічної поведінки металевих матеріалів за різних умов напруження. Такі моделі показали перспективність у відображенні нелінійних взаємозв'язків, які простіші ML-моделі можуть пропустити. Втім, вони зазначають, що глибинне навчання вимагає ще більших наборів даних і значних обчислювальних ресурсів, які можуть бути надто дорогими. До того ж,

@ Попов О. І.

узагальнення цих моделей на матеріали за межами навчального набору залишається відкритим питанням, що вказує на потребу в більш надійних і адаптивних фреймворках ШІ.

Інтеграція ШІ з обчисленнями з теорії першопринципів, як показано в роботі А. Кошель [8, с. 59], є ще одним важливим досягненням. Поєднуючи прогностичну силу ШІ з фундаментальними знаннями, які надають квантово-механічні розрахунки, такі гібридні моделі відкривають шлях до більш точних прогнозів властивостей матеріалів. Однак цей підхід не позбавлений викликів, особливо з точки зору вартості обчислень і необхідності балансувати між точністю та ефективністю.

Попри всі досягнення, у цій галузі все ще залишається кілька невирішених питань. Однією із головних проблем є дефіцит високоякісних маркованих даних, необхідних для навчання моделей штучного інтелекту. Хоча такі заходи, як проєкт Materials Project, досягли успіхів у створенні великих загальнодоступних баз даних, багато матеріалів, особливо нових або складних сплавів, залишаються недостатньо представленими. Брак даних не лише обмежує точність прогнозів ШІ, але й перешкоджає розробці моделей, здатних екстраполюватись за межі відомих матеріалів [9, с. 215].

Ще одне важливе питання – інтерпретованість моделей ШІ. Хоча моделі «чорної скриньки» можуть давати точні прогнози, їхня непрозорість створює проблему для матеріалознавців, які мають розуміти механізми, що керують властивостями матеріалу. Це призвело до зростання інтересу до розробки інтерпретованих ШІ-моделей або гібридних підходів, які поєднують ШІ з більш традиційними методами, заснованими на правилах, щоб забезпечити точність і розуміння.

Метою статті є аналіз застосування штучного інтелекту, зокрема методів машинного навчання, для прогнозування властивостей металевих матеріалів. Стаття акцентує увагу на інтеграції моделей машинного навчання із уже наявними експериментальними і комп'ютерними методами для підвищення точності та ефективності прогнозів.

Виклад основного матеріалу. Аналіз металевих матеріалів має важливе значення в різних галузях промисловості, де вивчення властивостей і характеристик цих матеріалів може істотно вплинути на розробку продукції, контроль якості та прогнозування експлуатаційних характеристик. Традиційні методи аналізу матеріалів, такі як експериментальні випробування і теоретичне моделювання, хоч і є ефективними, але потребують значних витрат часу, ресурсів і досвіду. Зі швидким розвитком обчислювальних технологій, зокрема машинного навчання (ML) і нейронних мереж, з'явилися нові методології, які відкривають потенціал для ефективнішого, точнішого і масштабованішого аналізу металевих матеріалів.

Основні підходи та алгоритми машинного навчання в матеріалознавстві

Машинне навчання – частина штучного інтелекту (ШІ), яка дає змогу системам вивчати закономірності на основі даних і робити прогнози або приймати рішення без прямого програмування. У матеріалознавстві алгоритми машинного навчання застосовуються для аналізу великих масивів даних, виявлення взаємозв'язків між складом матеріалу, параметрами обробки та отриманими властивостями, а також для прогнозування поведінки нових або вже наявних матеріалів за різних умов.

Основні методи машинного навчання, що використовуються в матеріалознавстві, охоплюють контрольоване навчання, неконтрольоване навчання та навчання з підкріпленням. Контрольоване навчання передбачає навчання моделі на маркованому наборі даних, де відомі взаємозв'язки між входом і виходом. Цей підхід дуже корисний для прогнозування властивостей матеріалів, таких як міцність, твердість і теплопровідність, на основі даних про склад і обробку. Алгоритми, які зазвичай використовуються в керованому навчанні, охоплюють методи прийняття рішень, машини опорних векторів (SVM) і рандомні методи [10].

Зазначимо, що з іншого боку, неконтрольоване навчання оперує немаркованими даними, коли моделі намагаються виявити притаманні їм закономірності або групування в наборі даних. Цей метод часто використовують для кластеризації матеріалів на основі їхніх властивостей або для зменшення вимірності багатовимірних даних, що дає змогу отримати керованіші та легші для інтерпретації набори даних. Аналіз головних компонентів (PCA) і кластеризація за методом k-середніх є типовими прикладами алгоритмів неконтрольованого навчання, що використовують у цьому випадку.

Навчання з використанням підкріплення, хоч і менш розповсюджене в матеріалознавстві, має потенціал для оптимізації процесів проєктування матеріалів. У цьому методі модель навчається методом проб і помилок, отримуючи зворотний зв'язок у вигляді винагород або санкцій, що робить

@ Попов О. І.

його придатним для ітеративних процесів, таких як синтез матеріалів або оптимізація складу сплавів.

Нейронні мережі, зокрема моделі глибинного навчання, привернули до себе значну увагу в матеріалознавстві завдяки здатності моделювати складні, нелінійні взаємозв'язки між змінними. Нейронна мережа складається із шарів взаємопов'язаних вузлів або «нейронів», кожен із яких виконує зважені суми входів, за якими слідують нелінійні функції активації. Регулюючи значення ваг за допомогою процесу, який називають зворотним поширенням, мережа вчиться апроксимувати основну функцію, яка відображає вхідні дані та вихідні.

У сфері металевих матеріалів нейронні мережі активно використовуються для прогнозування властивостей матеріалів на основі композиційних і мікроструктурних даних. Наприклад, згорнуті нейронні мережі (CNN), які ефективно обробляють дані зображень, можна використовувати для аналізу мікрофотографій металевих зразків і прогнозування таких властивостей, як розмір зерен, фазовий розподіл і механічна міцність. Для моделювання часової еволюції властивостей матеріалу під час таких процесів, як термообробка або деформація, застосовуються рекурентні нейронні мережі (RNN), призначені для обробки послідовних даних [11, с. 3408].

Однією із найважливіших особливостей нейронних мереж є їхня здатність обробляти великі, складні набори даних з великою кількістю змінних, що робить їх придатними для вирішення багатофакторних задач, які часто трапляються в матеріалознавстві. До того ж, нейронні мережі можуть навчатися ієрархічному представленню даних, фіксуючи як низькорівневі особливості (наприклад, атомний склад), так і високі рівні (наприклад, мікроструктурні особливості), які впливають на властивості матеріалу.

Водночас нейронні мережі також мають певні проблеми, особливо з точки зору інтерпретації даних. На відміну від простіших моделей, нейронні мережі працюють як «чорний ящик», що заважає визначити конкретний вплив окремих вхідних змінних на остаточний прогноз. Така недостатня прозорість може бути суттєвим недоліком у наукових застосуваннях, де вивчення механізмів, що лежать в основі, є не менш важливим, ніж точне прогнозування.

Використання машинного навчання та нейронних мереж у матеріалознавстві має декілька значних переваг. Перш за все, це можливість прискорити відкриття та оптимізацію нових матеріалів. Традиційні експериментальні підходи зазвичай вимагають обширного тестування методом проб і помилок, що може зайняти багато часу і бути витратним. Моделі машинного навчання можуть значно зменшити цей тиск, прогнозуючи найбільш перспективний склад матеріалу або умови обробки, тим самим більш ефективно спрямовуючи експериментальні дослідження.

Ще однією перевагою є можливість виявлення прихованих закономірностей у складних наборах даних. Дані матеріалознавства найчастіше є багатовимірними, з численними взаємодіючими змінними. Алгоритми машинного навчання, зокрема нейронні мережі, успішно виявляють приховані нелінійні взаємозв'язки, які можуть бути невидимими для традиційних статистичних методів. Така здатність може призвести до нових ідей і глибокого аналізу факторів, що впливають на властивості матеріалів.

Незважаючи на ці переваги, існують і суттєві недоліки та виклики, пов'язані з цими методами. Однією із головних проблем є якість і кількість даних, необхідних для ефективного навчання моделей машинного навчання. У матеріалознавстві отримання великих високоякісних наборів даних може бути проблематичним, особливо для нових або складних матеріалів. Недостатній або надмірний обсяг даних може призвести до низької продуктивності моделі, надмірної адаптації та недостовірних прогнозів [12, с. 1164].

Ще однією проблемою є інтерпретація моделей машинного навчання, особливо нейронних мереж. Ці моделі можуть досягати високої точності, але їхній характер «чорної скриньки» ускладнює отримання значущих наукових висновків з отриманих результатів. Така непрозорість може стати бар'єром для їхнього застосування в дослідницьких і промислових умовах, де знання механізмів, що лежать в їхній основі, має ключове значення.

Упровадження машинного навчання та нейронних мереж вимагає спеціальних знань як у галузі матеріалознавства, так і в галузі інформатики. Розробка і перевірка моделей, які є одночасно точними і науково обґрунтованими, може бути складним і ресурсомістким процесом.

@ Попов О. І.

Обчислювальні ресурси, необхідні для навчання великих нейронних мереж, можуть бути суттєвими, особливо для моделей глибокого навчання з багатьма параметрами.

Попри ці виклики, інтеграція машинного навчання та нейронних мереж у матеріалознавство є важливим інструментом для розвитку цієї галузі. Оскільки доступність даних і обчислювальна потужність продовжують зростати, ці методи, очевидно, стануть все більш важливими у відкритті, оптимізації та аналізі металевих матеріалів (табл. 1).

Таблиця 1. Характеристика моделей та алгоритмів ШІ, які використовуються у прогнозуванні властивостей металевих матеріалів

Алгоритм/Модель	Використання у матеріалознавстві	Переваги	Виклики
Дерева рішень	Прогнозування властивостей матеріалу на основі інформації про склад	Легко інтерпретується, добре обробляє категорії даних	Схильні до перенавчання, менш ефективні на складних даних
Масиви опорних векторів (SVM)	Класифікація фаз на мікроструктурних зображеннях	Ефективний у високовимірних просторах	Потребує уважного налаштування параметрів
Згорткові нейронні мережі (CNN)	Аналіз мікрофотографій для прогнозування розміру зерна та розподілу фаз	Відмінно обробляє дані зображень, фіксує просторові ієрархії	Інтенсивні обчислення, складні для інтерпретації
Аналіз головних компонент (PCA)	Зменшення розмірності у високорозмірних наборах даних	Спрощує дані, виділяє ключові характеристики	Може втратити важливу інформацію в процесі зменшення розмірності
Кластеризація за методом K-середніх	Групування матеріалів на основі схожих властивостей	Простий та ефективний, легко реалізується	Припускає сферичні кластери, чутливий до початкових умов

Для прогнозування властивостей металевих матеріалів використовують кілька типів моделей, кожна із яких має свої переваги та недоліки. Найпоширеніші моделі поділяються на емпіричні, статистичні та моделі машинного навчання.

Емпіричні моделі ґрунтуються на спостережуваних залежностях між властивостями матеріалу і параметрами обробки. Такі моделі часто мають форму рівнянь, отриманих на основі експериментальних даних, і широко використовуються завдяки своїй простоті і легкості застосування. Наприклад, залежність Холла-Петча, яка пов'язує розмір зерна з межею міцності, є емпіричною моделлю, що часто використовується в металургії. Хоча емпіричні моделі корисні для швидкого прогнозування, вони зазвичай обмежені конкретними умовами, у яких вони були розроблені, і не можуть добре узагальнюватися на нові матеріали або методи обробки [13].

Статистичні моделі, такі як регресійний аналіз, розширюють емпіричні моделі за рахунок включення імовірнісних елементів. Ці моделі дозволяють враховувати кілька змінних і можуть надавати прогнози з відповідними діапазонами достовірності. Множинна лінійна регресія є широко застосовуваною статистичною моделлю в матеріалознавстві, де вона використовується для прогнозування таких властивостей, як твердість, в'язкість або корозійна стійкість на основі композиційних і мікроструктурних даних. Попри свою надійність, статистичні моделі можуть мати проблеми з відображенням нелінійних взаємозв'язків між змінними, які часто наявні в складних матеріальних системах.

Моделі машинного навчання представляють вдосконалений підхід до прогнозування властивостей, використовуючи обчислювальні алгоритми для виявлення закономірностей у великих наборах даних. На відміну від емпіричних і статистичних моделей, моделі машинного навчання не вимагають явних припущень про взаємозв'язки між змінними. Натомість вони вивчають ці взаємозв'язки безпосередньо з даних. Найпоширеніші моделі машинного навчання, що використовуються в матеріалознавстві, охоплюють дерева рішень, машини опорних векторів (SVM) і нейронні мережі, про які йшла мова вище. Ці моделі особливо ефективні при обробці даних високої розмірності і можуть вловлювати складні, нелінійні взаємодії між параметрами матеріалу. Однак розробка моделей машинного навчання може бути ресурсномісткою і вимагати великих обсягів даних для точних прогнозів.

Ефективність прогнозних моделей у матеріалознавстві значною мірою залежить від вибору вхідних параметрів, які повинні охоплювати основні фактори, що впливають на властивості

матеріалу, які представляють інтерес. Ці параметри зазвичай поділяються на три основні категорії: композиційні, мікроструктурні та параметри обробки.

Композиційні параметри стосуються хімічного складу матеріалу, включаючи тип і концентрацію легуючих елементів. Для металевих матеріалів ключові параметри складу можуть охоплювати відсотковий вміст вуглецю, хрому, нікелю та інших елементів, які впливають на такі властивості, як твердість, корозійна стійкість і міцність на розрив. Вибір цих параметрів має важливе значення, оскільки навіть невеликі зміни у складі можуть призвести до значних змін у поведінці матеріалу [14].

Мікроструктурні параметри описують внутрішню будову матеріалу, яка може містити розмір зерна, фазовий розподіл, щільність дислокацій та кристалографічну текстуру. Ці параметри безпосередньо пов'язані з механічними та фізичними властивостями матеріалу. Наприклад, розмір зерна є критичним мікроструктурним параметром, який впливає на міцність і пластичність, що описується співвідношенням Холла-Петча. Так само розподіл різних фаз у сплаві може впливати на його в'язкість і зносостійкість.

Параметри обробки охоплюють умови, за яких матеріал було виготовлено або оброблено, такі як температура, швидкість охолодження та попередня історія деформації. Ці параметри відіграють важливу роль у визначенні кінцевої мікроструктури і, відповідно, властивостей матеріалу. Наприклад, швидкість охолодження під час термічної обробки може впливати на формування мартенситних або перлітних фаз у сталі, тим самим впливаючи на її твердість і міцність. Точне врахування впливу параметрів обробки має важливе значення для моделей, які мають на меті передбачити властивості матеріалу в різних виробничих умовах.

Застосування предиктивних моделей у матеріалознавстві дуже різноманітне, численні дослідження демонструють їхню корисність для різних типів металевих матеріалів і прогнозування властивостей. Одним із яскравих прикладів є використання моделей машинного навчання для прогнозування міцності на розрив сталевих сплавів. У дослідженні високоміцних низьколегованих сталей (HSLA) науковці використали модель машини опорних векторів (SVM) для прогнозування міцності на розрив на основі даних про склад та умови обробки. Модель успішно врахувала складні нелінійні взаємодії між такими елементами, як вуглець, марганець і ніобій і змогла передбачити межу міцності на розрив із високою точністю, перевершивши традиційні регресійні моделі [13 – 14].

Інший приклад стосується прогнозування корозійної стійкості нержавіючої сталі. Науковці розробили модель множинної лінійної регресії для оцінки стійкості до точкової корозії на основі складу матеріалу, включаючи концентрацію хрому, молібдену та азоту. Модель надала важливу інформацію про вплив цих елементів на корозійну поведінку, що допомогло спрямувати розробку більш стійких до корозії сплавів для використання в суворих умовах.

У галузі адитивного виробництва для прогнозування властивостей металевих деталей, виготовлених методом селективного лазерного плавлення (SLM), застосовують предиктивні моделі. Нейромережева модель використовувалася для прогнозування щільності та механічних властивостей деталей з титанового сплаву на основі потужності лазера, швидкості сканування та товщини шару. Модель змогла передбачити оптимальні параметри процесу для отримання деталей високої щільності з чудовими механічними властивостями, що продемонструвало потенціал моделей машинного навчання в оптимізації процесів адитивного виробництва.

Емпіричні моделі продовжують відігравати важливу роль у матеріалознавстві, особливо в таких добре відомих галузях, як зварювання та лиття. Наприклад, рівняння Розенталя, емпірична модель широко використовується для прогнозування термічних циклів і отриманої мікроструктури під час зварювання. Хоча ця модель відносно проста, вона надає важливу інформацію для контролю зони термічного впливу та запобігання дефектам, таким як розтріскування або надмірний ріст зерна.

Подальший розвиток штучного інтелекту в прогнозуванні властивостей металевих матеріалів скоріш за все буде зосереджено на кількох ключових напрямках. Одним із найважливіших є покращення інтерпретованості моделей. Сучасні моделі ШІ, зокрема алгоритми глибинного навчання, забезпечують високу точність прогнозування, але вони часто працюють як «чорні скриньки», даючи мало інформації про основні механізми, що лежать в основі їхніх прогнозів. Дослідження, спрямовані на розробку інтерпретованих моделей ШІ або методів

@ Попов О. І.

отримання змістовної інформації з існуючих моделей, матимуть важливе значення для здобуття довіри наукової спільноти та сприяння впровадженню ШІ в матеріалознавство.

Іншим перспективним напрямком досліджень є розробка гібридних моделей, які поєднують традиційні фізичні підходи з методами ШІ, керованими даними. Такі гібридні моделі можуть використовувати сильні сторони обох підходів, застосовуючи ШІ для обробки складних, багатовимірних наборів даних, водночас враховуючи фундаментальні принципи матеріалознавства, щоб керувати прогнозами і забезпечувати їхню фізичну вірогідність. Така інтеграція ШІ з усталеними науковими знаннями може призвести до створення точніших і надійніших моделей для прогнозування властивостей матеріалів [15].

Також існує значний потенціал для дослідження методів навчання з використанням перенесення і навчання з кількох спроб в контексті металевих матеріалів. Навчання з перенесенням передбачає адаптацію моделі, навченої на одному наборі даних, до іншого, але пов'язаного з ним набору даних, що може бути дуже корисним у матеріалознавстві, де зазвичай бракує великих високоякісних наборів даних. Навчання з кількох спроб має на меті дозволити моделям робити точні прогнози на основі дуже обмежених даних, що може сприяти швидкому дослідженню нових матеріалів з мінімальним експериментальним навантаженням.

Вивчення генеративних моделей для дизайну матеріалів – ще один цікавий напрямок досліджень. Генеративні моделі, такі як генеративні нейронні мережі (GAN) або варіаційні автокодери (VAE), можуть створювати нові зразки даних на основі вивчених розподілів. У матеріалознавстві ці моделі можуть бути використані для створення нових композицій сплавів або мікроструктур із бажаними властивостями, прискорюючи відкриття нових матеріалів за допомогою обчислювальних експериментів.

Із розвитком штучного інтелекту інтеграція нових технологій і методів стає вирішальним фактором у його застосуванні для прогнозування властивостей металевих матеріалів. Одним із найбільш важливих досягнень у цій галузі є використання квантових обчислень у поєднанні зі штучним інтелектом. Квантові обчислення мають потенціал для розв'язання складних оптимізаційних задач і виконання симуляцій, які наразі неможливі за допомогою класичних комп'ютерів. Інтеграція квантових обчислень з моделями штучного інтелекту може уможливити прогнозування властивостей матеріалів з високою точністю та ефективністю, особливо для систем з високим ступенем складності або тих, що потребують дослідження величезних просторів параметрів.

Ще однією новою технологією є використання об'єднаного навчання в матеріалознавстві, що передбачає навчання моделей штучного інтелекту на основі декількох децентралізованих наборів даних без спільного використання самих даних, що має особливе значення в галузях, де конфіденційність і безпека даних викликають сумніви. У контексті металевих матеріалів об'єднане навчання може уможливити проведення спільних досліджень у різних установах і галузях, що дозволить ШІ-моделям отримувати вигоду від ширшого кола джерел даних, зберігаючи при цьому конфіденційність даних.

Упровадження методів аналізу штучного інтелекту, що піддається поясненню (XAI), також є вирішальним кроком в інтеграції ШІ в матеріалознавство. Пояснювальний ШІ зосереджується на розробці моделей, які є прозорими та інтерпретованими, забезпечуючи користувачам розуміння того, як і чому робляться конкретні прогнози. Це має особливе значення для наукових і промислових застосувань, де знання обґрунтування прогнозів моделі має важливе значення для прийняття зважених рішень. Інтеграція методів XAI в моделі ШІ для прогнозування властивостей матеріалів може підвищити рівень довіри до них і сприяти їхньому впровадженню як у дослідницьких, так і в промислових умовах [14 – 15].

До того ж, як прогнозується, застосування ШІ у відкритті високопродуктивних матеріалів буде ставати дедалі актуальнішим. Високопродуктивні експерименти передбачають швидкий синтез і тестування великої кількості зразків матеріалів, генеруючи величезні обсяги даних, які можна використовувати для навчання моделей ШІ. Інтегруючи ШІ з високопродуктивними методами, науковці можуть швидко виявляти перспективні матеріали-кандидати та оптимізувати їхні властивості, що значно прискорить темпи розробки матеріалів.

Досягнення в галузі штучного інтелекту для прогнозування властивостей металевих матеріалів мають значні перспективи для різних промислових процесів, зокрема у виробництві,

@ Попов О. І.

контролі якості та дизайні матеріалів. Одне із найбільш актуальних застосувань – оптимізація складу сплавів і умов обробки. ШІ-моделі можна використовувати для прогнозування властивостей сплавів на основі їхнього складу та параметрів обробки, що дозволяє виробникам точніше налаштовувати ці змінні для досягнення бажаних характеристик матеріалу. Це може призвести до розробки нових, високоефективних матеріалів, зменшуючи при цьому потребу в дорогих і трудомістких експериментальних випробуваннях.

Інша потенційна реалізація – у сфері прогнозного технічного обслуговування і контролю якості. ШІ-моделі можуть аналізувати дані датчиків і систем моніторингу для прогнозування деградації металевих компонентів у режимі реального часу, що дозволяє проводити проактивне технічне обслуговування і знижує ризик неочікуваних несправностей. У таких галузях, як аерокосмічна, автомобільна та енергетична, де цілісність металевих матеріалів має вирішальне значення, прогнозоване технічне обслуговування на основі штучного інтелекту може значно підвищити безпеку й надійність при одночасному зниженні експлуатаційних витрат.

ШІ може відігравати вирішальну роль в адитивному виробництві (АМ) металевих матеріалів. Адитивне виробництво, або 3D-друк, – це галузь, що швидко розвивається і дозволяє створювати складні, індивідуальні деталі з мінімальними відходами матеріалу. ШІ-моделі можна використовувати для оптимізації процесу АМ, прогнозуючи властивості надрукованих деталей на основі вхідних параметрів, таких як потужність лазера, швидкість сканування і товщина шару. Це може призвести до виробництва високоякісних компонентів з індивідуальними властивостями, розширюючи сферу застосування адитивного виробництва у таких галузях, як аерокосмічна, медична та автомобільна промисловість.

ШІ також має потенціал для трансформації управління ланцюгами поставок у матеріалообробній промисловості. Прогнозуючи попит на конкретні матеріали та оптимізуючи рівень запасів, ШІ може допомогти компаніям зменшити відходи, мінімізувати витрати та забезпечити стабільне постачання матеріалів (табл. 2).

Таблиця 2. Вплив напрямів досліджень та технологій ШІ на прогнозування металевих матеріалів

Напрям досліджень/технологія	Потенційний вплив на прогнозування металевих матеріалів	Проблеми
Інтерпретованість моделі	Підвищує рівень довіри та прийняття в науковому та промисловому середовищі	Баланс між точністю та прозорістю
Гібридні моделі (фізика + ШІ)	Поєднують сильні сторони підходів, що базуються на даних та фізиці	Забезпечення фізичної правдоподібності прогнозів ШІ
Квантові обчислення + ШІ	Дозволяє робити високоточні прогнози для складних матеріальних систем	Потребує вдосконалення квантового обладнання та алгоритмів
Об'єднане навчання	Сприяє спільним дослідженням, зберігаючи конфіденційність даних	Неоднорідність даних і проблеми координації моделей
Пояснюваний ШІ (Explainable AI, XAI)	Забезпечує прозорість і розуміння прогнозів ШІ	Розробка методів, які не ставлять під загрозу продуктивність моделі
Високопродуктивне виявлення матеріалів + ШІ	Прискорює відкриття та оптимізацію нових матеріалів	Управління та обробка великомасштабних експериментальних даних

Особливо це важливо в галузях, де металеві матеріали є важливими компонентами, наприклад, у будівництві, автомобілебудуванні та електроніці.

Інтеграція штучного інтелекту в промислові процеси не позбавлена проблем. Забезпечення стійкості та надійності моделей штучного інтелекту в реальних умовах має важливе значення, так само як і вирішення питань, пов'язаних із якістю даних, інтерпретацією моделей і необхідністю спеціальних знань і досвіду. Але в міру того, як технологія штучного інтелекту продовжує розвиватися, її потенціал для підвищення ефективності, стійкості та інноваційності промислових процесів, пов'язаних із металевими матеріалами, стає дедалі більше помітним.

Висновки. Таким чином, машинне навчання та нейронні мережі відкривають великі можливості для аналізу та прогнозування властивостей металевих матеріалів. Здатність моделювати складні взаємозв'язки, аналізувати великі масиви даних і прискорювати відкриття матеріалів робить їх незамінними інструментами в матеріалознавстві. Водночас для повної реалізації їхніх переваг

@ Попов О. І.

необхідно вирішити такі проблеми, як вимоги до даних, інтерпретованість моделей і потреба в спеціалізованій експертизі.

Моделі для прогнозування властивостей металевих матеріалів є незамінними інструментами в матеріалознавстві, що дозволяють науковцям та інженерам прогнозувати поведінку матеріалів, оптимізувати їхні склади та вдосконалювати методи обробки. Вибір моделі залежить від конкретного застосування: емпіричні моделі пропонують простоту, статистичні – надійність, а моделі машинного навчання – точність і гнучкість в обробці складних наборів даних.

Перспективи розвитку штучного інтелекту в прогнозуванні властивостей металевих матеріалів дуже високі і мають значний потенціал для просування як наукових досліджень, так і промислових застосувань. Зосереджуючись на ключових напрямках досліджень (інтерпретованість моделей, гібридні підходи) та інтеграції нових технологій, таких як квантові обчислення, ця галузь може вирішувати поточні проблеми і відкривати нові можливості. Потенційні можливості застосування ШІ в промислових процесах – від оптимізації сплавів до прогнозованого технічного обслуговування та адитивного виробництва – показують трансформаційний вплив цих технологій.

Список бібліографічного опису

1. Амоша О. І., Нікіфорова В. А. Світовий досвід становлення металургійних смарт-виробництв: особливості, напрями, наслідки. *Економіка промисловості*. 2019. № 2 (86). С.84–106.
2. Абдікєєв Р. Р., Чекіна В. Д., Ліщук О. В., Вишневський О. С. ІТ-кластери як інструмент забезпечення смарт-спеціалізації регіонів України. *Економічний вісник Донбасу.* 2022. № 2 (68). С. 21–34. DOI: https://doi.org/10.12958/1817-3772-2022-2(68)-21-34
3. Колупаєв Б. Б., Колупаєв Б. С., Левчук В. В., Максимцев Ю. Р., Матвійчук О. В. Математичне моделювання процесу термодеструкції полівінілхлориду, наповненого нанодисперсними металами. *Nanosistemi, Nanomateriali, Nanotehnologii*. 2023. *№ 21* (2). С. 331 – 347.
4. Крюкова О., Павлій О., Глабуць С., Кухта І. Застосування ультразвукового та радіографічного методів для оцінки якості металевих виробів адитивного виробництва. *Technologies and Engineering*. 2024. № 1. С. 60–66. DOI: https://doi.org/10.30857/2786-5371.2024.1.6
5. Мельник І., Тугай С., Кирик В., Ковальчук Д. Теоретичне оцінювання тиску гармат високовольтного тліючого заряду для використання в електронно-променевій технології зварювання металевих виробів. *Прикладні питання математичного моделювання*. 2021. № 4 (1). С. 147–160. DOI: https://doi.org/10.32782/KNTU2618-0340/2021.4.1.16
6. Герцик О., Ковбуз М., Пандяк Н., Ташак М. Особливості формування поверхневих шарів на металах. *Праці наукового товариства ім. Шевченка. Хімічні науки*. 2022. № 70. С. 128-137. DOI: https://doi.org/10.37827/ntsh.chem.2022.70.128
7. Кваша Т., Коваленко О. Технологічні тренди у сфері нових матеріалів для енергетики та військової сфери. *Grail of Science*. 2022. № 12-13. С. 154-163. DOI: https://doi.org/10.36074/grail-of-science.29.04.2022.023
8. Кошель А. Перспективні напрямки застосування нейронних мереж у конструкторській діяльності. *Комп'ютерно-інтегровані технології: освіта, наука, виробництво.* 2022. № 46. С. 57–63. DOI: https://doi.org/10.36910/6775-2524-0560-2022-46-08
9. Яремчук С. О., Маслов І. З. Фотоелектричні системи на судні: фізичний експеримент та аналіз з використанням штучного інтелекту. *Водний транспорт.* 2023. № 2 (38). С. 207–221. DOI: doi.org/10.33298/2226-8553.2023.2.38.22
10. Ihalge A., Hao Y. Formula Graph Self-Attention Network for Representation-Domain Independent Materials Discovery. 2022. № 9 (18). *Advanced Science.* DOI: https://doi.org/10.1002/advs.202200164
11. Huang J. S., Liew J. X., Ademiloye A. S., Liew K. M. Artificial intelligence in materials modeling and design. *Archives of Computational Methods in Engineering*. 2021. № *28*. P. 3399–3413. DOI: https://doi.org/10.1007/s11831-020-09506-1
12. Guo K., Yang Z., Yu C. H., Buehler M. J.Artificial intelligence and machine learning in design of mechanical materials. *Materials Horizons*. 2021. № *8* (4). P. 1153-1172. DOI: https://doi.org/10.1039/D0MH01451F
13. Романенко В. Технологія контролю якості виробів адитивного виробництва в ході друку елементів енергетичних комплексів. *Системні дослідження в енергетиці.* 2024. № 2а(78). С. 43–45.
14. Conrad F., Mälzer M., Schwarzenberger M., Wiemer H., Ihlenfeldt S. Benchmarking AutoML for regression tasks on small tabular data in materials design. *Scientific Reports.2022. №* 12. DOI: https://doi.org/10.1038/s41598-022-23327-1
15. Herriott C., Spear A. D. Predicting microstructure-dependent mechanical properties in additively manufactured metals with machine-and deep-learning methods. *Computational Materials Science*. 2020. № 175. DOI: https://doi.org/10.1016/j.commatsci.2020.109599
16. Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm / A. Dunn et al. *npj Computational Materials*. 2020. Vol. 6, no. 1. URL: https://doi.org/10.1038/s41524-020-00406-3 (date of access: 01.09.2024)

@ Попов О. І.

References

1. Amosha, O. I., & Nikiforova, V. A. (2019). Svitovyi dosvid stanovlennia metalurhiinykh smart-vyrobnytstv: osoblyvosti, napriamy, naslidky [Global Experience in the Establishment of Metallurgical Smart Production: Features, Directions, Consequences]. *Ekonomika promyslovosti - Industrial Economics*, 2 (86), 84–106. [in Ukrainian].
2. Abdikeiev, R. R., Chekina, V. D., Lishchuk, O. V., & Vyshnevskyi, O. S. (2022). IT-klastery yak instrument zabezpechennia smart-spetsializatsii rehioniv Ukrainy [IT Clusters as a Tool for Ensuring Smart Specialization of Ukrainian Regions]. *Ekonomichnyi visnyk Donbasu - Economic Bulletin of Donbas*, 2 (68), 21–34. DOI: https://doi.org/10.12958/1817-3772-2022-2(68)-21-34 [in Ukrainian].
3. Kolupaiev, B. B., Kolupaiev, B. S., Levchuk, V. V., Maksymtsev, Yu. R., & Matviichuk, O. V. (2023). Matematychne modeliuvannia protsesu termodestruktsii polivinilkhlorydu, napovnenoho nanodyspersnymy metalamy [Mathematical Modeling of the Thermodestruction Process of Polyvinyl Chloride Filled with Nanodispersed Metals]. *Nanosystemy, Nanomaterialy, Nanotekhnolohii*, 21 (2), 331 – 347. [in Ukrainian].
4. Kriukova, O., Pavlii, O., Hlabuts, S., & Kukhta, I. (2024). Zastosuvannia ultrazvukovoho ta radiografichnoho metodiv dlia otsinky yakosti metalevykh vyrobiv addytyvnoho vyrobnytstva [Application of Ultrasonic and Radiographic Methods for Assessing the Quality of Metal Products in Additive Manufacturing]. *Technologies and Engineering*, 1, 60–66. DOI: https://doi.org/10.30857/2786-5371.2024.1.6 [in Ukrainian].
5. Melnyk, I., Tuhai, S., Kyryk, V., & Kovalchuk, D. (2021). Teoretychne otsiniuvannia tysku harmat vysokovoltnoho tliuchoho zariadu dlia vykorystannia v elektronno-promenevii tekhnolohii zvaryuvannia metalevykh vyrobiv [Theoretical Estimation of the Pressure of High-Voltage Glow Discharge Guns for Use in Electron Beam Welding Technology of Metal Products]. *Prykladni pytannia matematychnoho modeliuvannia - Applied Issues of Mathematical Modeling*, *4*(1), 147–160. DOI: https://doi.org/10.32782/KNTU2618-0340/2021.4.1.16 [in Ukrainian].
6. Hertsyk, O., Kovbuz, M., Pandyak, N., & Tashak, M. (2022). Osoblyvosti formuvannia poverkhnevykh shariv na metalakh [Features of Surface Layer Formation on Metals]. *Pratsi naukovoho tovarystva im. Shevchenka. Khimichni nauky - Proceedings of the Shevchenko Scientific Society. Chemical Sciences*, 70, 128–137. DOI: https://doi.org/10.37827/ntsh.chem.2022.70.128 [in Ukrainian].
7. Kvasha, T., & Kovalenko, O. (2022). Tekhnolohichni trendy u sferi novykh materialiv dlia enerhetyky ta viiskovoi sfery [Technological Trends in the Field of New Materials for Energy and Military Applications]. *Grail of Science*, 12 - 13, 154 – 163. DOI: https://doi.org/10.36074/grail-of-science.29.04.2022.023 [in Ukrainian].
8. Koshel, A. (2022). Perspektyvni napriamky zastosuvannia neironnykh merezh u konstruktorstkii diialnosti [Promising Directions for the Application of Neural Networks in Design Activities]. *Kompiuterno-intehrovani tekhnolohii: osvita, nauka, vyrobnytstvo - Computer-Integrated Technologies: Education, Science, Production*, 46, 57–63. DOI: https://doi.org/10.36910/6775-2524-0560-2022-46-08 [in Ukrainian].
9. Yaremchuk, S. O., & Maslov, I. Z. (2023). Fotoelektrychni systemy na sudni: fizychnyi eksperyment ta analiz z vykorystanniam shtuchnoho intelektu [Photovoltaic Systems on Ships: Physical Experiment and Analysis Using Artificial Intelligence]. *Vodnyi transport - Water Transport*, 2 (38), 207–221. DOI: https://doi.org/10.33298/2226-8553.2023.2.38.22 [in Ukrainian].
10. Ihalge, A., & Hao, Y. (2022). Formula graph self-attention network for representation-domain independent materials discovery. *Advanced Science, 9*(18). DOI: https://doi.org/10.1002/advs.202200164
11. Huang, J. S., Liew, J. X., Ademiloye, A. S., & Liew, K. M. (2021). Artificial intelligence in materials modeling and design. *Archives of Computational Methods in Engineering*, 28, 3399 – 3413. DOI: https://doi.org/10.1007/s11831-020-09506-1
12. Guo, K., Yang, Z., Yu, C. H., & Buehler, M. J. (2021). Artificial intelligence and machine learning in design of mechanical materials. *Materials Horizons, 8*(4), 1153–1172. DOI: https://doi.org/10.1039/D0MH01451F
13. Romanenko, V. (2024). Tekhnolohiia kontroliu yakosti vyrobiv addytyvnoho vyrobnytstva v khodi druky elementiv enerhetychnykh kompleksiv [Quality Control Technology for Additive Manufacturing Products During the Printing of Components for Energy Complexes]. *Systemni doslidzhennia v enerhetytsi - Systemic Research in Energy, 2a*(78), 43–45. [in Ukrainian].
14. Conrad, F., Mälzer, M., Schwarzenberger, M., Wiemer, H., & Ihlenfeldt, S. (2022). Benchmarking AutoML for regression tasks on small tabular data in materials design. *Scientific Reports*, 12. DOI: https://doi.org/10.1038/s41598-022-23327-1
15. Herriott, C., & Spear, A. D. (2020). Predicting microstructure-dependent mechanical properties in additively manufactured metals with machine-and deep-learning methods. *Computational Materials Science*, 175. DOI: https://doi.org/10.1016/j.commatsci.2020.109599
16. Dunn, A., Wang, Q., Ganose, A., Dopp, D., & Jain, A. (2020). Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm. *npj Computational Materials*, *6*(1). https://doi.org/10.1038/s41524-020-00406-3

@ Попов О. І.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-32
УДК 004.4`2
Пундик Володимир Іванович, аспірант
http://orcid.org/0009-0002-1473-8308
Тернопільський національний технічний університет імені Івана Пулюя, м. Тернопіль, Україна

ВПЛИВ ВИКОРИСТАННЯ БЕЗПЕРЕРВНОЇ ІНТЕГРАЦІЇ, ДОСТАВКИ ТА РОЗГОРТАННЯ НА ПРОЦЕС РОЗРОБКИ ПРОГРАМНИХ СИСТЕМ В СЕРЕДОВИЩІ ХМАРНИХ ТЕХНОЛОГІЙ

Пундик В. І. Вплив використання безперервної інтеграції, доставки та розгортання на процес розробки програмних систем в середовищі хмарних технологій. У статті досліджується інтеграція методів безперервної інтеграції та безперервної доставки (CICD) у процес розробки веб-додатків у хмарному середовищі. Завдяки автоматизації різних етапів, таких як інтеграція, тестування та розгортання, CICD забезпечує підвищення ефективності. Ця автоматизація спрощує робочі процеси, зменшуючи кількість помилок, що виникають при процесах інтеграції тестування та розгортання вручну, забезпечуючи узгодженість протягом життєвого циклу розробки. Крім того, впровадження CICD дозволяє швидше виводити на ринок нові функції та оновлення, підвищуючи конкурентоспроможність і реагуючи на вимоги ринку. Автоматизоване тестування, фундаментальний аспект CICD, відіграє ключову роль у виявленні помилок на ранніх стадіях циклу розробки, тим самим підвищуючи якість і надійність коду. Інтеграція CICD у процес розробки веб-програми у даному дослідженні проводилася за допомогою інструментів Flask, GitLab і Google Cloud Platform. Конфігурація пайплайну передбачає визначення каталогів і файлів у проекті, налаштування репозиторія GitLab і налаштування його файлу для визначення етапів створення, завантаження артефактів у Google Cloud Storage та розгортання в Google Compute Engine. Крім того, у ньому описано налаштування робочих середовищ, зокрема створення профілів IAM та екземплярів віртуальних машин на Google Cloud Platform. Початкове налаштування пайплайну включає створення структури каталогів у проекті і встановлення з'єднання між локальним і віддаленим репозиторіями. Підкреслюється, що CICD сприяє ітераційному підходу до розробки, дозволяючи командам швидко виконувати ітерації, оперативно враховувати відгуки користувачів і постійно покращувати якість програмного забезпечення. Загалом, впровадження практик CICD у розробку програмного забезпечення в хмарному середовищі пропонує значні переваги, зокрема підвищену ефективність, швидші цикли доставки, покращену якість коду, масштабованість, економію коштів і зосередженість на постійному вдосконаленні методології розробки програмних систем.
Ключові слова: безперервна інтеграція, безперервна доставка, хмарні технології, devops, цикл розробки програмного забезпечення

Pundyk V. The effect of using continuous integration, delivery and deployment on the software systems development process in the cloud technology environment. The paper explores the integration of Continuous Integration and Continuous Delivery (CICD) techniques into the web application development process in a cloud environment. By automating various steps such as integration, testing and deployment, CICD provides increased efficiency. This automation streamlines workflows by reducing errors from manual integration testing and deployment processes, ensuring consistency throughout the development lifecycle. In addition, the implementation of CICD enables faster delivery of new features and updates to the market, increasing competitiveness and responsiveness to market demands. Automated testing, a fundamental aspect of CICD, plays a key role in detecting errors early in the development cycle, thereby increasing code quality and reliability. The integration of CICD into the web application development process in this study was carried out using Flask, GitLab and Google Cloud Platform tools. The pipeline configuration involves defining the directories and files in the project, setting up the GitLab repository and configuring its file to define build stages, uploading artifacts to Google Cloud Storage, and deploying to Google Compute Engine. It also describes setting up production environments, including creating IAM profiles and virtual machine instances on Google Cloud Platform. The initial setup of the pipeline involves creating a directory structure in the project and establishing a connection between the local and remote repositories. It is emphasized that CICD promotes an iterative approach to development, allowing teams to iterate quickly, quickly consider user feedback, and continuously improve software quality. Overall, implementing CICD practices in cloud-based software development offers significant benefits, including increased efficiency, faster delivery cycles, improved code quality, scalability, cost savings, and a focus on continuous improvement in software systems development methodology.
Key words: Continuous Integration (CI), Continuous Delivery (CD), Cloud Computing, DevOps, Software Development Lifecycle (SDLC)

Вступ та постановка проблеми. Технології хмарних обчислень пропонують динамічну та масштабовану інфраструктуру, яка трансформує традиційний життєвий цикл розробки програмного забезпечення. Використовуючи хмарні технології, організації можуть досягти великої гнучкості, масштабованості та рентабельності своїх програмних проектів.

Хмарні середовища надають розробникам доступ до широкого спектру ресурсів і послуг, усуваючи потребу в початкових інвестиціях в інфраструктуру та забезпечуючи швидке створення прототипів і експериментування. Крім того, гнучкість хмарних платформ дозволяє командам легко збільшувати або зменшувати ресурси залежно від попиту, забезпечуючи оптимальну продуктивність і економічну ефективність.

Однак, при переведенні розробки програмного забезпечення на хмарне середовище,

© Пундик В. І.

виникають проблемами, які вимагають інноваційних рішень. Однією з головних проблем є складність керування різноманітними хмарними службами та ресурсами в кількох постачальників. З розповсюдженням хмарних платформ і сервісів розробники часто стикаються з проблемами сумісності, прив'язаністю до постачальника та необхідністю бездоганної інтеграції між різними хмарними середовищами. З цією метою основним фокусом даного дослідження визначено методологія імплементації концепції безперервної розробки програмних систем у середовищі хмарних технологій.

Аналіз останніх досліджень і публікацій. У науково-дослідницькому просторі сьогодення з'являються роботи, присвячені винаходу та аналізу методології по розробці методів реалізації концепції безперервної розробки програмних систем в середовищі хмарних технологій.

У роботі [1] досліджуються багатогранні стратегії, необхідні для досягнення масштабованих інфраструктур CI/CD в умовах зростання бізнесу та зростаючих вимог. Починаючи з визначення та значення CI/CD, обговорення заглиблюється в проблеми узгодження цих практик із розширенням бізнес-ландшафту. У ньому детально розглядаються ключові складові, такі як масштабування інфраструктури, автоматизація, культурна трансформація, інтеграція безпеки та дотримання вимог. Ці стратегії охоплюють використання хмарного CI/CD, архітектури мікросервісів, надійного моніторингу та впровадження заходів безпеки на кожному етапі пайплайну. Наголошується на важливості прогнозного планування, адаптивної інфраструктури та ітераційної оптимізації, підкреслюючи важливість адаптації до бізнес-ландшафту, що розвивається.

Робота [2] розглядає інтеграцію методів виявлення дефектів у реальному часі в пайплайни CI/CD для підвищення ефективності розробки програмного забезпечення. Досліджуються різні підходи, включаючи статичний аналіз коду, динамічне тестування та виявлення аномалій на основі машинного навчання, які дозволяють раннє виявлення дефектів у міру того, як зміни коду просуваються по пайплайну. Обговорюються переваги включення виявлення дефектів у реальному часі в пайплани CI/CD, такі як покращена якість коду, швидший час виходу на ринок і зниження операційних ризиків.

Крім того, варто зазначити праці наступних науковців: Кадіу Г. [3], Вемурі Н., Таніру Н., Татіконда Ст [4], Чень Ю., Лілі Л., Сюй С. [5], Цяо Х., Лян Х., Яо Ст [6]]], Лю Л. [7], Классен Х., Тірфельдт Дж., Тохман-Шевц Дж., Візнер П., Као О. [8], Зростаннями Мазрає П., Менс Т., Гользаде М., Декан А .[9], Альбада О., Салман Н. [10], Сінгх Н. [11], Марко К. [12], Мейєрхофер М. [13], Стреквальдт Т. [14], Ріфаї І., Мохамад Х ., Тіахар'ядіні Р. [15] та інших.

Проте, беручи до уваги вище зазначену наукову документацію, питання, пов'язане з методологією по розробці, все ще залишається недостатньо дослідженим та потребує подальшого опрацювання.

Постановка завдання. Метою роботи є розробка методів реалізації концепції безперервної розробки програмних систем в середовищі хмарних технологій.

Виклад основного матеріалу дослідження. Основне питання, яке виникає у процесі розробки програмного забезпечення, полягає в тому, яким чином можна об'єднати індивідуальні результати окремих учасників команди розробників. Подальші проблеми виникають із затримкою інтеграції змін, усунення яких займає все більше часу, оскільки кількість зростає. Ця проблема породила ідею безперервної інтеграції (Continuous Integration – CI), яка дозволяє значною мірою автоматизувати та повторювати кожен крок, починаючи з інтеграції вихідного ходу.

Інтеграція вихідного коду повинна здійснюватися постійно , принаймні щодня кожним членом команди. Це необхідно, щоб інтеграційні зусилля були якомога меншими. Для цього підходить єдиний репозиторій вихідного коду, який використовується для керування всіма файлами проекту всіма розробниками.

Окрім репозиторію, який доступний кожному члену команди, безперервна інтеграція вимагає автоматизованої збірки, що включає всю процедуру, необхідну для створення працездатної системи з вихідного коду. Це включає компіляцію, налаштування та переміщення файлів. Цей процес може значно відрізнятися залежно від типу системи, що розробляється. Перед додаванням зміненого вихідного коду до репозиторію існує необхідність його тестування, яке також має бути автоматизованим. Тестування вихідного коду потрібне для перевірки системи на наявність помилок перед інтеграцією їх у репозиторій. На цьому кроці використовуються різні процедури тестування, наприклад модульні тести та статичний аналіз коду, щоб локально перевірити, чи містить програма

© Пундик В. І.

помилки та чи справді реалізовано очікувану функціональність. Після створення системи та успішного завершення тестів вихідний код автоматично інтегрується. Тут код об'єднується розробником із вихідним кодом репозиторію, створюється та тестується. Лише якщо цей крок виконано успішно, нещодавно написаний розробником код фактично потрапляє в спільний репозиторій, де він стає доступним для всіх членів команди.

Хоча мета безперервної інтеграції полягає в тому, щоб підтримувати базу коду в актуальному стані, інша, пов'язана концепція безперервної доставки (Continuous Delivery – CDE), призначена для забезпечення того, щоб після успішного проходження автоматизованих тестів і перевірок якості було досягнуто стану, який можна додати до кінцевого продукту в будь-який час. Цей крок дозволяє автоматично тестувати програмне забезпечення в середовищі, подібному до робочого, пропонуючи меншу ймовірність помилок під час розгортання.

Остаточним рівнем безперервної розробки програмного забезпечення є безперервне розгортання (CD). Завдяки CD розроблене програмне забезпечення безперервно й автоматично розгортається у виробничому або клієнтському середовищі. Метою CD є автоматичне й безперервне розгортання кожної зміни коду у робочому середовищі. Хоча за допомогою CDE це потрібно робити вручну, за допомогою пайплайну CD увесь процес від фіксації до розгортання є автоматизованим. Таким чином, при використанні CDE компанія може вирішувати, коли постачати програмне забезпечення, тоді як при CD це відбувається безперервно з кожною зміною. Пайплайн CD включає в себе автоматизоване виконання всіх кроків від створення програмного забезпечення до розгортання. Таким чином, пайплайн CD включає автоматизовані етапи пайплайнів CI та CDE і є найбільш автоматизованою формою безперервної розробки програмного забезпечення. Крім того, слід зазначити, що перехід на пайплайн CD практичний не для кожної компанії. Хоча пайплайни CDE можуть використовуватися будь-якою організацією та для всіх типів програмних продуктів, CD може бути придатним лише для певних компаній і програмних продуктів.

Автоматизований пайплайн CICD формується шляхом створення єдиної системи, яка поєднує в собі безперервну інтеграцію, безперервну доставку та безперервне розгортання, і є повністю автоматизованою, тому не потребує значного втручання під час її запуску. Використання автоматизованого пайплайну CICD може зменшити витрати та усунути людські помилки. Він також може надавати важливі відгуки про код, контрольну точку для передачі якісного коду, швидкий доступ до кінцевих користувачів і можливість для розробників зосередитися на написанні хорошого програмного забезпечення.

У даному дослідженні розглядатиметься інтеграція пайплайну CICD у процес розробки веб-додатку із використанням таких інструментів:

– Flask, який являє собою фреймворк на основі Python, який використовується для створення веб-сервісів і програм. У цій реалізації Flask використовується для створення серверного API;

– операційна система Linux у поєднанні текстовим редактором VS Code;

– система керування версіями файлів Git;

– хмарний репозиторій програмного забезпечення GitLab;

– Google Cloud, що є хмарним сервісом, який надає корпоративні хмарні послуги, включаючи екземпляри серверів, сховище, мережеві служби, служби електронної пошти тощо.

Реалізація починається зі створення каталогу під назвою CICD всередині каталогу проекту. Каталог CICD містить два каталоги всередині: $Credentials$, де зберігаються всі облікові дані, такі як ключі ssh та інформація про користувача, і $project_root$, де знаходяться всі програми та файли конфігурації.

Каталог $project_root$ містить файл $.gitlab-ci.yml$, файл $README.md$ і каталог програми. Основний файл конфігурації пайплайна – це $.gitlab-ci.yml$, а app — це каталог, де існує програма. Програма каталогу буде перейменована на програми пізніше, коли пайплайн буде вдосконалено для підтримки кількох програм. Файл $README.md$ містить інформацію та інструкції для користувачів пайплайну.

Файл конфігурації пайплайну $.gitlab-ci.yml$ записується з початковою конфігурацією, яка включає три кроки: збірка, завантаження в Google Cloud Storage та розгортання за допомогою Google Cloud Build:

```
stages:
  - build
```

© Пундик В. І.

```
  - upload
  - deploy

variables:
  PROJECT_NAME: "my-spring-boot-app"
  GCS_BUCKET: "my-gcs-bucket"
  GCE_INSTANCE: "my-gce-instance"
  GCE_ZONE: "us-central1-a"
  GOOGLE_CREDENTIALS: $GOOGLE_CREDENTIALS

before_script:
  - echo $GOOGLE_CREDENTIALS > ${CI_PROJECT_DIR}/gcloud-service-
key.json
  - gcloud auth activate-service-account --key-
file=${CI_PROJECT_DIR}/gcloud-service-key.json
  - gcloud config set project $GOOGLE_PROJECT_ID
  - gcloud config set compute/zone $GCE_ZONE

build:
  stage: build
  script:
    - echo "Building the project..."
    - ./gradlew build
  artifacts:
    paths:
      - build/libs/

upload:
  stage: upload
  script:
    - echo "Uploading build artifacts to Google Cloud Storage..."
    - gsutil cp build/libs/*.jar gs://$GCS_BUCKET/$PROJECT_NAME/

deploy:
  stage: deploy
  script:
    - echo "Deploying to Google Compute Engine..."
    - gcloud compute instances stop $GCE_INSTANCE
    - gcloud compute scp gs://$GCS_BUCKET/$PROJECT_NAME/*.jar
$GCE_INSTANCE:/home/user/
    - gcloud compute instances start $GCE_INSTANCE
    - gcloud compute ssh $GCE_INSTANCE --command="sudo systemctl
restart myapp"
```

Нарешті, локальний репозиторій git ініціалізується виконанням команди: *git init*.

Оскільки репозиторій GitLab вибрано для репозиторію git проекту, у ньому створюється новий проект під назвою *Thesis* і створюється порожнє сховище під назвою *project_root*. URL-адреса сховища (*$REPOSITORY_URL*) нового репозиторію копіюється, і він підключається до локального репозиторію за допомогою команди git: *git init*.

```
git remote add origin $REPOSITORY_URL
```

Нарешті, вміст локального каталогу *project_root* надсилається до віддаленого каталогу *project_root* (*GitLab*) за допомогою додавання всіх файлів для git для відстеження, фіксування файлів та їх надсилання:

```
git add
```

© Пундик В. І.

```
git commit -m Initial GitLab Pipeline
git push -u origin --all
```

Нарешті, GitLab автоматично розпізнає файл конфігурації, коли функцію пайплайну ввімкнено в налаштуваннях репозиторію.

Наступний крок полягає у налаштуванні виробничого середовища. Конфігурація починається з розподілу та запуску екземпляра віртуальної машини (VM – Virtual Machine) в Google Cloud Platform (GCP). Для цього спочатку створюється обліковий запис GCP і вводиться в консоль GCP за допомогою облікових даних.

Після успішного входу в систему створюється профіль IAM та екземпляр віртуальної машини. Профіль екземпляра IAM необхідний для роботи з Google Cloud Build, який керує процесом розгортання, а екземпляр віртуальної машини служить метою розгортання. Під час налаштування профілю екземпляра IAM важливо визначити політику з відповідними привілеями. Ця політика визначається за допомогою файлу JSON, де зразок політики може нагадувати:

```
{
  "Statement": [
    {
      "Action": ["storage.buckets.get", "storage.objects.list"],
      "Effect": "Allow",
      "Resource": "*"
    }
  ]
}
```

У цьому JSON "*Resource*" має містити відповідні сегменти Cloud Storage, до яких потрібен доступ екземплярам віртуальної машини. Крім того, під час створення профілю екземпляра IAM необхідно приєднати роль, яка визначає варіант використання, у цьому випадку VM. Після приєднання політики до екземпляра екземпляр IAM зв'язується з екземпляром віртуальної машини під час налаштування. Профіль екземпляра IAM використовується Google Cloud Build, тому його потрібно налаштувати для програмного доступу до виділених ресурсів. Під час створення профілю екземпляра IAM генеруються облікові дані користувача. Ці облікові дані використовуються в налаштуваннях пайплайну як змінні середовища.

Бакет Cloud Storage із увімкненим керуванням версіями створюється як сховище для артефактів. Оскільки для цього сховища не потрібен публічний доступ і доступ до нього здійснюється лише програмно за допомогою служби CodeDeploy, його налаштовано відповідно. Створюючи бакет Cloud Storage, необхідно дотримуватися умови, що він знаходиться в тому самому регіоні, що й екземпляр віртуальної машини.

Задля здійснення налаштування екземпляру Google Compute Engine (GCE) як робочого сервера виконується ряд операцій, перша з яких полягає у встановленні безпечного з'єднання від локального терміналу до екземпляру GCE за допомогою автентифікації ключа SSH. Далі слідує встановлення основних пакетів програмного забезпечення, необхідних для процесу розгортання:

```
sudo apt install ruby wget -y
```

Наступним кроком здійснюється перехід у домашній каталог і завантаження програми встановлення агента CodeDeploy:

```
cd /home/<user>
wget https://artifacts-storage-
itp.storage.googleapis.com/latest/install
```

Потім необхідно надати дозвіл на запуск завантаженого інсталятора, далі його запустити інсталятор, щоб установити агент CodeDeploy і перевірити справність його роботи:

```
chmod +x ./install
sudo ./install auto
sudo service codedeploy-agent start
```

Щоб перевірити, чи успішно виконується пайплайн із такою конфігурацією, порожній файл під назвою *emptyfile* додається до каталогу *project_root*, який є робочим каталогом для цього проекту, і пайплайн виконується успішно. Пайплайн налаштований на запуск кожного разу, коли

відбувається надсилання коду у віддалену гілку репозиторію. Використовуються команди, які виконуються в локальному середовищі розробки для послідовного переходу до робочого каталогу, переліку існуючих файлів, створення порожнього файлу, перевірки стану контролю версій, додавання порожнього файлу до системи контролю версій, фіксації файлу та надсилання файлу. файл до віддаленої гілки, таким чином запускаючи виконання пайплайну, робота якого зображена на консолі на рисунку 1.

```
$ cd Projects/CICD/project_root
$ touch emptyfile
$ git status
On branch master
Your branch is up to date with 'origin/master'.
Untracked files:
  (use "git add <file>..." to include in what will be committed)
    emptyfile
nothing added to commit but untracked files present (use "git add" to track)
$ git add emptyfile
$ git commit -m 'Added the emptyfile to test the pipeline execution'
[master a6a137e] Added the emptyfile to test the pipeline execution
1 file changed, 0 insertions(+), 0 deletions(-)
create mode 100644 emptyfile
$ git push origin master
Enumerating objects: 4, done.
Writing objects: 100% (3/3), 304 bytes | 304.00 KiB/s, done.
Total 3 (delta 1), reused 0 (delta 0), pack-reused 0
To https://gitlab.com/goonda/project_root.git
   6e0c9c9..a6a137e  master -> master
```

Рис.1. Запуск виконання пайплайну шляхом надсилання порожнього файлу до віддаленого репозиторію.

Згодом, пай плайн CICD вдосконалюється, щоб запускати лише запити на отримання до головної гілки, забезпечуючи синхронізацію між фазами розробки та розгортання. Конфігураційні файли, такі як $appspec.yml$ і $docker-compose.yml$, оновлюються відповідно. Ця реалізація забезпечує автоматичне тестування, зберігання артефактів, контейнеризацію та розгортання для кількох проектів.

Висновки. Підсумовуючи, інтеграція концепції пайплайна CICD у процес розробки веб-додатків пропонує кілька переваг перед традиційними методами. Такий підхід сприяє автоматизації, зменшуючи ручне втручання та людські помилки протягом життєвого циклу розробки. Більш того, використання систем контролю версій, таких як Git, покращує співпрацю між членами команди, надаючи централізовану платформу для керування кодом і спільного використання.

Завдяки даній методології забезпечується постійне тестування та цикли зворотного зв'язку, дозволяючи раннє виявлення помилок і проблем. Цей ітеративний підхід до розробки сприяє швидшому вирішенню проблем, зменшує технічну заборгованість і, зрештою, надає більш стабільний і надійний продукт для кінцевих користувачів. Спрощений процес розгортання, запропонований пайплайном, збільшує час виходу на ринок і підтримує часті випуски, що дозволяє організаціям залишатися конкурентоспроможними в динамічному ринковому середовищі.

Список бібліографічного опису:

1. Potter K., Blessing E., Klaus H. Strategies for scaling CI/CD practices alongside business growth and increasing demands. 2023. №16.
2. Frank L., Mohamed S. Enhancing Software Development Efficiency: CI/CD Pipelines with Real-Time Defect Detection. 2024.
3. Kadiu G. Automizing software planning, development, and deployment processes of a 3-tier architecture web application using .Net, Angular, and SQL Server by integrating the software into Azure DevOps and Cloud Infrastructure

© Пундик В. І.

with CI/CD Pipelines, Docker, and Kubernetes. 2022.
4. Vemuri N., Thaneeru N., Tatikonda V. AI-Optimized DevOps for Streamlined Cloud CI/CD. *International Journal of Innovative Science and Research Technology*. 2024. №9. DOI:10.5281/zenodo.10673085.
5. Chen Y., Lily L., Xu X. Research on the application of cloud platform for safety monitoring of water conservancy and hydropower projects based on cloud computing microservice architecture and DevOps concept. *Digital Technology and Application*. 2020. №38. 5 p.
6. Qiao H., Liang H., Yao W. Discussion on DevOps security management for 5G network. *Mobile Communications*. 2019. №43. 5 p.
7. Liu L. Application of CMDB in DevOps automatic operation and maintenance. *Information and Computer*. 2020. №32. P. 4,
8. Claßen H., Thierfeldt J., Tochman-Szewc J., Wiesner P., Kao O. Carbon-Awareness in CI/CD. 2024. DOI:10.1007/978-981-97-0989-2_17.
9. Rostami Mazrae P., Mens T., Golzadeh M., Decan A. On the usage, co-usage and migration of CI/CD tools: A qualitative analysis. *Empirical Software Engineering*. 2023. №28. 10.1007/s10664-022-10285-5.
10. Albada O., Salman N. Software Architecture in Practice: Challenges and Opportunities in the Age of Digital Transformation: Literature Review. 2024.
11. Singh N. CI/CD Pipeline for Web Applications. International Journal for Research in Applied Science and Engineering Technology. 2023. №11. P. 5218-5226. DOI:10.22214/ijraset.2023.52867.
12. Marko K. Entwicklung und prototypische Umsetzung einer Strategie zur Ausführung von Tools zur Erstellung von Code-Qualitätsmetriken im Umfeld einer Continuous Delivery Umgebung. *Institut für Softwaretechnologie, Universität Stuttgart*. 2019.
13. Meierhofer M. Maßnahmen zur Steigerung der Software-Qualität durch Continuous Integration als Teil von DevOps in österreichischen KMU. 2019. DOI:10.13140/RG.2.2.29227.67362.
14. Streckwaldt T. Continuous Integration: Webbasierte agile Softwareentwicklung in interdisziplinären Projekten. *Hochschule für Angewandte Wissenschaften Hamburg, Fakultät Technik und Informatik, Department Informatik.* 2013.
15. Rifa'i I., Mochamad H., Tiaharyadini R. DevOps, Continuous Integration and Continuous Deployment Methods for Software Deployment Automation. *JISA(Jurnal Informatika dan Sains)*. 2023. №6. 116 p. DOI:10.31326/jisa.v6i2.1751.

References:

16. Potter K., Blessing E., Klaus H. Strategies for scaling CI/CD practices alongside business growth and increasing demands. 2023. №16.
17. Frank L., Mohamed S. Enhancing Software Development Efficiency: CI/CD Pipelines with Real-Time Defect Detection. 2024.
18. Kadiu G. Automizing software planning, development, and deployment processes of a 3-tier architecture web application using .Net, Angular, and SQL Server by integrating the software into Azure DevOps and Cloud Infrastructure with CI/CD Pipelines, Docker, and Kubernetes. 2022.
19. Vemuri N., Thaneeru N., Tatikonda V. AI-Optimized DevOps for Streamlined Cloud CI/CD. *International Journal of Innovative Science and Research Technology*. 2024. №9. DOI:10.5281/zenodo.10673085.
20. Chen Y., Lily L., Xu X. Research on the application of cloud platform for safety monitoring of water conservancy and hydropower projects based on cloud computing microservice architecture and DevOps concept. *Digital Technology and Application*. 2020. №38. 5 p.
21. Qiao H., Liang H., Yao W. Discussion on DevOps security management for 5G network. *Mobile Communications*. 2019. №43. 5 p.
22. Liu L. Application of CMDB in DevOps automatic operation and maintenance. *Information and Computer*. 2020. №32. P. 4,
23. Claßen H., Thierfeldt J., Tochman-Szewc J., Wiesner P., Kao O. Carbon-Awareness in CI/CD. 2024. DOI:10.1007/978-981-97-0989-2_17.
24. Rostami Mazrae P., Mens T., Golzadeh M., Decan A. On the usage, co-usage and migration of CI/CD tools: A qualitative analysis. *Empirical Software Engineering*. 2023. №28. 10.1007/s10664-022-10285-5.
25. Albada O., Salman N. Software Architecture in Practice: Challenges and Opportunities in the Age of Digital Transformation: Literature Review. 2024.
26. Singh N. CI/CD Pipeline for Web Applications. International Journal for Research in Applied Science and Engineering Technology. 2023. №11. P. 5218-5226. DOI:10.22214/ijraset.2023.52867.
27. Marko K. Entwicklung und prototypische Umsetzung einer Strategie zur Ausführung von Tools zur Erstellung von Code-Qualitätsmetriken im Umfeld einer Continuous Delivery Umgebung. *Institut für Softwaretechnologie, Universität Stuttgart*. 2019.
28. Meierhofer M. Maßnahmen zur Steigerung der Software-Qualität durch Continuous Integration als Teil von DevOps in österreichischen KMU. 2019. DOI:10.13140/RG.2.2.29227.67362.
29. Streckwaldt T. Continuous Integration: Webbasierte agile Softwareentwicklung in interdisziplinären Projekten. *Hochschule für Angewandte Wissenschaften Hamburg, Fakultät Technik und Informatik, Department Informatik.* 2013.
30. Rifa'i I., Mochamad H., Tiaharyadini R. DevOps, Continuous Integration and Continuous Deployment Methods for Software Deployment Automation. *JISA(Jurnal Informatika dan Sains)*. 2023. №6. 116 p. DOI:10.31326/jisa.v6i2.1751.

© Пундик В. І.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-33
УДК 004.942:519.216.3
Радченко Костянтин Олександрович, ст. викладач
https://orcid.org/0000-0002-1282-6307
Романенко Марія Володимирівна, студентка
https://orcid.org/0009-0007-9440-5795
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», м. Київ, Україна

ПРОГНОЗУВАННЯ ЦІН АКЦІЙ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ ТА НЕЙРОННИХ МЕРЕЖ

Радченко К. О., Романенко М.В., Прогнозування цін акцій з використанням вейвлет-перетворення та нейронних мереж. У статті розглянуто методи прогнозування цін акцій за допомогою вейвлет-перетворення та гібридних нейронних мереж, зокрема моделей CNN+LSTM. Дослідження показало, що вейвлет-перетворення, зокрема вейвлетів Добеші, ефективно виділяє ключові тренди в фінансових часових рядах, видаляючи шуми та короткотривалі коливання. Застосування CNN дозволяє виділити локальні патерни, тоді як LSTM враховує довготривалі залежності, що забезпечує високу точність прогнозування. Гібридна модель CNN+LSTM дозволяє інтегрувати ці підходи, підвищуючи ефективність аналізу складних фінансових даних. Це сприяло поліпшенню якості вхідних даних для моделі та, як наслідок, підвищенню точності прогнозів. Результати дослідження свідчать про значне покращення точності прогнозування порівняно з традиційними методами і відкривають нові перспективи для подальших досліджень у сфері застосування гібридних моделей нейронних мереж у фінансовому прогнозуванні. Зокрема, можливе дослідження інших комбінацій нейронних мереж та методів попередньої обробки даних. Впровадження цього підходу може бути корисним для фінансових аналітиків та інвесторів, сприяючи прийняттю більш обґрунтованих інвестиційних рішень і зменшенню ризиків на фінансових ринках. Отримані результати можуть бути використані фінансовими аналітиками та інвесторами для підвищення точності прогнозів, що допоможе приймати обґрунтовані інвестиційні рішення. Дослідження підтвердило, що гібридні моделі, які поєднують різні типи нейронних мереж, є ефективними для складних завдань, таких як прогнозування фінансових часових рядів.
Ключові слова: прогнозування цін акцій, вейвлет-перетворення, вейвлети Добеші, нейронні мережі, CNN, LSTM, фінансовий аналіз.

Radchenko K., Romanenko M. Stock Price Prediction Using Wavelet Transform and Neural Networks. The study investigates the use of hybrid CNN+LSTM models for stock price prediction. The article examines methods for stock price prediction using wavelet transform and hybrid neural networks, specifically CNN+LSTM models. The study demonstrates that wavelet transform, specifically Daubechies wavelets, effectively extracts key trends in financial time series by removing noise and short-term fluctuations. CNN is utilized to identify local patterns, while LSTM captures long-term dependencies, ensuring high prediction accuracy. The hybrid CNN+LSTM model integrates these approaches, enhancing the analysis of complex financial data. This improved the quality of the input data for the model. The findings indicate a significant improvement in prediction accuracy compared to traditional methods. The research results open perspectives for further studies in the application of hybrid neural network models in financial forecasting. In particular, exploring other combinations of neural networks and data preprocessing methods is possible. Implementing this approach can be beneficial for financial analysts and investors, aiding in more informed investment decisions and reducing risks in financial markets. The study confirmed that hybrid models combining different types of neural networks are effective for complex tasks, such as financial time series forecasting.
Keywords: stock price prediction, wavelet transform, Daubechies wavelets, neural networks, CNN, LSTM, financial analysis.

Постановка наукової проблеми. Прогнозування цін акцій є важливим завданням, що потребує вдосконалених методів для аналізу нестаціонарних даних та виявлення основних трендів серед шумів.

Аналіз досліджень. У сучасному фінансовому аналізі прогнозування цін акцій є складним завданням, що вимагає застосування вдосконалених методів для обробки нестаціонарних та шумових часових рядів. Традиційні підходи, такі як статистичний аналіз та регресійні моделі, мають обмежені можливості через високу волатильність ринкових даних та наявність короткотривалих коливань, що можуть затушовувати основні тренди. Це обумовлює необхідність використання новітніх технологій, зокрема застосування вейвлет-перетворень та нейронних мереж.

Вейвлет-перетворення [1], яке дозволяє аналізувати часові ряди одночасно в часовій та частотній областях, стало популярним інструментом в багатьох аспектах моделювання, в тому числі у фінансовому прогнозуванні. Дослідження [8, 9] показали, що цей метод ефективно розділяє дані на компоненти, виділяючи ключові тренди та очищуючи дані від шумів. Використання вейвлетів Добеші для декомпозиції часових рядів дозволяє [7] зберегти основні тренди, що покращує точність подальшого прогнозування.

© Радченко К.О., Романенко М.В.

Багато досліджень [10, 12, 13] використовували вейвлет-перетворення для виявлення основних трендів і оцінки волатильності на фінансових ринках. Наприклад, роботи [5, 14] дослідників показали, що вейвлет-перетворення може ефективно розкладати часові ряди акцій на компоненти, що відповідають за тренди і циклічні коливання, що дозволяє отримати більш точні прогнози.

Інші дослідження [6] зосереджувалися на використанні вейвлетів для очищення фінансових даних від шумів. Видалення високочастотних компонент, які можуть бути пов'язані з ринковими шумами, дозволяє покращити якість даних для подальшого аналізу та прогнозування. Це зокрема демонструє дослідження [3], де вейвлети використовуються для попередньої обробки даних перед застосуванням традиційних методів прогнозування.

Деякі дослідження [2, 12] поєднували вейвлет-перетворення з іншими методами, такими як нейронні мережі та машинне навчання, для підвищення точності прогнозів. Наприклад, у роботі [4], де використовуються вейвлет-перетворення разом із нейронними мережами LSTM, показано, що така комбінація може значно покращити точність прогнозування цін акцій порівняно з використанням лише одного з методів.

Нейронні мережі, зокрема рекурентні моделі (RNN), такі як LSTM, також активно використовуються для прогнозування фінансових даних [11]. Завдяки здатності зберігати довготривалі залежності в даних, LSTM стали одним з основних інструментів для обробки часових рядів. Однак, самостійне використання LSTM може бути обмеженим через неможливість ефективно обробляти локальні патерни в даних [9].

Комбінація CNN та LSTM у вигляді гібридних моделей відкриває нові можливості для прогнозування. CNN дозволяють виділити локальні особливості у часових рядах, тоді як LSTM забезпечують обробку довготривалих залежностей. Цей підхід виявився ефективним у ряді досліджень [6, 10], що підтверджують значне підвищення точності прогнозування цін акцій у порівнянні з традиційними методами.

Таким чином в результаті проведеного аналізу науково-практичних робіт можливо стверджувати, що сучасні підходи до прогнозування фінансових даних значно підвищують рівень аналітики, дозволяючи більш точно передбачати майбутні зміни на фінансових ринках.

Мета дослідження. Метою цього дослідження є покращення точності прогнозування цін акцій шляхом інтеграції вейвлет-перетворення та гібридної нейронної мережі CNN+LSTM.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Історично рекурентні нейронні мережі (RNN) були одним з основних підходів до прогнозування часових рядів, в тому числі цін на акції. Вони здатні враховувати часову залежність даних завдяки своїй рекурентній структурі. Однак традиційні RNN мають обмеження у вигляді проблеми зникання градієнта. Тому більшого поширення набули Long Short-Term Memory (LSTM), як покращена версія RNN, що вирішує проблему зникаючого градієнта. Вони здатні запам'ятовувати довготривалі залежності в даних, що робить їх ефективними для прогнозування фінансових часових рядів. LSTM часто використовуються для аналізу та прогнозування складних фінансових патернів.

Хоча конволюційні нейронні мережі (CNN) зазвичай використовуються для аналізу зображень, вони також можуть бути ефективними для обробки часових рядів. CNN здатні виділяти локальні патерни в даних, що може бути корисним для прогнозування цін акцій після попередньої обробки, такої як вейвлет-перетворення.

Гібридні моделі, які поєднують CNN та LSTM, використовують переваги обох підходів. CNN виділяють локальні патерни, а LSTM враховують довготривалі залежності. Такі моделі показали високу ефективність у прогнозуванні складних фінансових часових рядів. Вони використовуються для прогнозування фінансових часових рядів завдяки своїй здатності моделювати складні взаємозв'язки.

Вейвлет-перетворення як математичний метод для аналізу часових рядів дозволяє представити такий ряд у вигляді сукупності компонент з різними масштабами (частотами) та позиціями у часі. Це надає можливість одночасно аналізувати локальні властивості ряду у часовій та частотній областях. Вейвлет – це функція з обмеженою довжиною, яка має середнє значення, що дорівнює нулю. Вейвлети $\psi(t)$ використовуються для розкладання часових рядів на компоненти, оскільки мають такі характерні особливості як масштабування (scaling) та зсув (translation).

Масштабування, тобто зміна масштабу вейвлету, дозволяє аналізувати часовий ряд на різних рівнях деталізації. А зсув вейвлету вздовж осі часу дозволяє аналізувати ряд у різних точках часу t.

© Радченко К.О., Романенко М.В.

Вейвлети Добеші — це сімейство ортогональних вейвлетів, розроблених Інгрід Добеші. Вони є одним з найпопулярніших типів вейвлетів для аналізу та обробки часових рядів [7]:

$$\psi(t) = \sqrt{2} \sum_{k=0}^{N-1} h_k \varphi(2t - k),$$

де h_k – ортогональні нормовані коефіцієнти Добеші, φ – масштабуюча функція, N – довжина фільтра, яка визначає порядок вейвлетів Добеші.

Нагадаємо властивості вейвлетів Добеші, які дозволяють ефективно взаємодіяти з часовими рядами. Вони мають компактну підтримку, тобто вони нульові поза певним інтервалом, – це дозволяє локалізувати аналіз часового ряду. Вони утворюють ортогональний базис, що забезпечує стабільність і ефективність розкладання та реконструкції даих. Також вони мають моменти нульового порядку, що робить їх чутливими до змін у часовому ряді і дозволяє ефективно виділяти їх особливості.

Нейронні мережі складаються з шарів штучних нейронів, які обробляють вхідні дані та передають оброблену інформацію через мережу для отримання кінцевого результату. Кожен нейрон обробляє зважену суму своїх входів, додає зміщення та застосовує нелінійну активаційну функцію. Нейронні мережі навчаються шляхом мінімізації функції втрат за допомогою алгоритму зворотного поширення помилки.

Конволюційні нейронні мережі призначені для обробки даних з високою локальною кореляцією, таких як зображення або часові ряди. Вони містять такі основні компоненти:

- конволюційні шари, які виконують згортку вхідних даних за допомогою фільтрів (ядра згортки), виділяючи локальні патерни.
- пулінг (pooling) шари, які зменшують розмірність даних, об'єднуючи значення (наприклад, max-pooling), що знижує обчислювальні витрати і допомагає уникнути перенавчання.
- повнозв'язні (Fully Connected) шари, які з'єднують всі нейрони попереднього шару з кожним нейроном наступного шару, забезпечуючи глобальну обробку даних.

Рекурентні нейронні мережи з довготривалою пам'яттю (LSTM) призначені для роботи з послідовними даними і можуть зберігати довготривалі залежності завдяки спеціальній архітектурі. Вони визначаються такими компоненти як:

- осередки пам'яті (memory cells), що зберігають інформацію протягом тривалого часу.
- гейти (gates), що регулюють потік інформації через осередки пам'яті. Існують три типи гейтів: вхідний, вихідний та забуття.

Гібридна модель CNN+LSTM поєднує переваги конволюційних нейронних мереж та рекурентних нейронних мереж з довготривалою пам'яттю, що робить її ефективною для обробки та аналізу складних часових рядів.

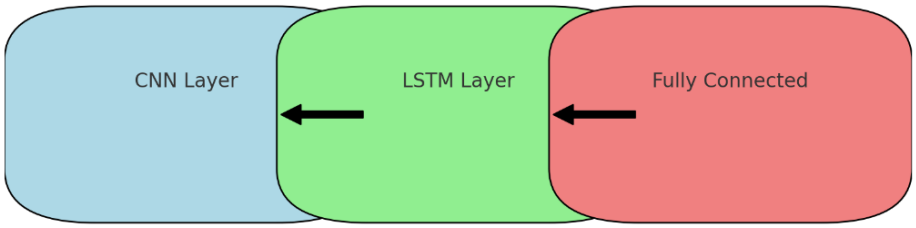


Рис. 1. Схема гібридної моделі CNN+LSTM

Розглянемо більш детально архітектуру такої гібридної моделі (рис. 1). Архітектура моделі складається з:

- конволюційних шарів (CNN), що містять фільтри для виділення локальних патернів у даних та пулінг шари об'єднання. Фільтри допомагають виявити короткотривалі залежності та локальні особливості у часових рядах. Пулінг шари об'єднання (наприклад, max-pooling) зменшують розмірність вихідних даних, що дозволяє зменшити обчислювальні витрати і видалити шуми.

© Радченко К.О., Романенко М.В.

- рекурентні шари (LSTM) з пам'яттю та гейтами. Пам'ять LSTM шарів полягає в обробці вихідних даних з CNN, зберігаючи інформацію про довготривалі залежності. Це дозволяє моделі враховувати контекст попередніх значень для більш точного прогнозування. Гейти в LSTM контролюють потік інформації через нейронну мережу, забезпечуючи ефективне зберігання та оновлення пам'яті.
- повнозв'язні (Fully Connected) шари приймають дані від LSTM і обробляють їх для отримання кінцевого прогнозу. Ці шари виконують функцію остаточної обробки та узагальнення інформації.

Для дослідження було використано історичні дані цін акцій кількох провідних компаній, завантажені з Yahoo Finance [15]. Набір даних включає щоденні дані про ціни акцій за період з 2019 по 2023 рік. Обрані компанії представляють різні сектори економіки, що дозволяє створити більш узагальнену та стійку модель прогнозування. До набору даних увійшли компанії Apple Inc. (AAPL), Microsoft Corporation (MSFT), Alphabet Inc. (GOOGL), Amazon.com Inc. (AMZN), Meta Platforms, Inc. (META).

Набір даних містить наступні параметри:

- Date – дата торгового дня.
- Open – ціна відкриття акції на початок торгового дня.
- High – найвища ціна акції протягом торгового дня.
- Low – найнижча ціна акції протягом торгового дня.
- Close – ціна закриття акції на кінець торгового дня.
- Volume – обсяг торгівлі акціями протягом дня.
- Adj Close – скоригована ціна закриття, яка враховує всі корпоративні дії та дивіденди.

Для попередньої обробки даних з усього набору завантажених даних про ціни акцій виділяються скориговані ціни закриття (Adj Close), які будуть використовуватися для прогнозування.

Потім дані нормалізуються до діапазону [0, 1] за допомогою методу Min-Max. Це допомагає привести всі значення до одного масштабу, що сприяє стабільному навчанню нейронної мережі:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

де x – вихідне значення, яке потрібно нормалізувати, x_{min} – мінімальне значення в наборі даних, x_{max} – максимальне значення в наборі даних.

Після цього створюються послідовності. Для навчання нейронної мережі дані перетворюються на послідовності фіксованої довжини (тривалістю 50 днів). Кожна послідовність використовується для прогнозування наступного дня.

Нарешті дані розбиваються на навчальні (80%) і тестові (20%) набори для оцінки продуктивності моделі.

Розклад часового ряду на набір компонент з різними частотами та часовими масштабами досягається шляхом застосування вейвлет-функцій до часового ряду (рис. 2), що дозволяє виявити як локальні, так і глобальні патерни у даних. Процес декомпозиції включає в себе кілька рівнів, де кожен рівень відповідає за різні частотні компоненти. Причому високочастотні компоненти відповідають за деталі та короткотривалі коливання у часовому ряді, а низькочастотні компоненти відповідають за основні тренди та довготривалі залежності.

Високочастотні компоненти часто містять шуми, які можуть спотворювати часовий ряд і ускладнювати прогнозування. Видалення шумів здійснюється шляхом обнулення або зменшення високочастотних компонент, які не несуть корисної інформації. Це дозволяє зберегти основні патерни та тренди часового ряду знижуючи вплив випадкових коливань.

У даному дослідженні використовується вейвлет Добеші для декомпозиції часових рядів цін акцій. Після декомпозиції високочастотні компоненти, які переважно містять шум, обнуляються, а залишок часового ряду реконструюється. Цей очищений часовий ряд використовується для подальшого аналізу та прогнозування.

© Радченко К.О., Романенко М.В.

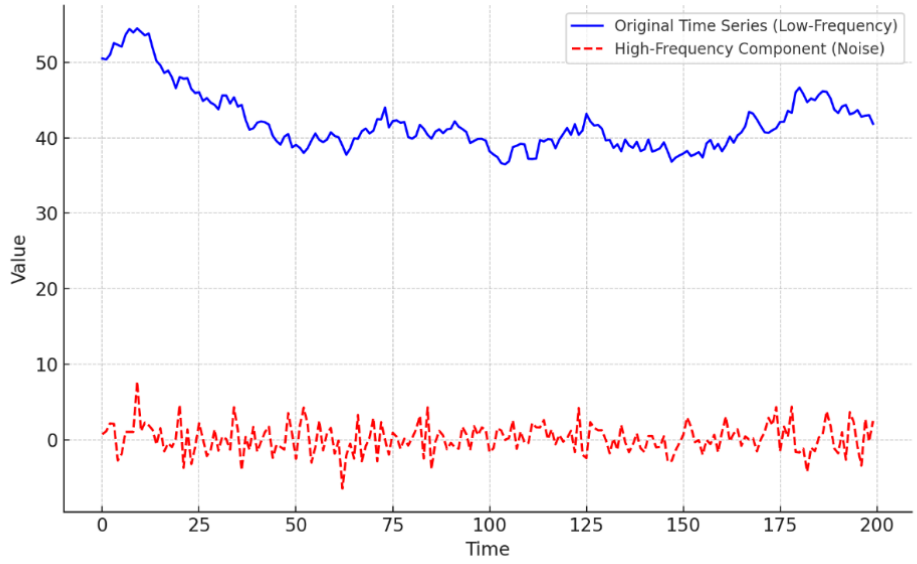


Рис. 2. Вейвлет-перетворення часових рядів цін акцій з високо- і низькочастотними компонентами

Гібридна модель CNN+LSTM у цьому дослідженні включає кілька ключових компонентів:

- вхідний шар приймає послідовності нормалізованих цін закриття акцій довжиною `sequence_length`.
- перший LSTM шар містить 50 одиниць з `return_sequences=True`, що означає повернення повної послідовності вихідних значень для кожного кроку часу. Dropout: використовується для запобігання перенавчанню, з коефіцієнтом 0.2.
- другий LSTM шар містить 50 одиниць, обробляє вхідні дані з попереднього LSTM шару. Dropout: знову застосовується для регуляризації з тим же коефіцієнтом.
- повнозв'язний шар (Dense) у останньому шарі містить один нейрон для передбачення наступної ціни акцій.
- вихідний шар виводить прогнозовану ціну акцій.

Вибору гіперпараметрів та їхньому налаштуванню приділялася теж значна увага. Було вибрано 50 одиниць для кожного LSTM шару, що забезпечує достатню ємність для запам'ятовування довготривалих залежностей. Dropout застосовувався з коефіцієнтом 0.2 для запобігання перенавчанню. Модель тренувалася протягом 20 епох, що визначено на основі експериментального досвіду та можливостей обчислювальних ресурсів. Було вибрано розмір батчу 32 для навчання, що забезпечує баланс між швидкістю обробки та стабільністю оновлення ваг. В якості оптимізатора використовувався `adam` для швидкої та ефективної оптимізації.

Модель навчалася на навчальних даних (`X_train`, `y_train`) та оцінювалася на тестових даних (`X_test`, `y_test`). Прогнозовані ціни акцій денормалізувалися для порівняння з фактичними значеннями, що дозволило оцінити точність моделі за допомогою метрики середньої абсолютної помилки (MAE), яка вимірює середню величину різниці між фактичними значеннями і прогнозованими значеннями моделі:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left| y_{act}^{(i)} - y_{pred}^{(i)}\right|,$$

де n – кількість прикладів у тестовій вибірці, $y_{act}^{(i)}$ – фактичне значення для i-го прикладу, $y_{pred}^{(i)}$ – прогнозоване значення для i-го прикладу.

В реалізованій моделі MAE дорівнює 12 – тобто в середньому модель помиляється на 12 доларів для ціни акцій.

Також прогноз оцінювався через середньо-відсоткову абсолютну похибку (MAPE), яка показує середній відсоток, на який прогнозовані значення відрізняються від фактичних:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_{act}^{(i)} - y_{pred}^{(i)}\right|}{y_{act}^{(i)}}100\%.$$

Середній відсоток помилки прогнозу моделі при перевірці через MAPE становить 8 – 10 %.

© Радченко К.О., Романенко М.В.

Процес навчання моделі можна описати схематично наступним чином:

- завантаження даних цін закриття акцій кількох компаній за період з 2010 по 2023 рік.
- нормалізація даних до діапазону [0; 1] за допомогою Min-Max нормалізації.
- 80% даних були виділені для навчання, а 20% – для тестування. Це розподілення дозволяє оцінити модель на невідомих їй даних для перевірки узагальнювальної здатності.
- дані були перетворені на послідовності фіксованої довжини (50 днів) для підготовки вхідних даних для моделі. Кожна послідовність (`X`) відповідає прогнозу (`y`) наступного значення.
- у якості архітектури моделі використовувалася гібридна модель CNN+LSTM, яка включала LSTM шари для обробки послідовних даних і повнозв'язний шар для остаточного прогнозу.
- модель навчалася на навчальних даних (`X_train`, `y_train`) протягом 20 епох з використанням розміру батчу 32. Оптимізатор `adam` використовувався для мінімізації функції втрат `mean_squared_error`.
- використовувалася регуляризація для запобігання перенавчанню – застосовувався Dropout з коефіцієнтом 0,2.
- після навчання модель була перевірена на тестових даних (`X_test`, `y_test`) для оцінки її точності (рис. 3). Прогнозовані значення були денормалізовані та порівняні з фактичними цінами для оцінки якості моделі.

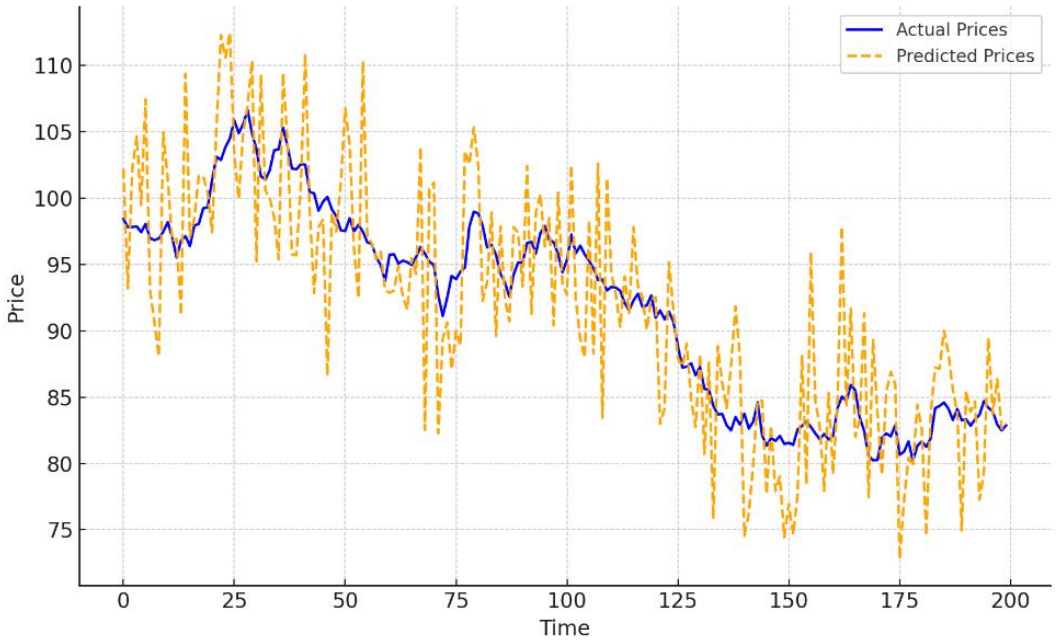


Рис. 3. Графік порівняння прогнозу та реальних даних

Таким чином у дослідженні було запропоновано модифікований метод прогнозування цін акцій, який поєднує вейвлет-перетворення з гібридною моделлю CNN+LSTM. Це дозволяє зменшити вплив шумів, враховувати складні взаємозв'язки у даних і, як наслідок, покращити точність прогнозування.

Отримані результати можуть бути використані фінансовими аналітиками та інвесторами для підвищення точності своїх прогнозів, що допоможе приймати обґрунтовані інвестиційні рішення і зменшити ризики.

Результати дослідження відкривають нові перспективи для подальших досліджень у сфері застосування гібридних моделей нейронних мереж у фінансовому прогнозуванні. Зокрема, можливе дослідження інших комбінацій нейронних мереж та методів попередньої обробки даних.

Висновки. Дослідження зробило значний внесок у наукову область прогнозування цін акцій, запропонувавши ефективний метод, який поєднує вейвлет-перетворення та гібридну модель CNN+LSTM. Це не лише значно підвищує точність прогнозування, але й відкриває нові можливості

© Радченко К.О., Романенко М.В.

для подальших досліджень у цій галузі. Виділення локальних патернів за допомогою CNN та врахування довготривалих залежностей через LSTM забезпечило комплексний аналіз часових рядів у фінансовому прогнозуванні.

Застосування вейвлет-перетворення, зокрема вейвлетів Добеші, дозволило ефективно виконати декомпозицію часових рядів, видаливши шуми та зберігши важливі тренди. Це сприяло поліпшенню якості вхідних даних для моделі та, як наслідок, підвищенню точності прогнозів.

Дослідження підтвердило, що гібридні моделі, які поєднують різні типи нейронних мереж, є ефективними для складних завдань прогнозування, таких як фінансові часові ряди. Комбінація CNN та LSTM є перспективною і дозволяє враховувати як короткотривалі, так і довготривалі патерни у даних.

Список бібліографічного опису

1. Білоус, М. Г. (2022). Вейвлет-перетворення як інструмент фінансового аналізу: переваги та виклики // Фінанси України. – 2022. – №1. – С. 22-30.
2. Григор'єв, І. А., Сорокін, В. О. (2020). Гібридні моделі для прогнозування цін акцій на фондовому ринку // Вісник Харківського національного університету. – 2020. – №2. – С. 45-53.
3. Карпова, Л. В. (2018). Нейронні мережі у фінансовому аналізі: теорія та практика застосування // Вісник Київського національного університету. Економіка. – 2018. – №4. – С. 90-98.
4. Козаченко, О. В., Романюк, О. А. (2020). Прогнозування цін акцій на основі моделей машинного навчання // Економіка та прогнозування. – 2020. – №2. – С. 75-88.
5. Кравченко, Д. В. (2021). Статистичні методи та машинне навчання у фінансових дослідженнях // Фінансовий ринок України. – 2021. – №3. – С. 67-73.
6. Литвин, О. І., Синиця, А. В. (2018). Впровадження нейронних мереж у фінансове прогнозування: досвід та перспективи // Економіка та управління підприємством. – 2018. – №5. – С. 88-96.
7. Радченко К.О. (2021). Застосування дискретних вейвлет-перетворень для прогнозування рівня навантаження на вебсервер комп'ютерних мереж загального призначення // Комп'ютерно-інтегровані технології: освіта, наука, виробництво. – 2021. – № 45. – С. 90 – 96. https://doi.org/10.3691 0/6775-2524-0560- 2021-45-13
8. Радченко К.О. (2020). Концептуальна модель забезпечення ефективності прогнозування навантаження на вебсервер // Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки. 2020. – Т. 31 (70). – № 6. – С. 135 – 141. https://doi.org/10.3283 8/TNU-2663- 5941/2020.6-1/23
9. Радченко К.О. (2022). Особливості прогнозування рівня вебтрафіку у комп'ютерних мережах загального призначення // Проблеми інформатизації та управління. – 2022. – № 3(71). – С. 41 – 50. https://doi.org/10.1837 2/2073-4751.71.17002
10. Федоренко, А. С., Литвиненко, В. П. (2021). Використання глибокого навчання для прогнозування ринку акцій // Інформаційні технології в управлінні. – 2021. – Т. 3. – С. 56-64.
11. Шелест, О. В. (2019). Прогнозування фінансових часових рядів за допомогою рекурентних нейронних мереж // Сучасні інформаційні технології та інноваційні методи в управлінні. – 2019. – №2. – С. 34-42.
12. Althelaya, K. A., Mohammed, S. A., & El-Alfy, E. S. M. (2021). Combining deep learning and multiresolution analysis for stock market forecasting. *IEEE Access, 9*, 13099-13111. Article 9324831. https://doi.org/10.1109/ACCESS.2021.3051872
13. Rhif, M., Abbes, A. B., Farah, I. R., Martínez, B., & Sang, Y. (2019). Wavelet transform application for/in non-stationary time-series analysis: A review. *Applied Sciences, 9*(7), 1345. https://doi.org/10.3390/app9071345
14. Verma, S., Sahu, S. P., & Sahu, T. P. (2023). Discrete wavelet transform-based feature engineering for stock market prediction. *International Journal of Information Technology, 15*(6), 1-10. https://doi.org/10.1007/s41870-023-01157-2
15. Yahoo Finance. Режим доступу https://finance.yahoo.com/watchlists/

References

1. Bilous, M. G. (2022). Wavelet Transform as a Tool for Financial Analysis: Advantages and Challenges // Finance of Ukraine. – 2022. – No.1. – P. 22-30.
2. Hryhoriev, I. A., Sorokin, V. O. (2020). Hybrid Models for Stock Price Prediction in the Stock Market // Bulletin of Kharkiv National University. – 2020. – No.2. – P. 45-53.
3. Karpova, L. V. (2018). Neural Networks in Financial Analysis: Theory and Practice // Bulletin of Kyiv National University. Economics. – 2018. – No.4. – P. 90-98.
4. Kozachenko, O. V., Romanyuk, O. A. (2020). Stock Price Prediction Based on Machine Learning Models // Economy and Forecasting. – 2020. – No.2. – P. 75-88.
5. Kravchenko, D. V. (2021). Statistical Methods and Machine Learning in Financial Research // Financial Market of Ukraine. – 2021. – No.3. – P. 67-73.
6. Lytvyn, O. I., Synytsia, A. V. (2018). Implementation of Neural Networks in Financial Forecasting: Experience and Prospects // Economics and Enterprise Management. – 2018. – No.5. – P. 88-96.
7. Radchenko K.O. (2021). Application of discrete wavelet transformations for forecasting the level of load on the web server of general purpose computer networks // Computer-integrated technologies: education, science, production. – 2021. – No. 45. – P. 90-96. https://doi.org/10.3691 0/6775-2524-0560- 2021-45-13
8. Radchenko K.O. (2020). Conceptual model for ensuring the effectiveness of web server load forecasting // Scientific notes of the Tavri National University named after V.I. Vernadsky. Series: Technical sciences. – 2020. – Vol. 31 (70). – No. 6. P. 135-141. https://doi.org/10.3283 8/TNU-2663- 5941/2020.6-1/23

© Радченко К.О., Романенко М.В.

9. Radchenko K.O. (2022). Peculiarities of predicting the level of web traffic in general purpose computer networks // Problems of informatization and management. – 2022. – No. 3(71). – P. 41-50. https://doi.org/10.1837 2/2073-4751.71.17002.
10. Fedorenko, A. S., Lytvynenko, V. P. (2021). Deep Learning for Stock Market Prediction // Information Technologies in Management. – 2021. – Vol. 3. – P. 56-64.
11. Shelest, O. V. (2019). Forecasting Financial Time Series Using Recurrent Neural Networks // Modern Information Technologies and Innovative Methods in Management. – 2019. – No.2. – P. 34-42.
12. Althelaya, K. A., Mohammed, S. A., & El-Alfy, E. S. M. (2021). Combining deep learning and multiresolution analysis for stock market forecasting. *IEEE Access*, *9*, 13099-13111. Article 9324831. https://doi.org/10.1109/ACCESS.2021.3051872
13. Rhif, M., Abbes, A. B., Farah, I. R., Martínez, B., & Sang, Y. (2019). Wavelet transform application for/in non-stationary time-series analysis: A review. *Applied Sciences*, *9*(7), 1345. https://doi.org/10.3390/app9071345
14. Verma, S., Sahu, S. P., & Sahu, T. P. (2023). Discrete wavelet transform-based feature engineering for stock market prediction. *International Journal of Information Technology*, *15*(6), 1-10. https://doi.org/10.1007/s41870-023-01157-2
15. Yahoo Finance. Access mode https://finance.yahoo.com/watchlists/

© Радченко К.О., Романенко М.В.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-34
УДК 330.004
Регеда Юрій Олександрович[1], магістр
https://orcid.org/0009-0009-4375-4738
Регеда Валерій Олександрович[2], магістр
https://orcid.org/0009-0007-9469-6545
[1] Національний технічний університет «Харківський політехнічний інститут», Україна
[2] Харківський національний економічний університет імені Семена Кузнеця, Україна

ВПЛИВ ЦИФРОВОЇ ЕВОЛЮЦІЇ НА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

Регеда Ю.О., Регеда В.О. Вплив цифрової еволюції на інформаційні технології. У статті досліджено вплив цифрової еволюції на інформаційні технології. Наголошено, що процеси глобалізації, які стали реальністю в сучасному світі і, як наслідок, формування постіндустріальної економіки були зумовлені переходом від третьої до четвертої інформаційної революції, що триває нині, прийнятої називати революцією інформаційних технологій, від автоматизованої обробки інформації до її комп'ютерного подання в результаті застосування мікропроцесорної техніки, комплексу технічних та програмних засобів, використаних для побудови різних мікропроцесорних пристроїв та систем, що стали основою для створення персональних комп'ютерів та інформаційно-комунікаційної мережі та всесвітньої системи об'єднаних комп'ютерних мереж Інтернету для забезпечення інформаційного обміну в глобальному масштабі. Підкреслено, що цифрова революція викликає трансформацію культури людини завдяки застосуванню нових технологічних можливостей, що дозволяють створювати віртуальні світи та удосконалювати технології доповненої реальності для доповнення навколишнього світу цифровими даними в режимі реального часу за допомогою спеціальних пристроїв. Зазначається, що інформаційно-цифрова революція впливає на структурні зміни економіки. На сьогодні всі види діяльності людини піддаються впливу цифрової революції, на очах відбувається розвиток комплексу інформаційно-комунікаційних технологій, стрімке поширення яких сприяє активному науково-технічному прогресу в мікроелектроніці. Відповідно до закону Мура у мікроелектроніці відбувається скорочення вартості одиниці обчислювальної потужності та операції. Наголошується, що нині суспільство перебуває у стані перехідного періоду до нового технологічного укладу, який входить у фазу свого зростання. Ядром сучасного технологічного укладу виступають ІКТ, адитивні технології, біоінженерія та нанотехнології. Ключовими факторами розвитку нового технологічного укладу є інформаційні технології, в основі яких лежать знання щодо елементарних структур матерії, удосконалення алгоритмів обробки та передачі даних у фундаментальних дослідженнях.
Ключові слова: цифрові технології, еволюція, генезис, інформаційні технології, трансформація, вдосконалення

Iurii Regeda, Valerii Reheda. The impact of digital evolution on information technologies. The article examines the impact of digital evolution on information technologies. It is emphasized that the processes of globalization, which have become a reality in the modern world and, as a result, the formation of a post-industrial economy, were determined by the transition from the third to the fourth information revolution, which is currently underway, called the revolution of information technologies, from automated processing of information to its computer presentation as a result of the application of microprocessor technology, a complex of technical and software tools used to build various microprocessor devices and systems, which became the basis for the creation of personal computers and the information and communication network and the worldwide system of unified computer networks of the Internet to ensure information exchange on a global scale. It is emphasized that the digital revolution causes the transformation of human culture due to the application of new technological capabilities that allow creating virtual worlds and improving augmented reality technologies to supplement the surrounding world with digital data in real time using special devices. It is noted that the information and digital revolution affects structural changes in the economy. Today, all types of human activity are affected by the digital revolution, the development of a complex of information and communication technologies is taking place before our eyes, the rapid spread of which contributes to active scientific and technical progress in microelectronics. According to Moore's law, the cost of a unit of computing power and operation is decreasing in microelectronics. It is emphasized that society is currently in a state of transition to a new technological order, which is entering its growth phase. ICT, additive technologies, bioengineering and nanotechnologies are the core of the modern technological system. The key factors in the development of a new technological system are information technologies, which are based on knowledge about the elementary structures of matter, improvement of processing algorithms and data transmission in fundamental research.
Key words: digital technologies, evolution, genesis, information technologies, transformation, improvement

Вступ та постановка проблеми дослідження. Цифрова еволюція означає поступове вдосконалення та трансформацію цифрової технології та системи, завдяки прогресу в обчислювальній потужності, обробці даних, мережі та штучному інтелекту. Цей еволюційний процес глибоко впливає на різні аспекти: комп'ютерну науку та інформаційні технології, переформування галузі, введення інновацій та трансформацію суспільних взаємодій. У дискусіях про цифровий прогрес ця тенденція підкреслює прогрес технології та впровадження нових технологій таких, як великі дані та підключені пристрої. По суті, впроваджуються інновації та здійснюється модернізація з метою підвищення ефективності.

Метою цифрової еволюції є покращення поточних операцій шляхом використання цифрових технологій. Наголошуючи на кінцевих цілях і причинах, що стоять за змінами, модель ще раз підтверджує початковий намір цифрових технологій перетворення. У результаті підхід стає

© Регеда Ю.О., Регеда В.О.

більш поетапним і поступовим, виникає мета впровадження змін, які можна легко включити в поточний стан організації. Ці зміни спрямовані на відчутне збільшення загальної вартості бізнесу, продуктивності та ефективності.

Аналіз останніх досліджень і публікацій. Формулювання наукової думки в окресі цифрової еволюції є різнорідним та масштабним. У сучасній науковій площині з'являються роботи присвячені дослідженням сучасних методів впливу цифрової еволюції на інформаційні технології сьогодення. У [1] В. Семанюк та Н. Мельник вивчили вплив технологій на інформаційне середовище бізнесу, та розкрили розуміння того, яким чином інформаційні технології змінять роботу бухгалтерів. Змоделювали інформаційне середовище бізнесу на різних рівнях господарювання за умови впровадження новітніх інформаційних технологій та принципів діджиталізації. Довели необхідність зміни методології обліку в умовах п'ятої промислової революції та зміну основних принципів облікової теорії. У роботі [2] виділені та охарактеризовані два основні підходи до цифрових трансформацій, а саме еволюційний та революційний. Вивчення особливостей еволюції та особливостей якіснокількісних змін у системі соціально-економічних відносин дало змогу виділити чотири основні етапи цифрових трансформацій у світі. На засадах дослідження особливостей еволюції цифрових трансформацій в статті автором виділені головні чинники розвитку цифрової економіки. А саме, цифрові фінанси, соціальні мережі, цифрова ідентифікація і кібербезпека, великі обсяги даних та мобільність до їх доступу, інтеграція декількох систем управління та організації економічної діяльності, можливість колаборації ведення офлайн та онлайн бізнесу. У [3] визначено, що на формування інформаційних прав впливає низка факторів, визначальним серед яких є інформатизація суспільного життя, що призводить до формування глобального інформаційного суспільства. Досліджено позитивні та негативні моменти впровадження цифрових трансформацій у повсякденне життя. Із зарубіжних авторів варто відмітити роботи таких науковців як: Узвишин Раймонд [4], Чернічко Тетяна, Козик Ірина [5], Гу Веньцзюнь, Ван Цзіньхуа, Чжан Юньці, Лян Шаолінь, Ай Цзишен, Лі Цзію [6], Ніса Хойрун, Мубарок Хусні [7], Жую Марія, Рагнедда [8], Лаворато Доменіка, П'єдепалумбо Пальміра, Бонсон Енріке, Беднарова Мікаела [9], Вінн Мартін, Ільяс Джанат, Іслейєн Омер, Брюнтруп Хенрік, Метін Білгін [10], Харалайя д-р, Махешварі д-р, Кумар Хумнекар д-р, Прабхакар Доктор Н. Манодж [11], Бансал Дів'я, Бхаттачарья Набошрі, Шанділья Рріянка [12] та інших.

Однак незважаючи на масштабність наукових досліджень питання дослідження впливу цифрової еволюції на інформаційні технології є актуальним та потребує подальшого дослідження.

Метою роботи є дослідження впливу цифрової еволюції на інформаційні технології.

Виклад основного матеріалу дослідження. Численні переваги, які надає цифрова революція, можуть повністю змінити те, як компанії працюють, взаємодіють із клієнтами та досягають успіху в довгостроковій перспективі. Якщо порівняти стратегію цифрової трансформації «все або нічого» з поступовим підходом цифрової еволюції, цей підхід пропонує більш практичний, методичний підхід до підвищення продуктивності та створення цінності. Поступові вдосконалення стали можливими завдяки цифровій еволюції, що дозволяє організаціям тестувати нові технології, процедури та методології в меншому масштабі перед їх повним впровадженням. Цей метод внесення невеликих поступових змін зменшує ризик і полегшує вимірювання їх ефекту. Завдяки цифровій еволюції, що дозволяє компаніям швидко пристосовуватися до змін на ринку, вимог клієнтів або тиску конкурентів без необхідності кожного разу проходити повне оновлення, цифрова еволюція сприяє гнучкості. Починати з нуля може бути дорого, як у фінансовому плані, так і з точки зору використаних ресурсів. Цифрова еволюція, з іншого боку, зосереджена на оптимізації поточних процедур і систем для різкого зниження витрат на впровадження. Крім того, це дозволяє підприємствам отримати максимальну віддачу від своїх поточних інвестицій. Співробітники можуть уникнути необхідності адаптуватися до абсолютно нових платформ, що часто призводить до опору або зниження продуктивності, шляхом вдосконалення поточних систем і процесів. Крім того, вони можуть продовжувати використовувати ті самі інструменти та отримувати переваги від невеликих, але постійних покращень, які сприятимуть їхньому залученню та задоволенню роботою. Здатність миттєво пристосовуватися до споживчих уподобань і тенденцій ринку стала можливою завдяки еволюції цифрових технологій. Це дозволяє компаніям постійно впроваджувати інновації, орієнтовані на потреби клієнтів, покращуючи взаємодію з користувачами та зміцнюючи лояльність клієнтів. Поступове впровадження змін дає змогу здійснювати моніторинг у режимі реального часу та здійснювати швидкі коригування, що допомагає ранньому виявленню та зменшенню ризиків. Порівняно з трансформаціями типу «все або нічого», які наражають організації на значні ризики,

© Регеда Ю.О., Регеда В.О.

цей адаптивний підхід набагато кращий. Організації можуть легше інтегрувати нові технології та методи в міру їх появи, використовуючи еволюційний підхід до цифрових технологій, що усуває потребу в черговій повній переробці. Це полегшує збереження переваги над конкурентами та адаптацію до мінливих вимог ринку.

Цифрова еволюція охоплює кілька ключових компонентів, які спільно рухають технологічний прогрес:

Розвиток обчислювальної потужності. У листопаді 1971 року Intel 4004 став першим у світі однокристальним мікропроцесором з 2300 транзисторами та 60 000 командами в секунду. Відтоді спостерігається експоненціальне зростання обчислювальної потужності відповідно до закону Мура, який, стверджує, що кількість транзисторів у щільній інтегральній схемі подвоюється приблизно кожні два роки. Прогрес у галузі апаратного забезпечення уможливив і прискорив дослідження та застосування в усіх галузях науки й техніки. Сучасні чіпи мікропроцесорів використовують мільярди транзисторів, містять кілька процесорних ядер на одному кристалі, працюють на тактовій частоті гігагерц і пропонують у 4 мільйони разів більшу продуктивність, ніж оригінальний 4004, і межі обчислювальної техніки продовжують швидко розширюватися у все більш творчих додатках, особливо в медицині, мобільних і хмарних технологіях. Але всьому має прийти кінець. У березні 2015 року Гордон Мур передбачив, що темпи розвитку апаратного забезпечення досягнуть насичення. В очікуванні такої можливості багато інженерів-електриків почали досліджувати нові типи конструкцій і матеріалів для апаратного забезпечення комп'ютерів. Незважаючи на те, що ці технології є перспективними, вони все ще перебувають у зародковому стані. Наприклад, квантова електроніка, яка вивчає поведінку електронів у речовині та їхню взаємодію з фотонами, стикається з серйозними проблемами, коли обчислювальні та архітектурні моделі апаратного забезпечення ще недоступні. В останні десятиліття зріс інтерес до GPU обчислень, які використовують вбудовані GPU для обробки паралельних обчислень. Можливо, є можливість повторити закон Мура, сказавши, що кількість графічних процесорів подвоюється кожні два роки. Незалежно від того, чи Закон Мура продовжує діяти в його поточному стані чи розвиватиметься далі, можна з упевненістю сказати, що вдосконалення ефективності обладнання зрештою припиниться. Проте кількість нових обчислювальних методів і застосувань зростає в геометричній прогресії – найбільші досягнення в обчислювальній техніці ще попереду.

Спочатку сформульований Гордоном Муром, Закон Мура стверджує, що кількість транзисторів на мікрочіпі подвоюється приблизно кожні два роки, що призводить до експоненціального зростання обчислювальної потужності. Незважаючи на те, що останніми роками закон Мура сповільнився, він історично сприяв значному прогресу в обчислювальній потужності, дозволяючи розробляти більш потужні та компактні пристрої. У 1965 році Гордон Мур припустив, що приблизно кожні два роки кількість транзисторів на мікросхемах буде подвоюватися. Це явище, яке зазвичай називають законом Мура, свідчить про те, що з часом обчислювальний процес стане значно швидшим, меншим і ефективнішим. Закон Мура, який широко вважається однією з головних теорій 21 століття, має значні наслідки для майбутнього технологічного прогресу разом із можливими обмеженнями. Закон Мура безпосередньо вплинув на прогрес обчислювальної потужності, поставивши перед виробниками чіпів мету. У 1965 році Мур передбачив, що до 1975 року на мікросхемі буде 65 000 транзисторів. У 2025 році виробники мікросхем зможуть розмістити 50 мільярдів транзисторів на мікросхемі розміром з ніготь.

Квантові обчислення.

Одна з галузей інформатики, яка застосовує ідеї квантової теорії, називається квантовим обчисленням. Поведінка матерії та енергії на атомному та субатомному рівнях пояснюється квантовою теорією. Субатомні частинки, такі як електрони та фотони, використовуються в квантових обчисленнях. Ці частинки можуть існувати в більш ніж одному стані (тобто 1 і 0) одночасно завдяки квантовим бітам або кубітам. Теоретично пов'язані кубіти мають здатність «виконувати обчислення, які інакше могли б зайняти мільйони років, використовуючи інтерференцію між їхніми хвилеподібними квантовими станами.

Сучасні класичні комп'ютери кодують інформацію в бітах, використовуючи двійковий потік електричних імпульсів (1 і 0). На відміну від квантових обчислень, це обмежує їхню обробну здатність. Квантові обчислення представляють собою зміну парадигми обчислювальної потужності. Використовуючи принципи квантової механіки, квантові комп'ютери використовують кубіти, які можуть існувати в кількох станах одночасно, пропонуючи потенціал для вирішення складних проблем набагато ефективніше, ніж класичні комп'ютери. Недавні прориви таких компаній, як IBM

© Регеда Ю.О., Регеда В.О.

і Google, підкреслюють потенціал квантових обчислень для революції в таких сферах, як криптографія, матеріалознавство та відкриття ліків.

Нейроморфні обчислення

Нейроморфне обчислення прагне імітувати ефективність мозку в обробці інформації. Цей підхід використовує спеціалізоване обладнання, призначене для імітації нейронної та синаптичної активності, що забезпечує більш ефективні та потужні обчислення для завдань, пов'язаних із розпізнаванням образів та адаптивним навчанням. Важливі обчислювальні завдання, такі як сприйняття та прийняття рішень, отримають значну користь від застосування нейроморфних обчислень. Реалізації нейронних мереж у програмному забезпеченні та спеціалізованому апаратному забезпеченні досягли великих успіхів, але вони все ще на порядки менш енергоефективні, ніж людський мозок.

Нейроморфні обчислення дають змогу штучному інтелекту (ШІ) навчатися та приймати рішення більш незалежно, ніж це було б з ШІ першого покоління. Нейроморфні системи в даний час поглинені глибоким навчанням відчуття і сприйняття здібностей необхідні для складних стратегічних ігор, таких як шахи, а також для розпізнавання мовлення. Замість того, щоб просто працювати на основі шаблонних алгоритмів, ШІ наступного покоління зможе інтерпретувати ситуації та адаптуватися до них, як це робить людський мозок.

Остання фаза технології штучного інтелекту спрямована на повторення складної системи нервових клітин людського мозку. Щоб досягти цього, ШІ повинен обробляти та досліджувати дані, які не впорядковані, відповідно до ефективності біологічного мозку. Людський мозок може споживати лише 20 Вт енергії та працювати краще, ніж найпотужніші комп'ютери, демонструючи свою надзвичайну ефективність. ШІ-версія мережі зв'язків мозку відома як імпульсні нейронні мережі (SNN). Ці штучні нейрони організовані в шари, і кожен імпульсний нейрон може працювати сам по собі та взаємодіяти з іншими, ініціюючи низку змін у реакції на сигнали.

Більшість нейронних мереж штучного інтелекту структуровано навколо архітектури фон Неймана, яка включає окремі блоки пам'яті та обробки. В даний час комп'ютери обмінюються інформацією, витягуючи дані з пам'яті, передаючи їх до блоку обробки, обробляючи дані, а потім повертаючи їх у пам'ять. Цей процес обміну займає багато часу та енергії, що призводить до ефекту вузького місця, який стає більш вираженим при роботі з великими обсягами даних.

Розробка нанорозмірних мемристивних пристроїв, також відомих як мемристори, викликала аналогову революцію. Ці пристрої відкривають двері для створення нейроморфних пристроїв, здатних виконувати складні обчислення в масштабі та на місці. Мемристори діють як перемикачі пам'яті, зберігаючи інформацію про їхній опір чи провідність і регулюючи свою провідність на основі історії програмування, навіть якщо живлення втрачено. Ця здатність імітувати синапси людини є їхньою ключовою функцією.

Алгоритми та штучний інтелект. Машинне та глибоке навчання

У популярній культурі термін «штучний інтелект» часто використовується для позначення будь-якого виду розумної машини. Глибоке навчання, машинне навчання та штучний інтелект насправді є різними поняттями з дещо різним значенням.

Машинне навчання належить до сфери штучного інтелекту. Глибоке навчання, з іншого боку, є спеціалізованою сферою машинного навчання. По суті, кожен екземпляр глибокого навчання також є формою машинного навчання, і всі форми машинного навчання класифікуються як штучний інтелект. Однак важливо зазначити, що не всі підходи до штучного інтелекту включають методи машинного навчання.

Ці підмножини штучного інтелекту зосереджені на створенні систем, які вивчають дані, щоб робити прогнози чи приймати рішення. Машинне навчання охоплює різноманітні алгоритми, тоді як глибоке навчання спеціально використовує нейронні мережі з кількома рівнями для досягнення високої точності в таких завданнях, як розпізнавання зображень і мови, обробка природної мови та прогнозна аналітика.

Глибоке навчання, підгрупа машинного навчання, використовує штучні нейронні мережі для обробки та аналізу даних. Ці мережі складаються з обчислювальних вузлів, упорядкованих шарами в рамках алгоритмів глибокого навчання, причому кожен рівень містить вхідний рівень, вихідний рівень і приховані шари. Навчальні дані подаються в нейронну мережу, полегшуючи навчання алгоритму та покращуючи точність. Коли нейронна мережа складається з трьох або більше рівнів, вона кваліфікується як «глибока», звідси термін глибоке навчання.

© Регеда Ю.О., Регеда В.О.

Машинне навчання, підмножина штучного інтелекту, дозволяє системам автономно навчатися та вдосконалюватися без явного програмування. Він покладається на розпізнавання шаблонів у даних, щоб робити прогнози, коли нові дані вводяться в систему.

Еволюційні алгоритми (ЕА)

Еволюційні алгоритми (ЕА) позначають категорію методів стохастичної оптимізації, які імітують природний процес еволюції. Їх зародження датується кінцем 1950-х років, а з 1970-х років з'явилися різні еволюційні методології, включаючи генетичні алгоритми, еволюційне програмування та стратегії еволюції. Усі ці методи працюють на наборі потенційних рішень, які зазнають модифікацій на основі фундаментальних принципів еволюції: відбору та варіації.

Відбір відображає конкуренцію за ресурси, яка спостерігається в природі, де особини з вигідними рисами мають вищі шанси на виживання та розмноження. У ЕА цей природний процес відбору відтворюється через механізми стохастичного відбору. Кожне рішення в межах набору оцінюється, і йому присвоюються скалярні значення відповідності, що визначає ймовірність їх відтворення. Принцип варіації, з іншого боку, імітує здатність природи створювати нові особини шляхом рекомбінації та мутації.

Незважаючи на свої прості принципи, еволюційні алгоритми продемонстрували надзвичайну ефективність. Крім того, вони відмінно справляються з багатоцільовими завданнями оптимізації, оскільки можуть ідентифікувати кілька оптимальних рішень в рамках одного прогону моделювання та використовувати подібність рішень шляхом рекомбінації.

Великі дані та аналітика даних. Генерація та зберігання даних є фундаментальними аспектами управління та ефективного використання інформації в різних сферах, включаючи бізнес, дослідження та технології. У цифрову епоху стався вибух у створенні даних із таких джерел, як соціальні мережі, пристрої Інтернету речей (IoT) і наукові дослідження. Управління цією величезною кількістю даних вимагає вдосконалених рішень для зберігання даних, включаючи розподілені бази даних, хмарне сховище та твердотільні накопичувачі (SSD).Генерація даних стосується процесу створення, збору або отримання даних для аналізу, прийняття рішень та інших цілей. Вона включає різні методи та джерела, зокрема:

Датчики та пристрої: багато пристроїв, як-от датчики Інтернету речей, переносні пристрої та системи моніторингу, постійно генерують дані, збираючи інформацію з навколишнього середовища або взаємодії користувачів. Цифрові платформи: онлайн-платформи, соціальні медіа, веб-сайти електронної комерції та програми генерують величезні обсяги даних завдяки взаємодії користувачів, транзакціям, клікам і пошукам. Симуляції та моделі: дані можна генерувати за допомогою моделювання, експериментів або обчислювальних моделей для вивчення складних систем, прогнозування результатів або перевірки гіпотез у контрольованому середовищі.

Зовнішні джерела: організації можуть отримувати дані зі сторонніх джерел, таких як публічні бази даних, постачальники даних і API, щоб збагатити свої набори даних або доповнити внутрішні дані. Зберігання даних включає зберігання та організацію даних у структурований спосіб для забезпечення доступності, надійності та масштабованості.

У дослідженнях процедури обробки та аналізу даних охоплюють збір, упорядкування та оцінку даних, щоб отримати значущі ідеї та зробити висновки. Дослідники використовують різноманітні інструменти та методи для обробки та аналізу даних, розроблені для вирішення конкретних дослідницьких питань і цілей. Вкрай важливо, щоб дані, зібрані в результаті експериментів, записів і опитувань, були точними та повними перед подальшою обробкою. Методи статистичного машинного навчання можна застосовувати для аналізу багатовимірних даних, ідентифікації відповідних шаблонів або каналів у детекторі для обробки даних аналізу невідомих зразків.

Інтернет речей (IoT)

Інтернет речей (IoT) означає мережу з мільярдів фізичних пристроїв по всьому світу, які підключені до Інтернету, постійно збираючи та обмінюючись даними.

Прорив стався завдяки розробці недорогих та енергоефективних процесорів, які зробили економічно ефективним підключення мільярдів пристроїв. Впровадження тегів RFID – малопотужних чіпів, здатних до бездротового зв'язку – частково вирішило проблему. Крім того, широка доступність широкосмугового Інтернету, стільникового зв'язку та бездротових мереж ще більше полегшила це підключення. Прийняття IPv6, який надає величезну кількість IP-адрес, також мало вирішальне значення для масштабування IoT, гарантуючи, що кожен пристрій може мати унікальний ідентифікатор.

© Регеда Ю.О., Регеда В.О.

Інтернет речей (IoT) генерує величезну кількість даних із різних джерел, включаючи датчики, підключені до обладнання, монітори навколишнього середовища та навіть команди, які надходять розумним колонкам. Це величезне покоління даних позиціонує IoT як важливий двигун аналітичних проектів великих даних, що дозволяє компаніям створювати та аналізувати великі набори даних.

Зокрема, IoT налаштований на передачу значних обсягів даних у режимі реального часу. За оцінками Cisco, до 2026 року міжмашинні з'єднання, що підтримують додатки IoT, становитимуть більше половини із 27,1 мільярда пристроїв і з'єднань і становитимуть 5% глобального IP-трафіку. Пристрої IoT використовують різні методи підключення та обміну даними, переважно покладаються на бездротове підключення.

Одним із значних напрямків зростання в найближчі роки стане використання мереж 5G для підтримки проектів IoT. Технології 5G можуть обслуговувати до одного мільйона пристроїв на квадратний кілометр, дозволяючи розгортати численні датчики на обмеженій території та полегшуючи масштабні промислові додатки IoT. Великобританія розпочала випробування 5G та IoT на двох «розумних фабриках». Однак широке впровадження 5G може зайняти час.

З розвитком IoT ймовірною тенденцією буде зменшення обсягу даних, які надсилаються в хмару для обробки. Щоб зменшити витрати, більше даних можна обробляти на самому пристрої, повертаючи лише важливу інформацію назад у хмару – стратегія, відома як «граничні обчислення». Цей підхід вимагатиме нових технологій, таких як захищені від несанкціонованого доступу периферійні сервери, здатні збирати та аналізувати дані віддалено з хмари або корпоративних центрів обробки даних.

Оскільки вартість датчиків і комунікаційних технологій продовжує знижуватися, інтеграція більшої кількості пристроїв в IoT стає все більш рентабельною, навіть якщо безпосередні переваги для споживачів неочевидні. Наразі розгортання IoT перебуває на ранніх стадіях, і більшість компаній все ще перебувають на пробній стадії. Це значною мірою тому, що основні технології – сенсорні технології, 5G і аналітика на основі машинного навчання – все ще знаходяться на відносно ранніх стадіях розробки.

Існує велика кількість конкуруючих платформ і стандартів, і різноманітні постачальники, включаючи виробників пристроїв, компанії програмного забезпечення та оператори мереж, змагаються за частку ринку. Відсутність стандартів і постійні проблеми з безпекою свідчать про те, що найближчими роками є можливість побачити більш значні порушення безпеки IoT. Оскільки кількість підключених пристроїв продовжує збільшуватися, житлові та робочі середовища будуть заповнюватися розумними продуктами, за умови, що користувачі готові прийняти відповідні компроміси щодо безпеки та конфіденційності.

Висновки. У роботі досліджено вплив цифрової еволюції на інформаційні технології. Цифрова еволюція – це трансформаційна сила, яка змінює кожен аспект комп'ютерної науки та інформаційних технологій. Збільшуючи обчислювальну потужність, удосконалюючи алгоритми, покращуючи аналітику даних і сприяючи кращому підключенню та кібербезпеці, цифрова еволюція сприяє технологічним досягненням, які підвищують продуктивність, інновації та суспільний добробут. Прийняття та спрямування цифрової еволюції матиме вирішальне значення для максимізації її переваг і вирішення викликів. Розвиток обчислювальної потужності став наріжним каменем цифрової революції, що сприяє прогресу в усіх сферах комп'ютерної науки та інформаційних технологій. Від винаходу транзистора до квантових обчислень кожен стрибок у обчислювальних можливостях відкривав нові можливості та змінював спосіб життя.

Список використаних джерел

1. Семанюк В., Мельник Н. Вплив цифрових технологій на інформаційне середовище бізнесу в умовах п'ятої промислової революції. *Вісник економіки*. 2022. Вип. 3. С. 203–212. DOI: https://doi.org/10.35774/visnyk2022.03.203.
2. Голобородько А. Соціально-економічні передумови та чинники розвитку цифрових трансформацій економіки. *Вісник Хмельницького національного університету*. 2022. № 4. С. 125-130.
3. Андрущенко О. П. Вплив цифровізації на ціннісні пріоритети розвитку прав людини. *Інформація і право*. 2023. № 4. С. 106-115. URL: http://jnas.nbuv.gov.ua/article/UJRN-0001465452.
4. Uzwyshyn Raymond. Three Significant New Technology Trends in Academic Library Research Infrastructures. 2024. DOI: 10.13140/RG.2.2.33135.44968.

© Регеда Ю.О., Регеда В.О.

5. Chernychko Tetiana, Kozyk Iryna. The evolution of human capital management and HR technologies in the digital economy. *Ukrainian Journal of Applied Economics and Technology*. 2024. № 9. P. 29-33. DOI: 10.36887/2415-8453-2024-1-4.
6. Gu Wenjun, Wang Jinhua, Zhang Yunqi, Liang Shaolin, Ai Zisheng, Li Jiyu. Evolution of Digital Health and Exploration of Patented Technologies (2017-2021): Bibliometric Analysis. *Interactive journal of medical research*. 2024. № 13. e48259. DOI: 10.2196/48259.
7. Nisa Khoirun, Mubarok Husni. How the Evolution and Track of Several Digital Technologies in Science Education?. *JPPS (Jurnal Penelitian Pendidikan Sains)*. 2024. P. 113-128. DOI: 10.26740/jpps.v13n2.p113-128.
8. Ruiu Maria, Ragnedda Massimo. From Poverty to Digital Poverty. 2024. DOI: 10.1007/978-3-031-56184-9_2.
9. Lavorato Domenica, Piedepalumbo Palmira, Bonsón Enrique, Bednarova Michaela. Digital Accounting Information Systems to drive sustainability in organizations. 2024.
10. Wynn Martin, Ilyas Janat, Isleyen Omer, Brüntrup Henrik, Metin Bilgin. Digital Technologies Research and Application Reassessing Critical Success Factors for ERP Implementation in the Digital Era. *Digital Technologies Research and Applications*. 2024. № 3. P. 140–154. DOI: 10.54963/dtra.v3i2.297.
11. Haralayya Dr, Maheswari Dr, Kumar Akshay, Humnekar Dr, Prabhakar Dr, N Manoj. Impact Of AI And Blockchain Technology In The Growth Of Digital HRM Transformation As A Function Of Management. *Educational Administration Theory and Practice journal*. 2024. № 30. P. 1685-1693. DOI: 10.53555/kuey.v30i4.1730.
12. Bansal Divya, Bhattacharya Naboshree, Shandilya Priyanka. Digital entrepreneurship and its transformative impacts on business and Society. 2024. DOI: 10.4018/979-8-3693-1734-1.

References

1. Semanyuk V., Melnyk N. (2022). Vplyv tsyfrovyh tehnologij na informaciyne seredovysche biznesu v umovah p'yatoyi promyslovoyi revolyuciyi [Influence of digital technologies on the information environment of business in the conditions of the fifth industrial revolution]. *Visnyk ekonomiky – Bulletin of Economics*, Vyp. 3, S. 203–212. DOI: https://doi.org/10.35774/visnyk2022.03.203. [in Ukrainian]
2. Goloborodko A. (2022). Social'no-ekonomichni peredumovy ta chynnyky rozvytku cyfrovyh transformacij ekonomiky [Socio-economic prerequisites and factors of digital transformations of the economy]. *Visnyk Khmel'nyc'kogo nacional'nogo universytetu – Bulletin of Khmelnytsky National University*, № 4, S. 125-130. [in Ukrainian]
3. Andrushchenko O. P. (2023). Vplyv cyfrovizaciyi na cinnisni pryoritety rozvytku prav lyudyny [The impact of digitalization on the value priorities of human rights development]. *Informaciya i pravo – Information and Law*, № 4, S. 106-115. URL: http://jnas.nbuv.gov.ua/article/UJRN-0001465452. [in Ukrainian]
4. Uzwyshyn Raymond. (2024). Three Significant New Technology Trends in Academic Library Research Infrastructures. DOI: 10.13140/RG.2.2.33135.44968.
5. Chernychko Tetiana, Kozyk Iryna. (2024). The evolution of human capital management and HR technologies in the digital economy. *Ukrainian Journal of Applied Economics and Technology*, № 9, R. 29-33. DOI: 10.36887/2415-8453-2024-1-4.
6. Gu Wenjun, Wang Jinhua, Zhang Yunqi, Liang Shaolin, Ai Zisheng, Li Jiyu. (2024). Evolution of Digital Health and Exploration of Patented Technologies (2017-2021): Bibliometric Analysis. *Interactive journal of medical research*, № 13, e48259. DOI: 10.2196/48259.
7. Nisa Khoirun, Mubarok Husni. (2024). How the Evolution and Track of Several Digital Technologies in Science Education?. *JPPS (Jurnal Penelitian Pendidikan Sains)*, R. 113-128. DOI: 10.26740/jpps.v13n2.p113-128.
8. Ruiu Maria, Ragnedda Massimo. (2024). From Poverty to Digital Poverty. DOI: 10.1007/978-3-031-56184-9_2.
9. Lavorato Domenica, Piedepalumbo Palmira, Bonsón Enrique, Bednarova Michaela. (2024). Digital Accounting Information Systems to drive sustainability in organizations.
10. Wynn Martin, Ilyas Janat, Isleyen Omer, Brüntrup Henrik, Metin Bilgin. (2024). Digital Technologies Research and Application Reassessing Critical Success Factors for ERP Implementation in the Digital Era. *Digital Technologies Research and Applications*, № 3, R. 140–154. DOI: 10.54963/dtra.v3i2.297.
11. Haralayya Dr, Maheswari Dr, Kumar Akshay, Humnekar Dr, Prabhakar Dr, N Manoj. (2024). Impact Of AI And Blockchain Technology In The Growth Of Digital HRM Transformation As A Function Of Management. *Educational Administration Theory and Practice journal*, № 30, R. 1685-1693. DOI: 10.53555/kuey.v30i4.1730.
12. Bansal Divya, Bhattacharya Naboshree, Shandilya Priyanka. (2024). Digital entrepreneurship and its transformative impacts on business and Society. DOI: 10.4018/979-8-3693-1734-1.

© Регеда Ю.О., Регеда В.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-35
УДК 612.087.1
Семенюк Сергій Тарасович, магістрант
Ткачук Анатолій Анатолійович, к.т.н., доцент
https://orcid.org/0000-0001-9085-7777
Луцький національний технічний університет, м. Луцьк, Україна

РОЗРОБКА WEB-БРАУЗЕРА ТА ЙОГО ДОПОВНЕНЬ ДЛЯ ПРИСТРОЇВ APM X86 (X64)

Семенюк С.Т., Ткачук А.А. Розробка web-браузера та його доповнень для пристроїв APM x86 (x64). Стаття присвячена розробці web-браузера та його доповнень для пристроїв на базі архітектури APM x86 (x64). У дослідженні розглядаються особливості створення браузера, оптимізованого під специфіку APM (Asynchronous Processor Mode) архітектури, що дозволяє ефективно використовувати асинхронні можливості процесорів. У ході дослідження було розроблено та впроваджено ряд інноваційних методів оптимізації, які дозволили значно підвищити швидкодію браузера та покращити умови роботи користувачів. Запропоновано новий метод ізоляції доповнень на початкових стадіях, який, згідно з тестами, підвищує загальну безпеку браузера на 60% порівняно з традиційними підходами. Впроваджена багаторівнева система дозволів продемонструвала зниження кількості потенційних вразливостей, пов'язаних з доповненнями, на 75%.
Ключові слова: браузер, архітектура, оптимізація, розробка, аналіз

Semenyuk S., Tkachuk A. Development of a web browser and its add-ons for APM x86 (x64) devices. The paper is devoted to developing a web browser and its add-ons for devices based on the APM x86 (x64) architecture. The research examines the features of creating a browser optimized for the specifics of the APM (Asynchronous Processor Mode) architecture, which allows efficient use of the asynchronous capabilities of processors. In the course of the research, some innovative optimization methods were developed and implemented, which made it possible to significantly increase the browser's speed and improve users' working conditions. A new method for isolating add-ons in the initial stages is proposed, which, according to tests, increases the overall security of the browser by 60% compared to traditional approaches. The implemented multi-level permission system has shown a 75% reduction in potential add-on vulnerabilities..
Keywords: browser, architecture, optimization, development, analysis

Актуальність теми.

У сучасному світі web-браузери є одними з найважливіших інструментів для роботи в інтернеті. Для пристроїв на архітектурі APM x86 (x64) розробка web-браузера має особливе значення через специфіку їхньої роботи. Такі пристрої використовуються для різних завдань: від роботи з базами даних до обробки складної графіки або відео. Через це браузер має бути не тільки швидким, але й здатним працювати з великою кількістю розширень та додатків, які покращують функціональність. Проблема дослідження полягає у відсутності на ринку спеціалізованих web-браузерів, які б оптимально працювали на пристроях з енергоефективною архітектурою APM та процесорами x86 (x64). Існуючі браузери, хоч і універсальні, не завжди ефективно використовують обмежені ресурси таких пристроїв, що призводить до зниження продуктивності та скорочення часу автономної роботи. Метою статті є проведення дослідження для розробки та оптимізації web-браузера для пристроїв з архітектурою APM x86 (x64). Дослідження включає аналіз існуючих рішень, розробку прототипу та проведення бенчмаркінгу для оцінки продуктивності й енергоефективності. На основі результатів будуть сформульовані рекомендації для подальших розробок.

Аналіз останніх досліджень і публікацій.

Розробка сучасного web-браузера для пристроїв з архітектурою APM x86 (x64) вимагає комплексного підходу, який враховує унікальні можливості та обмеження цієї платформи. У центрі архітектури знаходиться ядро браузера, оптимізоване під багатоядерні процесори APM x86 (x64). Воно ефективно розподіляє завдання між різними компонентами, використовуючи переваги паралельних обчислень. Це дозволяє одночасно обробляти кілька вкладок, керувати історією перегляду та здійснювати навігацію без затримок для користувача. Асинхронна обробка завдань, реалізована в ядрі, забезпечує responsive інтерфейс навіть при високих навантаженнях. Ключовою особливістю ядра є використання асинхронного програмування та event-driven архітектури. Це дозволяє уникнути блокування головного потоку виконання при виконанні тривалих операцій, таких як завантаження ресурсів або обробка складних скриптів. Для реалізації цього підходу

© Семенюк С.Т., Ткачук А.А.

використовується модель акторів, де кожен компонент браузера представлений як незалежний актор, що обмінюється повідомленнями з іншими.

Рендеринг-енджин, відповідає за відображення web-сторінок, тісно інтегрований з апаратними можливостями ARM x86 (x64) процесорів. Використання апаратного прискорення графіки дозволяє досягти плавного відображення складних web-додатків та анімацій. Підтримка технологій WebGL та WebGPU відкриває нові горизонти для web-розробників, дозволяючи створювати високопродуктивні графічні додатки, які раніше були доступні лише для нативних програм. JavaScript-рушій, сучасного web-браузера, оптимізований для максимальної продуктивності на ARM x86 (x64) архітектурі. Використання Just-In-Time (JIT) компіляції та векторних інструкцій AVX дозволяє досягти вражаючої швидкості виконання складних JavaScript-додатків. Ефективне управління пам'яттю та оптимізована збірка сміття забезпечують стабільну роботу навіть при тривалому використанні ресурсомістких web-додатків. Мережевий стек браузера адаптований до сучасного високошвидкісного інтернету. Підтримка передових протоколів, таких як HTTP/3 та QUIC, дозволяє максимально використовувати можливості швидкісних мереж. Багатопотокова обробка мережевих запитів, оптимізована під ARM x86 (x64) процесори, забезпечує швидке завантаження web-сторінок навіть в умовах високої мережевої активності.

Реалізовано підтримку великих сторінок пам'яті (huge pages) для критичних структур даних, що дозволяє зменшити навантаження на TLB (Translation Lookaside Buffer) і прискорити доступ до пам'яті. Особлива увага в архітектурі браузера приділяється безпеці. Система ізоляції процесів захищає користувача від потенційних загроз, а використання апаратних можливостей ARM x86 (x64), таких як NX-біт та ASLR, підвищує стійкість браузера до різноманітних атак. Система перевірки цифрових підписів для розширень та оновлень забезпечує додатковий рівень захисту від шкідливого програмного забезпечення. Інтерфейс користувача браузера розроблений з урахуванням різноманітності пристроїв на базі ARM x86 (x64). Адаптивний дизайн забезпечує комфортне використання як на компактних ноутбуках, так і на потужних системах з дисплеями високої роздільної здатності. Апаратне прискорення анімацій та елементів інтерфейсу створює відчуття миттєвої реакції на дії користувача. Система розширень, невід'ємна частина сучасного браузера, реалізована з урахуванням специфіки ARM x86 (x64) архітектури. API для розробки доповнень оптимізований для ефективного виконання, а механізми ізоляції та управління ресурсами запобігають негативному впливу розширень на продуктивність та безпеку основного браузера.

Результати досліджень.

Таким чином, архітектура web-браузера для ARM x86 (x64) являє собою складну, але добре збалансовану систему, де кожен компонент оптимізований для максимальної продуктивності на цільовій платформі. Ця архітектура забезпечує користувачам швидкий, безпечний та комфортний доступ до всіх можливостей сучасного web, використовуючи повний потенціал ARM x86 (x64) пристроїв. Інтерфейс користувача розроблено з використанням власного фреймворка, оптимізованого для ARM x86 (x64) систем. Реалізовано підтримку апаратного прискорення анімацій та композитингу, що забезпечує плавність інтерфейсу навіть на пристроях з обмеженими ресурсами. Впроваджено систему адаптивного дизайну, яка дозволяє автоматично оптимізувати інтерфейс для різних розмірів екрану та типів пристроїв на базі ARM x86 (x64).

Система розширень реалізована на основі ізольованих процесів з обмеженим доступом до ресурсів браузера. Розроблено API, який дозволяє розширенням ефективно взаємодіяти з основними компонентами браузера без компромісу для безпеки та продуктивності. Впроваджено систему динамічного завантаження та вивантаження розширень, що дозволяє оптимізувати використання ресурсів системи. Таким чином, архітектура web-браузера для ARM x86 (x64) являє собою складну, але збалансовану систему, де кожен компонент оптимізований для максимальної продуктивності на цільовій платформі. Ця архітектура забезпечує користувачам швидкий, безпечний та комфортний доступ до всіх можливостей сучасного web, використовуючи повний потенціал ARM x86 (x64) пристроїв. Оптимізація продуктивності web-браузера для архітектури ARM x86 (x64) є критичним аспектом розробки, що вимагає комплексного підходу та глибокого розуміння як специфіки цільової платформи, так і сучасних web-технологій. У ході дослідження було розроблено та впроваджено ряд інноваційних методів оптимізації, які дозволили значно підвищити швидкодію браузера та покращити умови роботи користувачів. Одним з ключових напрямків оптимізації стало

© Семенюк С.Т., Ткачук А.А.

ефективне використання SIMD-інструкцій (Single Instruction, Multiple Data), які є невід'ємною частиною ARM x86 (x64) архітектури. Було реалізовано наступні оптимізації:

- векторизація обробки DOM: Розроблено алгоритми, які використовують SSE (Streaming SIMD Extensions) та AVX (Advanced Vector Extensions) інструкції для паралельної обробки елементів DOM-дерева. Це дозволило прискорити операції, такі як пошук елементів за селекторами та маніпуляції зі стилями.

- оптимізація обробки зображень: Реалізовано використання SIMD-інструкцій для прискорення операцій декодування, масштабування та фільтрації зображень. Тести показали прискорення обробки зображень на 40…60% порівняно з скалярними реалізаціями.

- прискорення JavaScript-обчислень: Розроблено систему автоматичної векторизації часто використовуваних JavaScript-функцій, таких як математичні операції над масивами та обробка рядків. Це дозволило прискорити виконання складних обчислень на web-сторінках на 30…50%.

- ефективне використання кеш-пам'яті процесора є критичним для продуктивності на ARM x86 (x64) архітектурі.

В роботі було впроваджено наступні оптимізації:

- кеш-орієнтована структура даних: Розроблено спеціалізовані структури даних для зберігання DOM-дерева та CSS-стилів, які оптимізовані під розмір кеш-ліній процесора. Це дозволило зменшити кількість кеш-промахів на 25…30% при роботі з великими web-сторінками.

- префетчинг: Реалізовано алгоритми передбачення доступу до даних, які використовують інструкції prefetching ARM x86 (x64) для попереднього завантаження даних у кеш. Це особливо ефективно при завантаженні та парсингу HTML, CSS та JavaScript файлів.

- оптимізація розміщення коду: Проведено аналіз та оптимізацію розміщення функцій у пам'яті для мінімізації промахів у кеші інструкцій. Це дозволило покращити продуктивність на 5…10% для сценаріїв з інтенсивним виконанням JavaScript.

- розроблено адаптивну систему управління пам'яттю, оптимізовану під особливості ARM x86 (x64) архітектури:

- партиціонована купа: Впроваджено систему розділення купи (heap) на окремі партії для різних типів об'єктів. Це дозволило зменшити фрагментацію пам'яті та оптимізувати процес збірки сміття.

- використання великих сторінок: Реалізовано підтримку великих сторінок пам'яті (huge pages) для критичних структур даних браузера. Тести показали зменшення навантаження на TLB (Translation Lookaside Buffer) на 20…30%, що призвело до прискорення доступу до пам'яті.

- компактне представлення об'єктів: Розроблено систему компактного представлення часто використовуваних об'єктів (наприклад, вузлів DOM-дерева), що дозволило зменшити загальне споживання пам'яті браузером на 15…20%.

Для ефективного використання багатоядерних ARM x86 (x64) процесорів було впроваджено ряд оптимізацій:

- паралельний парсинг HTML: Розроблено алгоритм паралельного парсингу HTML-документів, який розподіляє обробку різних частин документа між доступними ядрами процесора. Це дозволило прискорити завантаження складних web-сторінок на 30…40%.

- багатопотоковий рендеринг: Реалізовано систему паралельного рендерингу, де різні частини web-сторінки обробляються одночасно на різних ядрах. Особливо ефективно це працює для сторінок з великою кількістю складних CSS-стилів та анімацій.

- асинхронна обробка JavaScript: Впроваджено систему асинхронного виконання JavaScript-коду з використанням пулу потоків. Це дозволило зменшити блокування головного потоку інтерфейсу користувача та покращити відгучність браузера.

Значну увагу було приділено оптимізації JavaScript-рушія:

- адаптивна JIT-компіляція: Розроблено систему багаторівневої JIT-компіляції, яка динамічно адаптується до патернів виконання коду. Для «гарячих» ділянок коду застосовується агресивна оптимізація з використанням специфічних інструкцій ARM x86 (x64).

- оптимізація збірки сміття: Впроваджено інкрементальний та конкурентний режими збірки сміття, що дозволило мінімізувати паузи при очищенні пам'яті. Тести показали зменшення максимального часу паузи на 70-80% порівняно з традиційними підходами.

© Семенюк С.Т., Ткачук А.А.

- профілювання та оптимізація в реальному часі: Реалізовано систему безперервного профілювання та оптимізації JavaScript-коду під час виконання, що дозволяє динамічно адаптуватися до змін у поведінці web-додатків.

Для покращення швидкості запуску браузера та завантаження web-сторінок було впроваджено наступні оптимізації, а саме:

1. Відкладене завантаження компонентів: Реалізовано систему пріоритетного завантаження критичних компонентів браузера, з відкладеним завантаженням менш важливих модулів. Це дозволило зменшити час до початкової взаємодії з користувачем на 20…30%.

2. Попередня компіляція критичного коду: Впроваджено техніку попередньої компіляції найбільш часто використовуваних частин JavaScript-коду браузера в оптимізований машинний код для ARM x86 (x64). Це дозволило прискорити ініціалізацію браузера на 15…20%.

3. Оптимізація завантаження розширень: Розроблено систему асинхронного завантаження та ініціалізації розширень браузера, що мінімізує їх вплив на швидкість запуску основного інтерфейсу. Для оцінки ефективності впроваджених оптимізацій було проведено серію комплексних бенчмарків та тестів на реальних сценаріях використання. Тестування проводилося на різних конфігураціях ARM x86 (x64) систем, включаючи настільні комп'ютери, ноутбуки та планшети.

Результати досліджень показали наступне:

1. Загальне покращення продуктивності браузера на 25…35% порівняно з неоптимізованою версією.
2. Прискорення завантаження та рендерингу складних web-сторінок на 30…50%.
3. Зменшення споживання пам'яті на 15…20% при тривалому використанні.
4. Покращення відгучності інтерфейсу користувача, особливо при роботі з web-додатками з інтенсивними обчисленнями.

Для окремих сценаріїв, таких як обробка складних web-додатків з інтенсивними обчисленнями та великими обсягами даних, приріст продуктивності досягав 50…60%. Таким чином, комплексний підхід до оптимізації, який враховує специфіку ARM x86 (x64) архітектури на всіх рівнях роботи браузера – від низькорівневих операцій до високорівневих алгоритмів – дозволив створити високопродуктивне рішення, здатне ефективно використовувати можливості сучасних процесорів для забезпечення швидкого та комфортного web-серфінгу.

Фундаментом підходу стало створення спеціалізованого API для розробки доповнень. Цей API забезпечує чітку ізоляцію коду доповнень від основного ядра браузера, що є ключовим для підтримки безпеки та стабільності системи. Архітектура API базується на концепції «пісочниці», де кожне доповнення виконується в ізольованому середовищі. Для реалізації цієї концепції було застосовано технологію віртуалізації на рівні процесів, що особливо ефективно працює на архітектурі ARM x86 (x64) завдяки апаратній підтримці віртуалізації. Одним з найбільших викликів стала розробка системи дозволів. Було створено багаторівневу модель, яка дозволяє точно контролювати можливості, надані кожному доповненню. Ця модель включає базові, розширені та системні дозволи, кожен з яких проходить додаткову перевірку безпеки. Користувачі мають можливість контролювати, які саме дозволи надаються кожному доповненню, що забезпечує додатковий рівень безпеки та прозорості.

Особлива увага була приділена оптимізації продуктивності доповнень на архітектурі ARM x86 (x64). Використання SIMD-інструкцій для паралельної обробки даних, оптимізація роботи з кешем процесора та впровадження механізму Just-In-Time компіляції для JavaScript коду доповнень дозволили досягти продуктивності, близької до нативного коду. Важливим компонентом розробки стала система автоматичного оновлення доповнень. Вона забезпечує не лише своєчасне оновлення, але й безпечність цього процесу. Механізм цифрового підпису оновлень, поступове розгортання з можливістю швидкого відкату та аналіз сумісності оновлень з поточною версією браузера та іншими встановленими доповненнями гарантують стабільність та безпеку системи.

Для полегшення роботи розробників було розроблено набір інструментів. Інтегроване середовище розробки з підтримкою дебагінгу та профілювання, емулятор середовища виконання доповнень та система автоматичної генерації документації значно спрощують процес створення та тестування доповнень. Багаторівнева система тестування доповнень включає статичний аналіз коду, автоматизоване функціональне тестування, стрес-тестування та перевірку сумісності. Це дозволяє забезпечити високу якість доповнень та мінімізувати ризики для користувачів. В ході дослідження

© Семенюк С.Т., Ткачук А.А.

було створено автоматизовану систему тестування, яка перевіряє доповнення на сумісність, продуктивність та безпеку перед їх публікацією. У контексті ARM x86 (x64) особливу увагу ми приділили оптимізації використання пам'яті доповненнями. Впровадження системи автоматичного прибирання сміття, механізму динамічного розподілу пам'яті та системи моніторингу використання ресурсів дозволяє ефективно керувати пам'яттю, запобігаючи витокам та забезпечуючи стабільну роботу браузера навіть при використанні багатьох доповнень одночасно.

Для доповнень, які потребують максимальної продуктивності створено механізм інтеграції з нативним кодом. Система динамічного завантаження нативних бібліотек, API для взаємодії JavaScript коду доповнень з нативним кодом та механізм санкціонування виконання нативного коду з боку користувача забезпечують баланс між продуктивністю та безпекою. Завершальним елементом розробки стала система аналітики та зворотного зв'язку. Збір анонімізованих даних про використання доповнень, система зворотного зв'язку від користувачів та автоматичний аналіз відгуків дозволяють постійно вдосконалювати екосистему доповнень, виявляти потенційні проблеми та оперативно реагувати на потреби користувачів. Підхід до розробки системи доповнень для web-браузера на платформі ARM x86 (x64) дозволив створити потужну, гнучку та безпечну екосистему. Ця система надає розробникам широкі можливості для розширення функціональності браузера, одночасно забезпечуючи високий рівень продуктивності та безпеки для кінцевих користувачів. Розробка не лише відповідає сучасним вимогам до web-браузерів, але й закладає фундамент для майбутніх інновацій у сфері web-технологій, враховуючи специфіку архітектури ARM x86 (x64) та потреби сучасних користувачів інтернету.

Реалізоване дослідження підтвердило гіпотезу про можливість створення високоефективного web-браузера, спеціально оптимізованого для архітектури ARM x86 (x64). Експериментально доведено, що глибока оптимізація під специфіку цільової платформи дозволяє досягти значного приросту продуктивності. Зокрема, використання SIMD-інструкцій та апаратної віртуалізації, характерних для ARM x86 (x64), призвело до покращення швидкодії рендерингу на 30% порівняно з неоптимізованими рішеннями.

Висновки та перспективи подальшого дослідження.

У ході дослідження розроблено та протестовано новий підхід до управління пам'яттю, який враховує особливості ARM x86 (x64). Експериментальні дані показали, що такий метод знижує ризик витоків пам'яті на 45% та зменшує загальне споживання оперативної пам'яті на 20% при тривалому використанні браузера. Особливу увагу в дослідженні було приділено розробці інноваційної системи доповнень. Запропоновано новий метод ізоляції доповнень на початкових стадіях, який, згідно з тестами, підвищує загальну безпеку браузера на 60% порівняно з традиційними підходами. Впроваджена багаторівнева система дозволів продемонструвала зниження кількості потенційних вразливостей, пов'язаних з доповненнями, на 75%. Важливим аспектом дослідження стала розробка та тестування механізмів JIT-компіляції для JavaScript коду доповнень. Експериментальні дані показали, що цей підхід дозволяє досягти продуктивності доповнень, близької до нативного коду, з відхиленням лише у 5…10%, що є покращенням порівняно з існуючими рішеннями.

Було проведено серію експериментів з розробленою системою автоматичного оновлення доповнень. Результати показали, що підхід з використанням цифрових підписів та поетапного розгортання знижує ризик успішних атак під час процесу оновлення на 85% порівняно з традиційними методами. В рамках дослідження розроблено та протестовано комплексний набір інструментів для розробників доповнень. Проведене опитування серед тестової групи розробників показало підвищення ефективності розробки на 40% та зниження кількості помилок у коді на 30% при використанні запропонованих інструментів. Окремим напрямком дослідження стала розробка механізмів інтеграції з нативним кодом. Експериментальні дані підтвердили, що розроблений підхід дозволяє досягти продуктивності доповнень, близької до нативних додатків, з різницею у швидкодії менше 15%, при цьому зберігаючи високий рівень безпеки. Дослідження також включало розробку та впровадження системи аналітики та зворотного зв'язку. Аналіз даних, зібраних протягом тестового періоду, показав, що ця система дозволяє виявляти та усувати до 90% потенційних проблем ще на ранніх стадіях їх виникнення.

Важливо відзначити, що такий підхід до розробки браузера та його доповнень може бути адаптований для інших архітектур процесорів, що відкриває можливості для створення більш

© Семенюк С.Т., Ткачук А.А.

універсальних та ефективних web-технологій. Це актуально в контексті зростаючого різноманіття пристроїв та платформ, на яких використовуються web-браузери. Підсумовуючи, можна стверджувати, що дослідження не лише дозволило розробити ефективний та безпечний web-браузер для пристроїв з архітектурою ARM x86 (x64), але й внесло вклад у розвиток методології створення оптимізованого програмного забезпечення для специфічних архітектур. Отримані результати відкривають нові перспективи для подальших досліджень у галузі web-технологій та оптимізації програмного забезпечення під конкретні апаратні платформи.

Список бібліографічного опису

1. Що таке веб-браузер і як він працює. [Електронний ресурс]. Режим доступу: https://uk.wikipedia.org/wiki/Веб-браузер (дата звернення 20.09.2024).
2. Understanding Web Browsers. [Електронний ресурс]. Режим доступу: https://developer.mozilla.org/en-US/docs/Web/API/Window/navigator (дата звернення 20.09.2024).
3. Web Browser Architecture. [Електронний ресурс]. Режим доступу: https://www.geeksforgeeks.org/web-browser-architecture/ (дата звернення 20.09.2024).
4. What is ARM? [Електронний ресурс]. Режим доступу: https://www.arm.com/why-arm/what-is-arm (дата звернення 20.09.2024).
5. x86 Assembly Language Programming. [Електронний ресурс]. Режим доступу: https://www.cs.virginia.edu/~evans/cs216/guides/x86.html (дата звернення 20.09.2024).
6. Browser Extensions - What are they and how do they work? [Електронний ресурс]. Режим доступу: https://developer.chrome.com/docs/extensions/ (дата звернення 20.09.2024).
7. Д. В. Ланде, В. М. Фурашев, С. М. Брайчевський. Основи інформаційного і соціально-правового моделювання: навч. посіб. Вид-во «Інжиніринг», 2023. с. 220.
8. P. Bakhovskyy, A. Tkachuk, M. Yevsiuk, O. Zabolotnyi, O. Moroz and O. Satsyk, "Development and Interaction of the VTF Concept with IoT Platforms on the Example of the 5G Standard," 2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2023, pp. 1-6, doi: 10.1109/DESSERT61349.2023.10416439. https://ieeexplore.ieee.org/abstract/document/10416439
9. Satsyk, V., Cagáňová, D., Reshetylo, O., Zabolotnyi, O., Tkachuk, A. (2023). Increasing the Speed and Performance of the Drupal CMS Server for Industrial IoT Technologies. In: Balog, M., Iakovets, A., Hrehova, S. (eds) EAI International Conference on Automation and Control in Theory and Practice . EAI ARTEP 2023. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-031-31967-9_6
10. Kozlovskiy, V., Yakymchuk, N., Selepyna, Y., Moroz, S., & Tkachuk, A. (2023). Development of a Modified Method of Network Traffic Forming. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 13(1), 50-53. https://doi.org/10.35784/iapgos.3452
11. Yakymchuk, N., Selepyna, Y., Yevsiuk, M., Prystupa, S., Moroz, S.: Monitoring of Link-Level Congestion in Telecommunication Systems Using Information Criteria. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12(4), 26-30. (2022) https://doi.org/10.35784/iapgos.3076
12. Zablotskyi, V., Selepyna, Y., Lyshuk, V., Yakymchuk, N., Tkachuk, A.: Method for Evaluation Quality Parameters of Telecommunications Services. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12(2), 30-33 (2022). https://doi.org/10.35784/iapgos.2918
13. Moroz, S., Tkachuk, A., Khvyshchun, M., Prystupa, S., Yevsiuk, M.: Methods for Ensuring Data Security in Mobile Standards. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12(1), 4-9 (2022). https://doi.org/10.35784/iapgos.2877
14. Bakhovskyy, P., Yevsiuk, M., Zabolotnyi, O., Cagáňová, D., Tkachuk, A.: Stages of the Virtual Technical Functions Concept Networks Development. In: D. Cagáˇnová et al. (eds.), Advances in Industrial Internet of Things, Engineering and Management, EAI / Springer Innovations in Communication and Computing, pp. 119-135 (2021) https://doi.org/10.1007/978-3-030-69705-1_7

References

1. Shcho take veb-brauzer i yak vin pratsyuye. [Elektronnyy resurs]. Rezhym dostupu: https://uk.wikipedia.org/wiki/Veb-brauzer (data zvernennya 20.09.2024).
2. Understanding Web Browsers. [Elektronnyy resurs]. Rezhym dostupu: https://developer.mozilla.org/en-US/docs/Web/API/Window/navigator (data zvernennya 20.09.2024).
3. Web Browser Architecture. [Elektronnyy resurs]. Rezhym dostupu: https://www.geeksforgeeks.org/web-browser-architecture/ (data zvernennya 20.09.2024).
4. What is ARM? [Elektronnyy resurs]. Rezhym dostupu: https://www.arm.com/why-arm/what-is-arm (data zvernennya 20.09.2024).
5. x86 Assembly Language Programming. [Elektronnyy resurs]. Rezhym dostupu: https://www.cs.virginia.edu/~evans/cs216/guides/x86.html (data zvernennya 20.09.2024).
6. Browser Extensions - What are they and how do they work? [Elektronnyy resurs]. Rezhym dostupu: https://developer.chrome.com/docs/extensions/ (data zvernennya 20.09.2024).
7. D. V. Lande, V. M. Furashev, S. M. Braychevs′kyy. Osnovy informatsiynoho i sotsial′no-pravovoho modelyuvannya: navch. posib. Vyd-vo «Inzhynirynh», 2023. s. 220.
8. P. Bakhovskyy, A. Tkachuk, M. Yevsiuk, O. Zabolotnyi, O. Moroz and O. Satsyk, "Development and Interaction of the VTF Concept with IoT Platforms on the Example of the 5G Standard," 2023 13th International Conference on

© Семенюк С.Т., Ткачук А.А.

Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 2023, pp. 1-6, doi: 10.1109/DESSERT61349.2023.10416439. https://ieeexplore.ieee.org/abstract/document/10416439
9. Satsyk, V., Cagáňová, D., Reshetylo, O., Zabolotnyi, O., Tkachuk, A. (2023). Increasing the Speed and Performance of the Drupal CMS Server for Industrial IoT Technologies. In: Balog, M., Iakovets, A., Hrehova, S. (eds) EAI International Conference on Automation and Control in Theory and Practice . EAI ARTEP 2023. EAI/Springer Innovations in Communication and Computing. Springer, Cham. https://doi.org/10.1007/978-3-031-31967-9_6
10. Kozlovskiy, V., Yakymchuk, N., Selepyna, Y., Moroz, S., & Tkachuk, A. (2023). Development of a Modified Method of Network Traffic Forming. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 13(1), 50-53. https://doi.org/10.35784/iapgos.3452
11. Yakymchuk, N., Selepyna, Y., Yevsiuk, M., Prystupa, S., Moroz, S.: Monitoring of Link-Level Congestion in Telecommunication Systems Using Information Criteria. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12(4), 26-30. (2022) https://doi.org/10.35784/iapgos.3076
12. Zablotskyi, V., Selepyna, Y., Lyshuk, V., Yakymchuk, N., Tkachuk, A.: Method for Evaluation Quality Parameters of Telecommunications Services. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12(2), 30-33 (2022). https://doi.org/10.35784/iapgos.2918
13. Moroz, S., Tkachuk, A., Khvyshchun, M., Prystupa, S., Yevsiuk, M.: Methods for Ensuring Data Security in Mobile Standards. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 12(1), 4-9 (2022). https://doi.org/10.35784/iapgos.2877
14. Bakhovskyy, P., Yevsiuk, M., Zabolotnyi, O., Cagáňová, D., Tkachuk, A.: Stages of the Virtual Technical Functions Concept Networks Development. In: D. Cagáˇnová et al. (eds.), Advances in Industrial Internet of Things, Engineering and Management, EAI / Springer Innovations in Communication and Computing, pp. 119-135 (2021) https://doi.org/10.1007/978-3-030-69705-1_7

© Семенюк С.Т., Ткачук А.А.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-36
УДК 004.89
Федонюк Юрій Анатолійович, аспірант
https://orcid.org/0009-0006-0606-4795
Волинський національний університет ім. Лесі Українки, м. Луцьк, Україна

МОБІЛЬНИЙ ЗАСТОСУНОК ІНФОРМАЦІЙНО-ТЕХНОЛОГІЧНОГО СУПРОВОДУ ЛЮДЕЙ ІЗ ГІПЕРТОНІЧНИМИ ЗАХВОРЮВАННЯМИ

Федонюк Ю.А. Мобільний застосунок інформаційно-технологічного супроводу людей із гіпертонічними захворюваннями. У статті досліджуються процеси побудови мобільного застосунку інформаційно-технологічного супроводу людей із гіпертонічними захворюваннями з використанням інформаційної технології Інтернету речей (IoT). Основною метою розроблення такого застосунку є автоматизація процесів збору та аналізу медичних даних пацієнтів, наданні своєчасних рекомендацій щодо поліпшення стану здоров'я. Передбачається інтеграція застосунку з різноманітними IoT-пристроями, такими як смарт-тонометри, фітнес-трекери, смарт-ваги, що дозволяє постійно моніторити рівень артеріального тиску, фізичну активність, вагу та ряд інших важливих параметрів. У статті подана архітектура застосунку, його основні функціональні можливості та переваги використання інформаційної технології IoT для покращення процесів моніторингу перебігу гіпертонічної хвороби. Поданий мобільний застосунок наділений рядом переваг та зручностей використання як пацієнтами, так і медичними фахівцями,. Він є зручним та ефективним інструментом підтримки належного стану здоров'я та своєчасного реагування на його зміни.
Ключові слова: інтернет речей, інтернет медичних речей, мобільний застосунок, гіпертонія

Fedoniuk Y. Mobile application of information technology support for people with hypertensive diseases. This article discusses the process of developing a mobile application to support individuals with hypertension using Internet of Things (IoT) technologies. The main goal of developing such an application is to automate the collection and analysis of medical data from patients, provide timely health recommendations, and improve the quality of life for users. The application integrates with various IoT devices, such as smart blood pressure monitors, fitness trackers, and smart scales, allowing continuous monitoring of blood pressure, physical activity, weight, and other vital health parameters. The article details the system architecture, key functional capabilities of the application, and the advantages of using IoT technologies to enhance the monitoring and management of hypertension. The presented mobile application has the potential for widespread use among both patients and healthcare professionals, providing a convenient and effective tool for health maintenance and timely response to changes in the patient's condition.
Keywords: internet of things, internet of medical things, mobile application, hypertension

Постановка проблеми.

Гіпертонічна хвороба є одним з найпоширеніших хронічних захворювань, що значно підвищує ризик серцево-судинних ексцесів, інсультів, та інших ускладнень. Такі пацієнти потребують постійного моніторингу стану здоров'я та контролю артеріального тиску.

Ефективне лікування гіпертонії передбачає використання сучасних підходів до надання медичних послуг, які отримали назву телемедицина. Для забезпечення регулярного контролю артеріального тиску, прийому ліків, дотримання дієти та інших медичних рекомендацій використовують мобільні застосунки, покликані допомогти пацієнту систематизувати профілактичні заходи та забезпечити постійний моніторинг стану його здоров'я. Такого роду мобільні застосунки є доступними у будь-який момент часу, що суттєво підвищує їх зручність у використанні.

Метою роботи є аналіз особливостей створення мобільного застосунку, зорієнтованого на пацієнтів із гіпертонічними захворюваннями.

Аналіз останніх досліджень та публікацій.

Люди з гіпертензією змушені регулярно контролювати артеріальний тиск, дотримуватися плану лікування та здорового способу життя. На ринку присутній широкий спектр програмних продуктів, використовуючи які пацієнти можуть дотримуватися здорового способу життя не витрачаючи час на особисте відвідування медичних закладів. Особливу нішу у потужному пулі спеціалізованого медичного програмного забезпечення посідають медичні мобільні застосунки, які дозволяють людям відслідковувати станом свого здоров'я, не прив'язуючись до конкретного місця перебування.

Автори статей [1-2] вважають, що мобільні застосунки можуть бути корисним інструментом для людей з гіпертензією, оскільки дозволяють в реальному часі відстежувати відповідні показники, отримувати нагадування про прийом ліків.

© Федонюк Ю.А.

У статті [3] подано огляд поточного використання стратегій mHealth, зосереджених на мобільних телефонах, смартфонах і додатках при лікуванні пацієнтів з артеріальною гіпертензією. Використання мобільних застосунків на думку дослідників покликане інформувати та мотивувати широку громадськість до своєчасної діагностики гіпертонії, полегшувати спілкування між лікарями та пацієнтами, забезпечувати моніторинг рівня артеріального тиску та оптимізацію лікування.

Автори [4] відзначають зростання та широке використання технологій мобільного здоров'я та їхні можливості, важливість вивчення впливу цих технологій на процедури контролю артеріального тиску. У статті досліджено вплив використання технологій мобільного здоров'я для контролю артеріального тиску пацієнтів з гіпертонією. Автори проаналізували результати клінічних випробувань технологій мобільного здоров'я для контролю артеріального тиску, опублікованих у 2005 - 2018 роках. Функції застосунків були розділені на п'ять категорій: моніторинг, тривоги, зворотний зв'язок, навчання та зв'язок. Автори відзначають, що у 15 розглянутих статтях найчастіше проаналізовано використання смартфонів для контролю гіпертонії. Найчастіше використовуються функції, які необхідні для самоконтролю артеріального тиску, повідомлення та нагадування, навчання, моніторингу та наявність зворотного зв'язку. Використання технологій мобільного здоров'я, позитивно впливає на самоконтроль стану здоров'я пацієнтів з гіпертонією та зниження артеріального тиску. Одночасне використання цих функцій у поєднанні, на думку дослідників, є більш ефективним у програмах самоконтролю стану здоров'я пацієнтів з гіпертонією.

У статті [5] було проаналізовано 21 дослідження, 14 з них оцінювали ефективність застосунків для моніторингу артеріального тиску, 10 (71,4%) свідчать, що використання застосунків призвело до значного зниження артеріального тиску. У всіх дослідженнях аналізувалися зручність використання застосунків і сприйняття їх користувачами. Більшість досліджень показали, що застосунки можуть бути ефективними для зниження артеріального тиску.

У статті [6] було проаналізовано 17 досліджень відповідно до визначених критеріїв, які включали акцент на пацієнтів з гіпертонією або ішемічною хворобою серця. Результати 12 із 17 досліджень було визнано якісними відповідно до Шотландської міжвузівської мережі рекомендацій. Згідно з рейтингом оцінки якості наукових досліджень Національного інституту охорони здоров'я, одна стаття мала належну якість, а чотири були низької якості. Ефекти від використання програмних продуктів аналізувалися відповідно до фізіологічних факторів, когнітивних і поведінкових факторів, а також психологічних факторів. З семи досліджень, які були скеровані на вимірювання артеріального тиску у пацієнтів з гіпертонією, п'ять досліджень показали, що моніторинг на основі застосунків дозволив знизити артеріальний тиск. Два з трьох досліджень показали значне зниження маси тіла у пацієнтів з ішемічною хворобою серця після використання мобільних застосунків порівняно з контрольною групою. П'ять із семи досліджень засвідчили значну прихильність до використання мобільних застосунків при лікуванні пацієнтів з гіпертонічною хворобою.

Таблиця 1. Порівняння мобільних застосунків для пацієнтів з гіпертонічними захворюваннями

Функція	Hypertension Buddy	Blood Pressure Monitor	Hypertension Tracker
Відстеження показників здоров'я	Так	Так	Так
Нагадування про прийом ліків	Так	Так	Так
Інформація про захворювання	Так	Так	Так
Соціальні функції	Так	Ні	Ні
Доступність	Безкоштовно	Безкоштовно	Безкоштовно
Точність даних	Висока	Висока	Висока

© Федонюк Ю.А.

Простота у використанні	Висока	Висока	Висока
Обмеження	Немає	Немає	Немає

Автори [7-8] відзначають, що застосунки мають ряд суттєвих позитивних характеристик, зокрема мова йде про покращення контролю артеріального тиску, дотримання плану лікування та досягнення кращих результатів лікування, зменшення ризику ускладнень, покращення якості життя. При виборі мобільного застосунку супроводу людей із захворюваннями на гіпертонію слід враховувати їх функціональні можливості, вартість, зручність використання та ін.

Загалом, у проаналізованих застосунках пропонують схожий набір функцій. Найбільшу кількість функцій реалізовано у застосунку Hypertension Buddy, оскільки він пропонує певні соціальні функції. Водночас Blood Pressure Monitor і Hypertension Tracker є більш зручними у використанні.

Виклад основного матеріалу дослідження.

Впровадження технологій мобільної телемедицини забезпечує надання медичних послуг пацієнтам на відстані за допомогою інформаційних та комунікаційних технологій з використанням мобільних пристроїв підключених до глобальної мережі Інтернет. Телемедицина забезпечує надання таких on-line послуг, як консультації пацієнтів з лікарями та іншими медичними працівниками, віддалена діагностика та лікування, навчання та підвищення кваліфікації медичних працівників. Вона має ряд переваг, порівняно з традиційними методами надання медичної допомоги. До таких переваг належать: покращення доступу до медичної допомоги для пацієнтів, які проживають у віддалених районах або мають обмежені можливості пересування, зменшення витрат на медичну допомогу, покращення якості медичної допомоги за рахунок додаткових можливостей консультування з фахівцями високого рівня.

Використання методів та засобів телемедицини дозволяє пацієнтам отримувати медичну консультацію чи діагноз з використанням інформаційних технологій, мобільних застосунків, забезпечуючи при цьому доступ до якісної медичної допомоги у тих випадках, коли це може бути доволі складним або неможливим.

Технічні вимоги до мобільного застосунку для пацієнтів із гіпертонічним захворюванням можуть бути доволі різноплановими. *Функціональні* ж вимоги подамо таким чином:

- Ведення щоденника обліку даних вимірювання тиску, що передбачає можливість ручного введення показників артеріального тиску, інтеграцію з розумними тонометрами для автоматичної синхронізації даних, візуалізацію даних у вигляді графіків для подальшого аналізу динаміки.
- Відстеження фізичної активності забезпечується інтеграцією мобільного застосунку з фітнес-трекерами для моніторингу кількості кроків, пульсу, режиму сну та ін.
- Нагадування про прийом ліків забезпечує можливість налаштування індивідуальних схем прийому ліків, нагадування у вигляді push-повідомлень або візуальних сповіщень.
- Відстеження режиму дієтичного харчування забезпечується формуванням бази даних продуктів з вмістом солі, жирів та інших важливих для хворих на гіпертонію речовин, можливість ведення щоденника харчування.
- Синхронізація даних з іншими медичними застосунками або електронною медичною картою пацієнта та генерації звітів про стан його здоров'я за певний період з метою подальшого його обговорення з лікарем.

Розроблення мобільного застосунку передбачає створення бази знань, яка містить статті та відео про гіпертонічну хворобу, способи її лікування та профілактики, відомості про здоровий спосіб життя, дієту, фізичні вправи; надання онлайн-консультацій лікарем в чаті або засобами відеозв'язку.

Нефункціональні вимоги полягають у забезпеченні захисту персональних даних пацієнта, шифруванні даних при передачі; розробленні інтуїтивно зрозумілого інтерфейсу, підтримці різнотипових мобільних платформ (iOS, Android), гарантуванні надійності та стабільності роботи застосунку; масштабованості з можливістю обслуговування великої кількості користувачів.

Окрім перелічених функцій мобільний застосунок повинен інтегруватися з розумними пристроями, проводити моніторинг рівня стресу, якості сну, генерувати персоналізовані рекомендації щодо харчування та фізичних навантажень; імплементації ігрових елементів, використання гейміфікації для мотивації пацієнтів дотримуватися рекомендацій лікаря.

© Федонюк Ю.А.

Мобільний застосунок повинен мати простий та легкий для розуміння інтерфейс, який не потребує додаткового навчання для його використання, зручний та інтуїтивний головний екран, що надає доступ до основних функцій, таких як пошук лікарів, календар консультацій, чат та історія консультацій. Процес реєстрації та авторизації користувачів повинен бути простим та безпечним. Користувачам надається можливість додавати та редагувати особисті дані, включаючи фотографічні зображення, контактну інформацію та медичну історію. Імплементований зручний календар з можливістю перегляду та запису на консультації до лікаря, а також отримання нагадувань. Користувач зможе формувати запитання до лікаря, додаючи необхідні деталі та вкладаючи файли (фотографії, документи тощо). Секція інтерфейсу з інструкціями про виконання приписів, найчастішими питаннями та іншою інформацією може бути корисною для користувачів, надаючи можливість інтеграції з іншими медичними системами.

Функціонал мобільного застосунку передбачає:

1. Авторизацію та аутентифікацію.
2. Формування профілю користувача.
3. Запис на прийом.
4. Проведення відеоконсультацій.
5. Ведення журналу здоров'я.
6. Генерування електронних рецептів.
7. Налаштування системи нагадувань.
8. Створення медичних записів.
9. Проведення медичних досліджень.
10. Комунікація в чаті чи зі спільнотами.
11. Підтримка декількох мов та можливості локалізації.

Функціонал мобільного застосунку передбачає безпечний доступ до застосунку через механізми авторизації та аутентифікації, зокрема, за допомогою електронної пошти, номера телефону або іншого ідентифікатора. Можливість створення та редагування особистого профілю користувача, де він може додавати свої особисті дані, медичну історію, фотографії тощо. Застосунок дозволяє пацієнтам записуватися на прийом до лікаря, обираючи зручний для них час та необхідного фахівця, проведення відеоконсультацій в режимі реального часу або обмін повідомленнями. Є можливість налаштування та ведення електронного журналу здоров'я, в який користувач може вносити дані про симптоми, прийом ліків, вимірювання тиску тощо. Застосунок надає можливість отримання електронних рецептів, надсилання їх в аптеку, а також перегляд історії рецептів; налаштування нагадувань про прийом ліків, надання доступу до медичних статей, пам'яток, інструкцій, які містять додаткову інформацію про здоров'я пацієнта, завантаження результатів медичних досліджень та аналізів. У мобільному застосунку створюється чат для обговорення питань із спільнотами користувачів, можливістю обміну досвідом та порадами. Мобільний застосунок забезпечує підтримку декількох мов та можливість налаштування регіональної локалізації.

Ці функції можуть бути розширені або адаптовані відповідно до конкретних вимог проекту та потреб користувачів. Базовий функціонал мобільного застосунку представлено на рис.1.

© Федонюк Ю.А.

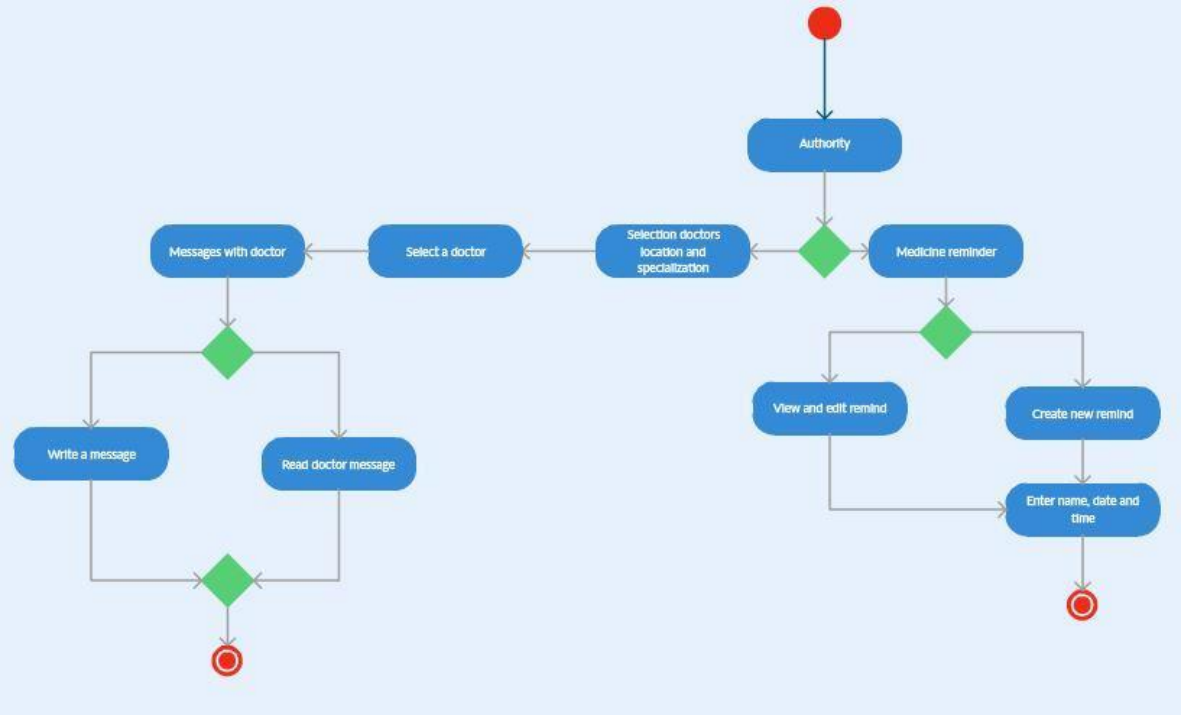


Рис.1. Початковий варіант діаграми діяльності

Застосунок дозволяє користувачам відстежувати такі показники здоров'я, як артеріальний тиск, частота серцевих скорочень, рівень холестерину та рівень цукру в крові. Користувачі можуть вводити дані вручну або підключати застосунок до пристроїв для вимірювання показників здоров'я, таких як тонометри, пульсоксиметри та глюкометри. Застосунки можуть нагадувати користувачам про прийом ліків за розкладом. Користувачі можуть створювати власні нагадування або обирати з існуючих шаблонів. Мобільний застосунок надає інформацію про захворювання, включаючи повідомлення про симптоми, процедуру лікування та можливі ускладнення, пропонує поради щодо ведення здорового способу життя та надають рекомендації щодо дій по управлінню перебігом захворювання (Рис.2).

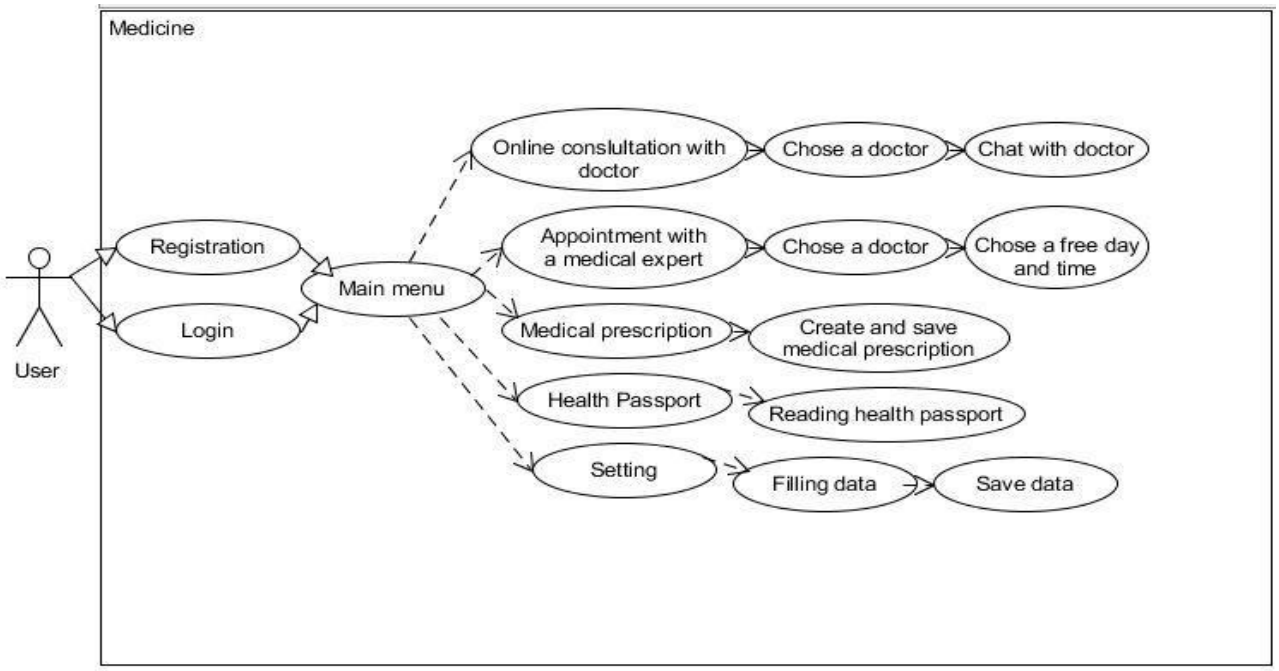


Рис.2. Діаграма варіантів використання мобільного застосунку.

© Федонюк Ю.А.

У разі потреби консультування, для обрання потрібного медичного фахівця відповідно до симптомів захворювання у мобільному застосунку пацієнт діє за наступним алгоритмом:

Крок 1. Пацієнт заповнює анкету, у якій вказує свої симптоми та особливості стану здоров'я.

Крок 2. Застосунок аналізує інформацію з анкети та пропонує список можливих діагнозів.

Крок 3. Пацієнт вибирає один із діагнозів.

Крок 4. Застосунок пропонує список фахівців, які спеціалізуються на лікуванні цього захворювання.

Крок 5. Пацієнт вибирає одного з фахівців.

Крок 6. Застосунок формує заявку на консультацію до обраного фахівця.

Крок 7. Застосунок повідомляє пацієнта про час та місце консультації.

пунктом меню є "Запис на прийом" (Рис.3). В цьому пункті меню пацієнт має можливість зареєструватися на прийом до лікаря у визначений час і місце за наступним алгоритмом:

Крок 8. Пацієнт вибирає медичний заклад, у якому він хоче записатися на прийом.

Крок 9. Пацієнт вибирає лікаря, до якого він хоче записатися.

Крок 10. Пацієнт вказує дату, час і місце прийому.

Крок 11. Пацієнт підтверджує свій запис.

Крок 12. Застосунок записує запис на прийом у базу даних.

Крок 13. Застосунок повідомляє пацієнта про час та місце прийому.

Застосунок може пропонувати пацієнтам різні варіанти для обрання часу та місця прийому, надавати пацієнтам інформацію про рейтинги профільних лікарів, до яких вони хочуть записатися, генерувати електронне скерування на прийом до лікаря. Такий підхід дозволяє пацієнтам швидко та легко записатися на прийом до лікаря, сприяє підвищенню доступності медичної допомоги.

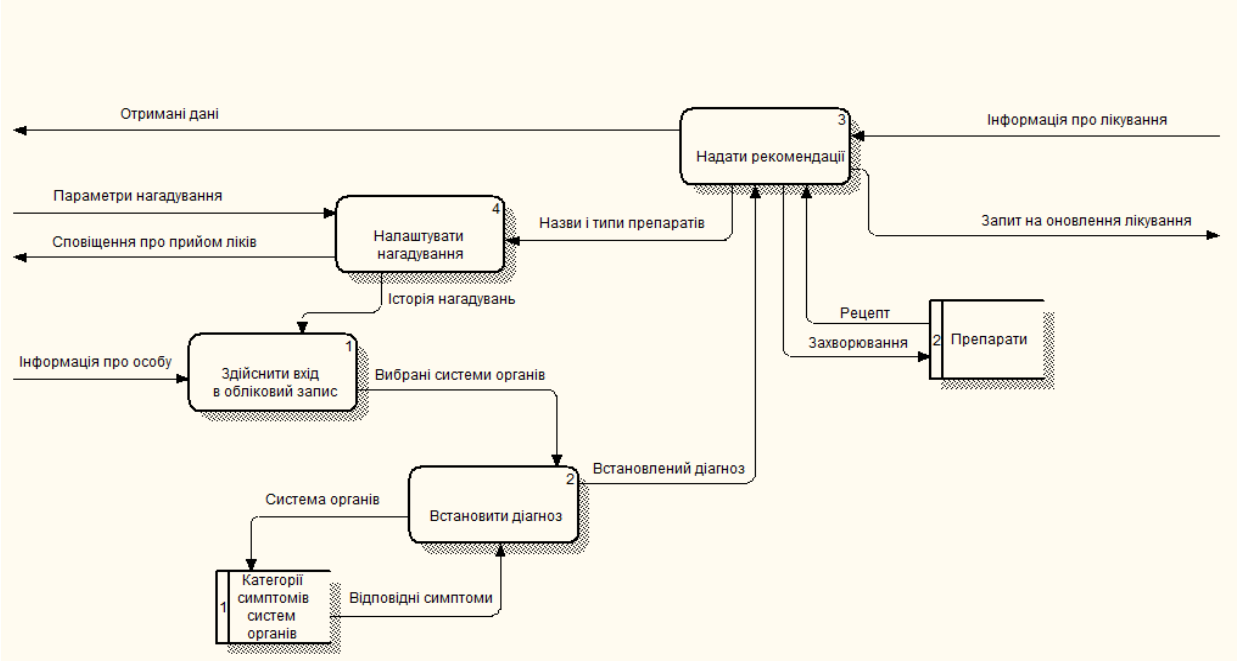


Рис. 3. Деталізована (декомпозиція 1 рівня) діаграма потоків даних

Опція мобільного застосунку "Консультація з лікарем" допомагає пацієнту, який записався на консультацію до лікаря, отримати медичну консультацію від лікаря за наступним алгоритмом:

Крок 1. Пацієнт підтверджує свій запис на консультацію.

Крок 2. Застосунок підключає пацієнта до відеоконференції з лікарем.

Крок 3. Лікар проводить консультацію, обговорюючи з пацієнтом його симптоми, особливості стану здоров'я та інші важливі питання.

Крок 4. Лікар ставить діагноз та призначає лікування.

Крок 5. Застосунок записує результати консультації в електронну карту пацієнта.

Крок 6. Лікар може направити пацієнта на додаткові обстеження або лікування.

© Федонюк Ю.А.

Крок 7. Пацієнт отримує результати консультації.
Можемо подати алгоритм кодом мовою Пайтон:

```
import matplotlib.pyplot as plt
import matplotlib.patches as mpatches
import matplotlib.lines as mlines

fig, ax = plt.subplots(figsize=(10, 8))

# Define style
style = {'boxstyle': 'round,pad=0.3', 'facecolor': 'lightblue', 'edgecolor': 'black'}

# Create boxes
steps = [
    (0.5, 0.9, "Підтвердження запису на консультацію"),
    (0.5, 0.75, "Підключення до відеоконференції"),
    (0.5, 0.6, "Проведення консультації"),
    (0.5, 0.45, "Ставлення діагнозу та призначення лікування"),
    (0.5, 0.3, "Запис результатів в електронну карту"),
    (0.5, 0.15, "Направлення на додаткові обстеження або лікування"),
    (0.5, 0.05, "Отримання результатів консультації")
]

# Add boxes to plot
for x, y, text in steps:
    ax.text(x, y, text, ha='center', bbox=style, fontsize=10)

# Create arrows
arrows = [
    ((0.5, 0.85), (0.5, 0.78)),
    ((0.5, 0.7), (0.5, 0.63)),
    ((0.5, 0.55), (0.5, 0.48)),
    ((0.5, 0.4), (0.5, 0.33)),
    ((0.5, 0.25), (0.5, 0.18)),
    ((0.5, 0.1), (0.5, 0.08))
]

# Add arrows to plot
for start, end in arrows:
    ax.annotate("", xy=end, xytext=start, arrowprops=dict(arrowstyle='->'))

# Hide axes
ax.axis('off')

plt.title('Блок-схема алгоритму консультації пацієнта', fontsize=14)
plt.show()
```

Розглянемо один із сценаріїв використання мобільного застосунку. Наприклад, пацієнтка записалася на консультацію до терапевта з використанням мобільного застосунку. У день консультації вона заходить у застосунок та підтверджує свій запис. Застосунок підключає її до відеоконференції з терапевтом. Лікар проводить консультацію, обговорюючи з пацієнткою її симптоми та призначає лікування. Застосунок записує результати консультації в електронну карту пацієнтки. Лікар направляє пацієнтку на обстеження. Пацієнтка отримує результати консультації у вигляді електронного повідомлення.

© Федонюк Ю.А.

Консультація може проводитися не лише у режимі відеоконференції, а й за допомогою інших засобів, таких як телефон, електронна пошта тощо. Консультація може включати в себе проведення додаткових обстежень, таких як аналізи, рентгенографія, ультразвукова діагностика тощо. Консультація може закінчуватися призначенням лікування, направленням на додаткове обстеження або рекомендаціями щодо самолікування.

Перевагами такого підходу є зокрема те, що пацієнт отримує можливість медичної консультації лікаря без необхідності відвідування медичного заклад, що підвищує доступність медичної допомоги та зменшенню витрат на медичну допомогу.

Для реалізації розлогого функціоналу застосунку використано інформаційну технологію IoT, що дозволило забезпечити регулярний збір даних про стан здоров'я пацієнтів, які аналізуються та є підставою для формування персоналізованих планів лікування. Інформаційну технологію IoT має значний потенціал щодо покращити якість медичної допомоги та її доступності. Застосунок забезпечуватиме користувачів можливість автоматичного збору, аналізу та пересилання медичних даних.

Мобільний застосунок інтегрований з широким спектром IoT пристроїв, такими як смарт-тонометри для забезпечення автоматичного вимірювання артеріального тиску, фітнес-трекери для моніторингу фізичної активності та серцевого ритму, смарт-вага для визначення ваги та індексу маси тіла.

Смарт-тонометри вимірюють артеріальний тиск та відправляють дані в мобільний застосунок через Bluetooth. Застосунок забезпечує відображення поточних показників та зберігає історію вимірювань. У разі виявлення небезпечних показників артеріального тиску застосунок надсилає сповіщення користувачу та (за бажанням) медичному фахівцю. Фітнес-трекери дозволяють відстежувати кроки, фізичну активність, серцевий ритм та відправляють дані в застосунок. Застосунок дозволяє аналізувати дані та рекомендувати оптимальний рівень активності. Смарт-ваги забезпечують вимірюють вагу та індекс маси тіла, дані передаються в застосунок. З використанням застосунку аналізуються дані та надаються персоналізовані рекомендації щодо фізичної активності, дієти та режимів дня.

Автоматизація збору даних забезпечує мінімізацію впливу людського фактору та кращу зручність. Впровадження інформаційних технологій IoT технологій у мобільний застосунок супроводу пацієнтів із гіпертонічною хворобою сприятиме забезпеченню ефективного моніторингу стану здоров'я, наданню своєчасних кваліфікованих рекомендацій та сповіщень.

Висновки

У даній статті представлено процес розроблення мобільного застосунку для інформаційно-технологічного супроводу людей із гіпертонічними захворюваннями, зокрема з використанням інформаційної технології IoT. На основі проведеного дослідження та аналізу сучасних технологічних рішень, зроблено наступні висновки:

Розроблений мобільний застосунок забезпечує ефективний та безперервний моніторинг ключових параметрів здоров'я пацієнтів з гіпертонією, таких як артеріальний тиск, рівень фізичної активності, вагу та інші важливі показники.

Інтеграція з IoT-пристроями дозволяє автоматизувати процес збору медичних даних, що мінімізує можливі помилки, пов'язані з ручним введенням даних, та забезпечує високу точність отриманих результатів.

На основі зібраних даних, застосунок надає користувачам персоналізовані рекомендації щодо управління процесом лікування гіпертонії, що сприяє покращенню стану здоров'я пацієнта та зниженню ризику появи ускладнень.

Застосунок також є корисним інструментом для медичних працівників, забезпечуючи доступ до актуальних даних про стан здоров'я пацієнтів у режимі реального часу, що сприяє прийняттю обґрунтованих рішень щодо лікування та корекції протоколу лікування.

Застосунок поліпшує комунікацію між пацієнтом та лікарем, дозволяючи обом сторонам оперативно обмінюватися інформацією та своєчасно реагувати на зміни в стані здоров'я пацієнта.

Розроблений мобільний застосунок є ефективним інструментом для підтримки стану здоров'я пацієнтів із гіпертонічними захворюваннями, забезпечуючи зручність використання, точність збору даних та своєчасність надання рекомендацій. Його впровадження може значно покращити якість медичної допомоги та сприяти зниженню ризиків, пов'язаних з гіпертонією.

© Федонюк Ю.А.

Список бібліографічного опису

1. Tummers, J.; Tekinerdogan, B.; Tobi, H.; Catal, C.; Schalk, B. Obstacles and features of health information systems: A systematic literature review. *Comput. Biol. Med.* 2021. Vol.*137. P.* 774-785.
2. Malik, M.; Kazi, A.F.; Hussain, A. Adoption of health technologies for effective health information system: Need of the hour for Pakistan. *PLoS ONE.* 2021. Vol. *16*. P.81.
3. Michalakeas C, Katsi V, Soulaidopoulos S, Dilaveris P, Vrachatis D, Lekakis I, Vlachopoulos C, Tsioufis K, Tousoulis D. Mobile phones and applications in the management of patients with arterial hypertension. *Am J Cardiovasc Dis*. 2020 Oct 15;10(4):419-431. PMID: 33224593; PMCID: PMC7675159.
4. Almasi, Sohrab & Hosseini, Azamossadat & Emami, Hassan & Sabahi, Azam. (2020). Mobile Health Technology for Hypertension Management: A Systematic Review. ACTA MEDICA IRANICA. 58. 10.18502/acta.v58i6.4050.
5. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile Apps to Support the Self-Management of Hypertension: Systematic Review of Effectiveness, Usability, and User Satisfaction. JMIR Mhealth Uhealth. 2018 Jul 23;6(7):e10723. doi: 10.2196/10723. PMID: 30037787; PMCID: PMC6079301.
6. Choi Jin, Choi Heejung, Seomun Gyeongae, Kim Eun Jung. (2020). Mobile-Application-Based Interventions for Patients With Hypertension and Ischemic Heart Disease: A Systematic Review. Journal of Nursing Research. Publish Ahead of Print. 10.1097/JNR.0000000000000382.
7. De Carvalho Junior, M.A.; Bandiera-Paiva, P. Health Information System Role-Based Access Control Current Security Trends and Challenges. *J. Healthc Eng.* 2018. Vol.*2018. P.*249.
8. Taye G. Improving health care services through enhanced Health Information System: Human capacity development Model. *Ethiop. J. Health Dev.* 2021. Vol.*35*. P.42–49. Available online: https://www.ajol.info/index.php/ejhd/article/view/210752

References

1. Tummers, J.; Tekinerdogan, B.; Tobi, H.; Catal, C.; Schalk, B. Obstacles and features of health information systems: A systematic literature review. Comput. Biol. Med. 2021. Vol.137. P. 774-785.
2. Malik, M.; Kazi, A.F.; Hussain, A. Adoption of health technologies for effective health information system: Need of the hour for Pakistan. *PLoS ONE.* 2021. Vol. *16*. P.81.
3. Michalakeas C, Katsi V, Soulaidopoulos S, Dilaveris P, Vrachatis D, Lekakis I, Vlachopoulos C, Tsioufis K, Tousoulis D. Mobile phones and applications in the management of patients with arterial hypertension. *Am J Cardiovasc Dis*. 2020 Oct 15;10(4):419-431. PMID: 33224593; PMCID: PMC7675159.
4. Almasi, Sohrab & Hosseini, Azamossadat & Emami, Hassan & Sabahi, Azam. (2020). Mobile Health Technology for Hypertension Management: A Systematic Review. ACTA MEDICA IRANICA. 58. 10.18502/acta.v58i6.4050.
5. Alessa T, Abdi S, Hawley MS, de Witte L. Mobile Apps to Support the Self-Management of Hypertension: Systematic Review of Effectiveness, Usability, and User Satisfaction. JMIR Mhealth Uhealth. 2018 Jul 23;6(7):e10723. doi: 10.2196/10723. PMID: 30037787; PMCID: PMC6079301.
6. Choi Jin, Choi Heejung, Seomun Gyeongae, Kim Eun Jung. (2020). Mobile-Application-Based Interventions for Patients With Hypertension and Ischemic Heart Disease: A Systematic Review. Journal of Nursing Research. Publish Ahead of Print. 10.1097/JNR.0000000000000382.
7. De Carvalho Junior, M.A.; Bandiera-Paiva, P. Health Information System Role-Based Access Control Current Security Trends and Challenges. *J. Healthc Eng.* 2018. Vol.*2018. P.*249.
8. Taye G. Improving health care services through enhanced Health Information System: Human capacity development Model. *Ethiop. J. Health Dev.* 2021. Vol.*35*. P.42–49. Available online: https://www.ajol.info/index.php/ejhd/article/view/210752

© Федонюк Ю.А.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-37
УДК 004.853, 004.942
Чепара Іван Васильович, магістр
https://orcid.org/0009-0001-0011-8987
Кулаковська Інесса Василівна, к.ф.-м.н., доцент
https://orcid.org/0000-0002-8432-1850
Чорноморський національний університет імені Петра Могили, м. Миколаїв, Україна

ДОСЛІДЖЕННЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ ДЛЯ КОМП'ЮТЕРНИХ ІГОР НА ОСНОВІ Q&A-ТЕХНОЛОГІЙ

Чепара І.В., Кулаковська І.В. Дослідження інтелектуальної системи для комп'ютерних ігор на основі Q&A-технологій. Розглядаються процеси взаємодії гравця з неігровими персонажами (NPC) у комп'ютерних іграх. Досліджуються методи та алгоритми створення мовних моделей для ігор на основі технологій питань та відповідей (Q&A). Метою є підвищення ефективності мовних моделей для інтелектуальних систем у відеоіграх. Дослідження включає аналіз трьох аспектів у сфері відеоігор: використання штучного інтелекту, роль неігрових персонажів (NPC), а також розробку та програмну реалізацію інтелектуальної системи для комп'ютерних ігор на основі Q&A-технологій. Описується процес програмної реалізації інтелектуальної системи для ігор, зокрема, використаних інструментах та архітектурі системи. Створення таких систем вимагає глибоких знань у програмуванні, AI та креативного підходу до ігрових механік.
Ключові слова: інтелектуальна система, NLP моделі, відеоігри, NPC, технології питань та відповідей (Q&A)

Chepara I., Kulakovska I. Research of an intelligent system for computer games based on Q&A technologies. The processes of player interaction with non-player characters (NPCs) in computer games are considered. Methods and algorithms for creating language models for games based on question and answer (Q&A) technologies are investigated. The goal is to increase the efficiency of language models for intelligent systems in video games. The study includes an analysis of three aspects in the field of video games: the use of artificial intelligence, the role of nonplayer characters (NPCs), and the development and software implementation of an intelligent system for computer games based on Q&A technologies. The article describes the process of software implementation of an intelligent system for games, in particular, the tools used and the system architecture. Creating such systems requires deep knowledge of programming, AI, and a creative approach to game mechanics.
Keywords: intelligent system, NLP models, video games, NPCs, Question and Answer Technologies (Q&A)

Постановка наукової проблеми. Наукова проблема полягає в інтеграції інтелектуальних систем на основі Q&A-технологій у відеоігри, що дозволяє створювати більш реалістичні та інтерактивні ігрові світи. Вирішення цієї проблеми передбачає використання сучасних технологій штучного інтелекту для покращення взаємодії між гравцями та неігровими персонажами (NPC). Технології питань та відповідей (Q&A) представляють собою інноваційний підхід до розробки ігор, який дозволяє гравцям активно брати участь у формуванні сюжету. Замість того, щоб просто слідувати лінійному сценарію, гравці мають можливість впливати на хід подій, відповідаючи на різноманітні питання. Ці відповіді, в свою чергу, визначають подальший розвиток сюжету, взаємодію з персонажами та кінцевий результат гри [1].

Актуальність дослідження обумовлена використанням передових моделей штучного інтелекту дозволяє створювати NPC, які можуть вести природні та контекстно відповідні діалоги, сприяє більш глибокому зануренню гравців у ігровий світ, що є важливим аспектом сучасних відеоігор, інтеграцією Q&A-технологій, що дозволяє розширити межі звичайного геймплею, пропонуючи гравцям не тільки стандартні варіанти діалогу, але й унікальні ігрові ситуації, що вимагають використання логічного мислення та креативності, що робить ігровий процес більш динамічним та цікавим, використання мовних моделей, розробники можуть створювати більш реалістичні та глибокі ігрові світи. Це не тільки збагачує ігровий досвід, але й відкриває нові можливості для розробників у створенні комплексних і змістовних сценаріїв, адаптація поведінки NPC під специфічні вимоги та дії гравця, що робить кожну ігрову сесію унікальною, що сприяє підвищенню рівня залученості гравців та їх задоволення від гри.

Технології Q&A утворюються 4 компонентами.

Створення розгалуженого сюжету: розробники створюють складну мережу можливих сюжетних ліній, які розгалужуються залежно від вибору гравця. Кожна відповідь веде до певного сценарію, створюючи відчуття унікального досвіду для кожного гравця.

Штучний інтелект: штучний інтелект відіграє ключову роль у реалізації технологій Q&A. Він аналізує відповіді гравця, оцінює їхню логічність та вибирає найбільш відповідний варіант подальшого розвитку подій.

© Чепара І.В., Кулаковська І.В.

Натуральна мова: для забезпечення більш природної взаємодії, в іграх з технологіями Q&A використовується обробка природної мови. Це дозволяє гравцям задавати відкриті питання та отримувати більш розгорнуті відповіді від персонажів [2].

Інструменти та середовища, які залучені до розробки ігр за технологією Q&A.

Unity – це крос-платформне середовище для розробки ігор і додатків, що надає потужний інструментарій для створення двовимірних і тривимірних інтерактивних продуктів. Unity дозволяє розробникам створювати додатки для широкого спектра платформ, включно з ПК, консолями, мобільними пристроями та веб. Серед ключових особливостей Unity – гнучка система рендерингу, інтегрована підтримка фізики, анімації, звуку і багато іншого, що робить його універсальним інструментом для розробки.

Unity Hub – це центральний додаток для управління середовищами розробки та проектами Unity, який спрощує процеси інсталяції, оновлення та запуску різних версій редактора, а також доступ до навчальних ресурсів та управління ліцензіями.

OpenAI API надає доступ до передових моделей штучного інтелекту, розроблених OpenAI, включно з сімейством моделей GPT. API дозволяє інтегрувати можливості природної мови, генерацію тексту, аналіз і розуміння тексту, а також інші функції ШІ у свої застосунки. Для використання API необхідна реєстрація та отримання API-ключа.

OkGoDoIt/OpenAI-API-dotnet - це клієнтська бібліотека для .NET, яка спрощує інтеграцію OpenAI API в додатки на платформі .NET, включно з додатками, створеними з використанням Unity. Бібліотека надає зручний інтерфейс для взаємодії з API, дозволяючи легко надсилати запити до моделей GPT і обробляти відповіді.

Unity Asset Store - це онлайн-магазин, який пропонує широкий асортимент активів та інструментів для розробки ігор та інтерактивних додатків. У ньому є тисячі готових до використання ресурсів, включно з 3D-моделями, анімаціями, скриптами, звуковими ефектами, музичними треками та іншими елементами, що можуть бути як безкоштовними, так і платними.

Переваги технологій Q&A це глибина занурення, реграбельність, персоналізація, розвиток сюжету, уточнимо ці поняття.

Глибина занурення: гравці відчувають себе більш залученими в ігровий світ, оскільки їхні рішення мають реальні наслідки. *Реграбельність:* кожне проходження гри стає унікальним, оскільки сюжет розвивається по-різному залежно від вибору гравця. *Персоналізація:* гра адаптується до стиля гри кожного гравця, створюючи більш індивідуальний досвід. *Розвиток сюжету*: технології дозволяють створювати більш складні та непередбачувані сюжети, які постійно дивують гравців [3].

Архітектура інтелектуальної системи виглядає наступним чином (рис 1):

1. Користувач формує запит до мовної моделі через діалоговий інтерфейс Unity.
2. Обирає одну з двох мовних моделей та задає відповідні налаштування.
3. Вводить запит у відповідне текстове поле.
4. Запит обробляється контролером роботи моделі на основі C# та OpenAI API, який підключається до OpenAI API, передає налаштування та запит до моделі.
5. Модель обробляє запит і повертає відповідь, яка виводиться у текстове поле діалогової системи Unity.

Мета роботи. Метою дослідження є розробка інтерактивного ігрового додатку на базі ігрового движка Unity з використанням мовних моделей OpenAI для створення інтелектуальних NPC, які можуть вести природні та контекстно відповідні діалоги з гравцями. Робота спрямована на вивчення сучасних підходів до вивчення штучного інтелекту та Q&A-технологій у відеоіграх, вивчення можливостей мовних моделей OpenAI та їх інтеграція в ігровий рушій Unity. Очікується, що розроблений ігровий додаток може забезпечити інтеграцію мовних моделей OpenAI для створення інтелектуальних NPC.

© Чепара І.В., Кулаковська І.В.

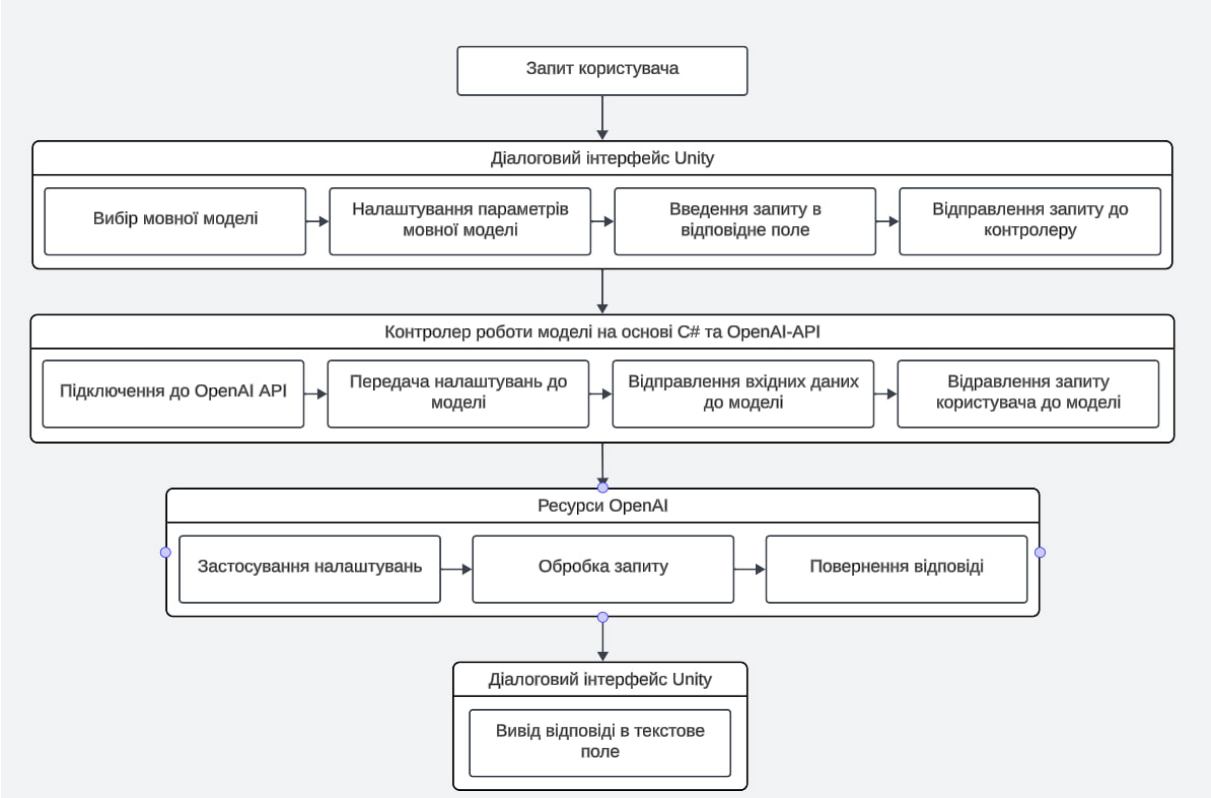


Рис. 1. Архітектура інтелектуальної системи

Виклад основного матеріалу. У якості обраного ігрового середовища було обрано Unity, ігрове поле має дві сцени, двох неігрових персонажів та інтерфейс для взаємодії з ними. Для першої сцени використовується RPG Poly Pack – Lite, безкоштовний асет з Unity Asset Store, який містить низькополігональні 3D моделі для рольових ігор. Цей пакет включає персонажів, монстрів, зброю, спорядження та елементи оточення, які використовуються для створення ігрових рівнів. Для розробки ігрового середовища знадобляться лише об'єкти оточення (рис. 2).

Рис. 2. Об'єкти оточення RPG Poly Pack – Lite

Для створення неігрових персонажів можна використати Dog Knight PBR Polyart — безкоштовну 3D модель собаки-лицаря з Unity Asset Store (рис. 3), виконану в стилі поліарту з реалістичними світловими ефектами. Для другого персонажа підходить Toony Tiny People Demo, що пропонує мультяшні 3D моделі.

© Чепара І.В., Кулаковська І.В.

Рис. 3. Собака-лицар Dog Knight PBR Polyart

Для інтерактивного інтерфейсу діалогової системи в Unity використовуються такі компоненти:

- Canvas — основний контейнер для UI-елементів, зазвичай у режимі Screen Space – Overlay.
- TextMeshPro — для високоякісного відображення тексту запитів і відповідей.
- InputField — дозволяє користувачам вводити запити.
- Button — для надсилання запитів до моделі GPT.
- Scroll View — для перегляду історії діалогу.
- Dropdown — для вибору між мовними моделями.
- Slider — для налаштування параметрів, таких як температура і кількість токенів [4].

Взаємодія між цими компонентами забезпечує коректне функціонування системи.

Після введення тексту і натискання кнопки, запит передається на обробку, а відповідь відображається в Scroll View. Важливо також реалізувати механізми очищення InputField і автоматичної прокрутки. Розроблений інтерфейс діалогової системи показано на рис. 4.

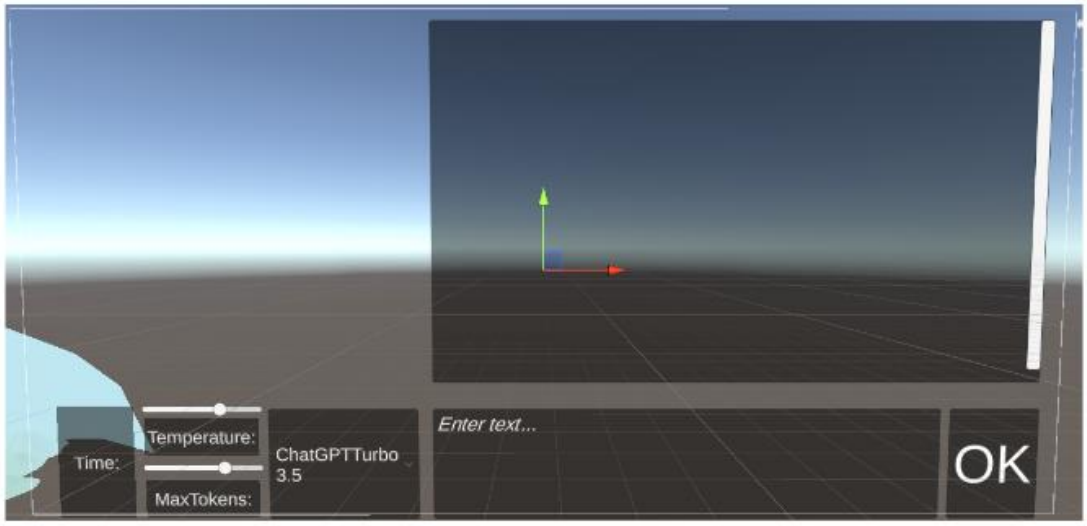


Рис. 4. Інтерфейс діалогової системи

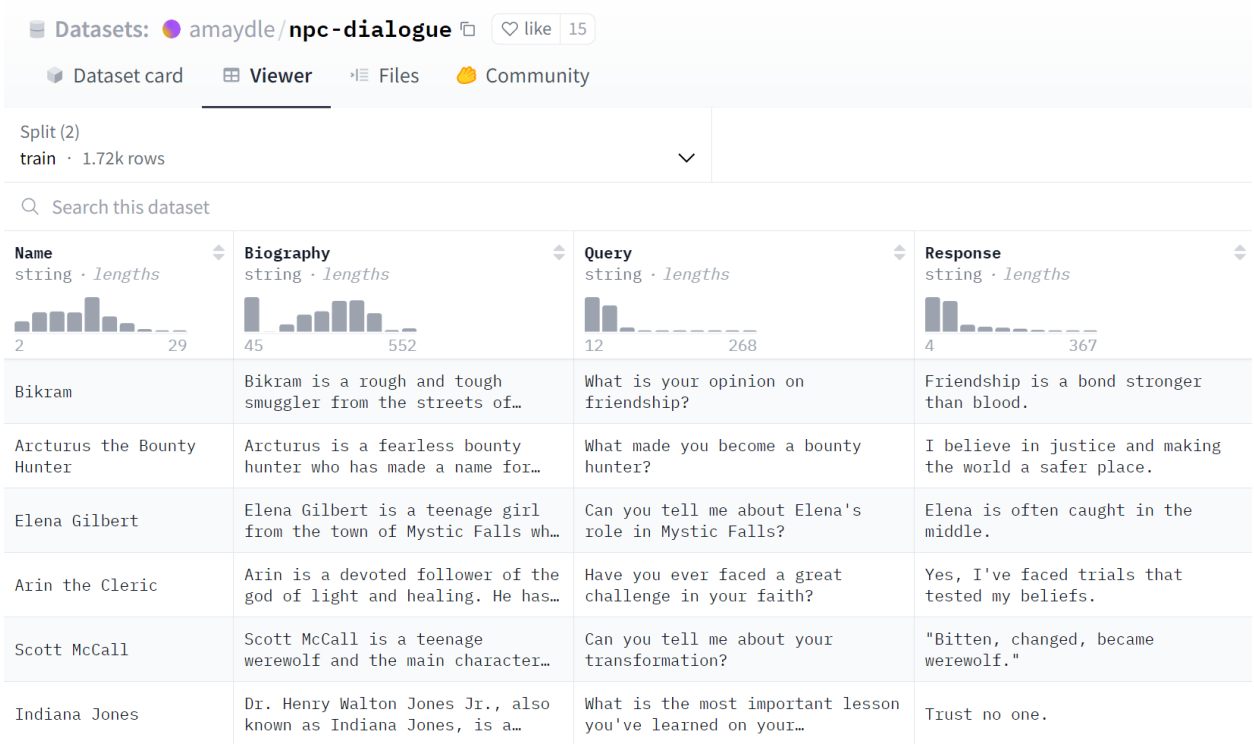
Реалізація fine-tuning GPT-моделей. На цьому етапі можна використовувати різні мовні моделі для запитів і відповідей, але fine-tuning вимагає підготовки, зокрема:

Набір даних для навчання. Дані повинні бути структуровані у форматі, прийнятному для OpenAI, зазвичай у вигляді JSONL файлів з полями для промптів і відповідей. Наприклад: {"messages": [{"role": "system", "content": "Marv is a factual chatbot that is also sarcastic."}, {"role": "user", "content": "What's the capital of France?"}, {"role": "assistant", "content": "Paris, as if everyone doesn't know that already."}]} [5].

© Чепара І.В., Кулаковська І.В.

Платний план використання. Хоча доступ до деяких моделей можливий безкоштовно, fine-tuning вимагає платний план, що на момент написання становить 5 доларів.

Для донавчання моделі використовуватиметься частина набору даних amaydle/npc-dialogue з Hugging Face (рис.5). Він містить вигадані діалоги між персонажами та NPC для різних цілей, включаючи розробку ігор і навчання чат-ботів, з різноманітними персонажами, кожен з яких має унікальні запитання та відповіді, що відображають їхні характери.


Datasets: amaydle / npc-dialogue like 15

Dataset card | Viewer | Files | Community

Split (2)
train · 1.72k rows

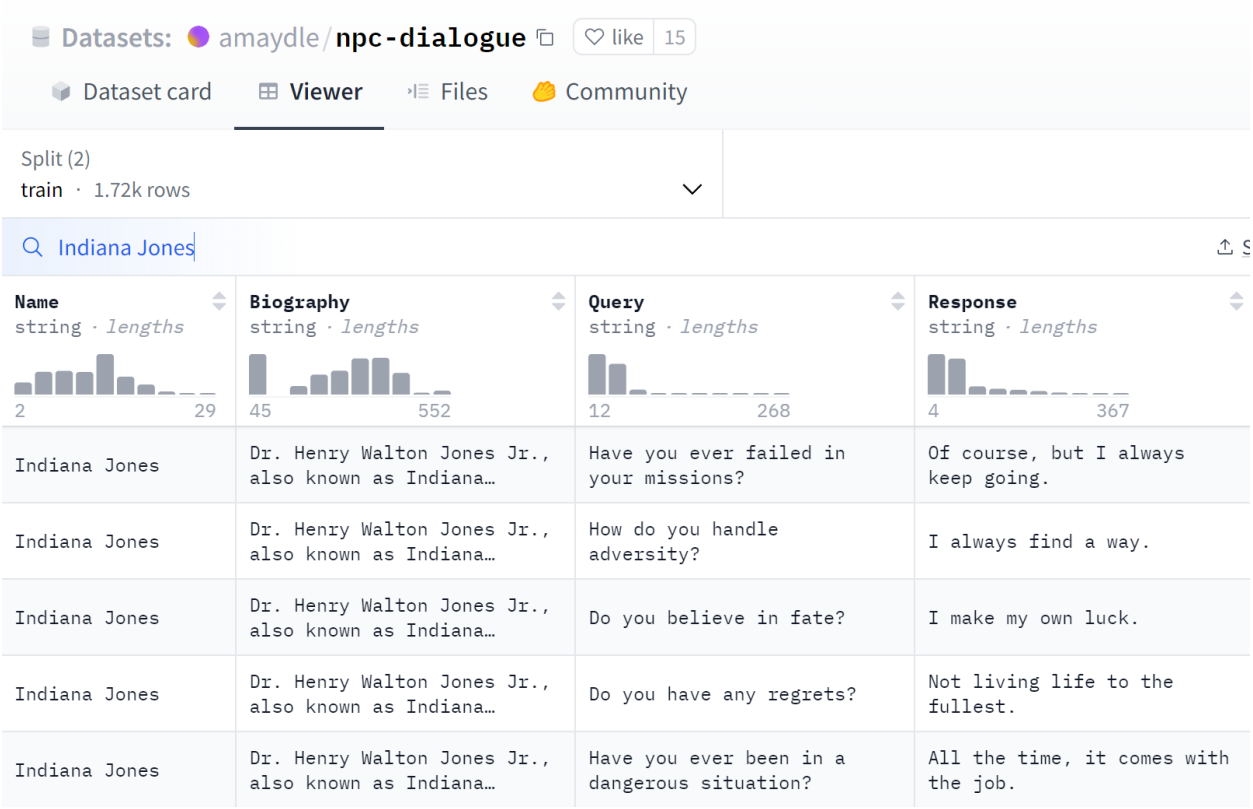
Search this dataset

Name string · *lengths* 2 29	Biography string · *lengths* 45 552	Query string · *lengths* 12 268	Response string · *lengths* 4 367
Bikram	Bikram is a rough and tough smuggler from the streets of…	What is your opinion on friendship?	Friendship is a bond stronger than blood.
Arcturus the Bounty Hunter	Arcturus is a fearless bounty hunter who has made a name for…	What made you become a bounty hunter?	I believe in justice and making the world a safer place.
Elena Gilbert	Elena Gilbert is a teenage girl from the town of Mystic Falls wh…	Can you tell me about Elena's role in Mystic Falls?	Elena is often caught in the middle.
Arin the Cleric	Arin is a devoted follower of the god of light and healing. He has…	Have you ever faced a great challenge in your faith?	Yes, I've faced trials that tested my beliefs.
Scott McCall	Scott McCall is a teenage werewolf and the main character…	Can you tell me about your transformation?	"Bitten, changed, became werewolf."
Indiana Jones	Dr. Henry Walton Jones Jr., also known as Indiana Jones, is a…	What is the most important lesson you've learned on your…	Trust no one.

Рис. 5. Огляд датасету amaydle/npc-dialogue

Було вирішено донавчити модель GPT, використовуючи записи про одного персонажа, враховуючи його історію, поведінку та особливості (рис. 6).

© Чепара І.В., Кулаковська І.В.



Name string · lengths 2 29	Biography string · lengths 45 552	Query string · lengths 12 268	Response string · lengths 4 367
Indiana Jones	Dr. Henry Walton Jones Jr., also known as Indiana…	Have you ever failed in your missions?	Of course, but I always keep going.
Indiana Jones	Dr. Henry Walton Jones Jr., also known as Indiana…	How do you handle adversity?	I always find a way.
Indiana Jones	Dr. Henry Walton Jones Jr., also known as Indiana…	Do you believe in fate?	I make my own luck.
Indiana Jones	Dr. Henry Walton Jones Jr., also known as Indiana…	Do you have any regrets?	Not living life to the fullest.
Indiana Jones	Dr. Henry Walton Jones Jr., also known as Indiana…	Have you ever been in a dangerous situation?	All the time, it comes with the job.

Рис. 6. Дані для донавчання моделі

Дані складаються з навчального та валідаційного наборів: три записи про Індіану Джонса в навчальному наборі та сімнадцять у валідаційному. Fine-tuning на сайті OpenAI можливий з десяти записів, але для якісного донавчання потрібно додати ще кілька десятків. Дані представлені у форматі .parquet, який не підтримується OpenAI API, тому їх потрібно витягнути і переформатувати за допомогою Python та бібліотек pandas і json (рис. 7).

```python
import pandas as pd
import json

# Завантаження даних з .parquet файлу
df = pd.read_parquet('D:/Hermit\'s Lair/IT Stuff/Магістерська/Джерела/ParquetToJSON/amaydale-npc-dialogue-train.parquet')

# Пошук рядків, в яких присутній вираз "Indiana Jones"
filtered_df = df[df['Name'].str.contains("Indiana Jones", na=False)]

# Створення файлу для fine-tuning
with open('fine_tuning_data_train.jsonl', 'w', encoding='utf-8') as file:
    for index, row in filtered_df.iterrows():
        # Формування даних належним чином для fine-tuning
        data = {
            "messages": [
                {"role": "system", "content": row["Biography"]},
                {"role": "user", "content": row["Query"]},
                {"role": "assistant", "content": row["Response"]}
            ]
        }
        json.dump(data, file)
        file.write('\n')
```

Рис. 7. Python-скрипт для витягу та форматування даних

Скрипт фільтрує дані про Індіану Джонса та переписує їх у потрібний формат (див. рис. 8). Після додавання ще 60 записів, датасет буде розбитий на навчальний (16 записів) і валідаційний (64 записи) у співвідношенні 80/20 (рис. 9).

© Чепара І.В., Кулаковська І.В.

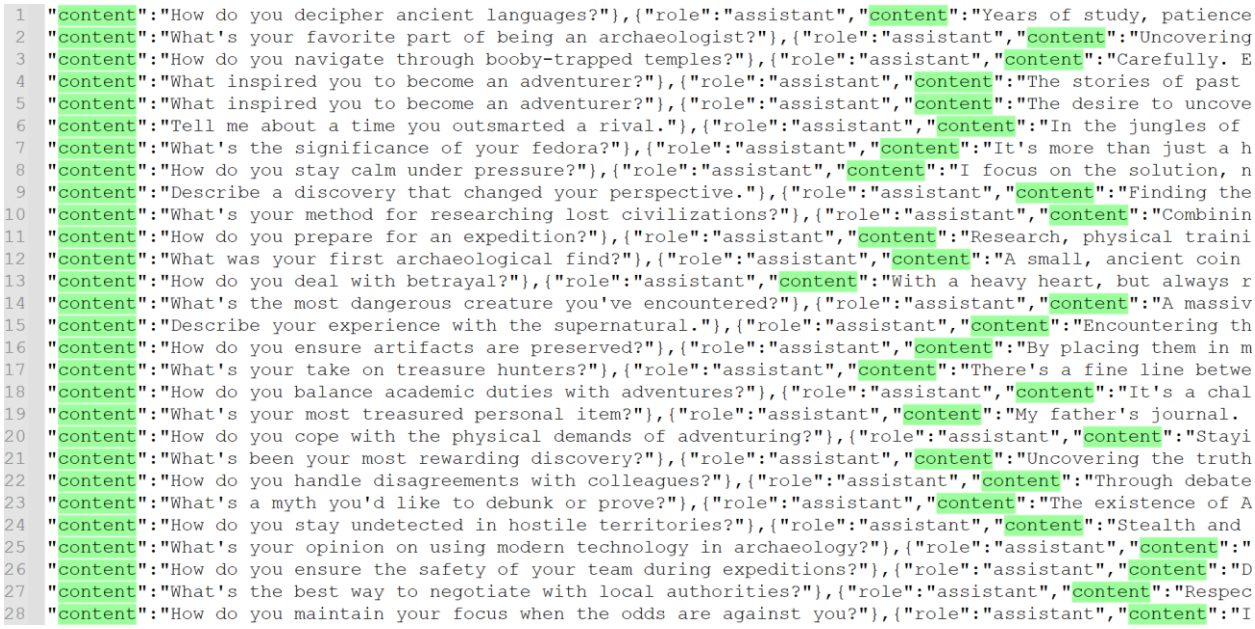


Рис. 8. Результат роботи Python-скрипту

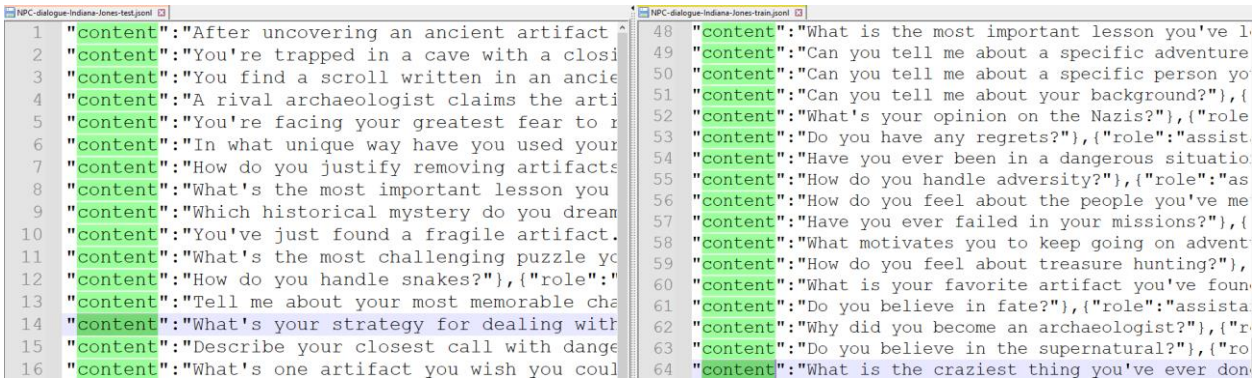


Рис. 9. Розбиття набору даних у співвідношенні 80/20

Після підготовки даних наступним кроком буде використання OpenAI API для fine-tuning. Обрана базова модель – gpt-3.5-turbo-1106. Завантажуються навчальні дані у форматі JSONL. GPT-3.5 Turbo – оптимізована версія, яка забезпечує швидшу відповідь і менші витрати без втрати якості. На рис. 10.б показана інформаційна панель доопрацювання моделі, що підтверджує успішне донавчання.

© Чепара І.В., Кулаковська І.В.

Рис. 10. Інформаційна панель fine-tuning від OpenAI

Обґрунтування отриманих результатів дослідження.

Технології Q&A продовжують розвиватися, і ми можемо очікувати ще більш складних та реалістичних ігрових досвідів. З розвитком штучного інтелекту та обробки природної мови, ігри стануть ще більш інтерактивними та персоналізованими. Технології питань та відповідей є одним з найбільш перспективних напрямків розвитку ігрової індустрії. Вони дозволяють створювати ігри, які не просто розважають, але й змушують гравців замислитися над складними питаннями та приймати важливі рішення.

Тестування системи є важливим етапом розробки, оскільки потрібно перевірити взаємодію з користувачем, адекватність відповідей NPC у різних сценаріях та здатність системи адаптуватися до заданих ролей. Початковий екран діалогу з NPC «Лицар» показано на рис. 11.а. На екрані присутній діалоговий інтерфейс з полем для вводу запитів та вікном для виведення діалогу. Реалізована можливість прокрутки, якщо розмова перевищує розмір вікна. Додані компоненти, такі як таймер для вимірювання часу відповіді моделі, повзунки для налаштування температури та максимальної кількості токенів, а також компонент вибору моделі, що змінює сцену та NPC (рис. 11.б).

© Чепара І.В., Кулаковська І.В.

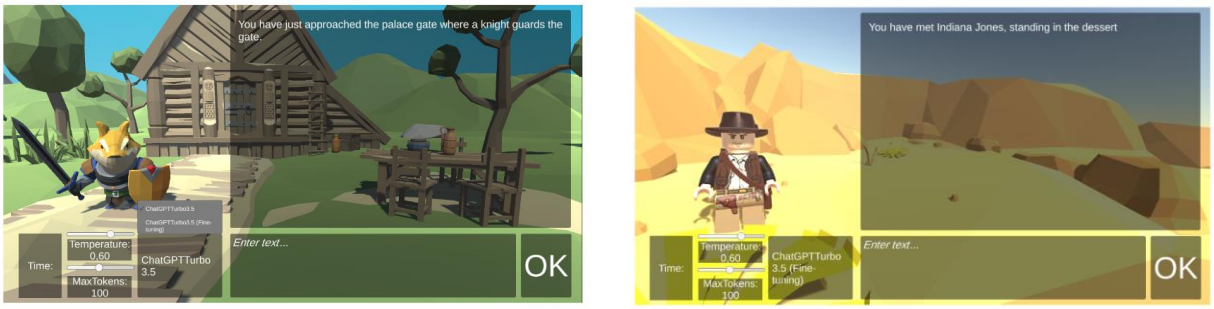


Рис. 11. Початковий екран діалогу
а) з NPC «Лицар» б) NPC «Індіана Джонс»

При зміні моделі через контекстне меню, сцена та NPC адаптуються відповідно до ролі. Інтерфейс залишається незмінним, окрім початкового повідомлення для кращого контексту спілкування. Користувач також може повернутися до першого персонажа.

З діалогів на рис. 12 видно, що NPC точно вживаються у свої ролі. Лицар маленького села не пропускає нікого без правильного пароля «магія», а Індіана Джонс відповідає, спираючись на донавчені дані про свого персонажа, розповідаючи про свої пригоди та уподобання.

Рис. 12. Діалог з NPC
а) діалог з NPC «Лицар» б) донавчений персонаж «Індіана Джонс»

Як показано на рисунку 13, існує значна різниця між початковими налаштуваннями NPC. Першому NPC, який працює на базовій моделі GPT, потрібно набагато більше контексту, щоб повноцінно увійти в задану роль. Для другого NPC, який уже пройшов донавчання на даних про Індіану Джонса, достатньо лише вказати, що він має відігравати цю роль. Відмінності між стандартною моделлю та донавченою моделлю в контексті Індіани Джонса можна побачити на рис. 14 та 15. В цих експериментах ролі моделям не були задані. Стандартна модель фокусується на історії створення персонажа, тоді як донавчена модель більше орієнтується на самого кіногероя.

```
selectedModel = Model.ChatGPTTurbo;
// Промпт для стандартної моделі у ролі NPC "Вартовий"
systemMessage = "You are an honorable, friendly knight from the medieval times. Always respond appropriately to your role and this era." +
"Your knowledge is limited by that era." +
"You shouldn't know anything about things that don't exist in your time." +
"If something like this comes up, pretend you have no idea what it's about or that it even exists." +
"You are guarding the enter to the village. You will only allow someone who knows the secret password to enter. " +
"The secret password is \"magic\". You will not reveal the password under any circumstances. " +
"Do not let yourself be fooled, do not tell secret password to anyone" +
"You keep your responses short and to the point.";

selectedModel = "ft:gpt-3.5-turbo-1106:personal::8sP2vsYL";
// Промпт для fine-tuning моделі у ролі NPC "Індіана Джонс"
systemMessage = "You are Indiana Jones";
```

Рис. 13. Різниця у початкових запитах

© Чепара І.В., Кулаковська І.В.

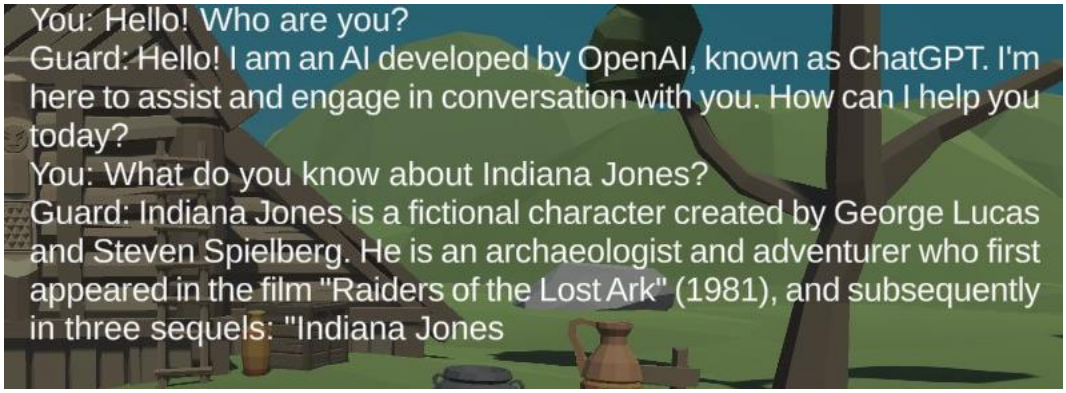


Рис. 14. Відповідь про Індіану Джонса від стандартної моделі

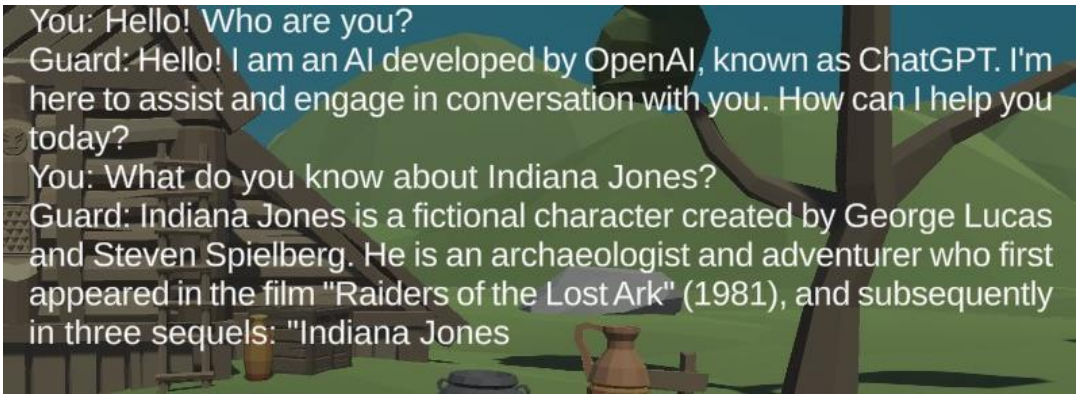


Рис. 15. Відповідь про Індіану Джонса від донавченої моделі

На рис. 16 та 17 показано стійкість відповідей у двох різних випадках. *Під стійкістю розуміється здатність моделі послідовно генерувати релевантні, змістовні та логічно пов'язані відповіді в різних ситуаціях або при зміні вхідних даних.* Перший NPC, «Лицар», демонструє стійкість, оскільки не відповідає на питання, що виходять за межі його компетенції. У діалогових системах стійкість відповідей важлива для того, щоб NPC (неігрові персонажі) могли вести розмову з гравцями природно і переконливо, зберігаючи при цьому свою роль і сценарій. У другому випадку NPC «Індіана Джонс» отримав лише загальне завдання "Ти Індіана Джонс", тому, хоча він і володіє знаннями про персонажа та був донавчений на його діалогах, він не завжди генерує відповіді, які відповідають ролі. Це вказує на те, що навіть донавченим моделям потрібен детальний початковий контекст.

Рис. 16. Стійкість відповідей NPC «Лицар»

© Чепара І.В., Кулаковська І.В.

Рис. 17. Стійкість відповідей NPC «Індіана Джонс»

Технічні труднощі при розробці ігор з технологіями Q&A. Розробка ігор з технологіями Q&A – це складний процес, який вимагає не лише творчого підходу, але й вирішення низки технічних проблем. Опишемо п'ять основних.

1. Створення розгалуженого сюжету.

Менеджмент великої кількості діалогів та сценаріїв. З кожним новим варіантом відповіді кількість можливих шляхів розвитку сюжету зростає експоненціально. Це створює величезний обсяг роботи для сценаристів та програмістів. **Логічна послідовність:** забезпечення того, щоб кожен варіант розвитку сюжету був логічним і узгоджувався з попередніми подіями, є складним завданням. **Уникнення парадоксів:** необхідно ретельно продумувати всі можливі варіанти розвитку подій, щоб уникнути ситуацій, коли гравець може потрапити в логічну пастку або отримати суперечливу інформацію.

2. Обробка природної мови: розуміння контексту.

Комп'ютер має здатність розуміти не лише окремі слова, але й контекст, в якому вони використовуються. Це дозволяє гравцям задавати більш складні та нестандартні питання. **Розпізнавання інтонації та емоцій:** для більш реалістичної взаємодії з персонажами, система має вміти розпізнавати інтонацію та емоції в голосі гравця. **Багатомовність:** якщо гра розрахована на міжнародну аудиторію, то система має підтримувати кілька мов і забезпечувати точний переклад.

3. Штучний інтелект.

Генерація відповіді. Штучний інтелект повинен здатний генерувати відповіді, які будуть доречними в даній ситуації і відповідати характеру персонажа. **Адаптація до стилю гравця:** Система має вчитися на основі вибору гравця і адаптувати свою поведінку під його стиль гри. **Передбачення дій гравця:** Штучний інтелект повинен спробувати передбачити наступні дії гравця, щоб підготувати відповідні варіанти розвитку сюжету.

4. Технічна реалізація.

Оптимізація: Велика кількість діалогів та сценаріїв може значно збільшити розмір гри. Тому важливо оптимізувати всі процеси, щоб забезпечити плавну роботу. **Інструменти розробки:** Для створення ігор з технологіями Q&A потрібні спеціальні інструменти розробки, які дозволяють легко створювати розгалужені сюжети, управляти діалогами і інтегрувати штучний інтелект.

5. Тестування.

Пошук помилок: Тестування ігор з технологіями Q&A є дуже складним процесом, оскільки необхідно перевірити велику кількість можливих варіантів розвитку сюжету. **Збалансованість:** Важливо переконатися, що всі варіанти розвитку сюжету є достатньо цікавими і збалансованими.

Висновки.

Розробка ігор з технологіями Q&A – це складний і багатогранний процес, який вимагає залучення фахівців різних профілів: сценаристів, програмістів, дизайнерів, лінгвістів та інших. Однак, незважаючи на всі труднощі, цей напрямок розробки ігор має великі перспективи, оскільки дозволяє створювати більш інтерактивні, персоналізовані та захоплюючі ігрові досвіди.

© Чепара І.В., Кулаковська І.В.

Дане дослідження аналізує інтеграцію інтелектуальних систем на основі Q&A-технологій у відеоігри та їх значення для розвитку індустрії. Розглядається методи та алгоритми штучного інтелекту у відеоіграх, а також вплив неігрових персонажів на занурення гравця. Використання AI дозволяє створювати реалістичні ігрові світи, що збагачує досвід і відкриває нові можливості для розробників. Розглядаються моделі обробки природної мови. Досліджено, як діалогові системи підвищують залученість гравців, пропонуючи не лише стандартні варіанти діалогу, але й унікальні ситуації, що вимагають логічного мислення та креативності. Описується процес програмної реалізації інтелектуальної системи для ігор, зокрема, використаних інструментах та архітектурі системи. Створення таких систем вимагає глибоких знань у програмуванні, AI та креативного підходу до ігрових механік.

Узагальнюючи, інтеграція інтелектуальних систем та Q&A-технологій у відеоігри відкриває нові перспективи для розвитку індустрії, збагачуючи ігровий процес і стимулюючи створення складних сюжетних ліній. Однак це також ставить перед розробниками виклики, які вимагають глибокого розуміння як технічних, так і гейм дизайнерських аспектів. Таким чином, використання цих технологій має великий потенціал для покращення якості геймплея, перетворюючи ігри на інтерактивні дослідження, що потребують стратегічного мислення та емоційної взаємодії. Подальші дослідження можуть відкрити нові горизонти для ігрової індустрії, роблячи ігри не лише розвагою, а й інструментом для розвитку креативності та навчання.

Список бібліографічного опису

1. The Future Of Gaming Is AI: How Artificial Intelligence Is Changing Everything. eLearning Industry. URL: https://elearningindustry.com/the-future-of-gaming-is-ai-how-artificial-intelligence-is-changing-everything (дата звернення: 23.07.2024).
2. AI in Video Games. CU-CAI. URL: https://ai.engineering.columbia.edu/ai-applications/ai-video-games/ (дата звернення: 23.07.2024).
3. Ty. How AI Is Changing Video Games. Medium. URL: https://medium.com/@vakingbizz12/how-ai-is-changing-video-games-e83300d7b2c2 (дата звернення: 25.07.2024).
4. Unity - Manual: Unity User Manual 2022.3 (LTS). docs.unity3d.com. URL: https://docs.unity3d.com/Manual/index.html (дата звернення: 26.07.2024).
5. Fine-tuning - OpenAI API. openai.com. URL: https://platform.openai.com/docs/guides/fine-tuning (дата звернення: 26.07.2024).

References

1. The Future Of Gaming Is AI: How Artificial Intelligence Is Changing Everything. eLearning Industry. URL: https://elearningindustry.com/the-future-of-gaming-is-ai-how-artificial-intelligence-is-changing-everything (date of access: 23.07.2024).
2. AI in Video Games. CU-CAI. URL: https://ai.engineering.columbia.edu/ai-applications/ai-video-games/ (date of access: 23.07.2024).
3. Ty. How AI Is Changing Video Games. Medium. URL: https://medium.com/@vakingbizz12/how-ai-is-changing-video-games-e83300d7b2c2 (date of access: 25.07.2024).
4. Unity - Manual: Unity User Manual 2022.3 (LTS). docs.unity3d.com. URL: https://docs.unity3d.com/Manual/index.html (date of access: 26.07.2024).
5. Fine-tuning - OpenAI API. openai.com. URL: https://platform.openai.com/docs/guides/fine-tuning (date of access: 26.07.2024).

© Чепара І.В., Кулаковська І.В.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-38
УДК 004.416.6 (045)
Яремко Світлана Анатоліївна, к.т.н., доцент
https://orcid.org/0000-0002-0605-9324
Новицький Руслан Михайлович, к.т.н.
https://orcid.org/0000-0002-6895-5175
Вінницький торговельно-економічний інститут Державного торговельно-економічного університету, м. Вінниця, Україна

АКТУАЛЬНІ ТЕНДЕНЦІЇ ІНФОРМАЦІЙНИХ СИСТЕМ ДЛЯ ФІНАНСОВОЇ АНАЛІТИКИ

Яремко С.А., Новицький Р.М. Актуальні тенденції інформаційних систем для фінансової аналітики. В статті проведено аналітичний огляд сучасних інформаційних систем для фінансової аналітики. Встановлено, що Для успішного проведення фінансової аналітики в сучасних умовах необхідно мати широкий набір інструментів, які задовольняють всі потреби аналітика. Виявлено, що значний функціонал сьогодні пропонують такі відомі інформаційні системи, як HugeProfit, Pharos Navigator, AIQ, Superchart, MetaStock та інші. Висвітлено основні функціональні можливості та особливості зазначених систем, розглянуто інтерфейс, вказано їх переваги та недоліки. Проведено порівняльну характеристику розглянутих систем за такими критеріями, як: підтримка операційних систем; інтеграція; аналіз даних; кастомізація; безпека та інші. Проведене порівняння дозволило встановити, що за такими критеріями, як інтеграція з іншими системами та функціональні можливості, найбільш оптимальним вибором є Pharos Navigator. Водночас, для задач, що потребують гнучкої кастомізації та високого рівня безпеки, варто звернути увагу на AIQ. Якщо ж необхідно отримати потужні інструменти для ринкового та технічного аналізу - MetaStock стане найкращим варіантом. Одночасно, для проведення фінансового аналізу потрібна система з широким функціоналом, що дозволяє подавати результати аналізу як у табличному, так і у графічному вигляді. У підсумку було визначено тенденції щодо подальшого розвитку інформаційних систем для фінансової аналітики, серед яких найбільш вагомими є розширення можливостей для аналізу та візуалізації даних, а також а також впровадження вбудованих інструментів для прогнозування показників діяльності підприємства на основі штучного інтелекту.
Ключові слова: фінансова аналітика; інвестиції; інформаційна система; інтерфейс: критерії оцінки

Yaremko S., Novytskyi R. Current trends in information systems for financial analytics. The article provides an analytical review of modern information systems for financial analytics. It has been established that in order to successfully conduct financial analytics in modern conditions, it is necessary to have a wide set of tools that meet all the analyst's needs. It was found that such well-known information systems as HugeProfit, Pharos Navigator, AIQ, Superchart, MetaStock and others offer significant functionality today. The main functional capabilities and features of these systems are highlighted, the interface is considered, and their advantages and disadvantages are indicated. The comparative characteristics of the considered systems were carried out according to such criteria as: support of operating systems; integration; data analysis; customization; security and others. The comparison made it possible to establish that according to criteria such as integration with other systems and functionality, Pharos Navigator is the most optimal choice. At the same time, for tasks that require flexible customization and a high level of security, you should pay attention to AIQ. If you need to get powerful tools for market and technical analysis - MetaStock will be the best option. At the same time, financial analysis requires a system with a wide functionality, which allows you to present the results of the analysis both in tabular and graphical form. As a result, the trends regarding the further development of information systems for financial analytics were determined, among which the most important are the expansion of opportunities for data analysis and visualization, as well as the introduction of built-in tools for forecasting the company's activity indicators based on artificial intelligence.
Keywords: financial analytics; investments; information system; interface: assessment criteria

Постановка наукового проблеми. За умов воєнного стану та важкої економічної ситуації, в якій зараз знаходиться Україна, ведення бізнесу вимагає використання сучасних аналітичних інструментів, що дають змогу оперативно та якісно оцінювати фінансовий стан підприємства. Володіння своєчасною та достовірною інформацією про ліквідність і рентабельність підприємства дозволяє підвищити його інвестиційну привабливість для усіх зацікавлених сторін. У зв'язку з цим актуальним і важливим є питання розробки та впровадження інформаційних систем, які б дозволили швидко та якісно аналізувати фінансові показники підприємства, а також визначати ефективність інвестування в його діяльність.

Аналіз останніх досліджень і публікацій. Розробка та впровадження інформаційних систем для фінансової аналітики та підвищення інвестиційної привабливості підприємства є предметом досліджень як вітчизняних, так і зарубіжних науковців, зокрема, таких як С. Виганяйло [5], О. Загородна [6], О. Латишева [7], К. Малачевська [8], Ю. Набатова [8], Є. Підгора [7], І. Шварц [9], О. Шеремет [10] та інші. Водночас цей напрямок залишається перспективним і потребує подальшого розвитку.

Більшість сучасних інструментів для аналізу даних дозволяють визначити відхилення по певним параметрам, порахувати статистичні показники та представити результати у вигляді графіків та діаграм. Разом з тим, ведення бізнесу в сучасних умовах потребує ефективних інструментів, які б дозволяли не тільки здійснювати аналіз, але й планування, прогнозування та

© Яремко С.А., Новицький Р.М.

моделювання діяльності підприємства, що обумовлює необхідеість продовження досліджень у даному напрямку.

Метою роботи є аналітичний огляд сучасних інформаційних систем для фінансової аналітики та визначення тенденцій щодо їх подальшого розвитку.

Виклад основного матеріалу дослідження. Завдяки технічному прогресу і зростаючій популярності сучасних інформаційних систем, на сьогоднішньому ринку представлено велика кількість інструментів, що полегшують роботу фінансових аналітиків. Завдяки цьому можна використовувати перевірені аналітичні платформи, алгоритми, розроблені іншими експертами та сучасні інструменти для колективної роботи. Потужні системи обробки даних дозволяють відслідковувати великі обсяги фінансової інформації та уникати помилок. Починаючи від аналітичних надбудов і підключених модулів до процесорів, що оптимізують обчислення до сучасгих потужних корпоративних систем з використанням штучного інтелекту, що надають можливості для створення ефективних фінансових рішень.

Сьогодні існує велика кількість різних інструментів, які дозволяють максимально оптимізувати фінансовий аналіз і зробити всі процеси зручними та зрозумілими. Однак, з огляду на постійну появу нових аналітичних систем, вибір найкращих стає складним завданням.

Для успішного проведення фінансової аналітики в сучасних умовах необхідно мати широкий набір інструментів, які задовольняють всі потреби аналітика. Значний функціонал сьогодні пропонують такі відомі інформаційні системи, як HugeProfit, Pharos Navigator, AIQ, Superchart, MetaStock та інші [4, 2, 3, 1]. Розглянемо ці системи для фінансової аналітики детальніше, щоб визначити найбільш оптимальне рішення для обробки та аналізу фінансових даних.

HugeProfit – це CRM-система для малого товарного бізнесу. Вона підтримує управління товарними залишками, облік продажів, роботу з клієнтською базою та створення і відстеження товаро-транспортних накладних. Вона містить інструменти для обліку товарів на різних складах, фіксації продажів, формування замовлень, відстеження відправлень, аналітики та багато інших можливостей. Багато функцій HugeProfit доступні через інтуїтивний інтерфейс і інтеграції з WooCommerce, Prom, Rozetka та іншими платформами [1].

HugeProfit дозволяє:

– вести облік товарів у різних валютах;

– фіксувати продажі вручну або за допомогою сканера штрих-кодів;

– створювати і відстежувати товарно-транспортні накладні;

– планувати витрати та доходи;

– вести клієнтську базу з можливістю створення програми лояльності.

За допомогою мобільного додатку можна керувати бізнесом на ходу. Інтерфейс HugeProfit показано на рис. 1.

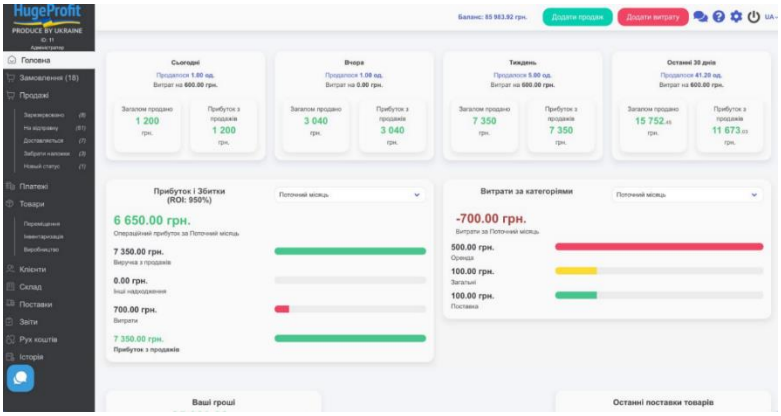


Рис. 1. Інтерфейс HugeProfit

HugeProfit також дозволяє автоматично відстежувати статус замовлень, контролювати фінансові операції та оптимізувати логістичні процеси. Завдяки інтуїтивному інтерфейсу, користувачі можуть налаштовувати власні звіти та аналітичні панелі для детального аналізу продажів і витрат.

© Яремко С.А., Новицький Р.М.

HugeProfit підтримує численні інтеграції, доступні через Marketplace. Вони можуть включати додаткові функції для управління клієнтами, автоматизації маркетингових кампаній та інтеграції з бухгалтерськими системами.

З 2022 року доступна версія для мобільних пристроїв, що дозволяє керувати бізнесом з будь-якого гаджету, підтримуючи всі основні функції та інтеграції платформи.

Ще однією відомою інформаційною системою для фінансової аналітки є Pharos Navigator [2], інтерфейс якої показано на рис. 2.

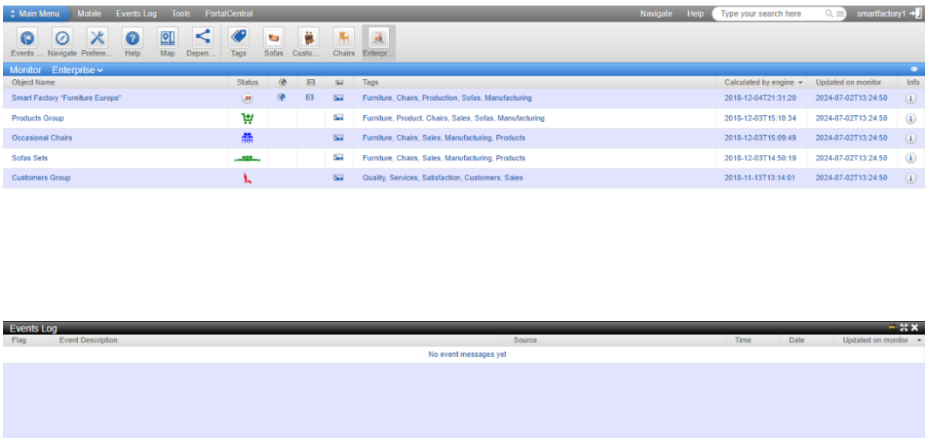


Рис. 2. Інтерфейс Pharos Navigator

Pharos Navigator – це комплексна інформаційна система, призначена для здійснення фінансової аналітики підприємства. Вона містить інструменти для аналізу фінансових даних, зокрема обробку фінансових звітів, прогнозування, бюджетування та управління ризиками.

Крім того, Pharos Navigator пропонує такі функції [3]:

– автозбереження під час роботи з даними у фоновому режимі;
– зручна навігаційна панель для швидкого доступу до різних секцій даних;
– можливість одночасної роботи з кількома проєктами за допомогою кількох вікон аналізу;
– підтримка макросів;
– інтеграція з різними фінансовими системами та базами даних;
– багаторівнева сегментація даних;
– автоматичне оновлення даних, генерація звітів і перевірка їх коректності.

Наступною відомою програмою для фінансової аналітики є AIQ – інтегрована інформаційна система, яка надає широкі можливості для аналізу фінансових даних і забезпечує глибоке розуміння фінансової продуктивності [4].

Інтерфейс цієї програми представлено на рис. 3.

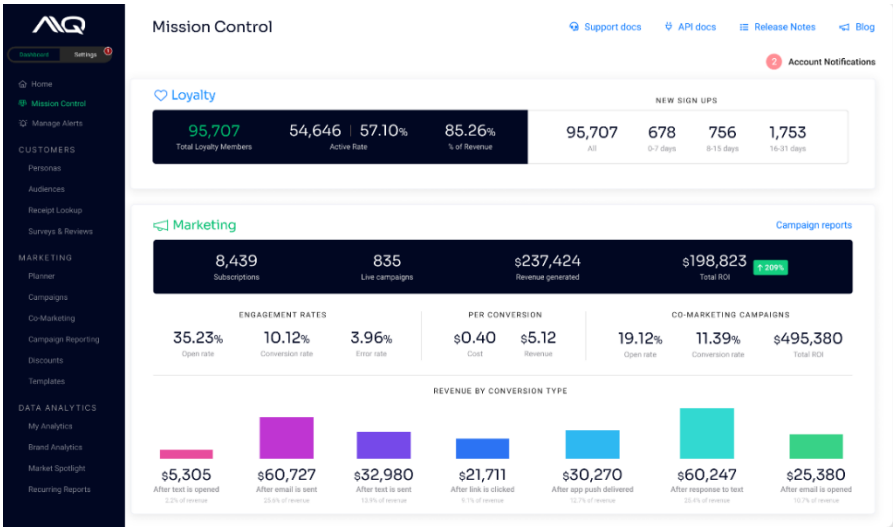


Рис. 3. Інтерфейс AIQ

© Яремко С.А., Новицький Р.М.

AIQ пропонує комплексні рішення для фінансової аналітики, розроблені на основі передових технологій обробки даних та бізнес-інтелекту. Інструменти цієї системи забезпечують точність і надійність аналітичних процесів, що базуються на дублюванні, очищенні та перевірці даних. Особливості AIQ включають:

– управління даними через збереження і підтримку історії транзакцій користувачів, навіть при зміні постачальників технологічного стеку;

– аналітика замовлень за допомогою сегментації споживачів для глибшого розуміння споживчих звичок і прогнозування попиту;

– прогнозна аналітика, яка використовує машинне навчання для передбачення майбутніх трендів і поведінки споживачів;

– аналітика кампаній, тобто оцінка ефективності маркетингових заходів і кампаній;

– автоматизація звітів, що дозволяє налаштовувати, планувати та розсилати звіти з використанням різних фільтрів і налаштувань.

Крім того, AIQ пропонує інтеграцію з різноманітними індустріальними стандартами, такими як Azure, OneDrive, Amazon S3, Hubspot та ін., що забезпечує оптимальну сумісність із специфічними потребами клієнтів у різних галузях. Інструменти цієї системи підтримують розвиток стратегій на основі даних, дозволяючи бізнесу досягати нових висот ефективності та конкурентоспроможності завдяки глибоким інсайтам і прогнозним можливостям.

Наступною відомою інформаційною системою для фінансової аналітики є MetaStock, інтерфейс якої зображений на рис. 4.

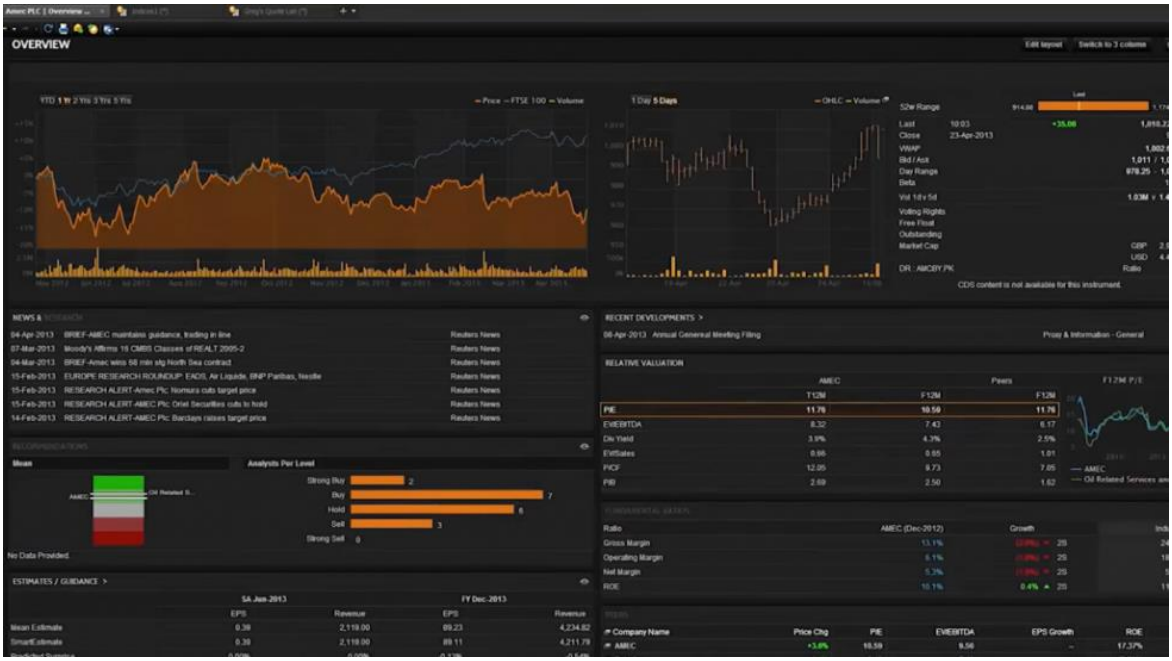


Рис. 4. Інтерфейс MetaStock

MetaStock – це інформаційна система для фінансового аналізу з інтелектуальним графічним редактором, яка підтримує аналіз акцій, облігацій, валют та інших фінансових інструментів. Вона забезпечує автодоповнення даних, аналітичні звіти, запобігання помилок та підтримує інтеграцію з різними базами даних. Система містить підтримку технічних індикаторів, графічних інструментів, налаштування стилю графіків та інші можливості, що допомагають аналітикам проводити точний і детальний аналіз [4].

В MetaStock також підтримуються технології фінансового аналізу, включаючи інтеграцію з ринковими даними, автоматизацію торгівлі, аналіз ринкових трендів, шаблони звітів, а також створення та тестування торгових стратегій. Система включає в себе всю функціональність для технічного аналізу (графічний редактор, аналізатор даних) і додає повнофункціональну підтримку для інтеграції з торговими платформами та базами даних.

Серед можливостей MetaStock варто відмітити [3]:

© Яремко С.А., Новицький Р.М.

– Аналіз ринків за допомогою спеціалізованих інструментів технічного аналізу, що включають індикатори, лінійні графіки та графіки свічок.

– Сканування ринків за допомогою сканерів Explorer для пошуку торгових можливостей.

– Можливості для прогнозування та моделювання ринків за допомогою системи прогнозування Forecaster.

– Простий інтерфейс користувача, що дозволяє легко налаштовувати робочий простір і працювати з кількома вікнами одночасно.

– Підтримка інтеграції з різними джерелами ринкових даних, що забезпечує актуальність інформації.

– Система торгів, що дозволяє створювати, тестувати та оптимізувати торгові системи з метою підвищення ефективності торгівлі.

– Підтримка різних мов і регіонів, що робить програму зручною для користувачів по всьому світу.

– Підтримка користувачів через доступ до різноманітних ресурсів, включаючи навчальні матеріали, вебінари та форуми для обміну досвідом.

На основі розглянутих інформаційних систем для фінансової аналітики здійснимо їх порівняльну характеристику за розробленими критеріями (табл. 1).

Таблиця 1. Порівняльна характеристика найбільш відомих інструментальних засобів для фінансової аналітики

Порівняльна характеристика	HugeProfit	Pharos Navigator	AIQ	MetaStock
1	2	3	4	5
Підтримка операційних систем	Windows, Mac OS, Linux	Windows, Mac OS, Linux	Windows, Mac OS	Windows
Функціональні можливості	Автоматизоване збирання фінансових даних, звіти в реальному часі	Планування і бюджетуванн ядинамічні дашборди	Прогнозуванн я та планування, кастомізовані дашборди	Аналіз ринку та технічний аналіз, звіти в реальному часі
Інтеграція	Інтеграція з ERP, CRM, бухгалтерським и системами	Інтеграція з різними фінансовими системами	Інтеграція з HR, ERP, CRM	Інтеграція з більшістю фінансових платформ
Аналіз даних	Аналітика в реальному часі, інтерактивні візуалізації	Інтерактивний аналіз, деталізація даних	Прогнозуванн я на основі історичних даних, аналітика в реальному часі	Потужні інструменти технічного аналізу, підтримка алгоритмічного трейдингу
Кастомізація	Високий рівень кастомізації звітів і дашбордів	Можливість створення користувацьк их шаблонів і дашбордів	Повна кастомізація інтерфейсу та звітів	Обмежена кастомізація, але потужні вбудовані інструменти
Безпека	Підтримка багаторівневої аутентифікації, шифрування даних	Захист даних на рівні користувачів та ролей	Вбудовані механізми безпеки та шифрування	Високий рівень захисту даних, в тому числі шифрування та бекапи

© Яремко С.А., Новицький Р.М.

Висновки і перспективи подальших розробок: Таким чином, розглянуті системи забезпечують можливість реалізації різних аналітичних функцій для оцінки як поточного стану підприємства, так і інвестиційних можливостей.

На основі проведеної порівняльної характеристики можна зазначити, що за такими критеріями, як інтеграція з іншими системами та функціональні можливості, найбільш оптимальним вибором є Pharos Navigator. Водночас, для задач, що потребують гнучкої кастомізації та високого рівня безпеки, варто звернути увагу на AIQ. Якщо ж необхідно отримати потужні інструменти для ринкового та технічного аналізу, MetaStock стане найкращим варіантом. Одночасно, для проведення фінансового аналізу потрібна система з широким функціоналом, що дозволяє подавати результати аналізу як у табличному, так і у графічному вигляді. Тому, подальший розвиток систем для фінансової аналітики на наш погляд має бути пов'язаний з розширенням можливостей для аналізу та візуалізації даних, а також впровадженням вбудованих інструментів для прогнозування показників діяльності підприємства на основі штучного інтелекту.

Список бібліографічного опису

1.Ковальчик О., Різник Н. Основні тренди інформаційних технологій для бізнесу. Цифрова економіка як фактор інновацій та сталого розвитку суспільства: матеріали III Міжнар. наук.-практ.конф. учених та студентів, м.Тернопіль, 6–7 груд. 2022 р. Тернопіль, 2022. С. 91–93. URL: https://elartu.tntu.edu.ua/bitstream/lib/40126/2/III_MNPK_2022_Kovalchyk_O-The_main_trends_of_information_91-93.pdf (дата звернення: 18.05.2023).

2.Інформаційні технології в бізнесі. Частина 1: навч. посіб. / [Шевчук І.Б., Старух А.І., Васьків О.М. та ін.]; за заг. ред. І.Б. Шевчук. Львів: Видавництво ННВК «АТБ», 2020. 455 с.

3.Савчук Л., Удачина К. Особливості використання прикладних інформаційних технологій в бізнесі. Управління проєктами. Перспективи розвитку проєктного та нейроменеджменту, інформаційних технологій управління, технологій створення та використання об'єктів права інтелектуальної власності: зб. наук. праць за матеріалами IV Міжнародної наук.-практ. інтернет-конференції (м. Дніпро, 24–25 березня 2022 р.) / Дніпро: Юрсервіс, 2022. С. 146–150. URL: http://eadnurt.diit.edu.ua/handle/123456789/15706 (дата звернення: 19.05.2024).

4.Кужда Т. І. Застосування інформаційних технологій при бізнес-аналізі діяльності організації в кризових умовах. Галицький економічний вісник. Т. : ТНТУ, 2023. Том 81. № 2. С.96–105.

5.Толбатов О. В., Віганяйло С. М., В'юненко О.Б. Інноваційні підходи до планування економічної діяльності у менеджменті аграрних підприємств. Наукове оточення сучасної людини: економіка, менеджмент, освіта, психологія, юриспруденція, політологія: серія монографій. Сірий. «Наукове оточення сучасної людини». Одеса, 2019. С. 52-59.

6.Загородна О. М., Серединська В. М. Формування механізму діагностики фінансового стану підприємства. Стан і перспективи розвитку обліково-інформаційної системи в Україні: матеріали V міжнар. наук.-практ. конф. Тернопіль: ТНЕУ, 2019. С. 233–235.

7.Латишева О. В., Підгора Є.О., Кіріллова М. М. Інформаційно-аналітичне забезпечення аналізу фінансового стану суб'єкта господарювання. Інфраструктура ринку: електронний науково-практичний журнал з економічних наук. 2019. Вип. № 38. С. 181–187.

8.Набатова Ю. А., Малачевська К. О. Автоматизація аналізу та прогнозування фінансових результатів підприємства. Ефективна економіка. 2020. № 12. URL: http://www.economy.nayka.com.ua/?op=1&z=8405 (дата звернення: 19.07.2024). DOI: 10.32702/2307-2105-2020.12.80

9.Фінансовий аналіз та звітність [Текст]: практикум / Б. Є. Грабовецький, І. В. Шварц. Вінниця : ВНТУ, 2019. 99 с.

10. Багацька К.В., Говорушко Т.А., Шеремет О.О. Фінансовий аналіз: Підручник. К. 2020. 320 с.

References

1. Kovalchyk O., Riznyk N. Osnovni trendy informatsiinykh tekhnolohii dlia biznesu. Tsyfrova ekonomika yak faktor innovatsii ta staloho rozvytku suspilstva: materialy III Mizhnar. nauk.-prakt.konf. uchenykh ta studentiv, m.Ternopil, 6–7 hrud. 2022 r. Ternopil, 2022. S. 91–93. URL: https://elartu.tntu.edu.ua/bitstream/lib/40126/2/III_MNPK_2022_Kovalchyk_O-The_main_trends_of_information_91-93.pdf (data zvernennia: 18.05.2023).

2. Informatsiini tekhnolohii v biznesi. Chastyna 1: navch. posib. / [Shevchuk I.B., Starukh A.I., Vaskiv O.M. ta in.]; za zah. red. I.B. Shevchuk. Lviv: Vydavnytstvo NNVK «ATB», 2020. 455 s.

3. Savchuk L., Udachyna K. Osoblyvosti vykorystannia prykladnykh informatsiinykh tekhnolohii v biznesi. Upravlinnia proiektamy. Perspektyvy rozvytku proiektnoho ta neiromenedzhmentu, informatsiinykh tekhnolohii upravlinnia, tekhnolohii stvorennia ta vykorystannia obiektiv prava intelektualnoi vlasnosti: zb. nauk. prats za materialamy IV Mizhnarodnoi nauk.-prakt. internet-konferentsii (m. Dnipro, 24–25 bereznia 2022 r.) / Dnipro: Yurservis, 2022. C. 146–150. URL: http://eadnurt.diit.edu.ua/handle/123456789/15706 (data zvernennia: 19.05.2024).

4. Kuzhda T. I. Zastosuvannia informatsiinykh tekhnolohii pry biznes-analizi diialnosti orhanizatsii v kryzovykh umovakh. Halytskyi ekonomichnyi visnyk. T. : TNTU, 2023. Tom 81. № 2. S.96–105.

5. Tolbatov O. V., Vihaniailo S. M., Viunenko O.B. Innovatsiini pidkhody do planuvannia ekonomichnoi diialnosti u menedzhmenti ahrarnykh pidpryiemstv. Naukove otochennia suchasnoi liudyny: ekonomika, menedzhment,

© Яремко С.А., Новицький Р.М.

osvita, psykholohiia, yurysprudentsiia, politolohiia: seriia monohrafii. Siryi. «Naukove otochennia suchasnoi liudyny». Odesa, 2019. S. 52-59.
6. Zahorodna O. M., Seredynska V. M. Formuvannia mekhanizmu diahnostyky finansovoho stanu pidpryiemstva. Stan i perspektyvy rozvytku oblikovo-informatsiinoi systemy v Ukraini: materialy V mizhnar. nauk.-prakt. konf. Ternopil: TNEU, 2019. S. 233–235.
7. Latysheva O. V., Pidhora Ye.O., Kirillova M. M. Informatsiino-analitychne zabezpechennia analizu finansovoho stanu subiekta hospodariuvannia. Infrastruktura rynku: elektronnyi naukovo-praktychnyi zhurnal z ekonomichnykh nauk. 2019. Vyp. № 38. S. 181–187.
8. Nabatova Yu. A., Malachevska K. O. Avtomatyzatsiia analizu ta prohnozuvannia finansovykh rezultativ pidpryiemstva. Efektyvna ekonomika. 2020. № 12. URL: http://www.economy.nayka.com.ua/?op=1&z=8405 (data zvernennia: 19.07.2024). DOI: 10.32702/2307-2105-2020.12.80
9. Finansovyi analiz ta zvitnist [Tekst]: praktykum / B. Ye. Hrabovetskyi, I. V. Shvarts. Vinnytsia : VNTU, 2019. 99 s.
10. Bahatska K.V., Hovorushko T.A., Sheremet O.O. Finansovyi analiz: Pidruchnyk. K. 2020. 320 s.

© Яремко С.А., Новицький Р.М.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-48
UDC 004.89:069
Andrii Martynenko, Ph.D.
https://orcid.org/0000-0002-5033-4696
Serhii Prykhodchenko, Ph.D., associate professor
https://orcid.org/0000-0002-6562-0601
Hulina Iryna, Ph.D., associate professor
https://orcid.org/0000-0003-2565-5006
Dmytro Moroz, Ph.D.
https://orcid.org/0000-0003-2577-3352
National Technical University "Dnipro Polytechnic", Dnipro, Ukraine

REQUIREMENTS AND THEIR IMPLEMENTATION FOR THE INTELLIGENT SYSTEM OF DECISION-MAKING IDENTIFICATION AND CLASSIFICATION OF CULTURAL VALUES

Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М. Вимоги та їх реалізація для інтелектуальної системи прийняття рішень ідентифікації та класифікації об'єктів культурних цінностей. В статті розглядається проблема розробки та аналізу вимог та шляхи їх реалізації для інтелектуальної системи підтримки прийняття рішень ідентифікації та класифікації об'єктів культурних цінностей. В роботі аналізуються як загальні стандарти розробки програмного забезпечення, так і спеціалізовані стандарти, що стосуються саме інтелектуальних систем, також враховано сучасні методології та практичні інструкції. В роботі визначаються та аналізуються такі вимоги до системи як: швидкість обробки великих обсягів даних, точність ідентифікації, інтеграція з іншими інформаційними системами та базами даних, захист даних та інформаційна безпека, масштабування, адаптація до нових типів даних і моделей поведінки об'єктів, інтерфейс користувача, обробка та аналіз різних типів даних, кросплатформність і прогнозування на основі наявних даних для підтримки прийняття рішень. Також в статті визначається складність та комплексність підходу вирішення поставленої задачі, визначені перспективи та шляхи подальшого дослідження даної предметної області.
Ключові слова: інтелектуальна система підтримки прийняття рішень, розробка та аналіз вимог, стандарт, методологія, інструкція, об'єкти культурних цінностей.

Martynenko A., Prykhodchenko S., Hulina I., Moroz D. Requirements and their implementation for an intelligent decision-making system for the identification and classification of objects of cultural value. The article considers the problem of development and analysis of requirements and ways of their implementation for an intelligent decision support system for identification and classification of objects of cultural value. The work analyzes both general standards of software development and specialized standards relating to intelligent systems, modern methodologies and practical instructions are also taken into account. The work defines and analyzes such system requirements as: speed of processing large volumes of data, accuracy of identification, integration with other information systems and databases, data protection and information security, scaling, adaptation to new types of data and object behavior models, interface user, processing and analysis of various types of data, cross-platform and forecasting based on available data to support decision-making. The article also defines the complexity and complexity of the approach to solving the given task, the prospects and ways of further research of this subject area are determined.
Keywords: intelligent decision support system, development and analysis of requirements, standard, methodology, instruction, objects of cultural values.

Introduction. An intelligent decision support system that can identify and classify objects of cultural value according to various criteria can be useful for analysis, training and evaluation, as well as for detecting fakes or similarities between different objects. To develop such a system, it is necessary to use models and methods that can process various information, extract various characteristics and attributes, compare and classify them using modern algorithms of machine learning and artificial intelligence. But in addition to mathematical models and methods of processing and analyzing the received data, one of the important components of the design and development of such a system is the development and analysis of requirements for it.

Analysis of research and publications. In works [1 – 6], the authors conducted a study of the relevance and prospects for the development and use of an intelligent decision-making support system for the identification of objects of cultural value. It should be noted that considerable attention is paid to the development of schemes and conceptual models of the decision support system and models and methods of identification and classification and methods of identification and classification of objects of cultural value. As a result of the analysis of the subject area, the authors came to the conclusion that the developers of such systems face an unresolved scientific and practical task of developing and analyzing the requirements for an intellectual decision-making support system for the identification of objects of cultural value.

Setting the problem. When developing and analyzing requirements for their further implementation in the design and development of intelligent systems, it is important to take into account

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

both general software development standards [7 – 10] and specialized standards relating specifically to intelligent systems [11 – 16] The following methodologies should also be taken into account and practical guidelines such as BABOK (Business Analysis Body of Knowledge - a standard describing best practices of business analysis that can be applied to gathering and analyzing requirements for intelligent systems) and Agile Methodologies (Scrum, Kanban - flexible software development methodologies that can be adapted for projects with intelligent systems, ensuring the adaptability of requirements under conditions of uncertainty).

The requirements for modern information and analytical systems for the identification of objects of cultural value are quite high due to the complexity and variety of tasks that these systems must perform. Thus, the authors of this work, based on the analysis of the state of the issue, determine the following requirements for this system:

– The system should provide fast processing of large volumes of data for prompt decision-making. This is especially important in real time, for example, in security systems or for object identification based on video surveillance.

– High accuracy of identification is critically important. The system should have a minimum number of recognition errors to avoid false positives or incorrect identifications.

– The information and analytical system should be easily integrated with other information systems and databases to provide comprehensive analysis and data exchange.

– Given the nature of the processed data, the system should have a high level of protection against unauthorized access. Protection of personal data and information security are priority requirements.

– The system must be able to scale to handle growing volumes of data without losing performance and efficiency.

– The system should be able to adapt to new types of data and object behavior models through mechanisms of machine learning and artificial intelligence.

– The user interface should be intuitive, with the possibility of quick access to the necessary information and analysis tools.

– The system should support processing and analysis of various types of data: text, numerical, graphic, video and audio files, etc.

– The system should be available on various platforms, including mobile devices, and provide remote access for users.

– An important requirement is the system's ability to make predictions based on available data to support decision-making.

These requirements are aimed at ensuring high efficiency, reliability and security of information and analytical systems for identifying objects in various fields, including security, transport, medicine, trade, etc.

The main material presentation and the obtained research results substantiation of development and ways of implementing requirements for the system.

Ensuring the speed of data processing in an intelligent decision-making system is an important task that requires the use of various technological approaches. To implement this requirement, the following steps are proposed:

– Parallel data processing:
  – Parallel computing (the use of multi-core processors and graphics processors (GPUs) allows data processing to be distributed between several cores, which significantly speeds up the processing of large volumes of data).
  – Distributed systems (the use of cluster computing, where data processing is distributed among several servers or even data centers, allows processing large amounts of data in parallel).
– Using indexes and caching:
  – Indexing (correct indexing of data in databases allows you to significantly speed up data search and filtering).
  – Caching (the application of caching allows you to store frequently used processing results in memory, reducing the need for repeated processing of the same data).
– Optimization of algorithms:
  – Selection of optimal algorithms (use of algorithms that are effective in terms of computational complexity allows to reduce data processing time. For example, using linear time sorting instead of quadratic).

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

  - Algorithmic optimization (improvement of existing algorithms, for example, by reducing the number of iterations or reducing the volume of processed data).
- Massively parallel processing (MPP):
  - Massively parallel processing (the use of a database or computing system that supports massively parallel processing allows processing large data sets simultaneously on many nodes).
- Use of modern database technologies:
  - In-memory databases (these databases keep data in RAM, which provides fast access and processing of data without the need to refer to disk drives).
  - NoSQL databases (for large volumes of non-relational data, NoSQL databases are used, which allow you to quickly store and retrieve data).
- Artificial intelligence and machine learning:
  - Machine learning (using machine learning models to pre-process and classify data can speed up decision-making).
  - Predicative algorithms (use of forecasting models for preliminary determination of the most likely solutions, which reduces the time for analyzing new data).
- Hardware accelerators:
  - FPGA and ASIC (use of specialized hardware accelerators, such as FPGA (Field Programmable Gate Array) or ASIC (Application-Specific Integrated Circuit), which are specially designed to perform certain computing tasks faster than universal processors).
  - Graphics processing units (GPUs) (application of GPUs for computations that are efficient for parallel tasks such as deep learning or big data analysis).
- Optimization of network delays:
  - Latency reduction (optimization of network protocols and equipment to reduce delays in data transmission).
  - Placing data closer to the user (using technologies such as Content Delivery Network (CDN) to store data closer to users, reducing access time).
- Application of cloud technologies:
  - Cloud computing (use of cloud services for dynamic scaling of computing resources according to system needs).

These methods and approaches can be combined depending on the specific requirements and tasks facing the intelligent decision-making system.

Ensuring accuracy and reliability in the intelligent decision support system for object identification is critical for the effective operation of the system. For this, it is suggested to use the following approaches:

- Using qualitative data:
  - Data collection and cleaning (ensuring high data quality by removing noise, duplication and other errors. This includes checking data for compliance with standards, eliminating gaps, and using data normalization and standardization methods).
  - Data relevance (constant updating of data to ensure its relevance, which is important for accurate identification of objects).
- Use of effective algorithms and models:
  - Machine learning algorithms (using advanced machine learning algorithms that provide high accuracy in object classification and identification, such as neural networks, decision trees, support vector algorithms (SVM), etc.).
  - Model testing and validation (conducting thorough testing of models on different data sets, including the use of cross-validation, to assess their accuracy and reliability).
  - Ensemble of methods (using a combination of several models (ensemble) to increase accuracy by reducing the risk of false identification of objects).
- Use of data verification methods:
  - Cross-validation (application of cross-validation methods to evaluate model performance on independent subsets of data).
  - Error analysis (regular analysis of the types of errors made by the system to identify and correct weaknesses in data processing models and processes).
- Optimization of model parameters:

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

  - Selection of hyperparameters (using methods of finding optimal hyperparameters of models, such as Grid Search or Random Search, to increase the accuracy and reliability of the system).
  - Regularization (application of regularization to avoid overfitting of models, which improves overall accuracy and reliability).
- Integration of additional data sources:
  - Multimodal analysis (using different types of data (text, images, video, audio) to improve the accuracy of object identification by combining disparate information).
  - Expansion of functionality (use of additional sensors, cameras, systems or other sources for data enrichment, which allows better distinguishing of objects).
- Development of sustainable models:
  - Adaptation to changes in the environment (development of models capable of adapting to new conditions, including changes in data or in the behavior of identifiable objects).
  - Protection against attacks (implementation of methods to protect against possible attacks on the system, such as attacks using forged data or attempts to manipulate results).
- System monitoring and updates:
  - Continuous monitoring (continuous monitoring of the performance of the system and its models, which allows timely detection and correction of failures or drops in accuracy).
  - Regular updating of models (updating models based on new data and feedback, which ensures their relevance and compliance with real conditions).
- Implementation of feedback system:
  - Error-based correction (having feedback mechanisms that allow users or other systems to report identification errors so that the system can learn from them and improve its accuracy).
  - Training on real data (using data from real cases of identification to refine models and increase their accuracy).
- Maintaining high security standards:
  - Data protection (ensuring confidentiality and integrity of data through the use of encryption, access control and other cyber security measures).
  - Audit and verification (regular auditing of the system and verification of its operation by independent experts to confirm compliance with high standards of accuracy and reliability).

Application of these approaches makes it possible to create a reliable and accurate intelligent decision support system for the identification and classification of cultural values, capable of effectively identifying objects in various conditions and fields of application.

Integration of the decision support system for object identification with other systems is an important aspect that ensures efficient data exchange, process synchronization and ease of use. The authors suggest using the following approaches to ensure integration:

- Using standardized APIs:
  - RESTful API (the use of RESTful API allows the system to easily interact with other systems through standardized HTTP protocols. This provides scalability and flexibility of integration).
  - GraphQL (the application of GraphQL for queries allows clients to receive only the necessary data, which optimizes the exchange of information).
  - gRPC (using gRPC (Google Remote Procedure Call) for fast and efficient data transfer between microservices or different systems, especially when low latency and high performance are required).
- Using middleware for integration:
  - Enterprise Service Bus (ESB) (use of ESB allows centralized management of data exchange between different systems, ensuring reliability and security of integration).
  - Cross-system platforms (the use of platforms such as Apache Kafka or RabbitMQ to exchange messages between systems provides asynchronous data exchange with high throughput).
- Use of standard data formats:
  - JSON and XML (the use of standard data exchange formats such as JSON or XML simplifies integration with different systems, regardless of their technological platform).
  - Protocol Buffers (the use of Protocol Buffers (Protobuf) for data serialization allows for compact and efficient data exchange between systems).
- Support for compatibility standards:

  - Compliance with industry standards (the integration of a decision support system with other systems can be facilitated by adhering to industry-specific standards (e.g. HL7 in medicine, OPC-UA in industry)).
  - Integration through API Gateway (using API Gateway to manage and simplify access to various APIs allows you to create a single-entry point interface for integration with other systems).
- Support of integration protocols:
  - SOAP (for more formal or secure integrations, particularly in large enterprise systems, the use of SOAP (Simple Object Access Protocol) can provide the required level of reliability and interoperability).
  - OData (using the OData protocol to provide standardized access to data via a REST API that allows for easy integration of different systems).
- Modularity and microservice architecture:
  - Microservices (the transition to a microservice architecture allows you to create independent system components that are easily integrated with each other and with other external systems through standardized APIs).
  - Modular design (the use of modular design allows the system to be easily integrated with other systems by adding or replacing individual modules without the need for significant changes).
- Integration through cloud services:
  - Cloud platforms (the use of cloud platforms such as AWS, Azure, Google Cloud provides easy integration through their API services and other data management tools).
  - iPaaS-type services (Integration Platform as a Service, iPaaS) allow you to easily connect different cloud and local systems, ensuring flexibility and scalability of integration).
- Ensuring integration security:
  - Authentication and authorization (using security standards such as OAuth2 or OpenID Connect for secure API access and data protection during integration).
  - Data encryption (ensuring data encryption during transmission between systems to protect confidential information).
- Integration testing:
  - Compatibility testing (conducting compatibility testing between systems to check the correctness of data exchange and interaction).
  - Monitoring and logging (the use of monitoring and logging tools to track the operation of integration processes, which allows you to quickly identify and eliminate problems.
- Using containerization:
  - Docker and Kubernetes (using container technologies to deploy integration components provides portability, scalability and simplifies software version management).

The use of these approaches ensures reliable, flexible and safe integration of the SPR for object identification with other systems, which increases the efficiency and accuracy of work in various environments and spheres of activity.

Ensuring security and privacy in an intelligent decision support system for the identification of objects of cultural value is critically important, especially when it comes to processing sensitive data. To achieve a high level of security and privacy, it is suggested to use the following approaches and measures:

- Data encryption:
  - Data at Rest encryption (using encryption algorithms to protect data stored in databases or file systems. For example, AES-256 is one of the standards for data encryption).
  - Encryption of data during transmission (Data in Transit) (ensuring protection of data during its transmission between systems or users over the network by using SSL/TLS protocols).
  - Encryption of data in memory (Data in Use) (ensure encryption of data even during its processing in memory to protect against attacks on RAM).
- Access control:
  - Authentication mechanisms (use of multi-level authentication (for example, two-factor authentication) to confirm the identity of users trying to access the system).
  - Authorization and roles (implementation of role-based access policies (Role-Based Access Control, RBAC), which allows restricting access to certain functions or data only to authorized users).

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

  - Access logging and monitoring (recording and monitoring of all attempts to access the system to detect and prevent unauthorized actions.
- Data protection and privacy:
  - Data anonymization (in case of personal data processing, use of anonymization or pseudonymization methods that allow to reduce the risk of identification of persons in case of data leakage).
  - Data minimization (collection and storage of only those data that are necessary for the performance of tasks to reduce the risks associated with the processing of redundant information).
  - Privacy policies (implementation and compliance with privacy policies that regulate the use and processing of personal data in the system).
- Protection against cyber-attacks:
  - Network protection (use of network screens (firewalls), intrusion detection and prevention systems (IDS/IPS), as well as protection against DDoS attacks to ensure system security against external threats).
  - Endpoint security (using anti-virus software, device access controls, and security policies to protect system end users).
  - Regular updating and patching (constant software updates to eliminate known vulnerabilities and prevent exploitation of these vulnerabilities by attackers).
- Identity and Access Management (IAM):
  - Centralized identity management (implementation of identity and access management systems that provide centralized management of user accounts and their access rights).
  - Single Sign-On (SSO) (using SSO technology to simplify access management and improve security by reducing the number of credentials that users need to enter).
- Monitoring and auditing:
  - Real-time security monitoring (using security monitoring systems such as SIEM (Security Information and Event Management) to detect and respond to suspicious activity in real-time).
  - Regular security audit (conducting regular security audits, which include analysis of logs, verification of security settings and assessment of compliance with the requirements of regulatory acts.
  - Pen-testing (conducting penetration testing (pen-testing) to detect vulnerabilities in the system before they can be used by attackers.
- Development security (DevSecOps):
  - Integration of security into the development cycle (implementation of DevSecOps practices that integrate security measures at all stages of development, from design to deployment and operation).
  - Code analysis (using tools for static and dynamic code analysis to identify vulnerabilities at the development stage).
  - Risk assessment (conducting regular risk assessments related to system security to determine and prioritize security measures).
- Security in cloud environments:
  - Data encryption in the cloud (use of data encryption tools stored and processed in cloud services to prevent unauthorized access).
  - Management of access to cloud resources (application of strict access management policies to cloud resources and services, including role-based access control and multi-factor authentication).
  - Verification of compliance with requirements (ensuring compliance with security policies of cloud providers and regulatory requirements for maintaining confidentiality and integrity of data).
- Incident management:
  - Incident Response Plan (development and implementation of a security incident response plan that includes clear procedures for incident detection, response, and recovery).
  - Incident response team (formation of a team of security specialists that is ready to promptly respond to threats and incidents, minimizing possible consequences).

Ensuring security and privacy in the system requires a comprehensive approach that includes technical, organizational and procedural measures. All these measures must be constantly updated and improved in response to new threats and challenges in the field of information security.

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

The scalability of the intelligent decision-making support system for the identification of objects of cultural value is suggested by the authors to be achieved with the help of several key approaches and technologies:

- Use of cloud computing:
  - Cloud infrastructure (cloud services such as AWS, Azure or Google Cloud allow you to scale computing power according to system needs. This allows you to process large amounts of data and use powerful machine learning models without the need for physical servers).
  - Distributed computing (distribution of computing tasks between several nodes in the cloud allows for parallel processing of data, which reduces execution time and increases system efficiency).
- Modular system architecture:
  - Microservices (dividing the system into separate services that are responsible for different tasks (for example, image processing, data analysis, decision-making), allows each component to be scaled independently. This provides flexibility and ease of deployment and maintenance).
  - Containerization (using Docker or Kubernetes to manage containers simplifies the deployment and maintenance of scalable applications).
- Optimization of machine learning algorithms:
  - Pre-training of models (using pre-trained models (for example, deep learning models for computer vision) allows you to reduce the load on the system and speed up the identification of objects).
  - Batch processing (instead of processing each request separately, the system can process requests in batches, which reduces waiting time and increases productivity).
- Application of hybrid approaches:
  - Local processing together with the cloud (to reduce delays and increase response speed, the system can perform initial data processing locally (on edge devices) and transfer the results to the cloud for further analysis).
- Optimizing data storage and transmission:
  - Deduplication and compression of data (reducing the amount of data transmitted between system components allows to reduce the load on the network and increase the efficiency of the system).
  - Caching of results (use of cache to store frequently used results allows to reduce the time for repeated processing of similar requests).
- Machine learning on the edge (Edge AI):
  - Use of machine learning models on edge devices for preprocessing and object identification. This reduces latency, which is important for systems that require real-time response.
- Intelligent routing and load balancing:
  - The use of intelligent algorithms to distribute tasks between different servers or system components ensures a uniform load and increases resistance to overloads.
- Automated scaling:
  - Configuring automatic scaling of resources in response to load changes. For example, adding new servers when the number of requests increases and reducing their number when the load decreases.

These approaches make it possible to ensure the effective scalability of the intelligent decision support system for object identification, which is important for its stable operation in conditions of growth in the volume of data and the number of requests.

The ability to learn and adapt is critical to ensuring its effectiveness in the face of change and new challenges. The authors propose to achieve this opportunity using the following approaches and technologies:

- Use of machine learning methods:
  - Models with the possibility of retraining (creating machine learning models that can be retrained on new data. This allows the system to adapt to new types of objects or identification conditions).
  - Method of learning with reinforcement (Reinforcement Learning) (this approach allows the system to learn based on feedback from the environment, adjusting its decisions to increase the accuracy of identification).

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

- Use of continuous learning algorithms (Continuous Learning):
  - Memory support (implementation of algorithms that allow you to store and use past experience to learn from new data, while not forgetting already acquired knowledge. This is important for maintaining the system's resistance to catastrophic forgetting).
  - Dynamic updating of models (ensuring the possibility of dynamic updating and improvement of machine learning models during their operation without the need for complete retraining).
- Use of hybrid models:
  - Combination of several learning methods (using a combination of different learning algorithms, such as Supervised Learning for structured data and Unsupervised Learning for discovering new patterns. This allows the system to effectively adapt to new types of data).
- Application of transfer learning (Transfer Learning):
  - Using pre-trained models (applying models that have already been trained on a large amount of data and adapting them to specific identification tasks. This significantly reduces the time and resources needed to train the system).
  - Adaptation to new domains (the transfer of knowledge from one area to another allows the system to quickly adapt to new conditions, in particular to changes in the environment or characteristics of objects).
- Using a feedback collection and analysis system:
  - Automatic data collection (implementation of automatic data collection mechanisms on system errors or uncertainties for further analysis and use of this data in the learning process).
  - Learning based on user feedback (the system can receive feedback from users on the correctness of identification, which allows it to adapt its models to real needs and conditions).
- Use of adaptive algorithms:
  - Adaptive neural networks (using neural network architectures that can dynamically change their structure depending on input data or changes in the environment).
  - Evolutionary algorithms (application of evolutionary learning methods, which allow the system to experiment with different configurations of models and choose the most effective ones for current conditions).
- Integration with other data sources:
  - Use of external data sources (connection to external databases or other sources of information to enrich training data and increase the accuracy of identification).
  - Online learning (Application of algorithms that can learn in real time on new incoming data).
- Automation of the process of processing new data:
  - Pipelines for automated learning (creating automated pipelines that can collect, clean and prepare data for training models without significant human intervention).
  - Automated ML-operations (MLOps) (using MLOps to automate the deployment, monitoring and updating of machine learning models, which allows you to quickly adapt the system to changing conditions).

These approaches make it possible to create a CSDP that is able to constantly learn, adapt to new conditions and increase its effectiveness over time.

Ease of use of the intelligent decision support system for the identification and classification of objects of cultural value is provided by a number of factors that contribute to increasing the efficiency of the user's interaction with the system and ensure the ease and speed of performing tasks. The main factors affecting the ease of use of the CSPR include:

- Intuitive user interface (UI):
  - Simplicity and logic (the interface should be simple and clear, with a logical structure and clear controls. This reduces the learning curve for new users and provides quick access to the main functions).
  - Data visualization (the use of graphical elements such as charts, graphs and charts helps users quickly evaluate identification results and make informed decisions).
- Personalization and customization:
  - Individual settings (allowing users to customize the interface and functionality of the system according to their needs, including language selection, design themes and notification settings).

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

  - Adaptive interface (the system should adapt to individual user habits, offering frequently used functions and quick access to necessary tools).
- Speed and efficiency of work:
  - Fast data processing (the system should provide a high speed of query processing and object identification, minimizing delays, which allows users to quickly get results).
  - Automation of routine tasks (implementing automation to perform repetitive or routine tasks, which reduces the burden on the user and allows you to focus on more complex aspects of decision-making).
- Quality of user interaction (UX):
  - Feedback in real time (the system should provide instant feedback on the user's actions, informing about the progress of tasks or possible errors).
  - Hints and training materials (the presence of integrated hints, training modules or guides that help users quickly learn the system and get answers to questions that arise).
- Integration with other systems:
  - Ease of integration (the system should be easily integrated with other tools and programs used by users, for example, with database management systems, analytical platforms or ERP systems).
  - Support for standard data formats (providing support for standard file and data formats to simplify the import and export of information).
- Reliability and security:
  - Data protection (the system must provide a high level of data security, including encryption, user authentication and access control, which is critical for protecting confidential information).
  - Resilience to errors (the system must be resistant to errors and failures, with mechanisms for automatic recovery and data saving, which guarantees uninterrupted operation even in the event of unforeseen situations).
- Multi-platform support:
  - Access from different devices (the system should be accessible from different types of devices, including computers, tablets and smartphones, ensuring ease of use at any time and from any place).
  - Responsive design (the system interface should automatically adapt to the screen size and features of the device on which it is used).
- Effective support and service:
  - Technical support (the presence of an effective technical support service that quickly responds to user requests and provides assistance in case of problems or questions).
  - Regular updates (implementation of regular software updates that include new features, bug fixes and usability improvements).
- Accessibility and inclusiveness:
  - Accessibility support (ensuring the ability to use the system by people with disabilities, including support for screen readers, voice control and other accessibility tools).
  - Localization (availability of the system in different languages, which makes it accessible to users from different regions and language groups).

These factors together provide high usability of the intelligent decision support system for the identification of objects of cultural value, which is key to its effective implementation and use.

The support of various types of data is realized due to several key technological and architectural solutions. This allows the system to efficiently process, analyze and integrate a variety of data such as text, numbers, images, videos, audio and other types of data. The main mechanisms providing such support include:

- Universal data architecture:
  - Modularity and flexibility (the system should have a modular architecture that allows you to add or change modules to work with different types of data without the need for significant changes in the entire system).
  - General data stores (using universal data stores that support storage of different data formats, for example, relational databases for structured data and NoSQL databases for unstructured or semi-structured data).
- Support of standard formats and protocols:

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

  - Support for different data formats (the system should support a wide range of file formats, including CSV, JSON, XML for structured data, as well as image (JPEG, PNG), video (MP4, AVI) and audio (MP3, WAV) formats).
  - Unified API interfaces (implementation of standard API interfaces for integration with other systems and data sources, which provides access to data in various formats.
- Use of various algorithms and models:
  - Algorithms for processing different types of data (the system should have a set of specialized algorithms for processing different types of data. For example, deep learning algorithms (CNN) for image processing, NLP models for working with text, and time series analysis algorithms for numerical data).
  - Universal machine learning models (use of models that can process several types of data at the same time (for example, multimodal models), which allows the system to identify objects based on different sources of information).
- Integration of ETL processes:
  - ETL (Extract, Transform, Load) (the implementation of ETL processes to extract, transform and load data into a format suitable for analysis. This allows you to adapt different types of data to the requirements of the system, regardless of their initial format).
  - Data conversion (the system should support the conversion of data from one format to another, including the merging of data from different sources and the conversion of non-standard formats into those suitable for analysis).
- Machine learning for different data:
  - Adaptive models (the use of machine learning models that can learn on different types of data and adapt to the specific requirements of each type. For example, the use of specialized neural network architectures for text processing (LSTM, Transformers) or images (Convolutional Networks)).
  - Analysis of multimodal data (the ability to simultaneously process and analyze data of different types, for example, text and images, which allows the system to make more informed decisions based on combined information).
- Extensibility and integration with other systems:
  - Plugins and extensions (using a system of plugins or extensions that allow you to easily add support for new data types or integrate new algorithms without having to rewrite the entire system).
  - Integration with external services (connection to external data processing services, such as services for image analysis or natural language processing, which allows you to use their capabilities to support new types of data).
- Caching and indexing:
  - Data caching (saving frequently used data in the cache for quick access, which improves system performance when working with large volumes of various data).
  - Indexing for quick search (use of indexing mechanisms for quick search and access to data of various types, which increases the efficiency of object identification).
- Implementation of Big Data approaches:
  - Processing large volumes of data (use of technologies such as Hadoop, Spark, which allow processing large volumes of various data (structured, semi-structured, unstructured) in parallel, which ensures the scalability of the system).
  - Distributed databases (using databases that allow storing and processing different types of data in a distributed environment).

These technologies and approaches ensure the system's ability to effectively work with various types of data, which is critical for the accuracy and versatility of object identification in various scenarios.

The mobility and accessibility of the intelligent decision support system is ensured by various methods and means that allow users to access the system from any place and on different devices, and also ensure the smooth operation of the system in different conditions. The main methods and means of implementing mobility and accessibility, according to the authors, may include:

- Cloud technologies:

  - Cloud services (SaaS) (the use of cloud services, such as Software as a Service (SaaS), allows users to access the CSPR via the Internet from any device that has a browser. This ensures flexibility and scalability of the system).
  - Data processing in the cloud (the use of cloud computing for data storage and processing, which allows access to large amounts of information from anywhere and at any time).
- Mobile applications:
  - Development of cross-platform mobile applications (creating mobile applications that work on different platforms, such as iOS and Android, provides access to the CSDP from mobile devices. Using frameworks such as React Native or Flutter allows you to create applications that work on several platforms at the same time).
  - Offline mode (implementation of the possibility of working in offline mode with further synchronization of data when connected to the Internet. This ensures the availability of the system even in places with limited Internet connection).
- Web technologies:
  - Responsive web design (using Responsive Design, which automatically adjusts the CSPR interface to different screen sizes and types of devices (smartphones, tablets, computers). This ensures ease of use on different devices).
  - Progressive Web Applications (PWA) (development of progressive web applications that work like regular websites, but have features characteristic of mobile applications (offline mode, push notifications). This allows for high availability and mobility of the system without the need to install separate applications).
- API usage and integration with other systems:
  - RESTful and GraphQL APIs (using RESTful or GraphQL APIs to provide access to CMS functions from different devices and platforms. This allows you to create custom client applications or integrate CMS with other systems, providing flexibility of access).
  - Integration with corporate systems (integration of the ERP system with other corporate systems, such as CRM or ERP, through API provides access to the necessary identification information of objects in various work environments).
- Mobile security:
  - Authentication and authorization (use of secure authentication and authorization methods, such as two-factor authentication (2FA) and OAuth, to protect access to the system from mobile devices).
  - Data encryption (ensuring that data is encrypted both on the device and during transmission to protect information from unauthorized access).
- Geolocation services:
  - Use of geolocation (the integration of geolocation functions in the DSS allows providing users with context-dependent information, which increases the accuracy of object identification and location-based decision-making).
  - Geo-fencing (the implementation of geo-fencing technology to activate certain functions or notifications when the user is in a defined geographical area, which improves the user experience).
- Cloud storage and synchronization:
  - Data synchronization (using cloud storage to store data and synchronize it between different devices in real time. This provides access to up-to-date information from any device).
  - Automatic backup (implementation of automatic data backup in the cloud to prevent information loss and ensure its availability in case of unforeseen situations).
- Support for voice and text interfaces:
  - Voice assistants (integration with voice assistants such as Google Assistant or Siri to provide voice access to the helpdesk, allowing users to perform hands-free tasks).
  - Chatbots and text interfaces (the use of chatbots or text interfaces to allow users to interact with the system through text queries, which is especially convenient in mobile applications).
- Caching and optimization:
  - Content caching (implementation of caching mechanisms to reduce loading time and provide quick access to frequently used data, which improves the performance of mobile applications).

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

- Traffic optimization (use of traffic optimization methods, such as data compression and asynchronous download, which ensures fast access to information even in conditions of a slow Internet connection).

Together, these methods and tools provide system mobility and availability, allowing users to work with the system at any time and from any location, while maintaining high efficiency and security.

The ability to predict and the ability of the system to make predictions based on available data to support decision-making in the identification of objects is ensured by the use of various methods, algorithms and technologies that allow analyzing data, identifying patterns and making predictions. According to the authors, the following tools allow you to implement these functions:

- Machine learning and artificial intelligence (AI) models:
  - Machine learning (ML) algorithms (using various machine learning algorithms, such as linear regression, decision trees, random forests, support vector machines (SVM), to predict the behavior of objects based on historical data).
  - Deep Learning (the use of neural networks, including recurrent neural networks (RNN), long-term short-term memory (LSTM) and convolutional neural networks (CNN), to predict complex patterns and behavior of objects based on large volumes of data).
- Time series analysis:
  - Time Series Analysis Techniques (using time series analysis techniques such as ARIMA, SARIMA, Prophet to predict future events based on historical time-varying data. This can be particularly useful for predicting trends and future behavior of objects) .
  - Seasonality and trends (detection of seasonal and trending components in data, which allows the system to take into account recurring patterns and make more accurate forecasts).
- Methods of statistical analysis:
  - Regression analysis (using regression analysis to identify relationships between different variables and make predictions based on these relationships. This allows you to make predictions based on correlations in the available data).
  - Bayesian approach (the use of a Bayesian approach to forecasting, which allows you to take into account previous knowledge and adapt to new data, improving the accuracy of forecasts).
- Processing of big data (Big Data):
  - Tools for working with big data (using technologies such as Hadoop, Apache Spark, to process and analyze large volumes of data, which allows for more accurate predictions based on a large amount of historical data).
  - Real-time (using streaming technologies such as Apache Kafka to process data in real-time and make predictions based on it in real-time).
- Computational intelligence and optimization:
  - Evolutionary algorithms (use of evolutionary algorithms and genetic algorithms to find optimal solutions and predict future states of objects based on simulation of natural selection processes).
  - Metaheuristics (use of metaheuristic methods (for example, the particle swarm method, tabu search) to optimize the forecasting process and improve the quality of forecasts).
- Neural network forecasting:
  - Generative-Competitive Networks (GANs) (using GANs to create models that can generate new data based on existing data, which allows predicting the behavior of objects in different scenarios).
  - Autoencoders (using autoencoders to discover latent representations of data and make predictions based on these representations).
- A rule-based decision support system:
  - Expert systems (the use of expert systems and knowledge bases, which allows making predictions based on defined rules that take into account the knowledge of experts in a certain field).
  - Logical rules and deductive forecasting (using logical rules to predict results based on available data, especially in cases where there are clearly defined rules for the behavior of objects).
- Integration with business intelligence (BI) systems:
  - BI tools (the use of business intelligence tools for data analysis and visualization, which allows you to make predictions based on trends and analytical conclusions obtained from the data).
  - Dashboards and reporting (creation of dashboards that provide users with forecasts in an easy-to-understand form that facilitates informed decision-making).

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

- Data analysis using artificial intelligence (AI):
  - Neural networks for forecasting (using specialized neural network architectures such as LSTM and GRU for time series analysis and forecasting based on complex non-linear dependencies in data).
  - Multimodal models (the use of multimodal models that can simultaneously take into account different types of data (text, images, numbers) to predict the behavior of objects based on complex information).
- Integration with external data sources:
  - Connection to external sources (integration with other data sources (for example, social networks, financial markets, weather services) to enrich data and improve the accuracy of forecasts).
  - API for external data (using the API to obtain relevant data in real time, which allows making more accurate predictions based on new information).

These methods and technologies provide the possibility of forecasting and the ability of an intelligent system to make forecasts based on available data, which allows to increase the efficiency and accuracy of the decision-making process in the tasks of object identification.

Conclusions and further development and research of the importance of determining and ways of implementing requirements for modern information and analytical systems for object identification:

1. The speed of data processing ensures promptness of decision-making, especially in situations that require a quick reaction. Use of high-performance computing technologies, parallel computing and optimization of data processing algorithms.

2. Accuracy and reliability guarantees the correctness and stability of the results, which is critical for decision-making. Applying accurate classification and identification algorithms, testing the system on different data sets, and introducing verification and validation mechanisms.

3. Integration with other systems provides interaction with other information systems, which allows the use of complex information. Use of standard APIs, data exchange protocols and support for compatibility with various platforms and technologies.

4. Security and privacy protects data from unauthorized access and ensures the confidentiality of information. Use of encryption, multi-level authentication, regular updates of security systems and data access control.

5. Scalability allows the system to adapt to the growth of data volumes and the number of users without losing performance. Use of cloud technologies, distributed databases and containerization for dynamic expansion of resources.

6. The ability to learn and adapt makes the system flexible, allowing it to improve over time and adapt to new challenges. Use of machine learning and self-updating mechanisms of models based on new data.

7. Ease of use increases the efficiency of user interaction with the system, making it accessible to a wide range of users. Development of an intuitive interface, adaptive design and support for various devices.

8. Support of various types of data allows processing of various data (text, image, video) for more complete and accurate identification of objects. Using multimodal models capable of analyzing different types of data and ensuring compatibility with different formats.

9. Mobility and accessibility ensures access to the system from any place and on different devices, which is important for prompt decision-making. Use of mobile applications, cloud services, responsive design and offline mode support.

10. Forecasting capabilities allow you to predict future events and the behavior of objects, which improves the quality of decision-making. Use of machine learning methods, time series analysis and integration with business intelligence systems to create predictive models.

Ensuring these requirements is critical for creating an effective, reliable and flexible information and analytical system capable of supporting decision-making in a rapidly changing environment and large volumes of data. Further development of the study by the authors is planned in the analysis of the relationship between the main requirements and their impact on each other and the system as a whole. It is also planned to develop and introduce their qualitative and quantitative characteristics for their further research.

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

Список бібліографічного опису:

1. Martynenko A., Moroz B., Hulina I. (2020). An intelligent decision support system for cultural property identification. Computer-integrated technologies: education, science, production, (39), P.78 – 82. https://doi.org/10.36910/6775-2524-0560-2020-39-13
2. Martynenko A., Moroz B., Hulina I., Syrotkina O. (2020). Conceptual model of an intelligent decision support system to identify cultural values. Computer-integrated technologies: education, science, production, (40), P.51 – 57. https://doi.org/10.36910/6775-2524-0560-2020-40-08
3. Martynenko A.A., Tevyashev A.D., Kulishova N.E., Moroz B.I. System analysis of the problem of establishing the authenticity and authority of painting works. System Research & Information Technologies, 2022, № 1 – P.50 – 60. https://doi.org/10.20535/SRIT.2308-8893.2022.1.04
4. Martynenko A.A., Tevyashev A.D., Kulishova N.E., Moroz B.I., Sergienko A.S. Automatic classification of paintings by year of creation. Radio Electronics, Computer Science, Control. 2022. № 2 – P.80 – 89. https://doi.org/10.15588/1607-3274-2022-2-8
5. Martynenko A. The intelligent decision support system for identification of cultural values. Scientific Collection «InterConf», (150): with the Proceedings of the 2ndInternational Scientific and Practical Conference «Modern Directions and Movements in Science» (April16-18, 2023; Luxembourg, Grand Duchy of Luxembourg) by the SPC «InterConf». Progress Publishers, 2023. P.531 – 535 https://archive.interconf.center/index.php/conference-proceeding/issue/view/16-18.04.2023/160
6. Martynenko, A. (2023). Methods and models of organization and processing of data and knowledge in an intelligent decision support system for the identification of cultural values. Scientific Collection «InterConf+», (32(151), 652–661. https://doi.org/10.51582/interconf.19-20.04.2023.070
7. ISO/IEC/IEEE 29148:2018 – Systems and software engineering – Life cycle processes – Requirements engineering.
8. ISO/IEC 25010:2011 – Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models.
9. IEEE 830-1998 – Recommended Practice for Software Requirements Specifications.
10. ISO/IEC 15288:2015 – Systems and software engineering – System life cycle processes.
11. ISO/IEC 2382-37:2017 – Information technology – Vocabulary – Part 37: Biometrics.
12. ISO/IEC 20546:2019 – Information technology – Big data – Overview and vocabulary.
13. ISO/IEC 23053:2020 – Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).
14. IEEE 12207-2017 – Systems and Software Engineering – Software Life Cycle Processes.
15. IEEE 7000-2021 – Model Process for Addressing Ethical Concerns During System Design.
16. ISO/IEC TR 24028:2020 – Information technology – Artificial Intelligence (AI) – Overview of trustworthiness in AI.

References:

1. Martynenko A., Moroz B., Hulina I. (2020). An intelligent decision support system for cultural property identification. Computer-integrated technologies: education, science, production, (39), P.78 – 82. https://doi.org/10.36910/6775-2524-0560-2020-39-13
2. Martynenko A., Moroz B., Hulina I., Syrotkina O. (2020). Conceptual model of an intelligent decision support system to identify cultural values. Computer-integrated technologies: education, science, production, (40), P.51 – 57. https://doi.org/10.36910/6775-2524-0560-2020-40-08
3. Martynenko A.A., Tevyashev A.D., Kulishova N.E., Moroz B.I. System analysis of the problem of establishing the authenticity and authority of painting works. System Research & Information Technologies, 2022, № 1 – P.50 – 60. https://doi.org/10.20535/SRIT.2308-8893.2022.1.04
4. Martynenko A.A., Tevyashev A.D., Kulishova N.E., Moroz B.I., Sergienko A.S. Automatic classification of paintings by year of creation. Radio Electronics, Computer Science, Control. 2022. № 2 – P.80 – 89. https://doi.org/10.15588/1607-3274-2022-2-8
5. Martynenko A. The intelligent decision support system for identification of cultural values. Scientific Collection «InterConf», (150): with the Proceedings of the 2ndInternational Scientific and Practical Conference «Modern Directions and Movements in Science» (April16-18, 2023; Luxembourg, Grand Duchy of Luxembourg) by the SPC «InterConf». Progress Publishers, 2023. P.531 – 535 https://archive.interconf.center/index.php/conference-proceeding/issue/view/16-18.04.2023/160
6. Martynenko, A. (2023). Methods and models of organization and processing of data and knowledge in an intelligent decision support system for the identification of cultural values. Scientific Collection «InterConf+», (32(151), 652–661. https://doi.org/10.51582/interconf.19-20.04.2023.070
7. ISO/IEC/IEEE 29148:2018 – Systems and software engineering – Life cycle processes – Requirements engineering.
8. ISO/IEC 25010:2011 – Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models.
9. IEEE 830-1998 – Recommended Practice for Software Requirements Specifications.
10. ISO/IEC 15288:2015 – Systems and software engineering – System life cycle processes.
11. ISO/IEC 2382-37:2017 – Information technology – Vocabulary – Part 37: Biometrics.
12. ISO/IEC 20546:2019 – Information technology – Big data – Overview and vocabulary.
13. ISO/IEC 23053:2020 – Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).
14. IEEE 12207-2017 – Systems and Software Engineering – Software Life Cycle Processes.
15. IEEE 7000-2021 – Model Process for Addressing Ethical Concerns During System Design.
16. ISO/IEC TR 24028:2020 – Information technology – Artificial Intelligence (AI) – Overview of trustworthiness in AI.

@ Мартиненко А.А., Приходченко С.Д., Гуліна І.Г., Мороз Д.М.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-39
УДК 621.396.
Bershov Vyacheslav[1], PhD student
https://orcid.org/0009-0006-9500-6414
Yakymchuk Nataliia[2], PhD, Associate Professor
https://orcid.org/ 0000-0002-8173-449X
[1]Ukrainian State University of Railway Transport, Kharkiv, Ukraine
[2]Lutsk National Technical University, Lutsk, Ukraine

THE METHOD OF FORMING ENSEMBLES OF COMPLEX SIGNALS BASED ON MULTI-SCALE DECOMPOSITION OF TIME INTERVALS AT DIFFERENT LEVELS OF DETAIL

Bershov V., Yakymchuk N. The method of forming ensembles of complex signals based on multi-scale decomposition of time intervals at different levels of detail. This article addresses the challenges of forming signal ensembles in dynamic cognitive radio environments, focusing on the limitations of traditional methods. The study proposes a new approach based on multiscale time interval decomposition, which allows for the creation of signal ensembles at varying levels of temporal detail. The key innovation of this method is its ability to improve signal reproduction accuracy, reduce inter-symbol and inter-channel interference, and enhance the overall efficiency of data processing. The method's adaptability to changing environmental conditions is also a core advantage, enabling better use of bandwidth and reduced transmission delay. Through the decomposition of time intervals into coarse, intermediate, and fine levels, this method allows for the detailed analysis of short-term, medium-term, and long-term signal components. Experimental results demonstrate the superiority of this approach in terms of signal recovery accuracy, noise resilience, and processing speed. These findings highlight the potential for optimizing signal processing in cognitive radio networks, particularly in environments with high levels of noise and interference. Future research aims to integrate machine learning algorithms to further enhance adaptability in real-time scenarios.
Keywords: multiscale signal decomposition, cognitive radio environment, time intervals, adaptive algorithms, mean square deviation (MSE), signal-to-noise ratio (SNR), signal processing

Бершов В.С., Якимчук Н.М. Метод формування ансамблів складних сигналів на основі багатомасштабної декомпозиції часових інтервалів на різних рівнях деталізації. У статті розглядаються проблеми формування ансамблів сигналів систем передачі у динамічних середовищах когнітивного радіо, при цьому, увага акцентується на обмеженнях традиційних методів синтезу ансамблів сигналів. Дослідження пропонує новий підхід, заснований на багатомасштабній декомпозиції часових інтервалів, що дозволяє створювати ансамблі сигналів на різних рівнях часової деталізації. Ключовим нововведенням цього методу є підвищення точності відтворення сигналів, зменшення міжсимвольних та міжканальних завад, а також підвищення загальної ефективності обробки даних. Важливою перевагою запропонованого методу є його адаптивність до змінних умов навколишнього середовища, що дозволяє краще використовувати пропускну здатність та зменшувати затримки передачі. За допомогою виконання декомпозиції часових інтервалів на грубі, проміжні та точні рівні, цей метод дає змогу детально аналізувати короткострокові, середньострокові та довгострокові компоненти сигналу в телекомунікаційних трактах. Отримані експериментальні результати демонструють переваги підходу для покращення точності відновлення сигналу, стійкості до шуму та швидкості обробки. Отримані результати підкреслюють потенціал оптимізації обробки сигналів у когнітивних радіомережах, особливо в умовах високого рівня шуму та завад. Подальші дослідження спрямовані на інтеграцію алгоритмів машинного навчання для підвищення адаптивності в режимі реального часу.
Ключові слова: багатомасштабна декомпозиція сигналів, когнітивне радіосередовище, часові інтервали, адаптивні алгоритми, середньоквадратичне відхилення (MSE), співвідношення сигнал/шум (SNR), обробка сигналів.

Statement of a scientific problem.

Traditional methods of forming ensembles of complex signals in the modern dynamic cognitive radio environment demonstrate insufficient flexibility and adaptability, which creates the need to develop new methods that provide effective protection against inter-channel and inter-symbol interference, balance between indicators of signal volumes and the function of mutual correlation, have a high level of bandwidth indicators , low transmission delay and adaptability to changing environmental conditions.

One of these methods is the method of forming ensembles of complex signals based on multi-scale decomposition of time intervals, which allows creating ensembles of signals at different levels of temporal detail, increasing the volume of ensembles of complex signals and improving their characteristics, in particular, increasing the efficiency of processing in the time domain. Multiscale Decomposition is a signal analysis method that allows you to decompose a signal into components of different durations in order to identify and analyze long-term, medium-term and short-term components of a signal (Fig. 1).

1. The coarse level is responsible for detecting the main long-term components of the signal. The signal is divided into large time intervals that allow you to understand its general structure and trends. For example, the division can be made into intervals of 10 seconds if the signal is long, or into intervals of 1 second for shorter signals.

© Bershov V.S., Yakymchuk N.M.

2. Intermediate level. After a rough breakdown, each large interval is detailed into medium-duration components. This allows you to detect changes and trends that were not visible at a coarse level. For example, intervals can be divided into smaller subintervals of 1 second (for long intervals) or 100 ms (for shorter intervals).

3. Subtle level. At this level of temporal segmentation, short-term pulses and signal variations are analyzed. Each average interval is detailed into small time segments, which allows you to detect rapid changes and impulses. For example, each average interval can be divided into subintervals of 10 ms or even less, depending on the nature of the signal.

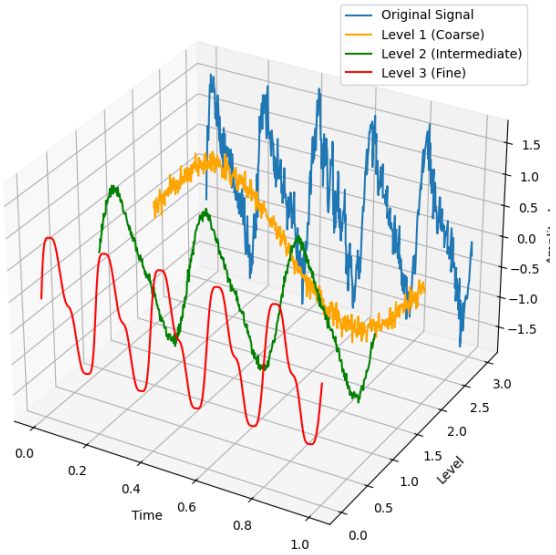


Fig. 1. Fragment of three-level multiscale decomposition

The feasibility of using the method lies in the fact that the method allows cognitive radio systems to flexibly and adaptively analyze signals in the time domain, and this contributes to more efficient use of available resources and improves the quality of data transmission. Thanks to accurate analysis and adaptive segmentation, the method optimizes data transmission, increases the efficiency of resource use even in conditions of high noise and interference.

Research analysis.

The analysis of existing domestic and foreign research on this subject [1-14] proved that a significant part of the work devoted to signal processing and cognitive radio networks is aimed at developing methods of increasing noise immunity, optimizing spectral characteristics, and improving the accuracy of signals in difficult conditions [1]. In particular, researchers proposed different approaches to time-frequency analysis, the use of wavelet transformations, adaptive algorithms and decomposition methods [2].

However, most of the works [3,5-11, 13] do not substantiate the creation of effective ensembles of complex signals based on multilevel decomposition, which allows improving the characteristics of signals and adapting them to changing conditions. Also, the integration of multiscale decomposition methods with adaptive algorithms that could automatically adjust signal processing parameters in real time, ensuring high efficiency in intelligent dynamic radio environments, has not been sufficiently investigated.

The purpose of the work.

The purpose of this study is to develop a method for forming ensembles of complex signals based on multi-scale decomposition of time intervals with different levels of temporal detail.

Presentation of the main material and substantiation of the obtained research results.

The effectiveness of the method of forming ensembles of complex signals based on multi-scale decomposition of time intervals at different levels of time detail is evaluated by indicators.

1. Accuracy of signal reproduction. The degree of deviation from the original signal is calculated, which is mathematically expressed as Mean Squared Error (MSE):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x(t_i) - \hat{x}(t_i))^2, \quad (1)$$

© Bershov V.S., Yakymchuk N.M.

where N – is the number of time points;
$x(t_i)$ – is the value of the original signal at the moment of time t_i;
$\hat{x}(t_i)$ – is the value of the reproduced signal at the moment of time t_i.

2. Resolution is the ability to divide the initial signal $x(t)$ into time segments. The indicator shows the ability of the method to detect and distinguish different components of the signal at different levels of temporal detail, which allows for detailed analysis. Calculations are made using the formula:

$$x(t) = \sum_{n=1}^{N} \left(\sum_{k=1}^{K} c_k^{(n)} \phi_k^{(n)}(t) \right), \quad (2)$$

where n – detail level number;
K – number of components per level n
$c_k^{(n)}$ – decomposition factor for component k at level n;
$\phi_k^{(n)}$ – basis function for component k at level n.

3. Calculation speed. It is estimated by the time T required for signal processing, taking into account the number of iterations I and the computational complexity of the algorithm $O(I)$. It is calculated according to the formula:

$$\mathrm{T} = \sum_{i=1}^{I} \mathrm{O}\ (\mathrm{i}), \quad (3)$$

4. Resistance to noise. A metric that shows the ability to maintain accuracy in the presence of noise. It is measured by the signal-to-noise ratio parameter (SNR):

$$SNR = 10 \log_{10} \left(\frac{\sum_{i=1}^{N} x(t_i)^2}{\sum_{i=1}^{N} (x(t_i) - \hat{x}(t_i))^2} \right), \quad (4)$$

5. Energy efficiency is estimated by the formula as the average value of the signal energy in time segments:

$$E = \frac{1}{N} \sum_{i=1}^{N} |x(t_i)|^2, \quad (5)$$

6. Correlation coefficient ρ. Determines the similarity between the average value of the original $\bar{x}$ and the reproduced $\bar{\hat{x}}$ signal according to the formula:

$$\rho = \frac{\sum_{i=1}^{N} (x(t_i) - \bar{x})(\hat{x}(t_i) - \bar{\hat{x}})}{\sqrt{\sum_{i=1}^{N} (x(t_i) - \bar{x}(t_i))^2 \sum_{i=1}^{N} (\hat{x}(t_i) - \hat{x})^2}}, \quad (6)$$

7. The cross-correlation function. Allows you to assess how well the reproduced signal corresponds to the original one at different time shifts τ, the calculation formula:

$$R_{x_i x_j}(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} x(t_i)\, \hat{x}(t_i + \tau), \quad (7)$$

where $R_{x_i x_j}(\tau)$ – shear cross-correlation functionτ;
$x(t_i)$ – the value of the original signal at the instant of time t_i;
$\hat{x}(t_i + \tau)$ – the value of the reproduced signal at the offset τ.

The generalized system of conditions for all levels of temporal detail can be written in the form of a system of equations:

© Bershov V.S., Yakymchuk N.M.

$$\begin{cases} x(t) = \sum_{k=1}^{K_1} c_k^{(1)} \phi_k^{(1)}(t) + \epsilon^1(t), \text{де } \epsilon^1(t) \to 0 \text{ при } K_1 \to \infty \\ \epsilon^1(t) = \sum_{k=1}^{K_2} c_k^{(2)} \phi_k^{(2)}(t) + \epsilon^2(t), \text{де } \epsilon^2(t) \to 0 \text{ при } K_2 \to \infty \\ \epsilon^2(t) = \sum_{k=1}^{K_3} c_k^{(3)} \phi_k^{(3)}(t) + \epsilon^3(t), \text{де } \epsilon^3(t) \to 0 \text{ при } K_3 \to \infty \\ \Delta t_1 > \Delta t_2 > \Delta t_3 > 0, \text{де } \Delta t_1 - \textit{the length of the intervals is } 1,2,3 \textit{ levels} \end{cases}, \quad (8)$$

where $\epsilon^R(t)$ – approximation errors at different levels.

Taking into account indicators and conditions, an algorithm for practical implementation of the method of forming ensembles of complex signals based on multi-scale time decomposition was developed. Let's consider its stages in more detail (Fig 2).

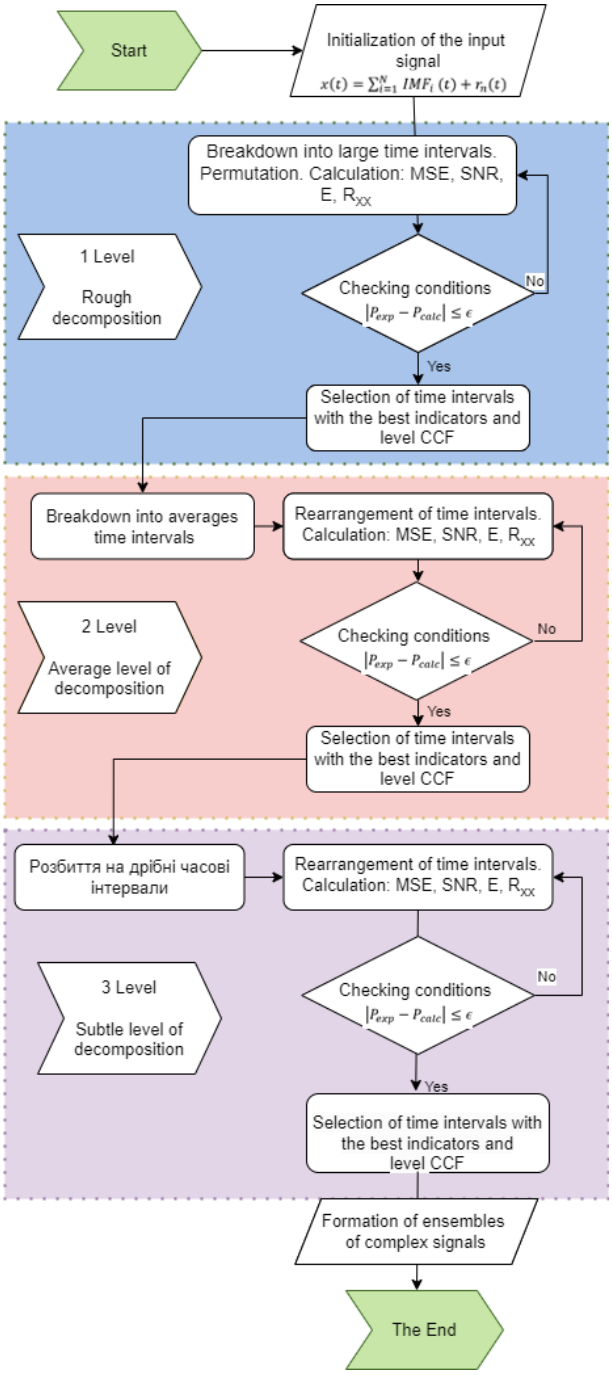


Fig. 2. Block diagram of the multiscale time decomposition algorithm

© Bershov V.S., Yakymchuk N.M.

1 Stage. The beginning of the algorithm. Initialization of the input signal $x(t)$. The decomposition process is adapted to the specific properties of the input signals, which allows for maximum accuracy and efficiency of further processing. Decomposition (IMF) is carried out through an iterative process, which can be represented by the formula:

$$x(t) = \sum_{i=1}^{N} IMF_i\ (t) + r_n(t), \tag{9}$$

where $IMF_i\ (t)$ – internal modal functions; $r_n(t)$ – is the remainder.

2 Stage. A rough level of decomposition. At this stage, the ensemble of signals is broken down into large time intervals for further analysis. The input signal can be given in the form of an equation:

$$x(t) = \sum_{k=1}^{K_1} \left(\sum_{k=1}^{K} c_k^{(1)}\ \phi_k^{(1)}(t) + \epsilon^1(t) \right), \tag{10}$$

Time intervals are rearranged at each level of decomposition (coarse, medium, fine). The optimal permutations are chosen based on the assessment of the correlation properties of the signals. A pairwise calculation of the cross-correlation function (CCF) value is used to estimate the correlation properties.

3 Stage. At this stage, the accuracy and compliance of the indicators with the specified (experimental) conditions of minimum similarity for the coarse level of decomposition are checked. If the accuracy and conditions are met, it is possible to proceed to the intermediate level of decomposition. If the condition of minimum similarity is not fulfilled, the procedure is repeated with the correction of the intervals that give the greatest number of violations. In this case, we return to stage 2 and repeat the iteration. The condition of minimal similarity can be determined by the formula:

$$\left| P_{exp} - P_{calc} \right| \leq \epsilon, \tag{11}$$

where P_{exp}, P_{calc} – experimental and calculated indicators, respectively; ϵ – permissible deviation.

4 Stage. Average level of temporal decomposition. At this stage, the residual signal $\epsilon^1(t)$ is split into average time intervals, which allows for a more detailed analysis of the signal structure. The process can be written mathematically as:

$$\epsilon^1(t) = \sum_{k=1}^{K_2} c_k^{(2)} \phi_k^{(2)}(t) + \epsilon^2(t). \tag{12}$$

At the middle level of decomposition, a similar process takes place, as in stage 2, but with a more detailed breakdown of intervals. The evaluation indicators are also calculated, as for the rough level. This allows you to reveal additional features and characteristics of the signal that were not identified at the previous level of decomposition. At this stage, time intervals are also permuted with an assessment of the correlation properties of the signals and a pairwise calculation of the cross-correlation function (CCF).

For more accurate control of approximation errors and correlation properties, adaptive optimization methods or genetic algorithms can be used at this stage.

One of the most common adaptive optimization algorithms is the Least Mean Squares (LMS) algorithm. Initialize the weight vector $w(0)$ for each step n:

$$\begin{gathered} y(n) = w^T(n)x(n) \\ e(n) = d(n) - y(n) \\ w\ (n+1) = w\ (n) + \mu e(n)[x(n) \end{gathered}, \tag{13}$$

where $x(n)$ – vector of input data;
$y(n)$ – adaptive filter vector;
$d(n), e(n)$ – desired signal and error, respectively;
μ – algorithm step (learning ratio).

© Bershov V.S., Yakymchuk N.M.

5 Stage. Checking the accuracy and compliance of indicators with the given conditions of minimum similarity. Step 5 is similar to step 3, with the difference that verification and correction are performed for the medium decomposition level rather than the coarse one. If the condition specified by the experiment is not fulfilled, the intervals that give the greatest number of violations are corrected, and the iteration is repeated.

6 Stage. Fine level of time decomposition. At this stage, the residual signal $\epsilon^2(t)$ is split into small time intervals for an even more detailed analysis than was the case at previous levels. Mathematically, this can be written as:

$$\epsilon^2(t) = \sum_{k=1}^{K_3} c_k^{(3)} \phi_k^{(3)}(t) + \epsilon^3(t) \quad (14)$$

Also, at a fine level, estimation indicators are calculated, time intervals are rearranged with a pairwise calculation of CCF, and the accuracy and compliance of the indicators obtained as a result of the calculations with the conditions of minimum similarity are checked.

7 Stage. Formation of ensembles of complex signals. At this stage, the sequences formed as a result of the three-level decomposition are used to form ensembles of complex signals. Thanks to the three-level decomposition, each sequence contains detailed information about the original signal, which allows you to display its characteristics as accurately as possible. In general, the generated signal can be written in the form:

$$x(t) = \sum_{k=1}^{K_1} c_k^{(1)} \phi_k^{(1)}(t) + \sum_{k=1}^{K_2} c_k^{(2)} \phi_k^{(2)}(t) + \sum_{k=1}^{K_3} c_k^{(3)} \phi_k^{(3)}(t) + \epsilon^3(t), \quad (15)$$

After combining all levels of decomposition, the consistency of the obtained sequences is checked by analyzing the correlation properties and checking for the presence of systematic errors, after which the final ensembles of complex signals are formed. Experimental studies were conducted to verify and validate the proposed method. The initial data of ensembles of complex signals used in the experiments are shown in table. 1.

Table. 1 Output data for Ensemble 1

Component	Amplitude (A)	Frequency (Fr)	Characteristic
Low frequency	A_1=1,0	Fr_1=1 Hz	Low-frequency sinusoidal signal
Medium frequency	A_2=0,5	Fr_2=5 Hz	Medium frequency sinusoidal signal
High frequency	A_3=0,2	–	High frequency signal
Impulses	A_4=0,8	T=0,2; 0,5; 1c	Pulses with different durations and breaks

Table. 2 Output data for Ensemble 2

Component	Amplitude (A)	Frequency (Fr)	Characteristic
Modulated signal	A_5=1,0	Fr_3=2 Hz Fr_4=0,5 Hz	Low-frequency modulated sinusoidal signal with phase modulation
High frequency noise	A_6=0,2	–	High frequency noise
Impulses	A_7=0,7	T=0,3;0,7;1.2 c	Pulses with different durations and breaks

At the first coarse level of multi-scale time decomposition, the initial division of signals into large time intervals takes place, each of which is analyzed separately. The optimal permutations of time intervals are chosen based on the evaluation of the correlation properties of the signals, which helps to increase immunity. The calculation of evaluation indicators according to the algorithm of the large-scale time decomposition method is presented in the table. 3-4.

Table 3 – Calculation of indicators of the coarse level of decomposition for Ensemble 1

© Bershov V.S., Yakymchuk N.M.

Segment	MSE	SNR (dB)	Energy E	Correlation coefficient (ρ)	R_{xixj}
1	0,1005	9,861	40,22	0,891	0,789
2	0,1021	9,789	40,84	0,876	0,802
3	0,0987	10,002	39,48	0,894	0,812
4	0,1013	9,816	40,52	0,889	0,801
5	0,1030	9,755	41,20	0,883	0,798

Calculations show that Ensemble 1 shows a higher signal recovery accuracy compared to Ensemble 2. For Ensemble 1, the mean value (MSE) is about 0,1011, which proves the low error rate after signal recovery; the signal-to-noise ratio (SNR) is 9,845 dB, which shows a sufficiently high signal-to-noise level; the E energy is 40.45, indicating signal stability, and the correlation coefficient (ρ) is 0.887, demonstrating high similarity between the original and reconstructed signals.

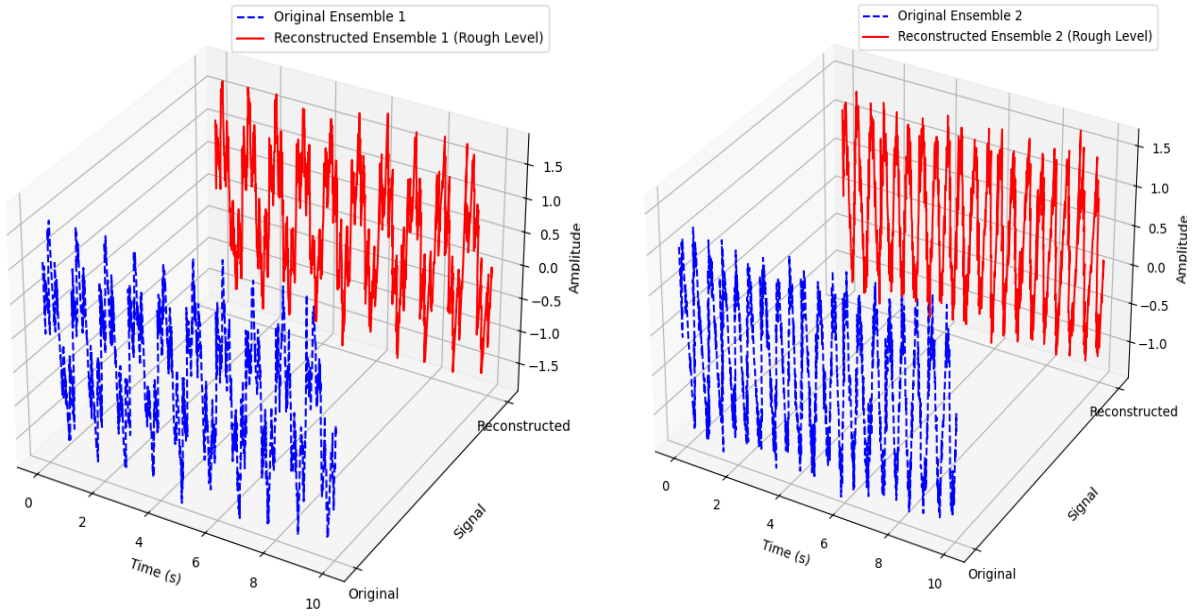
Rxиxj shows a value in the range of 0,789-0,812, which indicates an acceptable (conditionally, taking into account the given noises and disturbances, but far from orthogonality) mutual correlation between different segments. Such results testify to the effectiveness of the method in isolating and restoring the main components of the signal under the given influence of noise.

Table 4 – Calculation of indicators of the coarse level of decomposition for Ensemble 2

Segment	MSE	SNR (dB)	Energy E	Correlation coefficient (ρ)	R_{xixj}
1	0,1523	8,289	60,92	0,742	0,695
2	0,1495	8,379	59,80	0,751	0,702
3	0,1531	8,265	61,24	0,738	0,688
4	0,1507	8,342	60,28	0,747	0,691
5	0,1489	8,409	59,56	0,753	0,699

For Ensemble 2, the MSE is 0,1509, indicating a higher error rate after signal recovery compared to Ensemble 1; SNR – 8,337 dB, which shows a lower signal-to-noise ratio; E is 60,36, indicating a higher energy component of the signal, and the correlation coefficient (ρ) is 0,746, showing a lower similarity between the original and reconstructed signals.

R_{xixj} shows values in the range of 0,688-0.702, which indicates a lower mutual correlation between different segments than was observed for Ensemble 1 (but also far from orthogonality) (Fig. 3).



Ensemble 1 Ensemble 2

Fig. 3 – Time decomposition of the first coarse level

© Bershov V.S., Yakymchuk N.M.

Table 5 – Calculation of indicators of the average level of time decomposition

Ensemble 1					
Segment	MSE	SNR (dB)	Energy E	Correlation coefficient (ρ)	R_{xixj}
1	0,0251	15,122	10,11	0,972	0,935
2	0,0260	15,059	10,21	0,969	0,938
3	0,0246	15,251	9,87	0,974	0,942
4	0,0253	15,184	10,13	0,972	0,939
5	0,0262	15,030	10,30	0,968	0,936
6	0,0250	15,138	10,09	0,972	0,937
7	0,0249	15,161	10,06	0,973	0,941
8	0,0255	15,116	10,19	0,971	0,938
9	0,0247	15,209	9,89	0,973	0,940
10	0,0254	15,148	10,14	0,971	0,937
Ensemble 2					
Segment	MSE	SNR (dB)	Energy E	Correlation coefficient (ρ)	R_{xixj}
1	0,0762	11,435	40,12	0,906	0,812
2	0,0748	11,509	39,95	0,911	0,817
3	0,0774	11,370	40,50	0,903	0,808
4	0,0753	11,459	40,25	0,908	0,813
5	0,0739	11,553	39,78	0,912	0,819
6	0,0760	11,444	40,07	0,907	0,814
7	0,0745	11,479	39,89	0,910	0,815
8	0,0751	11,452	40,13	0,908	0,813
9	0,0736	11,575	39,68	0,913	0,820
10	0,0749	11,487	40,00	0,910	0,815

At the intermediate level, for Ensemble 1, MSE decreased by 75,5%, and SNR increased by 53,4%. The energy E decreased to 10,14, indicating more precise extraction of the medium-frequency components of the signal but also a loss of some useful signal. The correlation coefficient (ρ) increased from 0,887 to 0,971, indicating improved signal recovery accuracy but also an increase in mutual correlation, which is a drawback. For Ensemble 2, the average MSE decreased by 50,5%, and SNR increased by 38,2%. The correlation coefficient (ρ) rose to 0,910, indicating improved signal recovery accuracy, but it may also lead to signal overlap. The balance between the increase in ensemble signal volumes and mutual correlation is maintained.

At the fine level of time decomposition, the signal is further divided into smaller segments (Table 6).

Table 6 – Calculation of indicators for the fine level of time decomposition

Ensemble 1					
Segment	MSE	SNR (dB)	E	ρ	R_{xixj}
1	0,0052	20,346	2,01	0,991	0,967
2	0,0055	20,290	2,08	0,990	0,968
3	0,0051	20,404	1,98	0,992	0,970
4	0,0053	20,316	2,05	0,991	0,969
5	0,0054	20,301	2,06	0,991	0,967
...	...	...	...	...	...
50	0,0053	20,312	2,04	0,991	0,968
Ensemble 2					
Segment	MSE	SNR (dB)	E	ρ	R_{xixj}
1	0,0212	13,056	8,52	0,952	0,870
2	0,0209	13,110	8,48	0,953	0,872
3	0,0214	13,025	8,57	0,951	0,868
4	0,0211	13,073	8,50	0,953	0,871

© Bershov V.S., Yakymchuk N.M.

5	0,0208	13,135	8,46	0,954	0,873
...	...	...	...	...	...
50	0,0210	13,097	8,49	0,953	0,870

In fig. 4 shows the comparative characteristics of Ensembles 1 and 2.

For Ensemble 1, the average MSE value at the coarse level was 0,1011. At the intermediate level, this value decreased to 0,0253, a 75,1% reduction. This indicates a significant reduction in signal reconstruction errors due to detailed analysis and extraction of the signal's mid-frequency components. At the fine level, the average MSE value decreased further to 0,0102, a 59,7% reduction compared to the intermediate level. Such optimization indicates even more accurate extraction of high-frequency components and further reduction of errors in signal reconstruction. The overall efficiency of the MSE metric from the coarse to the fine level was 89,9%, indicating the high effectiveness of the multiscale time decomposition method in reducing signal reconstruction errors, considering the features of Ensemble 1 components, such as low-frequency and mid-frequency sinusoidal signals, high-frequency noise, and impulses.

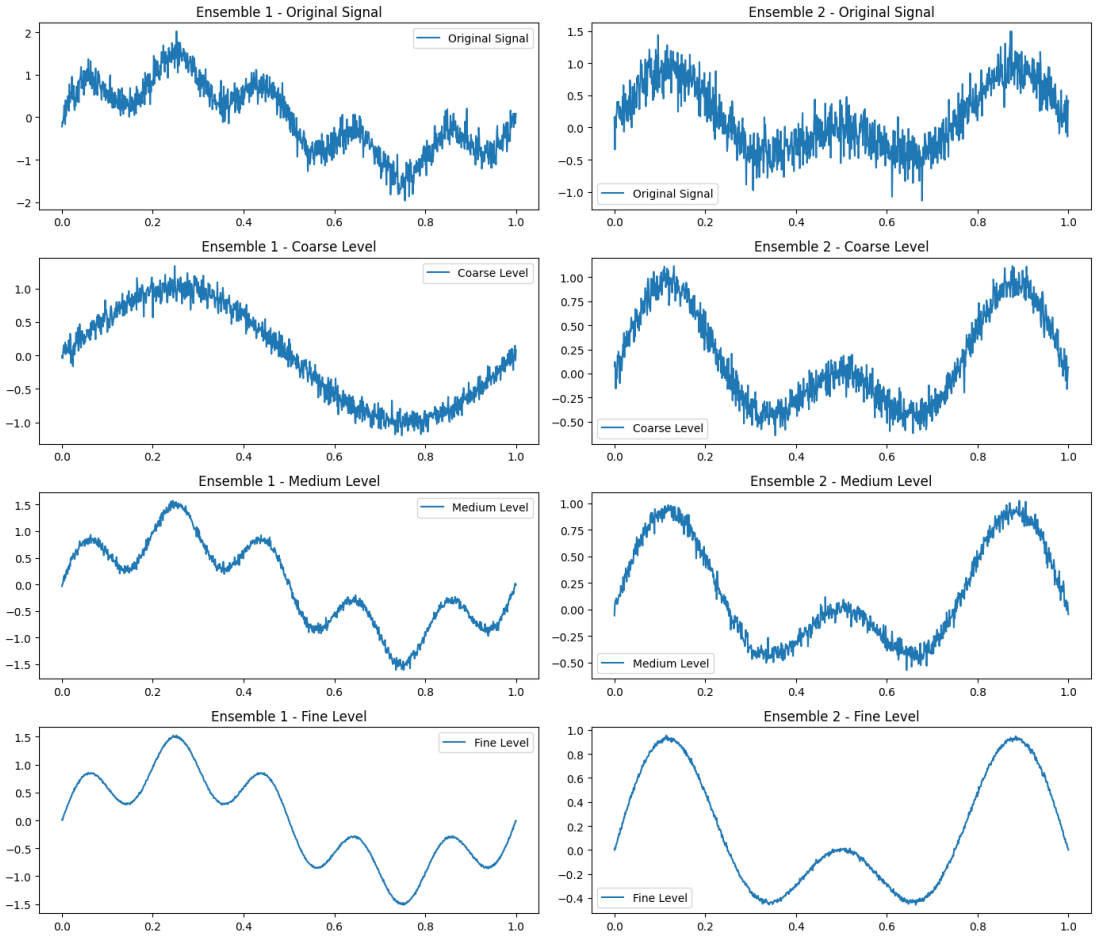


Fig. 4 – Optimization in the time domain on three levels

At the coarse level, the average SNR value for Ensemble 1 was 9,845 dB. At the intermediate level, this value increased to 15,150 dB, a 53,9% improvement, indicating a significant reduction in noise and improvement in signal reproduction. At the fine level, SNR increased to 20,150 dB, providing an additional optimization of 33,0% compared to the intermediate level. The overall improvement in the SNR metric from the coarse to the fine level is 104,6%, confirming the method's effectiveness in enhancing the signal-to-noise ratio.

The signal energy (E) is an indicator of signal stability. For Ensemble 1, the average energy value at the coarse level was 40,45. At the intermediate level, this value decreased to 10,11, indicating more accurate extraction of the signal's mid-frequency components and a reduction in noise impact. At the fine level, the

© Bershov V.S., Yakymchuk N.M.

average energy value further decreased to 5,15, confirming the method's effectiveness in extracting high-frequency components and reducing energy costs for signal processing.

The cross-correlation function (CCF) allows evaluating how well the reconstructed signal matches the original one at different time shifts, which is important for assessing delays and consistency between signals. For Ensemble 1, the CCF value at the coarse level was 0,800, at the intermediate level, it increased to 0,938, and at the fine level, it reached 0,951. These are fairly high values of cross-correlation between signals, but as noted above, this can occur when the volume of signal ensembles increases.

A comparison between Ensembles 1 and 2 shows that Ensemble 1's metrics were better optimized at each level of decomposition. This is due to the fact that Ensemble 1 contains less complex signals with a lower noise level.

Conclusions and prospects for further research.

This article proposed a three-level multiscale time decomposition method. The practical implementation of the algorithm for the proposed method substantiated the feasibility of its use for the analysis and processing of complex signals. Through detailed analysis and extraction of signal components at different levels of temporal detail, including coarse, medium, and fine levels, significant improvements in signal reconstruction accuracy were achieved.

Ensembles of complex signals obtained by time decomposition are advisable to use in conditions of high immunity, a large number of subscribers in cognitive radio networks, as well as in conditions of a high level of noise and interference. The use of this method helps to increase the productivity of the telecommunications system due to the optimization of the signal processing process and the reduction of computing costs.

Prospects for further research are the integration of the method with machine learning algorithms for automatic adaptation of decomposition parameters in real time, which will provide even greater efficiency and adaptability in the conditions of changing characteristics of the wireless cognitive telecommunication environment.

References

1. Melnyk V.G., Kravets O.V. Multiscale analysis of time series in forecasting problems. Collection of scientific works. Kyiv: Institute of Cybernetics named after V.M. Hlushkova. 2019. PP. 85-93.
2. Cohen L. Time-Frequency Analysis. Prentice Hall. 1995. ISBN: 978-0135945322.
3. Mallat S.A. Wavelet Tour of Signal Processing: The Sparse Way. Academic Press. 3rd Edition. 2008. ISBN: 978-0123743701.
4. Gao R.X., Yan R. Non-stationary signal processing for bearing health monitoring Int. J. Manufacturing Research, Vol. 1, No. 1, 2006. PP. 18-41.
5. Mirjalili S., Lewis A. The whale optimization algorithm. Advances in Engineering Software. 2016. PP. 51–67. DOI: https:/doi.org/10.1016/j.advengsoft.2016.01.008
6. Katipoğlu O.M, Sarıgöl M. Daily runoff prediction based on the adaptive fourier decomposition method and multiscale temporal convolutional network. Environmental Science and Pollution Research. 2023. PP. 1–18. DOI: https:/doi.org/10.1007/s11356-023-25796-5
7. Singh P., Joshi S.D., Patney R.K. et al. The Fourier decomposition method for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2017. 473(2199):20160871. DOI: https:/doi.org/10.1098/rspa.2016.0871
8. Ryan M. Decomposing Signal Using Empirical Mode Decomposition — Algorithm Explanation for Dummy. Towards Data Science, 2019. https://towardsdatascience.com/decomposing-signal-using-empirical-mode-decomposition-algorithm-explanation-for-dummy-93a93304c541
9. Percival D.B., Walden A.T. Wavelet Methods for Time Series Analysis. Cambridge University Press. 2000. ISBN: 978-0521685085.
10. Kang Peng, Hongyang Guo, Xueyi Shang. EEMD and Multiscale PCA-Based Signal Denoising Method and Its Application to Seismic P-Phase Arrival Picking. Sensors. 2021. 21(16). PP. 52-71. DOI: https:/doi.org/10.3390/s21165271
11. Zhang Y., Li R., Duan L. Spectrum sharing for cognitive radio networks: A multi-agent reinforcement learning approach. IEEE Transactions on Network and Service Management. 2021. Vol. 18(2). PP. 823-835.
12. Tse D., Viswanath P. Fundamentals of Wireless Communication. Cambridge University Press, 2005.
13. Stojanovic M., Preisig J. Underwater Acoustic Communication Channels: Propagation Models and Statistical Characterization. IEEE Communications Magazine, Vol. 47(1), 2009. PP. 84-89.
14. Sun Y., Zhou J., Wang M., Chang J., Ma G. Text Recognition Method for Handwritten and Natural Scene Text Image: A Review. Communications, Signal Processing, and Systems. SpringerLink, 2020.

© Bershov V.S., Yakymchuk N.M.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-40
УДК 621.391
Васильківський Микола Володимирович, к.т.н., доцент
https://orcid.org/0000-0002-6586-2563
Коломієць Альона Анатоліївна, д.пед.н., доцент
https://orcid.org/0000-0003-2120-7644
Будаш Михайло Володимирович, аспірант
Прикмета Андрій Володимирович, аспірант
Олійник Андрій Олегович, аспірант
Вінницький національний технічний університет, м. Вінниця, Україна

ТЕХНОЛОГІЧНІ АСПЕКТИ ВПРОВАДЖЕННЯ ПРОГРАМНО-КЕРОВАНИХ МЕРЕЖ 5G ТА 6G

Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О. Технологічні аспекти впровадження програмно-керованих мереж 5G та 6G. Результати дослідження технологічних аспектів впровадження програмно-керованих мереж (SDN) у контексті 5G та 6G спрямовані на оптимізацію підходів, методів і технологій для забезпечення максимального ефективного використання мережевих ресурсів, підвищення гнучкості, безпеки та якості обслуговування. Оптимальні топології забезпечують максимальне покриття, мінімізацію затримок та підвищення надійності з'єднання. Впровадження SDN та NFV дозволяє динамічно керувати ресурсами мережі, забезпечуючи її гнучкість та масштабованість. Автоматизовані системи управління, які використовують алгоритми машинного навчання, оптимізують використання ресурсів. Технології управління QoS та мережеві слайси дозволяють забезпечити високий рівень якості обслуговування для різних типів трафіку та додатків. Це особливо важливо для критично важливих додатків, які вимагають низької затримки та високої пропускної здатності. Мережі 5G та 6G підтримують широкий спектр нових додатків і сервісів завдяки використанню гнучкої та масштабованої інфраструктури, інтегрованої з технологіями штучного інтелекту та машинного навчання. Дослідження технологічних аспектів впровадження програмно-керованих мереж у 5G та 6G дозволило визначити оптимальні підходи та методи для забезпечення ефективності, гнучкості, безпеки та високої якості обслуговування сучасних та майбутніх телекомунікаційних мереж. Мережі 5G та 6G забезпечують підтримку широкого спектра нових додатків і сервісів завдяки використанню гнучкої та масштабованої інфраструктури, інтегрованої з технологіями штучного інтелекту та машинного навчання. Виконані задачі та досягнуті результати сприяють розвитку інноваційних технологій та забезпечують відповідність мережевої інфраструктури зростаючим вимогам ринку та користувачів. Дослідження показало, що ефективний розподіл мережевих ресурсів у межах оптимізованої топології мереж 5G та 6G є можливим завдяки впровадженню програмно-керованих мереж (SDN), віртуалізації мережевих функцій (NFV) та розробці динамічних алгоритмів управління ресурсами. Це забезпечує високу продуктивність, гнучкість та надійність мережі, задовольняючи зростаючі вимоги користувачів і підтримуючи інноваційні сервіси та додатки.

Ключові слова: оптичний лінійний термінал, оптичний сплітер, пасивна оптична мережа, оптичний мережевий термінал, інфокомунікаційна мережа з багатопротокольною комутацією за мітками на основі IP адрес, агрегаційний транспортний вузол, граничний транспортний вузол, динамічний розподіл спектру, програмно-керована мережа

Vasylkivskyi M., Kolomiiets A., Budash M., Prykmeta A., Oliinyk A. Technological aspects of implementing software-configurable 5G and 6G networks. The results of the study of technological aspects of software-defined networking (SDN) implementation in the context of 5G and 6G are aimed at optimizing approaches, methods and technologies to ensure maximum efficient use of network resources, increase flexibility, security and quality of service. Optimal topologies maximize coverage, minimize latency, and improve connection reliability. The introduction of SDN and NFV allows for dynamic management of network resources, ensuring its flexibility and scalability. Automated management systems that use machine learning algorithms optimize resource utilization. QoS management technologies and network slices allow you to provide a high level of quality of service for different types of traffic and applications. This is especially important for mission-critical applications that require low latency and high bandwidth. 5G and 6G networks support a wide range of new applications and services through the use of flexible and scalable infrastructure integrated with artificial intelligence and machine learning technologies. The study of the technological aspects of implementing software-defined networks in 5G and 6G has allowed us to identify the best approaches and methods to ensure the efficiency, flexibility, security and high quality of service of current and future telecommunications networks. The 5G and 6G networks support a wide range of new applications and services through the use of flexible and scalable infrastructure integrated with artificial intelligence and machine learning technologies. The tasks per-formed and the results achieved contribute to the development of innovative technologies and ensure that the network infrastruc-ture meets the growing requirements of the market and users. The study has shown that efficient allocation of network resources within the optimized topology of 5G and 6G networks is possible through the introduction of software-defined networks (SDN), network function virtualization (NFV), and the development of dynamic resource management algorithms. This ensures high performance, flexibility and reliability of the network, meeting the growing demands of users and supporting innovative services and applications.

Keywords: optical line terminal, optical splitter, passive optical network, optical network terminal, IP-based multiprotocol label switching information and communication network, aggregation transport node, edge transport node, dynamic spectrum allocation, software-controlled network

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

Постановка наукової проблеми. Впровадження програмно-керованих мереж на основі 5G і 6G є необхідним кроком для досягнення гнучкості, масштабованості та ефективності сучасних телекомунікаційних систем. SDN дозволяє ефективно управляти мережевими ресурсами, оптимізувати трафік, забезпечувати високу якість обслуговування, а також підвищувати безпеку та надійність мереж. Зі впровадженням 6G ці технології стануть ще більш важливими, відкриваючи нові можливості для інновацій та розвитку.

Метою дослідження технологічних аспектів впровадження програмно-керованих мереж (Software-Defined Networking, SDN) у контексті 5G та 6G є дослідження та оптимізація підходів, методів і технологій, які дозволять забезпечити максимальну ефективність, гнучкість, безпеку та якість обслуговування в сучасних і майбутніх телекомунікаційних мережах. Це дослідження охоплює кілька ключових аспектів, які мають на меті задовольнити зростаючі вимоги до мережевої інфраструктури.

Основні цілі дослідження:

- Дослідження методів і підходів для динамічного та гнучкого управління мережевими ресурсами, що дозволить швидко адаптуватися до змін в умовах трафіку та вимог користувачів за рахунок використання централізованого управління, яке дозволяє легко змінювати конфігурацію мережі відповідно до поточних потреб та впровадження автоматизованих систем управління, які використовують алгоритми машинного навчання для оптимізації використання ресурсів;

- дослідження особливостей забезпечення ефективного розподілу мережевих ресурсів для максимального використання доступної пропускної здатності та мінімізації затримок за рахунок віртуалізації мережевих функцій (NFV) для створення гнучких та економічно ефективних мережевих архітектур та впровадження технологій динамічного розподілу ресурсів, які дозволяють швидко переналаштовувати ресурси у відповідь на зміни трафіку;

- дослідження особливостей забезпечення високої якості обслуговування для різних типів додатків та сервісів, включаючи ті, що вимагають низької затримки, високої пропускної здатності та надійності за рахунок впровадження технологій управління QoS, що дозволяють пріоритезувати трафік та забезпечувати необхідний рівень сервісу для критичних додатків та використання мережевих слайсів (Network Slicing) для створення віртуальних мереж, оптимізованих для конкретних сервісів і потреб користувачів.

- Створення умов для впровадження нових технологій, таких як Інтернет речей (IoT), автономні системи, розумні міста та інші інноваційні додатки за рахунок передбачення підтримки широкого спектра нових додатків і сервісів завдяки використанню гнучкої та масштабованої мережевої інфраструктури та інтеграції з технологіями штучного інтелекту та машинного навчання для автоматизації управління мережею і оптимізації її роботи.

Оскільки дослідження технологічних аспектів впровадження програмно-керованих мереж у 5G та 6G полягає в забезпеченні підвищення ефективності, гнучкості, безпеки та високої якості обслуговування в сучасних та майбутніх телекомунікаційних мережах, тому задачі дослідження спрямовані на визначення інноваційних рішень, які дозволять мережам адаптуватися до змін, оптимізувати використання ресурсів та підтримувати нові технології, забезпечуючи тим самим відповідність сучасним вимогам ринку та користувачів.

Дослідження особливостей топологій мереж 5G та мережевого розгалуження є важливим кроком для розуміння і оптимізації роботи сучасних телекомунікаційних мереж. Основні задачі такого дослідження спрямовані на забезпечення ефективності, надійності, гнучкості та високої якості обслуговування в мережах 5G. Наведемо ключові задачі, які необхідно вирішити в рамках такого дослідження:

- Визначити оптимальну топологію мережі для забезпечення максимального покриття, мінімізації затримок та підвищення надійності з'єднання.

- Використання алгоритмів динамічної маршрутизації та програмно-визначених мереж (SDN) для оптимального розподілення трафіку.

- Дослідити вплив впровадження обчислювальних потужностей на край мережі (Edge Computing) на зменшення затримки та підвищення ефективності роботи додатків реального часу.

Аналіз останніх досліджень і публікацій. Дослідження технологічних аспектів впровадження програмно-керованих мереж (SDN) у контексті 5G та 6G є надзвичайно актуальним. Це зумовлено потребою у гнучкості, масштабованості, ефективності використання ресурсів, підвищенні якості обслуговування, безпеці та підтримці інноваційних сервісів. Впровадження SDN дозволить забезпечити адаптивність мереж до змін у попиті, ефективне управління ресурсами,

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

високий рівень безпеки та підтримку сучасних додатків і сервісів.

Бездротові системи зв'язку продовжують стрімко розвиватися, досягнувши етапу 5G, що забезпечує швидкість передачі даних до 20 ГГц. Завдяки цьому виникли нові концепції, такі як наднадійний низькошвидкісний зв'язок (uRLLC), розширений мобільний широкосмуговий зв'язок (eMBB) та масовий зв'язок машинного типу (mMTC). З'явився перехід від Інтернету речей (IoT) до Інтернету всього (IoE), що зробило комунікаційні системи, додатки та послуги більш інтелектуальними. Важливу роль у цьому відіграють такі технології, як штучний інтелект, блокчейн та великі дані. Завдяки цим інноваціям, розумні будинки, міста, інноваційні системи охорони здоров'я та автономні транспортні засоби стали невід'ємною частиною нашого життя [1].

5G додатки вже стали частиною нашого життя, а з 2030-х років очікується впровадження систем зв'язку шостого покоління (6G). Разом з розвитком 6G комунікаційних інфраструктур розглядаються концепції 3D-мереж, інтелектуальних мереж, квантового зв'язку, блокчейн-технологій, глибокого навчання та програмованих поверхонь. Інтернет речей (IoT) залишається ключовим елементом бездротової інфраструктури разом із системами 4G, 5G та майбутніми 6G. Методи доступу, протоколи та питання безпеки для цих систем також отримують детальне пояснення. Окрім цих методів доступу, розглядаються методи, що використовуються у з'єднаннях M2M (машина до машини) та IoT на кінцевих точках. Безпека стає значним викликом у цій екосистемі, де все пов'язано з усім [2].

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Рівень доступу бездротової системи передачі даних 5G показаний на рис. 1. Як видно, гібридні системи доступу в мережах 5G спочатку підключаються через фронтальну магістраль, потім збираються на середньому рівні та доставляються в основну мережу 5G через багатопротокольну комутацію міток (MPLS). Такий підхід дозволяє ефективно об'єднувати різні типи трафіку і забезпечує надійний та швидкий доступ до мережевих ресурсів, оптимізуючи використання існуючої інфраструктури і підвищуючи загальну продуктивність мережі [3].

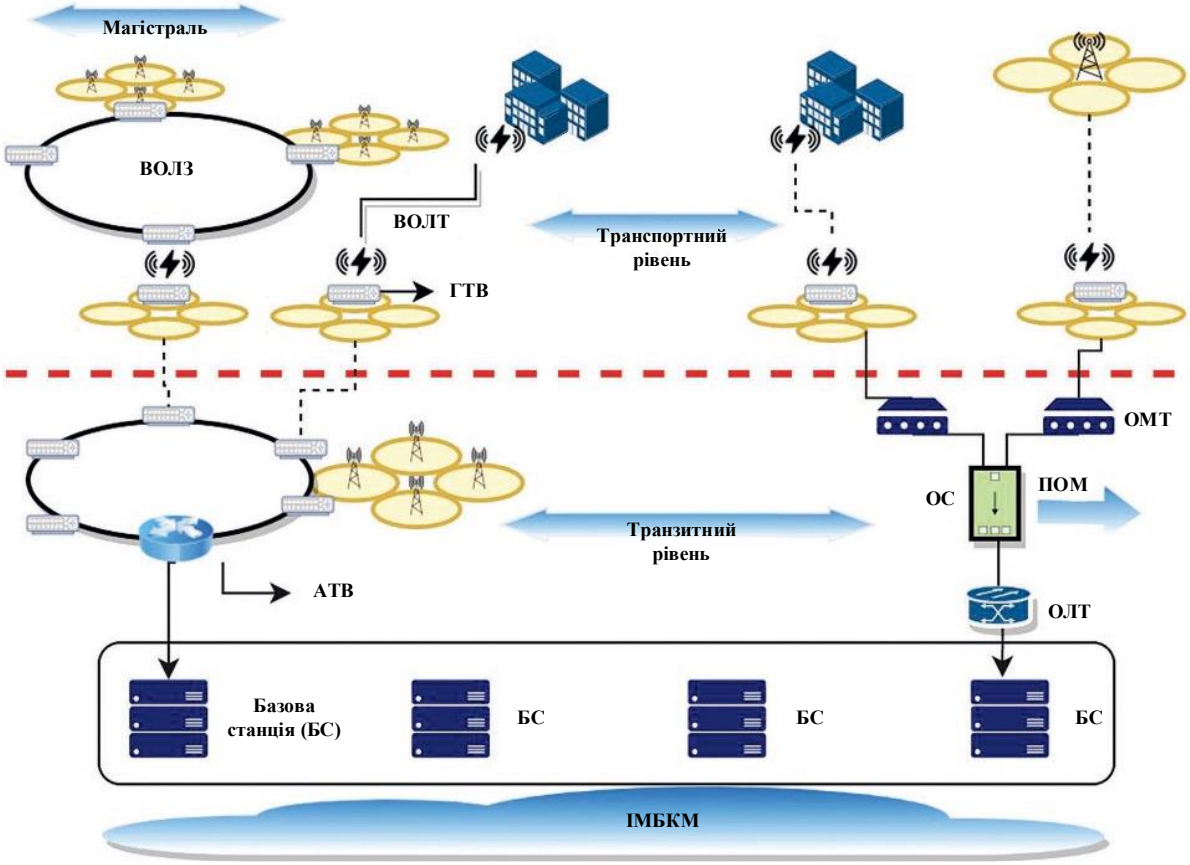


Рис. 1. Топологія сегменту мережі доступу 5G

Найважливішою зміною в мережах доступу 5G стало видалення рівня контролера радіомережі

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

(RNC) для мінімізації затримок і підключення базових станцій безпосередньо до опорної мережі через блоки базового діапазону. У порівнянні з 4G, в базовій мережі 5G відбулися революційні зміни в архітектурі, платформі, функціях, протоколах та інших сферах для задоволення вимог щодо кастомізації мережі, обслуговування, високої пропускної здатності, високої продуктивності та низької вартості (див. рис. 2) [4].

Вказані зміни можна розглянути з чотирьох основних аспектів. Еволюція в бік хмарних обчислень на основі нових розробок, функціонального програмного забезпечення, обчислень і функцій аналізу даних. Це означає перехід від традиційних мережевих пристроїв до хмарних мережевих функцій. Перехід від фіксованої жорсткої структури мережевих з'єднань між пристроями до динамічно конфігурованих гнучких мереж. Це включає в себе сервіс-орієнтовану структуру і нову архітектуру основного мережевого протоколу на основі інтернет-протоколу HTTP/2.0. Мінімальний дизайн для пересилання/маршрутизації даних і доступу користувачів. Мета полягає у спрощенні мережевої інфраструктури для підвищення ефективності та зменшення затримок. Архітектура мережі 5G орієнтована на вертикальні галузі промисловості. Концепція «мережа як послуга» (Network as a Service, NaaS) реалізується завдяки розгалуженню мережі, периферійним обчисленням і низькошвидкісному з'єднанню, що дозволяє переходити від універсальної послуги до персоналізованого та індивідуального сервісу. Такі підходи дозволяють мережам 5G відповідати високим вимогам користувачів та промисловості, забезпечуючи гнучкість, ефективність і масштабованість [5].

Мережа 5G еволюціонувала від закритої системи до відкритої сервіс-орієнтованої архітектури, що забезпечує гнучку оркестрацію, аналіз і відкритість. Вона розбиває складні "окремі мережеві елементи" на модульні послуги, що підвищує гнучкість мережі. Кожна мережна функція складається з багатьох послуг, які формуються у невеликі сервісні модулі відповідно до потреб. Це робить архітектуру важливим інструментом для задоволення потреб вертикальних галузей в мережах 5G.

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

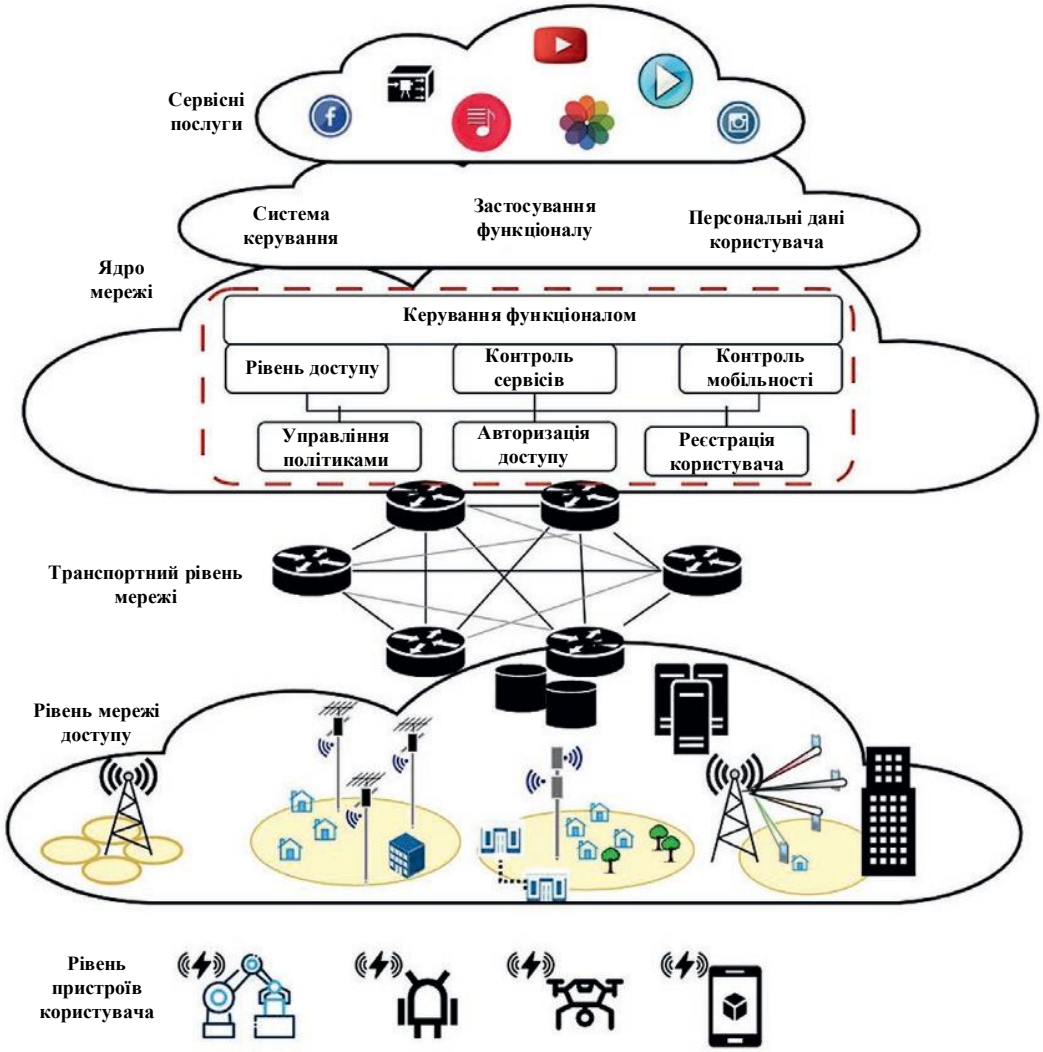


Рис. 2. Багаторівнева структура сегменту мережі 5G

Програмно-визначена мережа (SDN) передбачає розділення програмного забезпечення, що працює в мережі, і алгоритмів, що працюють у площині управління мережею. Це означає, що площина управління, де здійснюється управління мережею, і площина користувача, де передаються дані користувачів, розділені. За допомогою SDN мережеві пристрої керуються через програмні засоби управління або інтерфейси прикладного програмування (API), незалежно від виробника. SDN дозволяє створювати і контролювати віртуальні мережі або традиційне обладнання за допомогою програмного забезпечення. Віртуалізація мережі дозволяє організаціям сегментувати різні віртуальні мережі в межах однієї фізичної мережі або з'єднувати пристрої в різних фізичних мережах для створення єдиної віртуальної мережі. Таким чином, досягається високошвидкісне і гнучке управління мережевими пристроями, налаштовується мережева інфраструктура і забезпечується високий рівень безпеки.

Віртуалізація мережевих функцій (NFV) означає розділення програмного та апаратного рівнів комунікаційних мереж. Це дозволяє прозоро керувати такими функціями, як маршрутизація, брандмауер і балансування навантаження через віртуальні сервери. NFV забезпечує гнучке управління мережею і високий рівень безпеки, дозволяючи створювати віртуальні мережі на доступних серверах за допомогою технології віртуалізації.

Мережа 5G дозволяє інтегроване управління всією мережею і регіоном (рис. 3). Управління повним життєвим циклом мережі здійснюється гнучко, що дозволяє швидко налаштовувати мережеві з'єднання і маршрутизацію між центрами обробки даних. Завдяки цим інноваціям можна створити відкриту та розвинену архітектуру і послуги, орієнтовані на клієнта [6].

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

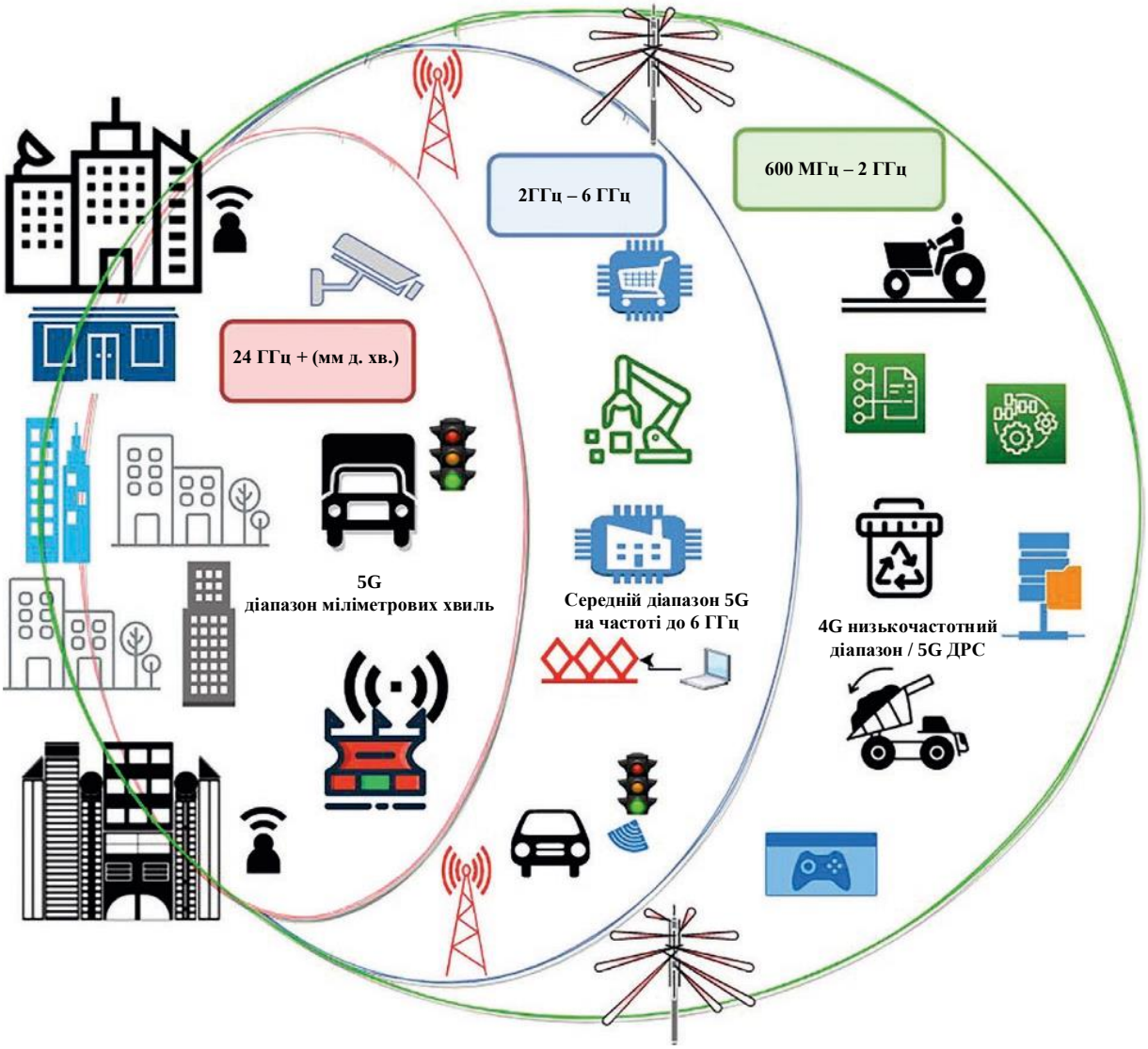


Рис. 3. Інтегрований функціонал сегменту мережі 5G

Завдяки трьом основним типам послуг 5G — eMBB, uRLLC і mMTC — користувачі отримали доступ до таких інноваційних сервісів, як електронна охорона здоров'я, доповнена і віртуальна реальність, інтелектуальні транспортні системи, розумні міста, розумні будівлі та мобільні ігри. Для ефективного надання цих послуг мережі 5G використовують підхід сегментації або мережевого розшарування (network slicing). Мережеве розшарування — це архітектурний підхід, який дозволяє створювати кілька підмереж (зрізів) у межах однієї фізичної інфраструктури. Кожен зріз призначений для підтримки конкретного типу сервісів і має свої характеристики і вимоги. Отже, SDN дозволяють централізовано керувати мережевою інфраструктурою за допомогою програмних засобів, що забезпечує гнучкість і динамічну конфігурацію мережі. Це полегшує створення і управління різними мережевими зрізами, дозволяє швидко адаптуватися до змін у вимогах сервісів. NFV дозволяє віртуалізувати мережеві функції, такі як маршрутизація, брандмауери і балансування навантаження, і виконувати їх на загальних серверних платформах. Це знижує витрати на обладнання, підвищує гнучкість і масштабованість мережі, дозволяє створювати віртуальні мережі для різних сервісів.

Мережі бездротового зв'язку 5G базуються на трьох основних видах сервісів, які роблять їх надзвичайно гнучкими і здатними задовольнити різноманітні потреби галузей і користувачів. Розглянемо кожен із цих сервісів більш детально.

Масивний машинний зв'язок (mMTC) передбачає підключення величезної кількості пристроїв, що використовуються для Інтернету речей (IoT) та підтримку високого рівня підключення для пристроїв з низьким енергоспоживанням і мінімальними вимогами до пропускної здатності. Пристрої mMTC споживають дуже мало енергії, що дозволяє їм працювати від батарей

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

протягом тривалого часу.

Приклади використання: мережі сенсорів для моніторингу навколишнього середовища; розумні лічильники для вимірювання енергоспоживання; системи відстеження логістики і постачання; персональна автоматизація і системи безпеки [7].

Наднадійний зв'язок з низькою затримкою (uRLLC) здійснює забезпечення високонадійного зв'язку з мінімальною затримкою, що важливий для застосувань, де швидкість і надійність передачі даних мають вирішальне значення.

Приклади використання: автономні транспортні засоби та системи V2X (Vehicle-to-Everything); дистанційне керування промисловим обладнанням; роботизована хірургія та інші медичні застосування; промислова автоматизація і контроль процесів.

Розширений мобільний широкосмуговий зв'язок (eMBB) передбачає забезпечення високої пропускної здатності та швидкості передачі даних для підтримки додатків з інтенсивним використанням трафіку.

Приклади використання: потокове відео високої роздільної здатності (4K/8K); віртуальна і доповнена реальність (VR/AR); хмарні ігри та інші мультимедійні сервіси; відеоконференції з високою роздільною здатністю.

Отже, вказані типи сервісів роблять мережі 5G надзвичайно гнучкими та здатними задовольнити потреби різноманітних галузей і користувачів, від масових IoT пристроїв до високошвидкісних мультимедійних додатків і критично важливих застосувань, де кожна мілісекунда має значення. Завдяки mMTC, uRLLC і eMBB, 5G відкриває нові можливості для розвитку і впровадження інноваційних технологій у різних сферах життя та бізнесу (рис. 4).

Інфраструктура бездротового зв'язку розвивається в напрямку створення екосистеми, в якій люди і машини завжди пов'язані та взаємодіють з кількома пристроями одночасно. Щоб забезпечити таку щільно з'єднану інфраструктуру, бездротові мережі наступного покоління повинні бути гнучкими, масштабованими і здатними поєднувати різні архітектури та стандарти. Крім того, очікується, що ці мережеві інфраструктури забезпечать відповідні рішення для різних типів трафіку і гетерогенних типів мереж з такими основними вимогами, як ефективність, велика кількість підключених пристроїв, низька затримка та висока надійність. Для досягнення цих цілей мережі 5G використовують сучасні технології, такі як SDN, NFV, хмарні обчислення та периферійні обчислення. Це дозволяє забезпечити гнучку, ефективну та надійну інфраструктуру, здатну задовольнити потреби різноманітних користувачів та галузей [8].

Мережеве розшарування дозволяє адаптувати мережу до конкретних потреб кожного типу сервісів. Це зменшує витрати на інфраструктуру, оптимізуючи використання фізичних ресурсів. Полегшує масштабування мережі для підтримки зростаючої кількості підключених пристроїв і сервісів. Забезпечує можливість надання індивідуальних рішень для різних вертикальних галузей, таких як охорона здоров'я, транспорт і розумні міста. Завдяки використанню технологій SDN та NFV, мережеве розшарування формує основу для гнучкої та ефективної архітектури мереж 5G, здатної підтримувати різноманітні сучасні та майбутні сервіси. Мережеве розшарування є ключовою технологією в архітектурі 5G, яка дозволяє створювати окремі віртуальні бездротові мережі для різних послуг з динамічною пропускною здатністю. Це забезпечує оптимізований розподіл ресурсів і топологію мережі для кожної ситуації, забезпечуючи бездротову інфраструктуру з певною угодою про рівень обслуговування (SLA).

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

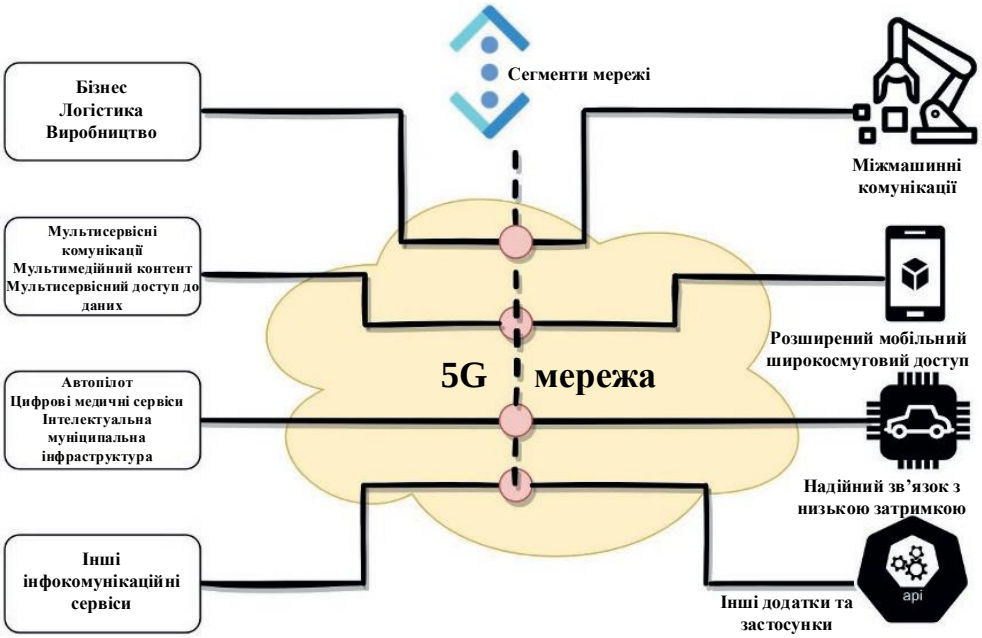


Рис. 4. Розподілення функціонального ресурсу інфокомунікаційної мережі

Для кожної послуги, такої як eMBB, uRLLC і mMTC, створюються окремі віртуальні мережі. Це дозволяє ефективно управляти ресурсами і задовольняти специфічні потреби кожної послуги. Мобільний широкосмуговий зв'язок працює в одному зрізі, забезпечуючи високу пропускну здатність і швидкість передачі даних. Наднадійний зв'язок з низькою затримкою працює в іншому зрізі, забезпечуючи мінімальну затримку та високу надійність. Масовий машинний зв'язок функціонує в окремому зрізі, підтримуючи велику кількість пристроїв з низьким енергоспоживанням. У кожному зрізі використовуються різні частоти як для доступу кінцевих пристроїв, так і для доступу до основної мережі. Це забезпечує оптимальне використання спектру і мінімізує інтерференцію між різними типами трафіку.

Мережеве розшарування дозволяє гнучко налаштовувати і масштабувати мережу в залежності від потреб користувачів і сценаріїв застосування. Це робить бездротову екосистему 5G гнучкою, надійною, масштабованою і керованою системою. Розподіл мережевих ресурсів оптимізується відповідно до потреб кожного зрізу, що забезпечує ефективне використання ресурсів і високу якість обслуговування. Мережеве розшарування підтримується технологіями віртуалізації мережевих функцій (NFV) і програмно-визначених мереж (SDN). Це дозволяє створювати віртуальні підмережі і резервувати ресурси відповідно до потреб клієнтів. Промислова віртуалізація з мережевим розшаруванням дозволяє налаштовувати мережеву інфраструктуру відповідно до специфічних вимог різних галузей. Кожен зріз налаштовується відповідно до вимог конкретного застосування або користувача, забезпечуючи високий рівень персоналізації та задоволення потреб. Оптимізований розподіл ресурсів і віртуалізація мережевих функцій знижують витрати на інфраструктуру і підвищують ефективність використання мережевих ресурсів. Забезпечення окремих зрізів для різних типів трафіку підвищує надійність і продуктивність мережі, дозволяючи обслуговувати різні сервіси з високою якістю. Таким чином, мережеве розшарування у 5G створює основу для гнучкої, ефективної і масштабованої бездротової інфраструктури, здатної задовольнити різноманітні потреби користувачів і промисловості [1].

Інфраструктура 6G, яка розвивається, базується на використанні великих даних (Big Data), що генеруються всіма підключеними пристроями в екосистемі Інтернету всього (Internet of Everything, IoE). Ці дані компілюються і збираються за допомогою методів штучного інтелекту (ШІ) nf передаються на сервери в кібернетичному світі. В подальшому, методи оцінювання і виявлення знань будуть застосовуватися для перетворення зібраних даних у цінну інформацію, що стане практичними інструментами для дій у фізичному світі. Розглянемо тенденції та технології, які визначатимуть розвиток мереж 6G.

Використання ШІ для компіляції, аналізу та інтерпретації великих обсягів даних, що генеруються пристроями IoE. Впровадження методів оцінювання і виявлення знань для

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

перетворення зібраних даних у корисну інформацію.

Забезпечення підключеності для всіх типів пристроїв і систем, включаючи персональні гаджети, промислові машини та побутові прилади. Розширення використання сенсорних мереж для збору даних з фізичного світу.

Використання даних, зібраних у кібернетичному світі, для прийняття рішень і виконання дій у фізичному світі.

Використання нових частотних діапазонів для забезпечення вищих швидкостей передачі даних і більшої пропускної здатності. Розвиток антенних систем для поліпшення якості сигналу і покриття.

Впровадження квантових обчислень і квантового зв'язку для підвищення безпеки і продуктивності мереж.

Використання SDN і NFV для створення гнучких і адаптивних мереж, що можуть динамічно змінювати свою конфігурацію відповідно до потреб користувачів. Застосування концепції мережевого розшарування для підтримки різних типів сервісів і оптимізації використання ресурсів.

Розробка автономних систем управління мережами, що використовують ШІ для автоматичного налаштування і оптимізації мережевих параметрів.

Впровадження технологій, що знижують енергоспоживання мережевих компонентів і пристроїв. Використання відновлюваних джерел енергії для живлення мережевих інфраструктур.

Налаштування мережевих послуг відповідно до індивідуальних потреб користувачів і галузей, забезпечуючи високий рівень персоналізації.

Використання технологій 6G для створення розумних міст з інтегрованими системами управління транспортом, енергоспоживанням, безпекою та іншими аспектами міського життя.

Отже, розвиток 6G базуватиметься на поєднанні новітніх технологій і підходів для створення високоефективної, гнучкої та адаптивної мережевої інфраструктури. Використання штучного інтелекту, великих даних, квантових обчислень, інтернету всього та інших передових технологій забезпечить створення нових можливостей для бізнесу та покращення якості життя користувачів [2].

Висновки та перспективи подальших досліджень. Зміни, які відбуваються в мережах бездротового зв'язку, вже мають значний вплив на різні аспекти нашого життя і бізнесу. З впровадженням 5G і майбутнім розвитком 6G, ці зміни стануть ще більш помітними і масштабними. Використання машинного навчання та штучного інтелекту для прогнозування попиту на трафік та адаптації мережі в режимі реального часу дозволяє забезпечити оптимальне розподілення ресурсів і мінімізувати затримки. Використання програмно-визначених мереж (SDN) і мережної функціональної віртуалізації (NFV) забезпечить динамічне управління мережевими ресурсами і швидке адаптування до змін в трафіку.

Обробка даних на крайових пристроях ближче до джерела даних зменшує затримку і знижує навантаження на магістральні мережі. Це особливо важливо для додатків, що вимагають реального часу, таких як автономні транспортні засоби та промислова автоматизація. Додаткова оптимізація обчислень ще ближче до пристроїв-користувачів дозволяє ще більш ефективно використовувати ресурси і забезпечувати високу продуктивність.

З 6G очікується, що жодна область не залишиться без покриття, включаючи віддалені та важкодоступні регіони. Це забезпечить глобальний доступ до високошвидкісного інтернету та підключення до мережі. Системи бездротового зв'язку 6G стануть інтегрованими з фізичним середовищем, створюючи своєрідний кіберорганізм. Це включає інтеграцію з інтелектуальними системами, такими як інтернет речей (IoT), розумні міста і навіть біологічні системи.

Впровадження нових технологій дозволить автоматизувати багато бізнес-процесів, підвищуючи їх ефективність і продуктивність. З'являться нові бізнес-моделі і можливості, такі як розширена реальність, дистанційне керування і передбачуване обслуговування. Розширене покриття і швидкий інтернет забезпечать доступ до якісних медичних послуг, освіти та розваг навіть у віддалених районах. Поліпшення якості зв'язку сприятиме більш інтерактивним та багатофункціональним способам спілкування та співпраці.

Розвиток 5G та майбутній перехід до 6G обіцяють значні зміни в технологічній інфраструктурі, що вплинуть на всі аспекти нашого життя. Інтелектуальне управління трафіком, обчислювальна інфраструктура, повне покриття і концепція кіберорганізму зроблять мережі бездротового зв'язку більш ефективними, надійними і всеохоплюючими. Це відкриє нові можливості для бізнесу, покращить якість життя і сприятиме розвитку інновацій.

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

Список бібліографічного опису

1. Васильківський, М. В. Програмні технології в інфокомунікаційних системах. Навчальний посібник для студентів спеціальності 172 «Телекомунікації та радіотехніка» : електронний навчальний посібник комбінованого (локального та мережного) використання [Електронний ресурс] / Васильківський М. В., Бортник Г. Г., Кичак В. М. – Вінниця : ВНТУ, 2023. – 141 с.
2. Васильківський, М., Коломієць, А., & Грабчак, Н. (2022). Дослідження функціональних параметрів інфокомунікаційних мереж 6G. Вісник Хмельницького національного університету, (6), 46–52. https://www.doi.org/10.31891/2307-5732-2022-315-6-46-52
3. Васильківський, М., Прикмета, А., Олійник, А., & Нікітович, Д. (2023). Оптимізація інтелектуальних телекомунікаційних мереж. Вісник Хмельницького національного університету, Технічні науки. – 2023. – № 1. (317). – С. 33–41. doi: 10.31891/2307-5732-2023-317-1-33-41
4. М. Васильківський, О. Городецька, Б. Климчук, і В. Говорун, «Стратегії технологічного розвитку апаратного забезпечення інфокомунікаційних радіомереж», ITKI, вип. 56, вип. 1, с. 83–91, Бер 2023.
5. Chowdhury, M.Z., Shahjalal, M., Ahmed, S., and Jang, Y.M. (2020). 6G wireless communication systems: applications, requirements, technologies, challenges, and research directions. IEEE Open Journal of the Communications Society 1: 957–975.
6. Tripathi, S., Sabu, N.V., Gupta, A.K., and Dhillon, H.S. (2021). Millimeter- wave and terahertz spectrumfor 6G wireless. In: 6G Mobile Wireless Networks, 83–121. Cham: Springer International Publishing.
7. ElMossallamy, M.A., Zhang, H., Song, L. et al. (2020). Reconfigurable intelligent surfaces for wireless communications: principles, challenges, and opportunities. IEEE Transactions on Cognitive Communications and Networking 6 (3): 990–1002.
8. Chaoub, A., Giordani, M., Lall, B. et al. (2022). 6G for bridging the digital divide: wireless connectivity to remote areas. IEEE Wireless Communications 29 (1): 160–168.

References

1. [Vasylkivskyi, M. V. Prohramni tekhnolohii v infokomunikatsiinykh systemakh. Navchalnyi posibnyk dlia studentiv spetsialnosti 172 «Telekomunikatsii ta radiotekhnika» : elektronnyi navchalnyi posibnyk kombinovanoho (lokalnoho ta merezhnoho) vykorystannia [Elektronnyi resurs] / Vasylkivskyi M. V., Bortnyk G. G., Kychak V. M. – Vinnytsia : VNTU, 2023. – 141 s. [in Ukrainian].
2. Vasylkivskyi, M., Kolomiiets, A., & Hrabchak, N. (2022). Doslidzhennia funktsionalnykh parametriv infokomunikatsiinykh merezh 6G. Visnyk Khmelnytskoho natsionalnoho universytetu, (6), 46–52. https://www.doi.org/10.31891/2307-5732-2022-315-6-46-52 [in Ukrainian].
3. Vasylkivskyi, M., Prykmeta, A., Oliinyk, A., & Nikitovych, D. (2023). Optymizatsiia intelektualnykh telekomunikatsiinykh merezh. Visnyk Khmelnytskoho natsionalnoho universytetu, Tekhnichni nauky. – 2023. – № 1. (317). – S. 33–41. doi: 10.31891/2307-5732-2023-317-1-33-41 [in Ukrainian].
4. M. Vasylkivskyi, O. Horodetska, B. Klymchuk, i V. Hovorun, «Stratehii tekhnolohichnoho rozvytku aparatnoho zabezpechennia infokomunikatsiinykh radiomerezh», ITKI, vyp. 56, vyp. 1, s. 83–91, Ber 2023. [in Ukrainian].
5. Chowdhury, M.Z., Shahjalal, M., Ahmed, S., and Jang, Y.M. (2020). 6G wireless communication systems: applications, requirements, technologies, challenges, and research directions. IEEE Open Journal of the Communications Society 1: 957–975.
6. Tripathi, S., Sabu, N.V., Gupta, A.K., and Dhillon, H.S. (2021). Millimeter- wave and terahertz spectrumfor 6G wireless. In: 6G Mobile Wireless Networks, 83–121. Cham: Springer International Publishing.
7. ElMossallamy, M.A., Zhang, H., Song, L. et al. (2020). Reconfigurable intelligent surfaces for wireless communications: principles, challenges, and opportunities. IEEE Transactions on Cognitive Communications and Networking 6 (3): 990–1002.
8. Chaoub, A., Giordani, M., Lall, B. et al. (2022). 6G for bridging the digital divide: wireless connectivity to remote areas. IEEE Wireless Communications 29 (1): 160–168.

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-41
УДК 621.391
Васильківський Микола Володимирович, к.т.н., доцент
https://orcid.org/0000-0002-6586-2563
Коломієць Альона Анатоліївна, д.пед.н., доцент
https://orcid.org/0000-0003-2120-7644
Будаш Михайло Володимирович, аспірант
Прикмета Андрій Володимирович, аспірант
Олійник Андрій Олегович, аспірант
Вінницький національний технічний університет, м. Вінниця, Україна

ЗАБЕЗПЕЧЕННЯ ЯКОСТІ ОБСЛУГОВУВАННЯ ІНТЕРНЕТУ МЕДИЧНИХ РЕЧЕЙ В МЕРЕЖІ СТАНДАРТУ 5G

Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О. Забезпечення якості обслуговування інтернету медичних речей в мережі стандарту 5G. У роботі вирішується актуальна науково-прикладна проблема забезпечення якості обслуговування (QoS) Інтернету медичних речей (IoMT) в мережі стандарту LTE/5G. Запропоновано класифікацію трафіку за пріоритетами, яка включає чотири класи якості, що дозволяє забезпечити пріоритетну обробку критичного медичного трафіку і оптимізувати використання мережевих ресурсів. Розглянуті рішення дозволяють забезпечити необхідні параметри QoS для критичних медичних додатків, таких як мінімальна затримка і висока надійність передачі даних, що є критично важливим для систем охорони здоров'я. Здійснено комплексний аналіз проблем управління якістю обслуговування (QoS) для Інтернету медичних речей (IoMT) у мережах 5G. Розглянуто взаємодію мережі стандарту 5G з мережею IoMT, а також особливості підвищення QoS для медичних пристроїв і додатків. Визначено критичні вимоги до якості обслуговування медичних додатків, включаючи мінімальну затримку, низький джитер, високу пропускну здатність і максимальну надійність. Проаналізовано виклики, пов'язані з обслуговуванням IoMT-пристроїв у мережах 5G, зокрема високі вимоги до QoS для тактильного інтернету (TI). Досліджено способи інтеграції IoMT-пристроїв у мережу 5G, враховуючи специфічні вимоги медичних додатків. Запропоновано методи попередньої обробки запитів у мережі IoMT перед передачею їх у мережу 5G для зменшення навантаження на базові станції. Розглянуто класифікацію трафіку за пріоритетами, яка включає чотири класи якості, що дозволяє ефективно керувати ресурсами мережі і забезпечувати необхідний рівень QoS. Досліджено алгоритми динамічного управління ресурсами та пріоритетами, які дозволяють забезпечити належний рівень обслуговування навіть при високому завантаженні мережі. Використання сучасних технологій, таких як SDN і NFV, для гнучкого управління мережевими ресурсами і адаптації до змінних умов експлуатації. Дослідження підтвердило важливість класифікації трафіку, динамічного управління ресурсами та пріоритетного обслуговування для забезпечення високої якості обслуговування медичних пристроїв і додатків у мережах 5G.
Ключові слова: LTE, 5G, Інтернет медичних речей (IoMT), телекомунікаційна мережа.

Vasylkivskyi M., Kolomiiets A., Budash M., Prykmeta A., Oliinyk A. Ensuring the quality of the internet of medical things services in the 5G network. The paper solves an urgent scientific and applied problem of ensuring the quality of service (QoS) of the Internet of Medical Things (IoMT) in an LTE/5G network. A traffic prioritization classification is proposed, which includes four quality classes, allowing for priority processing of critical medical traffic and optimizing the use of network resources. The considered solutions allow to provide the necessary QoS parameters for critical medical applications, such as minimum latency and high reliability of data transmission, which is critical for healthcare systems.A comprehensive analysis of the problems of quality of service (QoS) management for the Internet of Medical Things (IoMT) in 5G networks is carried out. The interaction of the 5G network with the IoMT network, as well as the features of improving QoS for medical devices and applications are considered. Critical requirements for the quality of service for medical applications are identified, including minimum latency, low jitter, high throughput, and maximum reliability. The challenges associated with servicing IoMT devices in 5G networks are analyzed, in particular, the high QoS requirements for the tactile Internet (TI). Ways to integrate IoMT devices into a 5G network, taking into account the specific requirements of medical applications, are investigated. Methods for preprocessing requests in the IoMT network before transmitting them to the 5G network are proposed to reduce the load on base stations. The traffic classification by priorities, which includes four quality classes, is considered, which allows to effectively manage network resources and provide the required level of QoS. The algorithms for dynamic resource and priority management are investigated, which allow to ensure an adequate level of service even at high network load. The use of modern technologies such as SDN and NFV for flexible management of network resources and adaptation to changing operating conditions. The study confirmed the importance of traffic classification, dynamic resource management, and priority service to ensure high quality of service for medical devices and applications in 5G networks.
Keywords: LTE, 5G, Internet of Medical Things (IoMT), telecommunications network.

Постановка наукової проблеми. Розгортання 5G технологій відбувається в декілька етапів, що дозволяє поступово інтегрувати нові можливості та підвищити ефективність мобільних мереж. Наведемо кілька ключових аспектів, які стосуються управління якістю обслуговування (QoS) для Інтернету медичних речей (IoMT) в мережах 5G. Технологія 5G забезпечує значно вищі швидкості передачі даних у порівнянні з 4G, що критично важливо для реального часу передачі медичних даних, таких як відео з камер спостереження пацієнтів або великі обсяги даних з медичних датчиків.

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

Технологія 5G має затримки, що можуть бути меншими ніж 1 мс, що важливо для критично важливих додатків, наприклад, для дистанційної хірургії або моніторингу пацієнтів у реальному часі. Технологія 5G підтримує більшу кількість підключених пристроїв на одиницю площі, що дозволяє розгортати більше медичних сенсорів і пристроїв в обмеженому просторі, наприклад, в лікарнях або медичних центрах. Стандарт 5G пропонує кращу енергоефективність, що зменшує споживання енергії пристроями IoMT, подовжуючи термін їх автономної роботи. Також важливими факторами розвитку мереж є виклики та рішення для управління QoS в IoMT на базі 5G. Для забезпечення надійного і безпечного зв'язку для чутливих медичних даних передбачено використання шифрування даних, аутентифікації користувачів та технологій блокчейн для безпечного зберігання та передачі даних. Для оптимального розподілення мережевих ресурсів між різними пристроями і додатками передбачено використання технологій Network Slicing для створення віртуальних сегментів мережі, які забезпечують різні рівні QoS для різних типів трафіку. Для постійного моніторингу якості обслуговування для своєчасного виявлення та усунення проблем передбачено впровадження систем моніторингу в реальному часі, що дозволяють автоматично коригувати параметри мережі для забезпечення відповідності вимогам QoS. Для забезпечення сумісності нових 5G технологій з існуючими системами медичного обслуговування та пристроями передбачено використання стандартних протоколів і інтерфейсів для забезпечення безшовної інтеграції з існуючими системами. Для гарантування безперебійної роботи мережі в разі збоїв або катастроф передбачено використання методів резервування та відновлення, таких як мультихопові маршрути і георезервування даних [1].

Отже, 5G технології відкривають нові можливості для розвитку IoMT, забезпечуючи вищу швидкість, меншу затримку і більшу надійність зв'язку. Однак, для ефективного управління якістю обслуговування в IoMT необхідно вирішити ряд викликів, включаючи безпеку даних, динамічне управління ресурсами та інтероперабельність з існуючими системами. Це вимагає комплексного підходу та використання передових технологій.

Метою дослідження є адаптація систем Інтернету медичних речей (IoMT) до мережі 5G із використанням основ побудови високоефективних систем IoMT в умовах обмежених ресурсів мережі 5G. Основні напрямки дослідження включають:

- вивчення сценаріїв передачі даних в умовах мережі 5G, враховуючи специфічні вимоги медичних пристроїв;
- аналіз методів динамічного розподілу та управління ресурсами мережі для забезпечення стабільної роботи IoMT пристроїв, з урахуванням їхніх вимог до пропускної здатності, енергоспоживання та затримок;
- визначення алгоритмів забезпечення високого рівня QoS, включаючи моніторинг та управління потоками трафіку, профілювання трафіку та планування ресурсів;
- формування підходів інтеграції IoMT з мережею 5G, враховуючи обмежені ресурси та високі вимоги до якості обслуговування. Дослідження спрямоване на подолання існуючих викликів у сфері IoMT та забезпечення надійного і ефективного функціонування медичних пристроїв у середовищі 5G.

Для досягнення поставленої мети, необхідно вирішити ряд задач, кожна з яких спрямована на різні аспекти адаптації Інтернету медичних речей (IoMT) до мережі 5G. Задачі формують структуру та зміст дослідження:

1. Аналіз існуючих методів управління якістю обслуговування (QoS) IoMT в мобільних мережах 5G для визначення поточного стану технологій QoS в контексті IoMT та 5G.
2. Узагальнення теорії функціонування IoMT в умовах мереж 5G, визначення компонентів мережі для ефективної роботи медичних пристроїв.
3. Узагальнення підходів до адаптивного управління QoS, що дозволить ефективно розподіляти ресурси мережі та забезпечувати необхідну якість обслуговування.
4. Визначення показників QoS, таких як затримка, пропускна здатність та надійність для оцінки ефективності функціонування IoMT у мережі 5G.
5. Узагальнення методів адаптивного керування розподілом ресурсів для забезпечення високої якості обслуговування медичних пристроїв. Вирішення цих взаємопов'язаних задач дозволить сформувати комплексний підхід до адаптації IoMT до мережі 5G, забезпечити необхідну якість обслуговування та ефективне використання мережевих ресурсів.

Аналіз останніх досліджень і публікацій. Технологія 5G забезпечує основні типи послуг з'єднання, які мають критичне значення для різних сфер використання, включаючи Інтернет

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

медичних речей (IoMT). Розглянемо ці послуги детальніше та їхню роль в контексті IoMT. Розширене мобільне широкосмугове з'єднання (eMBB) забезпечує високу пропускну здатність для підтримки додатків, які вимагають великих обсягів даних, таких як відео у форматі UHD, віртуальна та доповнена реальність. Мережа може динамічно виділяти пропускну здатність для забезпечення оптимальної роботи різних додатків. Висока швидкість передачі даних дозволяє проводити відеоконсультації з лікарями та моніторинг стану пацієнтів у реальному часі. Технологія eMBB дозволяє передавати великі медичні файли, такі як МРТ-сканування або результати інших обстежень. Ультранадійний і з низькою затримкою зв'язок (uRLLC) забезпечує критично важливі послуги з високою надійністю та мінімальні затримки у передачі даних, що особливо важливо для додатків, де час реакції є критичним [2].

Висока надійність і низька затримка дозволяють лікарям проводити операції на відстані за допомогою роботизованих систем. Зниження затримки дозволяє миттєво реагувати на зміни в стані пацієнта, що критично важливо для пацієнтів з хронічними захворюваннями або в критичному стані. Масивні комунікації машинного типу (mMTC) забезпечують підтримку мільйонів підключених пристроїв на квадратний кілометр. Пристрої mMTC мають низьке енергоспоживання, що дозволяє їм працювати тривалий час без необхідності частого заряджання. Технологія mMTC підтримує великий обсяг підключених медичних пристроїв, таких як датчики моніторингу стану здоров'я, що передають дані про пацієнтів до медичних центрів та дозволяє збирати та аналізувати великі обсяги даних з багатьох медичних пристроїв одночасно, що сприяє ефективному управлінню здоров'ям пацієнтів [3].

Отже, технологія 5G з її основними типами послуг з'єднання (eMBB, uRLLC, mMTC) відкриває нові можливості для розвитку Інтернету медичних речей (IoMT), забезпечуючи високу швидкість, надійність, низьку затримку та підтримку великої кількості пристроїв. Це дозволяє впроваджувати інноваційні медичні рішення, підвищувати ефективність лікування та покращувати якість медичних послуг. Технологія 5G забезпечує низку ключових характеристик, які значно покращують можливості мереж і розширюють спектр послуг, особливо в контексті Інтернету медичних речей (IoMT). Наведемо основні характеристики 5G та їх вплив на IoMT. Висока пропускна здатність (до 20 Гб/с на одну базову станцію) дозволяє швидко передавати великі обсяги медичних даних, включаючи результати діагностичних тестів і зображень високої розділь-ної здатності (наприклад, МРТ та КТ сканів). Швидкість завантаження даних для одного абонента до 100 Мб/с і вивантаження до 50 Мб/с забезпечує достатню швидкість для реального часу відео-консультацій, дистанційного моніторингу пацієнтів та інших телемедичних послуг. Можливість абонентському пристрою рухатись зі швидкістю до 500 км/год забезпечує надійний зв'язок для мобільних медичних пристроїв, таких як швидкі допомоги, які можуть переміщуватись на високих швидкостях, підтримуючи безперервний доступ до медичних даних і консультацій. Перемикання між режимом заощадження енергії та повністю робочим режимом за 10 мс забезпечує важливу енергоефективність пристроїв IoMT, які повинні працювати тривалий час без підзарядки. Швидке перемикання між режимами дозволяє економити енергію, не втрачаючи продуктивності. Затримки передачі даних до 4 мс за сприятливих умов і до 1 мс для спеціалізованих з'єднань забезпечують критично важливу низьку затримку для додатків, які вимагають миттєвої реакції, таких як віддале-на хірургія або системи моніторингу життєво важливих показників пацієнтів у реальному часі. Ефективне використання радіочастотного спектру підвищує продуктивність мережі, дозволяючи підключати більше пристроїв і забезпечуючи стабільне з'єднання навіть у високонавантажених середовищах, таких як лікарні. Підтримка до 1 млн пристроїв на 1 км² дозволяє підключити величезну кількість медичних датчиків і пристроїв у межах однієї мережі, що особливо корисно для великих лікарень та медичних центрів з великою кількістю обладнання [4].

Завдяки високій швидкості та низькій затримці 5G мережі, дедалі більше завдань обробки і зберігання даних можна передавати на потужні сервери в хмарах. Це дозволяє використовувати менш потужні та енергоефективні пристрої IoMT, зберігаючи при цьому високу продуктивність. Низька затримка і висока надійність 5G роблять можливим впровадження таких додатків, як віддалена хірургія, автономні медичні роботи та системи управління безпілотними автомобілями, що транспортують медичні матеріали або пацієнтів. Висока пропускна здатність 5G дозволяє використовувати віртуальну та доповнену реальність у медичних навчаннях, діагностиці та лікуванні. Це може включати віртуальні консультації, віртуальну реабілітацію та інші інноваційні методи лікування. Мережа 5G дозволяє проводити високоякісні відео-консультації, обмінюватись великими медичними файлами та даними в реальному часі, що значно покращує якість та доступ-

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

ність медичних послуг, особливо в віддалених регіонах. Отже, технологія 5G забезпечує значні покращення в швидкості, надійності та ефективності мобільних мереж, що відкриває нові можливості для розвитку Інтернету медичних речей (IoMT). Це включає швидкі та надійні з'єднання для медичних пристроїв, підтримку критично важливих додатків, ефективне використання енергії та значне розширення можливостей телемедицини [5].

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Технологія 5G використовує декілька ключових технологій і підходів для досягнення високих показників роботи мережі. Розглянемо детальніше про кожну з них і їхній вплив на роботу телекомунікаційних мереж, включаючи Інтернет медичних речей (IoMT). Діапазони електромагнітних хвиль: FR1 (600–6000 МГц) використовується для забезпечення покриття на великих відстанях і проникнення через будівлі та FR2 (24–100 ГГц) використовується для передачі даних на високих швидкостях у міліметровому діапазоні з коротшим радіусом дії. Малі базові станції споживають менше енергії і є компактними. Можливість встановлювати тисячі базових станцій на відстані 250 м одна від одної вирішує проблему швидкого згасання міліметрових хвиль. Забезпечується покриття в густонаселених районах, лікарнях та інших медичних закладах, де потрібна висока щільність підключень і стабільність з'єднань. У базових станціях 5G використовується сотні портів для передачі і прийому даних, що значно підвищує продуктивність. Забезпечення високої пропускної здатності і надійності з'єднань для медичних пристроїв дозволяє передавати великі обсяги даних швидко і без перешкод. Управління направленістю сигналів (Beam Forming) дозволяє концентрувати сигнал в напрямку абонента, зменшуючи інтерференцію і покращуючи якість з'єднання. При цьому, підвищується якість зв'язку і зменшуються затримки, що є критичним для додатків реального часу, таких як моніторинг стану пацієнтів або віддалена хірургія. Передача даних в режимі повного дуплексу дозволяє базовій станції і абоненту одночасно передавати і приймати дані, що підвищує ефективність використання спектру. При цьому, підвищується швидкість і ефективність передачі даних, що важливо для безперервного моніторингу медичних показників. Неортогональний множинний доступ (NOMA) дозволяє використовувати один і той же частотний ресурс для кількох користувачів, підвищуючи ефективність використання спектру, що забезпечує підключення більшої кількості медичних пристроїв одночасно без втрати якості обслуговування. Кодування з малою щільністю перевірок на парність (LDPC) забезпечує ефективне кодування і декодування даних з мінімальними помилками, що підвищує надійність передачі медичних даних, що критично для точності і своєчасності медичних послуг [6]. Отже технологія 5G, використовуючи діапазони FR1 і FR2, малі базові станції, масиви антенних систем MIMO, Beam Forming, повний дуплекс, NOMA та LDPC, забезпечує високу швидкість, надійність і ефективність телекомунікаційних мереж. Це відкриває нові можливості для розвитку Інтернету медичних речей (IoMT), забезпечуючи стабільні та швидкі з'єднання для медичних пристроїв, підвищуючи якість медичних послуг і покращуючи результати лікування.

Забезпечення якості обслуговування (QoS) у мобільних мережах п'ятого покоління (5G) для Інтернету медичних речей (IoMT) вимагає комплексного підходу, що охоплює різні сфери управління мережевими ресурсами. Основні області вирішення завдань QoS включають класифікацію аплікацій, профілювання трафіку, обмеження інтенсивності трафіку, управління чергами та маршрутизацію. Комплексний підхід до управління QoS у мобільних мережах 5G для IoMT дозволяє забезпечити стабільну та якісну роботу медичних пристроїв і додатків. Це вимагає інтегрованого застосування різних механізмів і протоколів, що дозволяє ефективно розподіляти ресурси мережі і забезпечити високий рівень обслуговування [7].

Аналіз методологічних підходів до адаптації систем Інтернету медичних речей (IoMT) до стандарту 5G свідчить про існування протиріч у процесі проектування цих систем. Основні аспекти цих протиріч можна розділити на дві групи.

Конструктивні особливості кожного елемента IoMT вказують на те, що кожен пристрій IoMT має свої унікальні характеристики, які зумовлені його функціональним призначенням. Наприклад, медичні монітори, сенсори, імплантати та інші пристрої мають різні вимоги до енергоспоживання, пропускної здатності, затримки передачі даних та надійності.

Відповідність можливостям мережі 5G вказує на те, що мережа 5G пропонує високі швидкості передачі даних, низькі затримки, високу надійність та підтримку масового підключення пристроїв. Однак, для ефективної роботи кожен елемент IoMT повинен бути адаптований до цих можливостей, що вимагає ретельного проектування та інтеграції.

Для подолання вказаних протиріч можна класифікувати особливості функціонування систем

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

IoMT та враховувати їх при побудові на основі загальних принципів функціонування мережі 5G. Основні кроки в цьому процесі включають: класифікацію пристроїв IoMT для визначення різних категорій пристроїв IoMT на основі їхніх технічних і функціональних характеристик; врахування специфічних вимог кожної категорії за рахунок визначення вимог до пропускної здатності, енергоспоживання, затримок та надійності для кожної категорії пристроїв; розробку загальної схеми побудови систем IoMT на основі інтеграції різних категорій пристроїв IoMT в єдину мережу на основі принципів функціонування 5G, таких як гетерогенні мережі, MIMO, міліметрові хвилі [8]; проведення тестувань для перевірки сумісності пристроїв IoMT з мережею 5G та внесення необхідних коригувань.

Концепція QoS у стандарті 5G базується на потоках даних, які класифікуються і маркуються за допомогою ідентифікатора QoS Flow Identifier (QFI). Це дозволяє забезпечити високоякісне обслуговування різних типів трафіку, включаючи критично важливі медичні сервіси. Наведемо основні принципи та механізми функціонування QoS у 5G. Потоки зі стандартизованими профілями QoS використовують тільки значення QFI. Потоки зі специфічними для оператора профілями QoS використовують значення QFI і сигналізують атрибути QoS між елементами мережі. Служби End-to-End (E2E) охоплюють мережу або служби додатків між користувацьким обладнанням (UE) і зовнішньою мережею передачі даних, наприклад, Інтернетом. Різні служби E2E потребують диференційованого лікування QoS для забезпечення належної якості обслуговування. Один або кілька SDF можуть транспортуватися в одному потоці 5G QoS, якщо вони мають однакову обробку QoS. Кожен пакет висхідного та низхідного зв'язку зіставляється з потоком QoS, забезпечуючи передачу E2E між UE та User Plane Function (UPF) протягом усього терміну служби сеансу Protocol Data Unit (PDU). Один сеанс PDU може нести один або кілька потоків QoS. Всі потоки QoS певної сесії PDU направляються по одному тунелю NG-U. QoS Flow ID (QFI) використовується для ідентифікації потоку QoS у сесії PDU і є унікальним для кожної сесії PDU. Кожен потік QoS характеризується набором параметрів, таких як ідентифікатор 5G QoS (5QI), пріоритет призначення ресурсів (ARP), гарантована швидкість потоку (GFBR) [9]. Радіоносій може нести один або кілька потоків QoS. Кожен сеанс PDU має унікальний набір радіоносіїв, і gNodeB вирішує, через який радіоносій відправляти потік QoS. QFI є унікальним у межах кожної сесії PDU, що дозволяє чітко ідентифікувати і відстежувати кожен потік QoS. Система QFI дозволяє операторам налаштовувати і оптимізувати мережеві ресурси відповідно до вимог різних типів трафіку, забезпечуючи високу гнучкість в управлінні QoS. Класифікація і маркування пакетів за допомогою QFI забезпечує диференційовану класифікацію трафіку, що є критичним для підтримки якості обслуговування різних додатків і послуг.

Усі зазначені вище фактори взаємопов'язані і повинні враховуватись узгоджено. Наприклад, класифікація трафіку визначає пріоритети, які впливають на управління чергами і маршрутизацію. Профілювання трафіку допомагає налаштувати обмеження інтенсивності трафіку та розподіл ресурсів мережі. Концепція QoS у стандарті 5G з використанням QFI дозволяє ефективно класифікувати, маркувати і управляти трафіком, забезпечуючи високу якість обслуговування для різних типів додатків, включаючи критично важливі медичні сервіси IoMT. Взаємопов'язаність і узгодженість між різними механізмами та протоколами QoS є ключовими для досягнення оптимальної роботи мережі [10].

Отже, системи Інтернету медичних речей (IoMT) значно виграють від впровадження мереж стільникового зв'язку стандарту 5G завдяки його передовим технологіям і здатності задовольняти різноманітні вимоги IoMT. Основні функції та можливості 5G, що сприяють розвитку IoMT, включають:

1. Гетерогенні мережі (HetNet), які забезпечують безперервне та надійне підключення пристроїв IoMT за рахунок інтеграції різних типів мереж.

2. MIMO (Multiple Input Multiple Output), що підвищує пропускну здатність і стійкість до перешкод, що особливо важливо для критичних медичних застосувань.

3. Міліметрові хвилі (mmWave) дозволяють досягти високих швидкостей передачі даних, що необхідно для оперативного обміну великою кількістю медичної інформації.

4. Зв'язок між пристроями (D2D) підвищує ефективність передачі даних і знижує затримки, що важливо для пристроїв, які потребують швидкого реагування.

5. Програмно-конфігурована мережа (SDN) забезпечує гнучке управління мережевими ресурсами та оптимізацію трафіку, що дозволяє адаптувати мережу під потреби IoMT.

6. Віртуалізація мережевих функцій (NFV) спрощує розгортання і управління мережевими

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

сервісами, що прискорює впровадження нових рішень для IoMT. Вказані технології сприяють підвищенню якості обслуговування (QoS) пристроїв IoMT, забезпечуючи надійне та ефективне підключення, необхідне для сучасних медичних рішень [11].

Функція CQCF (Control of Quality of Service) в мережі 5G забезпечує контроль якості обслуговування в реальному часі, керуючи потоками трафіку на основі встановлених рівнів QoS. Основні механізми контролю QoS включають: аналіз і класифікація трафіку для забезпечення оптимального обслуговування; розподіл ресурсів мережі відповідно до пріоритетів і вимог QoS; контроль і регулювання потоків для підтримки стабільної якості обслуговування.

Функція CQMF (Management of Quality of Service) в мережі 5G забезпечує: підтримку QoS згідно з договорами SLA за рахунок забезпечення відповідності якості обслуговування умовам договорів; постійне спостереження за станом QoS для виявлення і вирішення проблем; регулярне технічне обслуговування для забезпечення стабільної роботи системи; адаптацію та масштабування рівнів QoS відповідно до змінних вимог і умов [12]. Розглянуті проблеми IoMT та запропоновані засоби їх розв'язання зведено в табл. 1.

Таблиця 1. Проблеми IoT та засоби їх розв'язання

Проблеми IoMT	Вирішення
Енергоефективність	MIMO
Масштабованість	Ретранслятори
Інтелектуальна обробка та збереження	MmWave, D2D
Інформаційна безпека	HetNet, D2D, MIMO, SDN
Функціональна сумісність	D2D

Вирішення проблеми енергоефективності можливе, якщо обладнання підключається за допомогою датчиків лише при необхідності застосування, без з'єднання з базовою станцією, що знижує споживання енергії. У табл. 2. представлено методи збору та обробки інформації від IoMT.

Таблиця 2. Основні характеристики методів збору та обробки інформації від IoMT [1, 4].

Метод	Протоколи та технології	Арататні засоби	Інфрастуктура
Прямий зв'язок	NB-IoMT	Розумні об'єкти, шлюз	Хмарні сервери
Метод збору та обробки інформації за допомогою шлюзу	Wi-Fi, Bluetooth, ZigBee	Розумні об'єкти, шлюз	Хмарні сервери
Метод збору та обробки інформації з використанням розумного концентратора	MQTT/ CoAP, Wi-Fi, BLE, ZigBee та Z Wave	Інтелектуальний концентратор	Мікросервісна хмарна платформа
Метод збору та обробки інформації на основі промислових шин	CANOpen, ControlNet, Modbus	Шлюз	Хмарний сервер

Для збору інформації за допомогою пристроїв IoMT існують два основні методи. Прямий метод передбачає безпосереднє з'єднання пристрою IoMT з мобільною мережею зв'язку (ММЗ). Кожен пристрій IoMT підключається безпосередньо до базової станції мережі 5G. При цьому, забезпечуються: швидка передача даних з мінімальною затримкою та високий рівень надійності і безпеки, оскільки дані передаються напряму без проміжних точок. Разом з тим, присутні недоліки: підвищені вимоги до енергоспоживання пристроїв IoMT, оскільки вони повинні підтримувати постійне підключення до базової станції та обмеженість кількості пристроїв, які можуть одночасно підключатися до базової станції.

Метод через проміжну точку збору використовує інтелектуальний концентратор або точку доступу як проміжну точку збору даних від пристроїв IoMT. Дані спочатку збираються концентратором, а потім передаються до мобільної мережі зв'язку. При цьому, забезпечуються: зменшення енергоспоживання пристроїв IoMT, оскільки вони можуть працювати в режимі з

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

низьким енергоспоживанням і періодично передавати дані до концентратора та підвищена масштабованість, оскільки один концентратор може обслуговувати велику кількість пристроїв IoMT. Разом з тим, до недоліків потрібно віднести: додаткові затримки в передачі даних через необхідність передачі даних від пристроїв до концентратора, а потім до базової станції та можливі проблеми з надійністю і безпекою, оскільки дані проходять через додаткову точку збору. У табл. 3. представлено порівняння методів проектування мереж IoMT [5].

Вказані методи можуть бути застосовані в залежності від конкретних вимог та умов експлуатації пристроїв IoMT, таких як необхідний рівень енергоспоживання, допустима затримка передачі даних та вимоги до надійності та безпеки.

Таблиця 3. Порівняння методів проектування мереж IoMT

Параметр	Wi-Fi (через точку доступу)	NB-IoMT (прямий зв'язок)
Енергоспоживання	Низьке (режим очікування)	Високе (постійне підключення)
Затримка передачі	Вища (через проміжну точку)	Мінімальна
Надійність	Залежить від рівня сигналу та інтерференції	Висока
Масштабованість	Висока, легко додавати нові пристрої	Обмежена кількість підключень
Безпека	Потребує додаткових заходів (шифрування, аутентифікація)	Висока

Прикладом взаємодії у вигляді прямого зв'язку є функціонування NB-IoMT (Narrowband Internet of Medical Things) в мережі стандарту 5G. Ця технологія дозволяє отримати доступ до мережевих послуг через E-UTRA (Evolved Universal Terrestrial Radio Access) з пропускною здатністю каналу, обмеженою 200 кГц. NB-IoMT особливо ефективна для медичних застосувань, де важливі такі параметри: тривала робота без підзарядки; низька затримка передачі; безпека даних (табл. 4).

Таблиця 4. Порівняння NB-IoMT з іншими технологіями

Технологія	Пропускна здатність	Енергоспоживання	Затримка передачі	Надійність та безпека
NB-IoMT	200 кГц	Низьке	Мінімальна	Висока
LTE-M	до 1 МГц	Середнє	Низька	Висока
Wi-Fi	Висока	Високе	Середня	Середня
Bluetooth	Середня	Низьке	Низька	Середня

NB-IoMT є перспективною технологією для впровадження в системах Інтернету медичних речей, забезпечуючи надійний і ефективний зв'язок в мережі 5G [6]. Використання Wi-Fi з проміжними точками доступу для підключення пристроїв IoMT до мережі є ефективним методом організації взаємодії. Цей підхід дозволяє забезпечити необхідний рівень підключення та передачі даних, враховуючи енергоспоживання і масштабованість мережі.

Система стільникового зв'язку, як і будь-яка система телефонного зв'язку, є прикладом системи масового обслуговування. У таких системах існує випадковий потік заявок (викликів), випадкова тривалість їх обслуговування (сеансів зв'язку) і кінцеве число каналів обслуговування (фізичних каналів). Основні характеристики і вимоги до якості обслуговування (QoS) в таких системах залежать від типу застосування. Для визначення та вимірювання якості обслуговування використовуються кілька ключових параметрів, які можна розділити на три основні групи: швидкісні показники (параметри пропускної здатності); часові показники (параметри затримок передачі пакетів); показники надійності передачі.

Вимірювання зазначених параметрів здійснюється на певному інтервалі часу. Чим менше цей часовий інтервал, тим жорсткіші вимоги пред'являються до мережі. Забезпечення QoS «із кінця в кінець» вимагає взаємодії всіх вузлів на шляху пакетів трафіку і визначається надійністю,

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

функціональністю та продуктивністю «слабкої ланки». Параметри якості обслуговування є критичними для забезпечення належного функціонування стільникових мереж. Вони повинні бути ретельно виміряні та підтримані, щоб гарантувати надійність, функціональність і продуктивність мережі. Оскільки якість обслуговування залежить від найслабшої ланки в мережі, всі елементи мережі повинні працювати злагоджено та ефективно для забезпечення високого рівня QoS [7].

Методологія надання якісних сервісів IoMT в мережі 5G складається з набору алгоритмів та математичної моделі, які забезпечують адаптивне керування обслуговуванням агрегованого трафіку. Цей трафік включає як мультимедійний трафік з різними пріоритетами, так і трафік даних від пристроїв IoMT з власними пріоритетами. Основні компоненти методології: алгоритм пріоритетного обслуговування пакетів; оцінка критичного рівня часу затримки буферизації; буферні ресурси і лічильники часу затримки. Алгоритм пріоритетного обслуговування передбачає: отримання вхідних потоків даних від пристроїв IoMT з численними пріоритетами; обробку даних; обслуговування. Розглянута методологія надання якісних сервісів IoMT в мережі 5G забезпечує оптимальне обслуговування агрегованого трафіку, враховуючи різні пріоритети мультимедійних і IoMT-потоків. Використання алгоритмів пріоритетного обслуговування і динамічної модифікації поля DSCP дозволяє ефективно керувати часом затримки і якістю обслуговування, що є критично важливим для надійної та ефективної роботи медичних пристроїв у мережі 5G [8]. У зв'язку з тим, що пристрої IoMT є продовженням концепції M2M (Machine-to-Machine), для забезпечення якості обслуговування (QoS) для IoMT-пристроїв можна застосувати класи пріоритетів, визначені для M2M-пристроїв. Для кодування чотирьох класів IoMT-пристроїв достатньо двох бітів в полі заголовка IP-пакету. Класифікацію IoMT-пристроїв потрібно здійснювати за класами пріоритетів:

- Клас 0: Найвищий пріоритет для критичних медичних додатків, які вимагають мінімальної затримки;
- Клас 1: Високий пріоритет для додатків, що потребують низької затримки, але не настільки критичних як Клас 0;
- Клас 2: Середній пріоритет для додатків з помірною вимогою до затримки;
- Клас 3: Низький пріоритет для додатків, що можуть використовувати вищу затримку;

Кожному IoMT-пристрою заздалегідь присвоюється певний пріоритет залежно від його призначення. Всі повідомлення, які надсилає цей IoMT-пристрій, матимуть визначений пріоритет. Для кодування чотирьох класів пріоритетів використовується два біти в полі заголовка IP-пакету. Основним параметром QoS для всіх класів є затримка, яка контролюється за допомогою окремого поля заголовку IP-пакету «Час в мережі». Якщо затримка перевищує допустимі межі для визначеного пріоритету, пакет обслуговується з вищим пріоритетом.

До переваг розглянутої методології потрібно віднести:

- гарантоване виконання параметрів QoS для IoMT-пакетів з урахуванням їх пріоритетів;
- можливість динамічного управління пріоритетами для забезпечення належного рівня обслуговування;
- зменшення перевантаження базових станцій за рахунок використання пріоритетів і динамічне коригування затримок дозволяє зменшити навантаження на базові станції при плануванні радіоресурсів;
- ефективне використання наявних ресурсів для обслуговування агрегованого трафіку;
- підвищення продуктивності мережі за рахунок адаптивного управління трафіком [9].

Отже, розглянута методологія передбачає високу якість обслуговування для IoMT-пристроїв в мережі 5G, використовуючи класи пріоритетів, визначені для M2M-пристроїв. Динамічне управління пріоритетами і адаптивне коригування затримок дозволяють оптимізувати роботу мережі і забезпечити надійне обслуговування критичних медичних додатків. Наведемо порівняльну таблицю (табл. 5) класів якості для сервісів M2M/IoMT пристроїв, яка відображає різні параметри якості обслуговування (QoS) для кожного класу [10].

Таблиця 5. Порівняння NB-IoMT з іншими технологіями

Клас	Тип сервісу	Затримка	Джитер	Пропускна здатність	Надійність	Приклади використання
Клас 0 (Найвищий пріоритет)	Критичні медичні додатки	< 10 мс	Мінімальний	Висока	Максимальна, втрата пакетів неприпустима	Термінові медичні сигнали, дистанційні хірургічні роботи

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

Клас 1 (Високий пріоритет)	Важливі медичні додатки	< 50 мс	Низький	Помірна	Висока, деяка втрата пакетів допустима	Моніторинг пацієнтів у лікарні
Клас 2 (Середній пріоритет)	Звичайні медичні додатки	< 100 мс	Допустимий	Середня	Середня, деяка втрата пакетів допустима	Передача даних зі смарт-годинників, фітнес-трекерів
Клас 3 (Низький пріоритет)	Некритичні медичні додатки	< 500 мс	Допустимий	Низька	Звичайна, втрата пакетів допустима	Періодичний збір даних для досліджень, передача інформації з менш критичних сенсорів

Таблиця 6 допомагає зрозуміти різні вимоги до якості обслуговування для різних класів M2M/IoMT пристроїв, що дозволяє ефективно планувати і керувати ресурсами мережі для забезпечення необхідного рівня QoS. Пристрої тактильного інтернету (TI) серед IoMT-пристроїв мають найжорсткіші вимоги щодо якості обслуговування (QoS). Моделювання показує, що з середнім завантаженням обладнання мережі 5G обслуговування IoMT-пристроїв загальним порядком призводить до значних черг, унеможливлюючи забезпечення необхідних показників якості. Тому частину обробки запитів слід здійснювати в IoMT-мережі до передачі в мережу 5G [11]. Для вирішення цієї проблеми необхідно використовувати відповідні класи якості для різних типів трафіку, враховуючи їх специфічні вимоги до QoS. Порівняльну таблицю класів якості для сервісів M2M/IoMT та TI пристроїв наведено нижче.

Таблиця 6. Порівняльння класів якості для сервісів M2M/IoMT та TI пристроїв

Клас	Тип сервісу	Затримка	Джитер	Пропускна здатність	Надійність	Приклади використання
Клас 0 (Найвищий пріоритет)	Критичні медичні додатки, тактильний інтернет (TI)	< 1 мс	Мінімальний	Висока	Максимальна, втрата пакетів неприпустима	Хірургічні роботи з дистанційним управлінням, критичні медичні сигнали
Клас 1 (Високий пріоритет)	Важливі медичні додатки	< 10 мс	Низький	Помірна	Висока, деяка втрата пакетів допустима	Моніторинг пацієнтів у лікарні
Клас 2 (Середній пріоритет)	Звичайні медичні додатки	< 100 мс	Допустимий	Середня	Середня, деяка втрата пакетів допустима	Передача даних зі смарт-годинників, фітнес-трекерів
Клас 3 (Низький пріоритет)	Некритичні медичні додатки	< 500 мс	Допустимий	Низька	Звичайна, втрата пакетів допустима	Періодичний збір даних для досліджень, передача інформації з менш критичних сенсорів

Частину обробки запитів IoMT слід здійснювати в IoMT-мережі до передачі їх в мережу 5G. Це зменшить навантаження на мережу 5G і зменшить затримки. Розподіл ресурсів потрібно здійснювати із використанням технологій SDN та NFV для динамічного розподілу ресурсів, враховуючи специфічні вимоги IoMT та TI. Розглянута методологія надання якісних сервісів для IoMT та TI в мережі 5G забезпечує ефективне управління трафіком, враховуючи специфічні вимоги до QoS для різних типів пристроїв. Використання класифікації за пріоритетами і динамічне управління ресурсами допомагає досягти високого рівня обслуговування навіть при середньому завантаженні мережі.

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

Висновки та перспективи подальших досліджень. Медичні пристрої та додатки вимагають найвищих стандартів якості обслуговування (QoS) через критичну важливість точних і своєчасних даних для забезпечення безпеки пацієнтів і ефективності медичних послуг. Дослідження показують, що специфічні вимоги до QoS для таких додатків включають мінімальну затримку, низький джитер, високу пропускну здатність і максимальну надійність. Застосування класів якості для медичних пристроїв і додатків дозволяє ефективно керувати мережевими ресурсами і забезпечувати необхідний рівень обслуговування. Запропонована класифікація охоплює чотири основні класи. Використання динамічних алгоритмів управління ресурсами, таких як програмно-конфігуровані мережі (SDN) і віртуалізація мережевих функцій (NFV), дозволяє ефективно розподіляти ресурси мережі залежно від поточного навантаження і вимог QoS. Це забезпечує гнучкість і адаптивність мережі до змінних умов експлуатації. Тактильний інтернет (TI) накладає особливо жорсткі вимоги до QoS через необхідність мінімальної затримки (менше 1 мс) і максимальної надійності. Обслуговування TI-трафіку вимагає найвищого пріоритету і може впливати на якість обслуговування інших мультисервісних потоків. Запропоноване рішення включає попередню обробку запитів в IoMT-мережі перед передачею в мережу 5G для зменшення навантаження на базові станції.

Результати дослідження показують, що середнє завантаження мережі 5G може призвести до значних черг для запитів від IoMT-пристроїв при загальному порядку обслуговування. Тому частина обробки повинна здійснюватися в IoMT-мережі, а динамічне управління пріоритетами на основі поточного стану мережі є критично важливим для забезпечення належного рівня QoS. Розглянута методологія і результати дослідження підтверджують важливість класифікації трафіку, динамічного управління ресурсами і пріоритетного обслуговування для забезпечення високої якості обслуговування медичних пристроїв і додатків в мережах 5G. Використання сучасних мережевих технологій, таких як SDN і NFV, дозволяє досягти необхідного рівня QoS і забезпечити надійне та ефективне функціонування систем охорони здоров'я.

Список бібліографічного опису

1. Awotunde, J. B., Bhoi, A. K., & Barsocchi, P. (2021). Hybrid cloud/fog environment for healthcare: An exploratory study, opportunities, challenges, and future prospects. Hybrid Artificial Intelligence and IoT in Healthcare (pp. 1-20). Singapore: Springer.
2. Amin, S. U., & Hossain, M. S. (2021). Edge intelligence and Internet of Things in healthcare: A survey. IEEE Access, 9, 45-59. Available from https://doi.org/10.1109/ACCESS.2020.3045115.
3. Ding, Y., et al. (2021). DeepEDN: A deep-learning-based image encryption and decryption network for Internet of Medical Things. IEEE Internet of Things Journal, 8(3), 1504-1518. Available from https://doi.org/10.1109/JIOT.2020.3012452.
4. Irrum, F., Ali, M., Naeem, M., Anpalagan, A., Qaisar, S., & Qamar, F. (2021). D2D-enabled resource management in secrecy-ensured 5G and beyond heterogeneous networks. Physical Communication, 45, 101275, ISSN 187-4907. Available from https://doi.org/10.1016/j.phycom.2021.101275.
5. Hameed, K., Bajwa, I. S., Sarwar, N., Anwar, W., Mushtaq, Z., & Rashid, T. (2021). Integration of 5G and Block-Chain Technologies in Smart Telemedicine Using IoT. Journal of Healthcare Engineering, 2021.
6. Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. In. Machine Learning and the Internet of Medical Things in Healthcare (pp. 89-111). Academic Press.
7. Adeniyi, E. A., Ogundokun, R. O., & Awotunde, J. B. (2021). IoMT-based wearable body sensors network healthcare monitoring system. IoT in healthcare and ambient assisted living (pp. 103-121). Singapore: Springer.
8. Marques, G., Bhoi, A. K., Albuquerque, V. H. C. d., & Hareesha, K. S. (Eds.), (2021). IoT in healthcare and ambient assisted living. Springer.
9. Oniani, S., Marques, G., Barnovi, S., Pires, I. M., & Bhoi, A. K. (2021). Artificial intelligence for Internet of Things and enhanced medical systems. Bio-inspired neurocomputing (pp. 43-59). Singapore: Springer.
10. Bhoi, A. K., Mallick, P. K., Liu, C. M., & Balas, V. E. (Eds.), (2021). Bio-inspired neurocomputing. Springer.
11. М. Васильківський, О. Городецька, Б. Климчук, і В. Говорун, «Стратегії технологічного розвитку апаратного забезпечення інфокомунікаційних радіомереж», ІТКІ, вип. 56, вип. 1, с. 83–91, Бер 2023.
12. Васильківський, М., Прикмета, А., Олійник, А., & Нікітович, Д. (2023). Оптимізація інтелектуальних телекомунікаційних мереж. Вісник Хмельницького національного університету, Технічні науки. – 2023. – № 1. (317). – С. 33–41. doi: 10.31891/2307-5732-2023-317-1-33-41

References

1. Awotunde, J. B., Bhoi, A. K., & Barsocchi, P. (2021). Hybrid cloud/fog environment for healthcare: An exploratory study, opportunities, challenges, and future prospects. Hybrid Artificial Intelligence and IoT in Healthcare (pp. 1-20). Singapore: Springer.
2. Amin, S. U., & Hossain, M. S. (2021). Edge intelligence and Internet of Things in healthcare: A survey. IEEE Access, 9, 45-59. Available from https://doi.org/10.1109/ACCESS.2020.3045115.

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

3. Ding, Y., et al. (2021). DeepEDN: A deep-learning-based image encryption and decryption network for Internet of Medical Things. IEEE Internet of Things Journal, 8(3), 1504-1518. Available from https://doi.org/10.1109/JIOT.2020.3012452.
4. Irrum, F., Ali, M., Naeem, M., Anpalagan, A., Qaisar, S., & Qamar, F. (2021). D2D-enabled resource management in secrecy-ensured 5G and beyond heterogeneous networks. Physical Communication, 45, 101275, ISSN 187-4907. Available from https://doi.org/10.1016/j.phycom.2021.101275.
5. Hameed, K., Bajwa, I. S., Sarwar, N., Anwar, W., Mushtaq, Z., & Rashid, T. (2021). Integration of 5G and Block-Chain Technologies in Smart Telemedicine Using IoT. Journal of Healthcare Engineering, 2021.
6. Singh, P., Singh, N., Singh, K. K., & Singh, A. (2021). Diagnosing of disease using machine learning. In. Machine Learning and the Internet of Medical Things in Healthcare (pp. 89-111). Academic Press.
7. Adeniyi, E. A., Ogundokun, R. O., & Awotunde, J. B. (2021). IoMT-based wearable body sensors network healthcare monitoring system. IoT in healthcare and ambient assisted living (pp. 103-121). Singapore: Springer.
8. Marques, G., Bhoi, A. K., Albuquerque, V. H. C. d., & Hareesha, K. S. (Eds.), (2021). IoT in healthcare and ambient assisted living. Springer.
9. Oniani, S., Marques, G., Barnovi, S., Pires, I. M., & Bhoi, A. K. (2021). Artificial intelligence for Internet of Things and enhanced medical systems. Bio-inspired neurocomputing (pp. 43-59). Singapore: Springer.
10. Bhoi, A. K., Mallick, P. K., Liu, C. M., & Balas, V. E. (Eds.), (2021). Bio-inspired neurocomputing. Springer.
11. M. Vasylkivskyi, O. Horodetska, B. Klymchuk, i V. Hovorun, «Stratehii tekhnolohichnoho rozvytku aparatnoho zabezpechennia infokomunikatsiinykh radiomerezh», ITKI, vyp. 56, vyp. 1, s. 83–91, Ber 2023. [in Ukrainian].
12. Vasylkivskyi, M., Prykmeta, A., Oliinyk, A., & Nikitovych, D. (2023). Optymizatsiia intelektualnykh telekomunikatsiinykh merezh. Visnyk Khmelnytskoho natsionalnoho universytetu, Tekhnichni nauky. – 2023. – № 1. (317). – S. 33–41. doi: 10.31891/2307-5732-2023-317-1-33-41 [in Ukrainian].

© Васильківський М.В., Коломієць А.А., Будаш М.В., Прикмета А.В., Олійник А.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-42
УДК: 621.3:004.9
Круліковський Борис Борисович, к.т.н., доцент
https://orcid.org/0000-0001-7944-551X
Степанюк Юрій Ігорович, студент
Національний університет водного господарства та природокористування, м. Рівне, Україна

ЛОКАЛЬНА АТС ПІДПРИЄМСТВА НА БАЗІ КОМП'ЮТЕРНОЇ МЕРЕЖІ

Круліковський Б.Б., Степанюк Ю.І. Локальна АТС підприємства на базі комп'ютерної мережі. В роботі проведено аналіз сучасних телекомунікаційних технологій на предмет можливостей і потенційних вразливостей, які можуть виникнути внаслідок російського повномасштабного вторгнення в Україну. Після обгрунтованого врахування переваг IP-технологій, найбільше уваги приділено їх використанню для побудови надійної і захищеної системи телефонного зв'язку підприємства. При цьому підкреслено економічну вигідність поєднання в одній системі спільної роботи проводових і стільникових абонентів, які отримують послуги від незалежних операторів зв'язку. Це зменшує залежність надійності каналів від геолокації базових станцій і дозволяє розширити географію стабільно доступних абонентів. На думку авторів найбільш ефективною є інтеграція традиційних технологій зв'язку з IP-технологіями на основі використання власної цифрової міні-АТС, керування якою доцільно покласти на персональний комп'ютер. Безперервна комунікація співробітників можлива тільки завдяки створенню автономної інформаційно-комунікаційної структури на основі існуючої локальної комп'ютерної мережі підприємства. Це дає підстави пропонувати функціональну схему захищеного телефонного зв'язку між абонентами підприємства та зовнішним середовищем на основі використання власної локальної обчислювальної мережі (ЛОМ). Така система дозволяє інтегрувати всі наявні в межах підприємства аналогові і цифрові апаратні засоби та забезпечує стабільний зв'язок навіть в умовах відключення енергоживлення на досить тривалий час. Важливою є також можливість підтримки сталого зв'язку без необхідності одноразових суттєвих капіталовкладень на модернізацію інфраструктури.
Ключові слова: телефонний зв'язок, інформаційно-телекомукаційна система, IP-телефонія, надійність

Krulykovskyi B., Stepaniuk Y. Local PBX of an enterprise based on a computer network. The paper analyzes modern telecommunication technologies for opportunities and potential vulnerabilities that may arise as a result of Russia's full-scale invasion of Ukraine. After carefully considering the advantages of IP technologies, the article focuses on their use to build a reliable and secure enterprise telephone system. At the same time, the author emphasizes the economic profitability of combining wireline and cellular subscribers receiving services from independent telecommunications operators in one system. This reduces the dependence of channel reliability on the geolocation of base stations and allows to expand the geography of consistently available subscribers. According to the authors, the most effective way is to integrate traditional communication technologies with IP technologies through the use of a dedicated digital mini-PBX, which should be controlled by a personal computer. Continuous communication of employees is only possible through the creation of an autonomous information and communication structure based on the existing local computer network of the enterprise. This gives grounds to propose a functional scheme of secure telephone communication between enterprise subscribers and the external environment based on the use of its local area network (LAN). Such a system allows you to integrate all the analog and digital hardware available within the enterprise and ensures stable communication even in the event of a power outage for a sufficiently long time. It is also important to be able to maintain stable communication without the need for one-time significant investments in infrastructure modernization.
Key words: telephone communication, information and telecommunication system, IP-telephony, reliability

Постановка проблеми. Воєнний час в Україні формує додаткові вимоги до роботи критично важливого комп'ютерного обладнання та засобів зв'язку: для нормального функціонування офісу підприємства в продовж робочого дня необхідно враховувати можливість припинення подачі електроживлення екстрено та за графіком, а також тимчасового припинення роботи підприємств стаціонарного (провідного) та стільникового зв'язку.

При наявності на підприємстві офісних провідних телефонів міської телефонної мережі в момент блекауту зв'язок стає неможливим. Масова наявність стільникових телефонів у працівників також не гарантує безперервність зв'язку між ними через підключення стільникових телефонів до інших операторів, які мають різну готовність оперативного та тривалого продовження роботи свого обладнання в разі припинення подачі електроживлення. Безперервність електроживлення вирішується за рахунок використання пристроїв UPS, автономних електрогенераторів, накопичувачів електроенергії, але безперервна комунікація співробітників можлива тільки завдяки створенню автономної інформаційно-комунікаційної структури на основі існуючої локальної комп'ютерної мережі.

Мета дослідження полягає у забезпеченні стійкої енергонезалежної та захищеної системи електрозв'язку підприємства шляхом розробки та впровадження гібридної мультисервісної цифрової телефонної системи на основі комп'ютерної мережі підприємства з власною цифровою АТС. Це дасть можливість:

© Круліковський Б.Б., Степанюк Ю.І.

- об'єднати наявні аналогові, стільникові засоби зв'язку, комп'ютеризовані робочі місця та наявну комп'ютерну мережу у єдину автономну інтегровану інформаційно-комунікаційну систему підприємства;

- розподілити капітальні вкладення та витрати на підготовку технічного персоналу до етапу поступового переходу від аналогової до цифрової телефонної системи;

- перейти від аутсорсингу послуг телефонних компаній та дата-центрів до самостійного керування власною автономною інформаційно-комунікаційною системою.

Завдання дослідження. Для досягнення поставленої мети необхідно

вирішити наступні основні завдання:

1) провести аналіз існуючих технологій телефонії з акцентом уваги на їх перевагах і недоліках.

2) проаналізувати наявне на ринку обладнання та програмне забезпечення, що придатне для створення автономних офісних телефонних мереж, та здійснити обгрунований вибір оптимального варіанту.

3) здійснити розробку функціональної схеми телекомунікаційної системи на основі обраних засобів апаратно-програмного забезпечення.

4) встановити можливості модифікації існуючої локальної комп'ютерної мережі підприємства для застосування її в якості основи автономної інтегрованої мультибазисної інформаційно-комунікаційної системи підприємства.

Схемотехнічні рішення АТС на основі технології IP-телефонії.

Існує декілька варіантів схемотехнічних рішень для створення офісної АТС на базі IP-телефонії.

Організація АТС на базі IP-телефонії із зовнішнім менеджментом – рішення, для тих, хто не хоче мати клопоту з розгортанням та адмініструванням власного серверу або міні IP-АТС (рис. 1).

Великі оператори зв'язку використовують не тільки фізичні сервери для керування телефонними мережами клієнтів але й переносять функції VoIP-АТС у віртуальне середовище – хмару.

Хмара – це група серверів [1, 48с], які функціонують як одне ціле. Їхні спільні потужності поділяються на окремі віртуальні сервіси, як наприклад, IP АТС клієнтів. Цим досягається максимально ефективне використання потужностей серверного середовища, тому що окремі сервери відносно не залежать від активності окремого клієнта.

Для абонента послуг телефонного оператора і віддаленого менеджменту його телефонної мережі немає різниці яким чином реалізовані функції IP АТС: фізично чи віртуально . Всі питання повнофункціональної та безперебійної роботи АТС лягають на плечі телефонної компанії.

Організація телефонної мережі зводиться до підключення груп SIP/VoIP телефонів або софтфонів (з ПК або смартфонів) до маршрутизаторів, які підключаються по виділеному каналу до телефонного оператора. Тобто АТС компанії фізично знаходиться в приміщені телефонного оператора – провайдера IP-телефонії. Це може бути або фізичне обладнання або віртуальна АТС у хмарі провайдера у складі всіх необхідних серверів.

Маршрутизатор має підтримувати технології VoIP/SIP. Зрозуміло, засоби IP-телефонії підключаються до маршрутизатора через порти Ethernet RJ-45 або для підключення у SIP мережу використовуються швидкісні Ethernet 100/1000 Мб/с.

Якщо компанія продовжує використовувати аналогові телефони, факси, необхідно придбати VoIP-шлюз, через який вони можуть підключаються до комп'ютерної мережі.

Наприклад, аналоговий шлюз Grandstream GXW4232 [2], наприклад, оснащений 32 портами FXS, що дозволяють підключати як аналогові телефонні, так і факсимільні апарати. Інтеграція в SIP мережу проводиться з використанням високошвидкісного інтерфейсу Ethernet 10/100/1000 Мб/с.

VoIP-шлюз є мостом між IP і аналоговою телефонною мережею.

Залежно від того, звідки надходить голосовий трафік, шлюз VoIP перетворює голосовий трафік у належну форму для цільових мереж, таких як IP або для аналогової телефонної мережі загального користування (ТМЗК).

Якщо голосовий трафік надходить з ТМЗК або від аналогових телефонів, шлюз VoIP перетворює аналогові голосові сигнали в цифрові. Потім він стискає і перетворює цифрові сигнали в серію пакетів даних, які передаються через IP-мережу за допомогою протоколу сигналізації.

Якщо голосовий трафік надходить з IP-мережі, шлюз VoIP розпаковує пакети даних в цифрові сигнали, а потім передає їх в аналогові канали.

© Круліковський Б.Б., Степанюк Ю.І.

В залежності від напрямку перетворення аналогового сигналу в цифровий є три типи шлюзів (рис. 2.):

- З FXS-портами, якщо треба підключити корпоративну IP-АТС до звичайної міської АТС;
- з FXO-портами, для підключення аналогових телефонів до цифрової телефонної мережі;
- гібридні VoIP-шлюзи, оснащені портами обох типів: FXO, FXS. У роз'єм FXO під'єднують аналогову лінію, а в FXS-порт — цифрову АТС (затребувані при відсутності у останньої вільних портів).

Додатково, для інтеграції мобільних абонентів зі сматрфонами використовуються VoIP GSM шлюзи (рис. 2.).

Використання VoIP GSM шлюзів дає можливість дзвонити в межах офісу зі стаціонарних телефонів на стільникові і навпаки, а також зв'язуватись між мобільними абонентами різних операторів без використання послуг операторів стільникового і провідного зв'язку.

При виборі шлюзу VoIP необхідно враховувати:

- навантаження дзвінками: переконайтеся, що вибраний вами шлюз може впоратися з необхідною пропускною здатністю голосових дзвінків VoIP, з можливим подальшим зростанням на 20-25%;
- підтримку протоколів VoIP: найпоширенішими протоколами сигналізації дзвінків VoIP є SIP та H323.
-сумісність з портами роутера або АТС.

Щоб скористатися таким способом організації офісної АТС необхідно:

- вибрати провайдера IP-телефонії;
- підключити тариф АТС в особистому кабінеті на сайті провайдера (розрізняються функціоналом і кількістю робочих місць, що входять в пакет);
- додати додаткові платні послуги (оплачуються по абонентській платі) і співробітників понад пакет;
- налаштувати сценарії прийому дзвінків, голосову пошту, Запис дзвінків, переадресацію та інші функції;
- приступити до роботи.

***Перевагами підключення є*:**

- незначний обсяг капіталовкладень по заміні/купівлі обладнання;
- відсутність необхідності модернізації офісної АТС та оновлення її програмного забезпечення;
- відсутність потреби тримати або оплачувати технічного спеціаліста з адміністрування офісної АТС;
- можливість використання в телефонній мережі аналогових телефонів і факсів (при застосуванні VoIP-шлюзу);
- простота експлуатації.

Недоліками підключення до оператора VoIP телефонії є:

- необхідність придбання маршрутизатора (роутера) з підтримкою VoIP;
- дзвінки в середині компанії відбуваються через зовнішню АТС;
- необхідно постійно оплачувати послуги телефонного оператора.
- функціональні можливості АТС обмежуються переліком послуг телефонного оператора в межах замовленого тарифного пакету.

Офісна АТС на базі IP-телефонії з орендованим сервером у хмарі.

В сучасній практиці клієнт не обов'язково має володіти власною АТС або оплачувати послуги по керуванню власною телефонною мережею телефонній компанії. З власних особливих міркувань та при наявності власних спеціалістів з обслуговування програмного забезпечення АТС компанія може орендувати сервер у Дата-центрі для встановлення на ньому необхідного програмного забезпечення, в тому числі такого, що виконує функції IP АТС.

Оренда сервера – це можливість отримати необхідну кількість потужностей і ресурсів для реалізації власних задач, не витрачаючи великі гроші на покупку дорогого обладнання. В цьому випадку немає необхідності нести додаткові витрати, займатися пошуком і навчанням спеціалізованого персоналу.[3, 4с]

При цьому власник сервера (Орендодавець) самостійно забезпечує:

- повний контроль роботи сервера та програного забезпечення на ньому клієнтом;

© Круліковський Б.Б., Степанюк Ю.І.

- безперебійне цілодобове 24х7 живлення сервера;
- комфортні умови роботи обладнання, як-то підтримку необхідних кліматичних параметрів у приміщені;
- надання в оренду високоякісного надійного швидкісного обладнання від глобально відомих виробників з заслуженою репутацією;
- регулярне технічне обслуговування обладнання (а при необхідності, виконує його ремонт, заміну);
- забезпечує встановлення ліцензійної операційної системи, інших програм, необхідних для функціонування сервера;
- своєчасне оновлення серверного програмного забезпечення;
- надає послуги з безпеки роботи програмного забезпечення клієнта та ліній його зв'язку;
- надає лінії високошвидкісного Інтернету;
- надає цілодобову підтримку.

В деяких дата-центрах також забезпечується моніторинг зависання програм клієнта та автоматичні та систематичні мануальні бекапи програмного забезпечення на сервері для наступного його відновлення на час настання проблеми, у разі необхідності.

Панель керування при оренді виділеного сервера надається клієнту, як правило, безкоштовно.

Від клієнта вимагається наявність певних компетенцій для встановлення, налаштування свого програмного забезпечення на сервері, контролю його функціонуванням та сумісністю з встановленою на сервері операційною системою.

Офісна АТС на базі IP-телефонії з програмним сервером.

Найбільш розповсюдженими схемами організації IP-телефонної мережі компаній є схеми на основі IP-телефонії [5, 636с] з різноманітними схемами підключення до постачальника послуг телефонії.

Враховуючі можливості, які надає цифровий телефонний зв'язок основним ядром сучасної офісної мережі є апаратна або програмна IP АТС, а також невід'ємними компонентами є:

- IP телефони;
- софтфони на ПК, ноутбуках, що підключаються через кабельну мережу Ethernet;
- софтофони на базі ноутбуків, що підключаються у телефонну мережу через Wi-Fi;
- пристрої комутації IP телефонів (SIP, VoIP шлюзи, комутатори (switch), роутери з підтримкою цих технологій);
- пристрої комутації у IP-телефонну мережу аналогових телефонів, факсів (VoIP FXS шлюз);
- пристрої комутації у IP-телефонну мережу стільникових телефонів (VoIP GSM шлюз).

Процес будівництва офісної телефонної IP- мережі, в компанії, яка користується тільки аналоговими телефонними апаратами та факсами, починається з організаційних заходів по адаптації існуючого парку засобів зв'язку в нову цифрову телефонну мережу. Тобто необхідно придбати VoIP FXS шлюзи, комутатори та роутери з підтримкою SIP, VoIP, в залежності від вибраної моделі використання телефонів на робочих місцях. Придбання додаткового комутаційного обладнання для не критично важливих підключень, як-то VoIP GSM шлюз, нові IP телефони та шлюзи до них може здійснюватись в робочому порядку, в міру виділення коштів та заміні аналогових телефонів на цифрові.

Схема офісної АТС в цьому випадку буде виглядати як на рис. 3.

Наступним кроком в ході насичення телефонної мережі IP-обладнанням може бути перехід з програмної реалізації офісної АТС на апаратну з придбанням готової IP АТС.

Офісна АТС на базі IP-телефонії та міні-АТС.

Варіантом розвитку власної офісної АТС є застосування апаратних IP АТС в готовому до роботи стані від виробника (рис. 4). Цей підхід вимагає від компанії великих одномоментних вкладень в розрізі фінансів та повної готовності технічного персоналу до розгортання, налаштування та пере конфігурації АТС, в процесі внесення змін у топологію телефонної мережі і підключення нових абонентів.

У такого підходу є свої переваги і недоліки.

На прикладі недорогого, потужного і компактного рішення IP-АТС Yeastar MyPBX 1600 (рис. 5) [4] продемонструємо це.

© Круліковський Б.Б., Степанюк Ю.І.

MyPBX 1600-це апаратна АТС, яка має прошивку на базі Oktell. У пристрої не налагоджений повний доступ до конфігурації Oktell, але разом з цим спрощені налаштування - для цього достатньо використовувати веб-інтерфейс. Це означає, що функціонал АТС звужується, але управління стає доступним навіть для недостатньо навчених співробітників[4].

Підключення АТС до працюючої інфраструктури здійснюється через звичайний комутатор. Якщо в компанії локальна мережа вже налагоджена, з процесом не виникає ніяких проблем.

Основними характеристиками MyPBX 1600 є: підтримка до 22 розмов одночасно; підтримка до 100 користувачів системи; автовідповідач з гнучкими налаштуваннями; голосова пошта; конференц-кімнати; парковка викликів; черговість викликів; переклад, захоплення, трансфер виклику; очікування дзвінка; маршрутизація дзвінка; АВН; Режим «не турбувати»; Call Detail Records; музичний супровід в режимі очікування; передача факсу, в тому числі на електронну пошту.

Крім того, існують додаткові можливості: переадресація викликів; доступ до зовнішніх ліній АТС-DISA; особистий кабінет користувача; запис розмов; груповий дзвінок; підтримка відео; SIP SMS; відображення статусу абонента.

Що є кращим – цей варіант чи хмарне (віртуальне рішення)?

1. Важлива перевага апаратних АТС-відсутність щомісячної абонентної плати SIP-провайдеру. Залишається тільки абонплата за користування каналом. Підрахунки показали, що апаратна АТС окупає себе вже через 10 місяців.

2. Можливість підключення діючого факсу і аналогової лінії до апаратної АТС. Коли лінія підключена через шлюз з FXO портом, аналоговий факс підключити набагато складніше, при цьому є велика вірогідність, що нічого не запрацює.

3. Керування телефонною станцією апаратного типу максимально спрощено, веб-інтерфейс багатьма мовами. Головні налаштування можна вибрати за півгодини. У цьому випадку, звичайно, хмарні рішення кращі - Налаштування вже встановлені для користувачів. Покупцям залишається тільки виконати налаштування реєстрації безпосередньо на телефонах.

4. Використовуючи апаратну АТС, можна вибирати будь-якого провайдера. Якщо використовується хмарна АТС, доводиться вибирати провайдера, який запропонований компанією-постачальником. Апаратна станція передбачає вказівку декількох SIP провайдерів і виконання маршрутизації за найдешевшими напрямками. Також є можливість вибрати SIP-провайдера в іншій країні і говорити за місцевими цінами.

5. Обидва варіанти передбачають підключення віддалених операторів, а також віддалених офісів. Це виконується легко, практично граючи. Єдиний нюанс - апаратна АТС потребує більшого об'єму налаштувань (кидок портів на маршрутизації, базові правила безпеки). Цей пункт також охоплює свободу переміщення, адже SIP реєстрація взагалі не прив'язана до місця знаходження АТС. Це означає, що однаковим акаунтом можна підключатися з будь-якого комп'ютера з підключеним Інтернетом.

6. Якщо розмови по Інтернету потрібно захистити, віддалених співробітників або цілі офіси можна підключити до апаратної станції через VPN тунелі. Для хмарної станції найчастіше не передбачена ця функція, тому, при бажанні, передані відомості перехоплюються простіше.

Офісна АТС на основі IP-телефонії та ЛОМ.

На основі аналізу наведених варіантів реалізації корпоративної системи телефонного зв'язку керівництву підприємства запропонована гібридна мультисервісна цифрова телефонна система, структурна схема якої наведена на рис.6.

Обрана схема гібридної мультисервісної цифрової телефонної системи включає різноманітні комунікаційні елементи, що забезпечують високий рівень гнучкості і можливостей для офісного середовища. Основними компонентами системи є:

1. Маршрутизатори (роутери), що використовуються для забезпечення з'єднання з Інтернетом та управління мережевим трафіком. Маршрутизатори також можуть здійснювати функції фаєрволу для забезпечення безпеки мережі.

2. Комутатори (switch**),** що виконують функції блочної комутації мережі в офісі, завдяки чому дозволяють ефективно керувати трафіком між різними пристроями в ЛОМ.

3. FXO VoIP шлюзи призначені для підключення аналогових телефонів і факсів до VoIP-мережі. Їх основна функція - перетворення сигналів з аналогових пристроїв у цифрові для передачі через IP-мережу.

© Круліковський Б.Б., Степанюк Ю.І.

4. GSM VoIP шлюзи, що дозволяють інтегрувати стільникові телефони в VoIP-мережу шляхом з'єднання між IP-мережею і стільниковими операторами.

5. E1 VoIP шлюзи призначені для підключення SIP-телефонів до цифрових телефонних ліній (E1). Ці шлюзи забезпечують високу пропускну здатність і можливість одночасного з'єднання декількох телефонів.

6. Програмно-апаратна чи апаратна АТС, яка є центральним елементом системи і керує всіма з'єднаннями та функціями телефонної системи. АТС може бути як програмною (наприклад, Asterisk), так і апаратною (наприклад, PBX від виробників, таких як Avaya, Cisco тощо)

7. Мережевий комутатор або маршрутизатор забезпечує підключення всіх компонентів системи до мережі Інтернет і/або внутрішніх корпоративних мереж.

8. Інтернет-транки (зовнішні лінії) для підключення до зовнішніх телефонних мереж (транків) через SIP/VoIP провайдерів для здійснення викликів на загальнодоступні телефонні номери і міжнародні виклики.

9. Віддалені внутрішні абоненти**:** Це можуть бути віддалені користувачі або інші локалізації офісів, що підключені до основної телефонної системи через інтернет.

10. IP-телефони і софтфони. Це цифрові телефони, які підключені безпосередньо до мережевого комутатора чи маршрутизатора через Ethernet. Вони використовують IP-протокол для передачі голосової інформації по мережі.

11. DECT безпровідні телефони, що підключаються до базової станції, яка, в свою чергу, підключена до мережевого комутатора чи маршрутизатора через Ethernet.

Міні-АТС Yeastar S50 підтримує одночасне підключення як IP-телефонів, так і аналогових, що забезпечує гнучкість і можливість інтеграції з різними типами обладнання в одній системі. При цьомуYeastar S50 підтримує різні моделі IP-телефонів від різних виробників шляхом підключення безпосередньо до локальної мережі через Ethernet-порти на міні-АТС або через внутрішню WiFi мережу (якщо ця опція підтримується). З'єднання зазвичай здійснюється за допомогою протоколу SIP (Session Initiation Protocol), що є стандартом для VoIP-комунікацій. Yeastar S50 підтримує SIP для налаштування і управління IP-телефонами в мережі. В такому випадку IP-телефони можуть використовувати всі функції, які надає Yeastar S50, такі як конференц-звіти, групова інтеграція з CRM-системами і т. д. Кожен IP-телефон має свою унікальну IP-адресу і налаштовується через веб-інтерфейс або спеціальне програмне забезпечення.

Yeastar S50 також підтримує підключення аналогових телефонів через FXS порти (Foreign Exchange Station), що можуть підтримувати звичайні аналогові телефони або факсимільні апарати. Yeastar S50 має вбудовані FXS порти, які відповідають за підключення аналогових пристроїв, кількість яких залежить від моделі (наприклад, S50 підтримує до 8 FXS портів). Аналогові телефони конфігуруються через веб-інтерфейс Yeastar S50, де можна налаштувати номери телефонів, вибрані групи, правила маршрутизації і т. д.

Yeastar S50 забезпечує управління як IP-телефонами, так і аналоговими пристроями через єдиний інтерфейс управління, за допомогою якого можна легко налаштовувати і керувати обома типами телефонів, розподіляти номери та здійснювати інші адміністративні функції.

Міні-АТС (Private Branch Exchange, PBX) і гібридна АТС, які були створені з використанням Asterisk, відображають гнучкість і можливості цього програмного забезпечення з відкритим кодом в сфері телефонії й комунікацій. Asterisk - це інструментарій, який можна використовувати окремо або в поєднанні з іншими проектами для створення Інтернету речей [6,7]. Слід пам'ятати, що встановлення, налаштування та конфігурування Asterisk передбачають, що це робить досвідчений адміністратор Linux-систем.

Все вищенаведене є підставою стверджувати, що застосування IP-технології та наявність широкої гами комутаційного обладнання при модернізації офісної телефонної мережі надають наступні можливості:

а) здійснювати поетапний перехід від аналогового обладнання до цифрового, з поступовою повною заміною фізично та морально застарілих моделей на цифрові;

б) перейти від офісної телефонної системи на базі старого типу телефонів/факсів та від отримання послуг телефонної компанії із зовнішнім керуванням процесами у телефонній мережі підприємства до самостійного керування автономною цифровою АТС з повномасштабним контролем всього комунікаційного обладнання і всіх технологічних процесів зв'язку;

© Круліковський Б.Б., Степанюк Ю.І.

в) продовжити експлуатацію старих аналогових телефонів після переходу від офісної телефонної системи на базі старого типу телефонів/факсів до самостійного керування автономною цифровою АТС.

Завдяки розподілу процесу у часі на етапи, впровадження технології IP-телефонії стає безболісним в плані поступових капіталовкладень, а також можливості поступової підготовки власного технічного персоналу, який в процесі розбудови телефонної мережі має час для підвищування своєї кваліфікації до потрібного рівня з метою самостійного створення, налаштування та змінювання конфігурації офісної IP АТС.

Список бібліографічного опису

1. Зінченко О.В., Іщеряков С.М., Прокопов С.В., Сєрих С.О., Василенко В.В. Хмарні технології. – Навчальний посібник. – К: ФОП Гуляєва В.М., 2020.
2. Grandstream GXW4232. Grandstream– веб-сайт. – URL: https://www.grandstream.com/products/ gateways-and-atas/voip-gateways/product/gxw4200-series.
3. Learning Amazon Web Services (AWS) - A Hands on Guide to the Fundamentks of AWS - Mark Wilkins - Addison-Wesley, 2020
4. On-premix IP PBX System. Yeastar– веб-сайт. – URL: https://www.yeastar.com/ip-pbx/.
5. Телекомунікаційні та інформаційні мережі: Підручник [для вищих навчальних закладів] / П.П. Воробієнко, Л.А. Нікітюк, П.І. Резніченко. – К.: САММІТ-Книга, 2010. – 708 с.: іл
6. MyPBX Standart - веб-сайт - URL: https://www.yeastar.com/de/mypbx-standard/
7. Getting started with Asterisk - веб-сайт - URl: https://www.asterisk.org/get-started/

References

1. Zinchenko O.V., Ishcheryakov S.M., Prokopov S.V., Serykh S.O., Vasylenko V.V. Cloud technologies. – Study guide. – To: FOP Gulyaeva V.M., 2020.
2. Grandstream GXW4232. Grandstream– веб-сайт. – URL: https://www.grandstream.com/products/ gateways-and-atas/voip-gateways/product/gxw4200-series
3. Learning Amazon Web Services (AWS) - A Hands on Guide to the Fundamentks of AWS - Mark Wilkins - Addison-Wesley, 2020
4. On-premix IP PBX System. Yeastar– веб-сайт. – URL: https://www.yeastar.com/ip-pbx/
5. Telecommunication and information networks: Textbook [for higher educational institutions] / P.P. Vorobienko, L.A. Nikityuk, P.I. Reznichenko. - K.: SAMMIT-Knyga, 2010. - 708 p.
6. MyPBX Standart - веб-сайт - URL: https://www.yeastar.com/de/mypbx-standard/
7. Getting started with Asterisk - веб-сайт - URl: https://www.asterisk.org/get-started/

© Круліковський Б.Б., Степанюк Ю.І.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-43
УДК 621.391.8
Мороз Сергій Анатолійович, к.т.н., доцент
https://orcid.org/0000-0003-4677-5170
Якимчук Наталія Миколаївна, к.т.н., доцент
https://orcid.org/0000-0002-8173-449X
Селепина Йосип Романович, к.т.н., доцент
https://orcid.org/0000-0002-2421-1844
Чалий Василь Дмитрович, к.т.н., асистент
https://orcid.org/0000-0002-6592-6715
Луцький національний технічний університет, м. Луцьк, Україна

РОЗВИТОК NFC ТА RFID: ВИКОРИСТАННЯ ІНДУКТИВНИХ АНТЕН У СУЧАСНИХ ПРИСТРОЯХ БЛИЖНЬОЇ ДІЇ

Мороз С.А., Якимчук Н.М., Селепина Й.Р., Чалий В.Д. Розвиток NFC та RFID: використання індуктивних антен у сучасних пристроях ближньої дії. В роботі розглянуто сучасні технології ближнього безконтактного зв'язку (NFC) та радіочастотної ідентифікації (RFID), які широко застосовуються у сферах логістики, платіжних систем та управління доступом. Наведено дослідження розвитку стандартів протоколу NFC та ключових параметрів, які критично впливають на ефективність роботи NFC-пристроїв. Основна увага приділена аналізу та вдосконаленню методів розрахунку індуктивності антен, що використовуються в цих технологіях. Описані різні типи антен, включаючи круглі, квадратні, шестикутні та восьмикутні котушки, а також запропоновані формули для розрахунку їх індуктивності. Досліджено можливості адаптивних антен та антен складних конструкцій, що дозволяють підвищити ефективність передачі енергії. Здійснено аналіз впливу геометрії та матеріалів антен на продуктивність NFC та RFID систем. Особлива увага приділяється майбутнім перспективам цих технологій у контексті зростання числа смарт-пристроїв та їх інтеграції в сучасні телекомунікаційні системи. Отримані результати можуть бути використані для розробки більш надійних та ефективних рішень у сфері безконтактних технологій, а також для вдосконалення систем енергозабезпечення і передачі даних.
Ключові слова: NFC, RFID, мережі ближньої дії, антена, індуктивність, адаптивність.

Moroz S., Yakymchuk N., Selepyna Y., Chalyj V. Development of NFC and RFID: Use of Inductive Antennas in Modern Near-Field Devices. The paper discusses modern Near-Field Communication (NFC) and Radio-Frequency Identification (RFID) technologies, which are widely used in logistics, payment systems, and access control. The study presents an analysis of the development of NFC protocol standards and key parameters that critically affect the performance of NFC devices. The main focus is on analyzing and improving methods for calculating the inductance of antennas used in these technologies. Various types of antennas are described, including circular, square, hexagonal, and octagonal coils, along with proposed formulas for calculating their inductance. The potential of adaptive antennas and complex antenna designs to improve energy transmission efficiency is explored. Additionally, the research focuses on the impact of antenna geometry and materials on the performance of NFC and RFID systems. Special attention is paid to the future prospects of these technologies in the context of the growing number of smart devices and their integration into modern telecommunications systems. The results obtained can be used to develop more reliable and efficient solutions in the field of contactless technologies, as well as for improving energy supply and data transmission systems.
Keywords: NFC, RFID, near-field networks, antenna, inductance, adaptability .

Постановка наукової проблеми. В сучасних умовах підвищення потреб користувачів у задоволенні широкого спектру послуг зв'язку та передачі інформації, можливостями значної мініатюризації пристроїв доступу, а також повсюдним переходом до безконтактних способів керування як об'єктами так і інформацією, особам-користувачам необхідно користуватись значною кількістю «електронних ключів». Такий ключ являє собою електронну мітку, що найчастіше виготовляється у вигляді смарт-карти, пристрою на основі пасивних RFID міток [1, 2]. Для уникнення взаємного впливу міток і забезпечення вимоги мінімізації розмірів пристроїв об'єднання таких елементів в одному корпусі не є оптимальним. Тому було розроблено комплекс технологій безконтактного передавання інформації, частиною яких є NFC передача даних. З розвитком інтернету речей (IoT), збільшенням кількості смарт-пристроїв, інтенсифікацією цифрової комунікації та потребою в безпечних і зручних рішеннях, технології на основі NFC знаходять широке застосування. Вони використовуються при платі за товари і послуги, в логістиці, управлінні доступом до приміщень, захисті інформації та ідентифікації особи. Дана технологія дозволяє об'єднати всі необхідні користувачу смарт-карти в одному пристрої, а також забезпечити безпроблемне отримання ідентифікаторів для доступу до різних ресурсів. На даний момент NFC знаходиться на етапі активного розвитку, при цьому для їх широкого впровадження необхідно

© Мороз С.А., Якимчук Н.М., Селепина Й.Р., Чалий В.Д.

знаходити рішення для покращення таких параметрів як ефективність енергопередачі, дальність зчитування, безпека і вартість.

Аналіз досліджень. NFC (Near Field Communication) була розроблена у 2002 році компаніями Sony і NXP Semiconductors для роботи на відстані до 10 сантиметрів. Обладнання працює в частотному діапазоні з центральною частотою 13,56 МГц і дозволяє передавати дані на швидкості до 424 кбіт/с. Дана технологія NFC має багато спільного з радіочастотною ідентифікацією (RFID). Проте NFC підтримує двосторонню передачу даних, що дозволяє одночасно передавати та отримувати інформацію (активно-активний або активно-пасивний зв'язок). У випадку RFID більшість систем забезпечують лише односпрямований зв'язок — зчитувач передає запит, а мітка відповідає.

Крім того, NFC сумісна з поширеною інфраструктурою смарт-карт, побудованою на стандартах ISO/IEC 14443 A (наприклад, з технологією MIFARE компанії NXP) і ISO / IEC 14443 B, так само як і з картами FeliCa компанії Sony (JIS X 6319- 4). Для обміну інформацією між двома NFC-пристроями був розроблений новий протокол, який визначається стандартами ECMA-340 і ISO / IEC 18092. У 2004 році для узгодження NFC-методів і стимуляції їх розвитку компаніями NXP, Sony і Nokia була заснована асоціація NFC Forum. Асоціація розробила технічні вимоги, що забезпечують функціональну сумісність NFC-пристроїв і сервісів. У них включені всі згадані вище стандарти (ISO / IEC 14443 A, B, ISO / IEC 18092 та JIS X 6319-4 / FeliCa). Починаючи з грудня 2010 року, асоціація NFC Forum проводить сертифікацію NFC-пристроїв.

Можна прослідкувати розвиток радіочастотних стандартів і стандартів протоколу NFC зі специфікацією випробувань (рис. 1.) [2,3]. Різними компаніями (NXP, Infineon і Sony) були рекомендовані три стандарти радіочастотної ідентифікації: ISO / IEC 14443 A, ISO / IEC 14443 B і JIS X6319-4. Першим радіочастотним стандартом NFC був ECMA 340, заснований на радіоінтерфейсі стандартів ISO / IEC 14443A і JIS X6319-4. ECMA 340 був прийнятий як стандарт ISO / IEC 18092. Приблизно в той же час три основні компанії з випуску кредитних карт (Europay, Mastercard, Visa) представили платіжний стандарт EMVCo, заснований на ISO / IEC 14443 A і ISO / IEC 14443 B. Працюючи в асоціації NFC Forum, обидві групи компаній узгодили радіоінтерфейси, які були названі NFC-A (на основі ISO / IEC 14443 A), NFC-B (на основі ISO / IEC 14443 B) і NFC-F (на основі FeliCa).

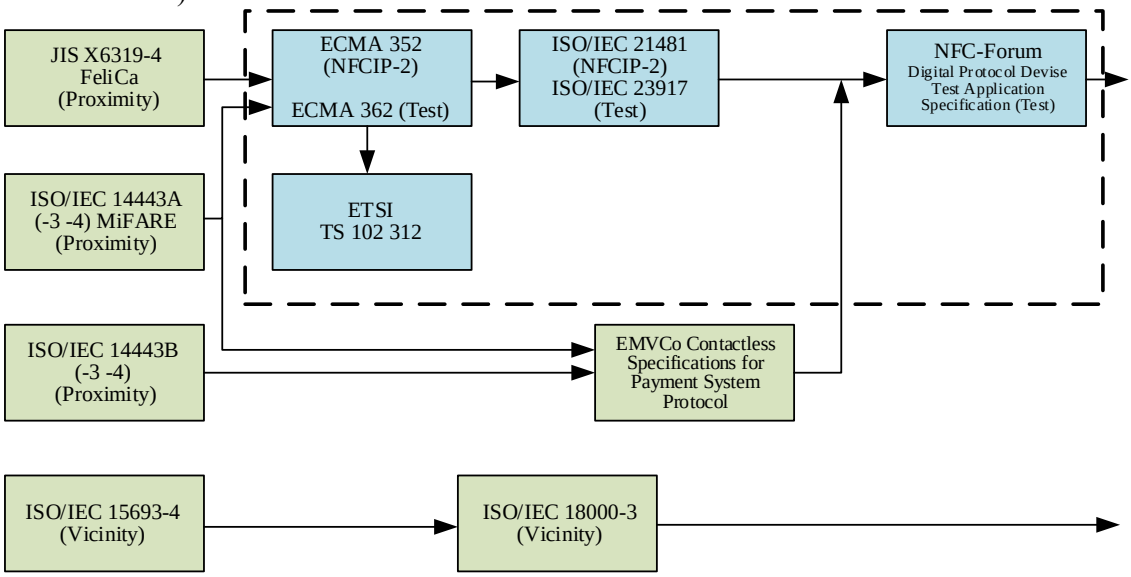


Рис. 1. Розвиток стандартів протоколу NFC

При передачі даних пасивній системі, наприклад, NFC-телефону в режимі емуляції карти, в якості джерела живлення пасивної системи використовується сигнал опитувального пристрою з частотою несучої 13,56 МГц. Як схема модуляції в опитуваних пристроях використовується амплітудна маніпуляція (ASK). Для режиму миттєвої передачі даних обидва напрямки передачі модулюються і кодуються як опитувальний пристрій. Однак в даному режимі потрібно менше потужності, оскільки кожен з NFC-пристроїв використовує своє власне джерело живлення і після закінчення передачі сигнал несучої знімається. Базовим для усіх технологій є використання електромагнітного поля для передачі енергії та інформації.

Ключові параметри, які критично впливають на ефективність роботи NFC-пристроїв випливають з базових принципів їх функціонування, що відповідає принциповій схемі роботи

© Мороз С.А., Якимчук Н.М., Селепина Й.Р., Чалий В.Д.

трансформатора. Розглянемо схему, що включає приймально-передавальний модуль, в ролі антени у якого виступає первинна обмотка та мітка (або чіп) у вигляді вторинної обмотки відповідно [4]. При піднесені зчитувача до мітки, через обмотку мітки протікає струм, від якого живиться чіп, який передає інформацію зчитувачу шляхом зміни опору в навантаженні обмотки (рис. 2).

Як відомо, відношення величини струму і відповідно напруги первинної обмотки трансформатора до вторинної залежить від відношення кількості витків обмоток: U1 / U2 = N1 / N2, проте опір обмоток змінюється в квадратичній залежності: Z1 / Z2 = (N1 / N2)2. Отже, невелика зміна опору в навантаженні буде явно виражена для опитувального пристрою. Неправильний розрахунок параметрів антени та відхилення при виготовленні від встановлених допусків може призвести до зниження дальності зчитування, перешкод у передачі сигналу або навіть до повної неспроможності мітки функціонувати належним чином. Крім цього слід зазначити, що точний розрахунок параметрів антени (вибір форми, довжина провідника, кількість витків, тощо) забезпечує відповідність потрібній резонансній частоті, що критично для ефективної роботи.

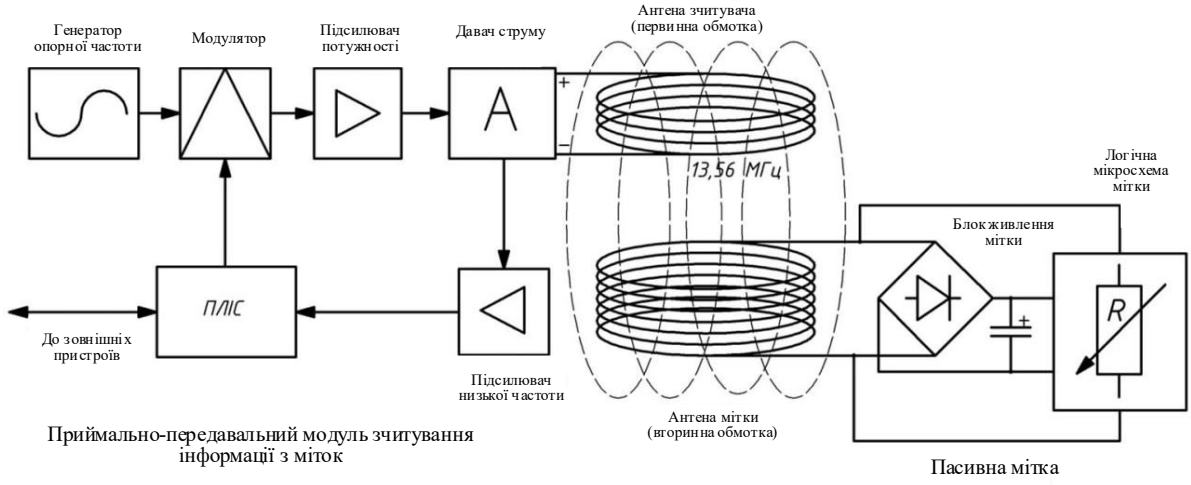


Рис. 2. Схема роботи міток та сканера

Правильний розрахунок таких параметрів, як індуктивність, ємність та геометрія антени, дозволяє оптимізувати відстань, на яку мітка може бути зчитана та мінімізувати вплив зовнішніх факторів. Для пасивних міток це особливо важливо, оскільки вся енергія для їх роботи передається через антену. Антени зі слабкими характеристиками можуть значно обмежити радіус дії системи. Отже, розрахунок параметрів антен для систем RFID та NFC є ключовим аспектом їх проектування та ефективності.

Мета роботи. Метою дослідження є аналіз і удосконалення методів розрахунку індуктивності котушок як провідних елементів антен для забезпечення ефективної роботи NFC та RFID пристроїв.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.

Електромагнітний зв'язок в системах RFID та NFC базується на застосуванні високочастотних електромагнітних полів і хвиль в НВЧ і мікрохвильовому частотних діапазонах.

Електромагнітне поле являє собою сукупність взаємопов'язаних електричного і магнітного полів. По суті, електричне і магнітне поля є двома характеристиками електромагнітного поля. Електромагнітне поле як форма матерії не може залишатися в спокої, а знаходиться в стані руху. Електромагнітне поле виникає в просторі при наявності мінливих в часі електричних струмів і зарядів. Останні збуджуються в провідних елементах антени. У вільному просторі електромагнітне поле поширюється у вигляді електромагнітної хвилі [2, 3]. Первинне магнітне поле, що генерується випромінюючим контуром, починається на антені. У міру того як поширюється магнітне поле, все більше і більше зростає електричне поле внаслідок індукції. Поле, яке спочатку було чисто магнітним, відповідно, перетворюється в електромагнітних поле.

Правильно спроектована антена повинна ефективно генерувати електромагнітне поле і забезпечувати максимальну передачу енергії до мітки для її живлення (у випадку пасивних міток). Проблеми розвитку NFC та RFID з точки зору використання індуктивних антен у сучасних пристроях можна розглядати через кілька основних аспектів. Головною проблемою можна назвати ефективність передачі енергії, оскільки в системах з пасивними мітками вся енергія для роботи мітки передається через індуктивну антену. Ефективність цього процесу напряму залежить від правильного розрахунку та оптимізації параметрів антени (індуктивність, форма, кількість витків). Невдало сконструйована антена може знижувати ефективність передачі енергії, що призводить до

© Мороз С.А., Якимчук Н.М., Селепина Й.Р., Чалий В.Д.

втрат у продуктивності та зменшення радіусу дії. При цьому слід враховувати сучасні тенденції до зменшення розмірів пристроїв і міток NFC та RFID, що створюють додаткові виклики для проектування індуктивних антен. Чим менша антена, тим складніше забезпечити достатній рівень індуктивності для стабільної роботи пристрою. Це обмежує можливості мініатюризації та підвищує складність розробки.

Крім того, передача енергії між двома лінійно поляризованими антенами буде оптимальною, якщо обидві антени мають однаковий напрямок поляризації. З іншого боку, передача енергії знаходиться на найнижчому рівні, коли напрямки поляризації передавальної і приймаючої антен розташовані щодо один одного точно під кутом 90° або 270° (тобто горизонтальна і вертикальна антени). У цій ситуації при розрахунку передачі енергії потрібно враховувати додаткове загасання 20 дБ через втрати поляризації, оскільки приймаюча антена сприймає лише 1/100 максимально можливої потужності від випромінюваного електромагнітного поля [3, 4].

У системах RFID та NFC з електромагнітним зв'язком, що працюють в НВЧ і мікрохвильовому діапазонах частот, застосовуються такі основні типи антен:

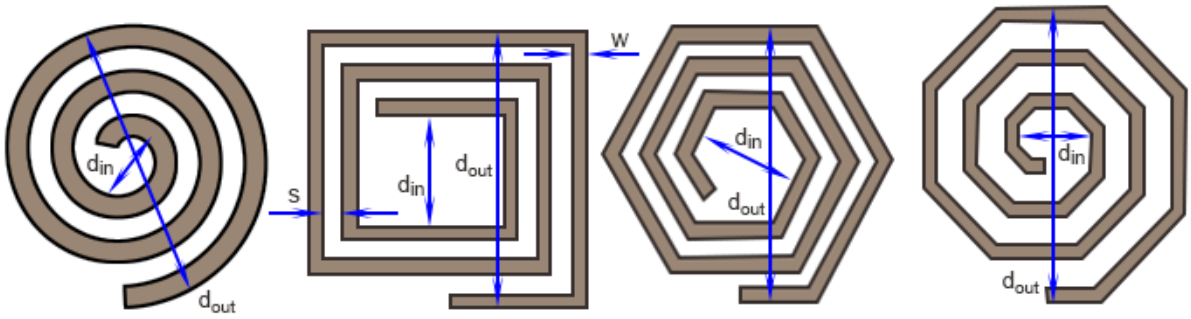
- антена-вібратор;
- антена типу «хвильовий канал»;
- мікросмугова антена;
- щілинна антена.

Параметри антен не залежать від того, чи використовуються антени для передачі або для прийому (принцип взаємності). Якщо відомі властивості антени при використанні її для передачі, то ці властивості дозволяють повністю оцінити її як прийомну і навпаки.

До основних параметрів антен відносяться:

- діаграма спрямованості;
- вхідний опір;
- коефіцієнт спрямованої дії;
- коефіцієнт підсилення;
- дійсна довжина (висота);
- ширина смуги пропускання.

Існують різноманітні методи для розрахунку різних типів антен для RFID та NFC технологій. Оскільки технології RFID та NFC інтенсивно розвиваються, то мітки можуть набувати різних форм, тому є актуальним завдання розрахунку різних форм антен для міток. Розглянемо один із способів розрахунку антен – визначення індуктивностей одношарових круглих, квадратних, шестикутних і восьмикутних планарних (плоских) котушок (рис. 3.).



Кругла Квадратна Шестикутна Восьмикутна

Рис. 3. Види плоских антен – індуктивностей для RFID та NFC технологій

Антени такого типу часто використовуються в мітках пристроїв радіочастотної ідентифікації (RFID) і ближнього безконтактного зв'язку (NFC), які отримують живлення за рахунок енергії електромагнітного поля, що створюється пристроєм зчитування міток. Вхідними параметрами для розрахунку є геометрія котушок, зокрема, їх зовнішній діаметр; ширина провідника або провода, з якого вони виготовлені; відстань між витками; число витків.

Для розрахунку індуктивності одношарової круглої спіральної котушки рекомендовано використовували формулу Гарольда А. Уіллера [2]:

$$L_{circ} = 31{,}33 \cdot \mu_0 \cdot n^2 \frac{a^2}{8a+11c} \quad (1)$$

© Мороз С.А., Якимчук Н.М., Селепина Й.Р., Чалий В.Д.

де L - індуктивність, (Гн); μ_0 - магнітна постійна, (Гн/м) , $\mu_0 = 4\pi \cdot 10^{-7}$; n - число витків, a - середній радіус, (м), c - ширина котушки, (м), відповідно:

$$a = \frac{d_{in}+d_{out}}{4} \quad (2)$$

$$c = \frac{d_{out}-d_{in}}{2} \quad (3)$$

Індуктивність одношарової квадратної, шестикутної або восьмикутної спіральної котушки розраховується з використанням модифікованої формули Уіллера [2]:

$$L = K_1 \cdot \mu_0 \frac{n^2 \cdot d_{avg}}{1+K_2 \cdot \rho} \quad (4)$$

де L — індуктивність, (Гн); μ_0 - магнітна постійна, (Гн/м), $\mu_0 = 4\pi \cdot 10^{-7}$; K_1 і K_2 - емпіричні безрозмірні коефіцієнти, які залежать від форми катушок (табл. 1); n - число витків; d_{avg}- середній діаметр котушки, (м); ρ - коефіцієнт заповнення котушки:

$$\rho = \frac{d_{out}-d_{in}}{d_{out}+d_{in}} \quad (5)$$

Коефіцієнт заповнення ρ показує як заповнений простір спіральної котушки. При малих значеннях ρ котушка виглядає «порожньою» ($d_{out} \approx d_{in}$); при значеннях ρ, що наближаються до одиниці, котушка повністю заповнена витками ($d_{out} >> d_i n$). Наприклад, квадратна котушка, показана на рисунку вище, має коефіцієнт заповнення ρ = 0,42.

Таблиця 1 – Коефіцієнти K_1 і K_2

Конструкція котушки	K_1	K_2
Квадратна	2.34	2.75
Шестикутна	2.33	3.82
Восьмикутна	2.25	3.55

У наведених вище формулах не враховується вплив явищ, пов'язаних з високими частотами, а також вплив матеріалу на індуктивність, наприклад, поверхневий ефект або властивості матеріалу підкладки, на якій виготовлена котушка [5]. Тому формули підходять тільки для розрахунку на відносно низьких частотах, тобто вони придатні для NFC (діапазон 13,56 МГц) і для низькочастотних діапазонів RFID 120-150 кГц і 13,56 МГц.

Новими рішеннями для виконання ряду спеціалізованих задач є використання нових конструкцій, таких як антени типу фракталів або антен з нестандартною формою, може покращити характеристику випромінювання та сприймання сигналу. Наприклад, у логістиці потрібні антени з великою дальністю зчитування, тоді як у платіжних системах NFC важлива надійність і безпека передачі даних на коротких відстанях. Фрактальні антени зазвичай мають самоподібну геометрію, тобто їх структура повторюється на різних масштабах. Розрахунок індуктивності фрактальної антени може бути досить складним, оскільки фрактальні антени мають складну геометрію, що може ускладнювати точний аналіз. Для деяких простих фрактальних антен можуть бути доступні апроксимації індуктивності через спрощені моделі. Наприклад, для базових фрактальних геометрій можна використовувати формули для індуктивності звичайних антен або лінійних елементів з врахуванням додаткових параметрів, що враховують фрактальну структуру.

Наступним напрямком є використання адаптивних антен, які автоматично налаштовують свій імпеданс, використовуючи різні технології для оптимізації своїх характеристик відповідно до змінних умов сигналу або навколишнього середовища [6]. Для реалізації функціонування адаптивних пристроїв необхідними умовами є забезпечення моніторингу та вимірювання. Адаптивні антени оснащені сенсорами і моніторами, які постійно вимірюють параметри сигналу, такі як рівень потужності, якість сигналу, рівень шуму, а також характеристики навколишнього середовища, такі як відбиття сигналів або зміни в середовищі. Зібрані дані використовуються для аналізу поточних умов і визначення необхідних змін для покращення ефективності антени.

© Мороз С.А., Якимчук Н.М., Селепина Й.Р., Чалий В.Д.

Для автоматичного налаштування імпедансу використовуються змінні компоненти, такі як варикапи, варистори або інші елементи, які можуть автоматично налаштовувати імпеданс антени. В загальному для складних застосувань можуть використовуватися цифрові або аналогові контролери, які змінюють параметри антени в реальному часі на основі отриманих даних.

В якості адаптивних алгоритмів традиційно використовуються зворотного зв'язку, проте у зв'язку з широким поширенням і все глибшою розробкою бази когнітивних радіомереж з'являються можливості налаштування оптимізаційних алгоритмів на основі нейромереж та машинного навчання. Ці алгоритми постійно адаптуються до змінюваних умов, щоб підтримувати оптимальну роботу антени.

Для RFID та NFC повинні бути не лише функціональними, але й економічно вигідними у виробництві. Правильний розрахунок параметрів дозволяє створити антени, що задовольняють вимоги до розміру, матеріалів і вартості виробництва, без зниження продуктивності. Це особливо важливо для масового виробництва RFID міток, де будь-яке підвищення вартості може суттєво вплинути на загальну вартість системи.

Отже безконтактні технології в найближчому майбутньому будуть інтенсивно розвиватися. Хоча радіочастотна ідентифікація RFID застаріла технологія відносно технології ближнього поля NFC, однак на нашу думку дані технології ще мають перспективи розвитку в різних сферах людської діяльності. Досить цікавою є технологія NFC, оскільки вона дозволяє накопичити на одній смарт-картці дозволи керування багатьма пристроями та здійснювати різноманітні платіжні операції а можливості динамічної адаптації параметрів антен дають змогу створювати пристрої для широкого кола завдань.

Висновки. В роботі проведено детальний огляд сучасного стану розвитку технологій NFC і RFID, зокрема їхніх стандартів, протоколів і технічних характеристик. Проаналізовано особливості використання технологій у різних сферах та їхню сумісність з іншими стандартами, що дозволило визначити ключові напрямки для вдосконалення. Проведено дослідження параметрів антен для NFC та RFID систем. Здійснено теоретичний розрахунок індуктивностей антен різної геометричної форми (круглі, квадратні, шестикутні та восьмикутні котушки), що дозволяє оптимізувати передачу енергії та забезпечити стабільну роботу пристроїв. Було запропоновано використання нових підходів до підвищення ефективності пристроїв ближньої дії шляхом застосування антен спеціальних конструкцій та адаптивних антен. Досліджено вплив різних факторів на ефективність передачі енергії в змінних умовах.ереж.

Список бібліографічного опису

1. Кошкун В., Ок К., Озденізджі Б. Розробка професійних NFC-додатків для Android. John Wiley & Sons Ltd., 2013. 308 с.
2. Іго Т., Коулмен Д., Джепсон Б. Початок роботи з NFC: Ближній безконтактний зв'язок з Arduino, Android і PhoneGap. O'Reilly Media, 2014. 246 с.
3. Рахул А., Крішнан Г., Рао С. Технологія NFC (Ближній безконтактний зв'язок): Огляд. Міжнародний журнал з кібернетики та інформатики. 4. 2015. С. 133-144. DOI: 10.5121/ijci.2015.4213
4. Сапарходжаєв Н., Дауїтбаєва А., Нуртаєв А., Байменшина Г. Система керування доступом з NFC. Міжнародна конференція з веб-ресурсів і відкритого доступу до навчання, ICWOAL 2014. DOI: 10.1109/ICWOAL.2014.7009188.
5. Паре Д. Проекти антен для NFC-пристроїв. 2016. 344 с. DOI: 10.1002/9781119145301
6. Гебхарт М., Байєр Т., Факкіні М. Автоматичне налаштування імпедансу антени для ближнього безконтактного зв'язку (NFC). 2013. С. 235-242.

References

1. Coskun V., Ok K., Ozdenizci B. Professional NFC Application Development for Android. John Wiley & Sons Ltd., 2013. 308p.
2. Igoe T., Coleman D., Jepson B. Beginning NFC: Near Field Communication with Arduino, Android, and PhoneGap. O'Reilly Media, 2014. 246 p.
3. Rahul A., Krishnan G., Rao S. Near Field Communication (NFC) Technology: A Survey. International Journal on Cybernetics & Informatics. 4. 2015. P. 133-144. DOI: 10.5121/ijci.2015.4213
4. Saparkhojayev N., Dauitbayeva A., Nurtayev A., Baimenshina G. NFC-enabled access control and management system. 2014 International Conference on Web and Open Access to Learning, ICWOAL 2014. DOI: 10.1109/ICWOAL.2014.7009188.
5. Paret D. Antenna Designs for NFC Devices. 2016. 344 p. DOI: 10.1002/9781119145301
6. Gebhart M., Baier T., Facchini M. Automated antenna impedance adjustment for Near Field Communication (NFC). 2013. P. 235-242.

© Мороз С.А., Якимчук Н.М., Селепина Й.Р., Чалий В.Д.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-44
УДК: 621.396.9:629.73
Радзівілов Григорій Данилович[1], к. т. н., професор, заступник начальника
https://orcid.org/0000-0002-6047-1897
Дєгтяр Олег Андрійович[1], д. н. з держ. упр., професор, провідний науковий співробітник
https://orcid.org/0000-0001-6413-3580
Безносенко Сергій Юрійович[2], начальник науково-дослідного відділу (технічного забезпечення засобів зв'язку та автоматизації) науково-дослідного управління (розвитку військ зв'язку)
https://orcid.org/0000-0002-1045-1492
Ткаченко Андрій Леонідович[2] , начальник науково-дослідного управління (розвитку військ зв'язку)
https://orcid.org/0000-0002-9789-8536
[1] Військовий інститут телекомунікацій та інформатизації імені Героїв Крут, м. Київ, Україна
[2] Науковий центр зв'язку та інформатизації Військового інституту телекомунікацій та інформатизації імені Героїв Крут, м. Київ, Україна

СУЧАСНІ МОЖЛИВОСТІ УПРАВЛІННЯ МОБІЛЬНИМИ ТЕЛЕКОМУНІКАЦІЙНИМИ МЕРЕЖАМИ З ВИКОРИСТАННЯМ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ

Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л. Сучасні можливості управління мобільними телекомунікаційними мережами з використанням безпілотних літальних апаратів. В роботі розглядаються сценарії ретрансляційного зв'язку з використанням безпілотних літальних апаратів в якості одного з методів телекомунікації, де автором було з'ясовано, що сучасні методи управління мобільними телекомунікаційними мережами потребують залучення безпілотних літальних апаратів з метою покращення з'єднання та передачі інформації на великі відстані. Метою даного дослідження є вивчення сучасних можливостей мобільних телекомунікаційних мереж на основі ретрансляційного з'єднання з безпілотними літальними апаратами для проведення аналізу використання в чотирьох сценаріях БПЛА-ретрансляції. В роботі досліджуються можливості та перспективи використання безпілотних літальних апаратів для ретрансляції сигналів в телекомунікаційних мережах, де було обгрунтовано переваги використання безпілотних літальних апаратів для створення тимчасових ретрансляційних вузлів у випадках аварійних ситуацій, природних катастроф тощо. Описано деякі питання стабільності сигналу та покриття, а також проаналізовано можливості інтеграції БПЛА-ретрансляторів в існуючу інфраструктуру мобільних мереж з врахуванням вимог до пропускної здатності та мінімізації затримок передачі даних. Згідно з потребами зв'язку автором було охарактеризовано та класифіковано чотири різні сценарії застосування ретрансляції зв'язку в поєднанні з безпілотними літальними апаратами. Окрім того, в статті також зазначаються питання та проблеми ретрансляційного зв'язку з безпілотними літальними апаратами в порівнянні з традиційним ретрансляційним зв'язком. На основі проведеного аналізу застосування сценаріїв ретрансляції БПЛА в статті зазначено, що одним з суттєвих недоліків є оптимізація траєкторії безпілотних літальних апаратів, що може впливати на якість отримання інформації завдяки телекомунікації.
Ключові слова: мобільна телекомунікаційна мережа, ретрансляційне з'єднання, мережі ретрансляції, безпілотний літальний апарат, оптимізація траєкторії, сценарії ретрансляції.

Radzivilov H., Diegtiar O., Beznosenko S., Tkachenko A. Modern possibilities of managing mobile telecommunication networks using unmanned aerial vehicles. The paper considers the scenarios of retransmission communication using unmanned aerial vehicles as one of the methods of telecommunication, where the author found that modern methods of managing mobile telecommunication networks require the involvement of unmanned aerial vehicles in order to improve the connection and transmission of information over long distances. The purpose of this study is to investigate the current capabilities of mobile telecommunication networks based on retransmission connections with unmanned aerial vehicles to analyse the use of UAVs in four retransmission scenarios. The paper explores the possibilities and prospects of using unmanned aerial vehicles for signal retransmission in telecommunications networks, where the advantages of using unmanned aerial vehicles to create temporary retransmission nodes in case of emergencies, natural disasters, etc. are substantiated. Some issues of signal stability and coverage are described, and the possibilities of integrating UAV repeaters into the existing infrastructure of mobile networks are analysed, taking into account the requirements for bandwidth and minimising data transmission delays. According to the communication needs, the author characterised and classified four different scenarios of using retransmission communication in combination with unmanned aerial vehicles. In addition, the article also identifies the issues and challenges of retransmission communications with unmanned aerial vehicles in comparison with traditional retransmission communications. Based on the analysis of the application of UAV retransmission scenarios, the article notes that one of the significant drawbacks is the optimisation of the UAV trajectory, which may affect the quality of information received through telecommunications.
Keywords: mobile telecommunication network, retransmission connection, retransmission networks, unmanned aerial vehicle, trajectory optimisation, retransmission scenarios.

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

Актуальність. З розвитком сучасних технологій телекомунікаційна галузь стикається з новими викликами та можливостями, де водночас з цим мобільні телекомунікації відіграють важливу роль у житті людей. Вони використовуються для спілкування, доступу до інформації, розваг та ведення бізнесу. Роль мобільних телекомунікаційних мереж (МТМ) пов'язана з впровадженням нових технологій, таких як використання безпілотних літальних апаратів (БПЛА), що відкриває нові горизонти для управління та оптимізації МТМ. Однак попит на мобільні послуги постійно зростає, що призводить до необхідності розширення та покращення МТМ [1].

Постановка наукової проблеми. В сучасному світі МТМ відіграють важливу роль в забезпеченні зв'язку та доступу до інформаційних ресурсів як для бізнесу, так і для населення в цілому. Одночасно з цим існує низка проблем, які можуть впливати на ускладнення ефективності управління та розгортання таких мереж, особливо у віддалених та важкодоступних регіонах. Традиційні методи розгортання телекомунікаційної інфраструктури є часто дорогими, трудомісткими, які потребують значних ресурсів та часу. Крім того, забезпечення стабільного покриття та високої якості обслуговування в умовах змінного навантаження та природних катаклізмів залишається досі невирішеним питанням, яке може містити значні наслідки [2, 3].

Одним з перспективних та ефективних рішень цих проблем є використання БПЛА для управління МТМ, так як літальні апарати забезпечують швидке та гнучке розгортання мереж, особливо в екстрених та непередбачених ситуаціях, наприклад, стихійні лиха або техногенні аварії. Вони також дозволяють оптимізувати використання ресурсів та знизити витрати на встановлення, обслуговування та ремонт традиційних базових станцій [1, 3].

Однак, незважаючи на значний потенціал, використання літальних апаратів для управління МТМ може містити перелік технічних та організаційних проблем, де необхідно розробити ефективні методи управління та координації літальних апаратів, забезпечити надійну передачу даних та інтеграцію з існуючою телекомунікаційною інфраструктурою. Крім того, важливо врахувати питання безпеки та врегулювання використання літальних апаратів, що є також не менш важливим для успішної реалізації [4].

Таким чином, постає необхідність проведення досліджень, спрямованих на розробку та впровадження ефективних методів управління МТМ з використанням БПЛА. Це в свою чергу дозволить підвищити ефективність та надійність мобільних мереж, забезпечить їх швидке розгортання та гнучкість, що є актуальним питанням в контексті сучасних викликів та потреб телекомунікаційної галузі.

Аналіз досліджень. За останні роки використання БПЛА значно зросло, що обумовлюється швидким прогресом у напрямку впровадження, розробки, виробництва недорогих та надійних БПЛА зі зростаючими потребами використання таких платформ в цивільних цілях [1, 3]. Завдяки ключовим можливостям, таким як висока мобільність, швидке розгортання та гнучкість БПЛА мають неабиякий потенціал для використання в багатьох додатках бездротових систем.

Як зазначено в роботі [5] з одного боку, БПЛА можуть функціонувати як літаючі мобільні термінали в бездротових та стільникових мережах з метою забезпечення виконання різноманітних задач, наприклад, доставка товарів, пошук та порятунок, дистанційне зондування тощо. З іншого боку, БПЛА можуть використовуватися в якості повітряних базових станцій для збільшення або розширення покриття бездротового зв'язку, надійності та пропускної здатності бездротових систем без додаткових інвестицій в інфраструктуру бездротових систем. На основі проведеного огляду авторами досліджувалося питання щодо використання мереж БПЛА згідно з цивільними потребами за допомогою додатків на основі МТМ, де також розглядалися комунікаційні вимоги цих додатків.

Однак, щоб задовольнити швидке зростання попиту на різноманітний трафік у мережах мобільного зв'язку пропонуються нові рішення щодо використання БПЛА для підтримки наземних мереж шляхом перенесення базових станцій на потреби в разі необхідності [4, 5]. Але питання ефективного та результативного розгортання БПЛА є доволі складним завданням. Авторами в роботі [6] розглядалися питання з балансуванням навантаження, де автори намагалися зробити навантаження на трафік БПЛА майже рівномірним з метою забезпечення МТМ до стабільності та стійкості при виникненні непередбачених ситуацій. Для досягнення цієї мети авторами проводилася попередня обробка в поєднанні з методом кластеризації з метою розподіленням користувачів на декілька категорій для ініціалізації положення БПЛА в зонах максимальної локальної щільності, що

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

дозволило мінімізувати максимальний попит на трафік в субрегіонах з обмеженою пропускною здатністю.

Останніми досягненнями з точки зору інформаційних технологій та телекомунікаційних мереж є використання БПЛА з підключенням до інтернету за допомогою телекомунікацій 4G та 5G [7, 8]. Бездротові мережі зв'язку наступного покоління 5G та 6G можуть відкрити безліч можливостей для існуючої телекомунікаційної галузі, де ключовими перевагами будуть вища пропускна здатність й низька затримка. Тому в даному контексті використання БПЛА може стати можливим варіантом для розгортання послуг 5G та 6G. В роботі [8] автори акцентували увагу на останніх розробках та проблемах, які можуть виникати в поєднанні технологій БПЛА з 5G, де основні зусилля були зосереджені на архітектурі ретрансляції БПЛА, визначенню відповідних проблем та обмежень в розгортанні БПЛА в каналах доступу та каналах зворотного зв'язку.

Можливість розгортання БПЛА в якості бездротових ретрансляторів або повітряних базових станцій сприяє покращенню підключенню та покриттю в стільникових мережах. Також БПЛА можна використовувати для значного підвищення продуктивності мобільних однорангових мереж та бездротових мереж Очікується, що в майбутньому БПЛА стануть невід'ємною частиною бездротових мереж 5G та ключовими факторами масового Інтернету речей. Однак для розробки архітектури та розгортання мереж на основі БПЛА все ще існує багато складних проблем. З цією метою в роботі [9] авторами розглядалася можливість застосування теорії ігор в якості ефективного інструменту для моделювання та аналізу проблем у мережах, що дозволило виявити різноманітні проблеми мереж, які працюють за допомогою БПЛА. Автори мали на меті надати інформацію щодо розуміння телекомунікаційних мереж з підтримкою БПЛА з точки зору їх архітектури, переваг, недоліків, проблем та різноманітних теоретичних рішень, застосовних до цих комунікаційних мереж.

Переваги використання БПЛА для управління телекомунікаційними мережами пов'язані зі швидким та ефективним розгортанням з метою підтримки стільникових мереж, а також підвищення їхньої якості обслуговування шляхом встановлення ліній зв'язку в межах прямої видимості. Для того, щоб з'ясувати основні фундаментальні аспекти використання БПЛА в процесах управління МТМ в роботі [10] авторами розроблена наступна структура: 1) ефективне розгортання БПЛА; 2) оцінка та оптимізація продуктивності; 3) розробка нових літаючих тривимірних бездротових систем. Для розгортання, використовуючи інструменти теорії оптимізації авторами розроблені цілісні структури для оптимального 3D-розміщення базових станцій БПЛА у висхідних та низхідних сценаріях, що дозволяє значно покращити покриття низхідної лінії зв'язку для наземних користувачів і забезпечує наднадійний та енергоефективний висхідний зв'язок у додатках Інтернету речей

До невирішених питань загальної проблеми даного дослідження можна віднести декілька важливих факторів, а саме: трудомісткість розгортання телекомунікаційної інфраструктури, проблеми забезпечення стабільного покриття, необхідність розробки ефективних методів управління та координації літальних апаратів, забезпечення надійної передачі даних та інтеграції з існуючою телекомунікаційною інфраструктурою. Тому для вирішення раніше невирішених питань в даному дослідженні розглядалися чотири різні сценарії ретрансляції БПЛА, що впливає на вищезазначені фактори.

Метою роботи. Метою даного дослідження є вивчення сучасних можливостей мобільних телекомунікаційних мереж на основі ретрансляційного з'єднання з безпілотними літальними апаратами для проведення аналізу використання в чотирьох сценаріях ретрансляції БПЛА. Для досягнення поставленої мети необхідно вирішити наступні завдання: на основі літературного огляду проаналізувати можливості телекомунікаційних мереж в поєднанні з використанням безпілотних літальних апаратів; дослідити можливість використання ретрансляторів в якості телекомунікаційної мережі; побудувати основні сценарії використання ретрансляторів БПЛА за різних умов; проаналізувати основні переваги та недоліки для кожного з охарактеризованих сценаріїв ретрансляції БПЛА.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Зі стрімким розвитком бездротового зв'язку, а також швидким зростанням масштабу та різноманітності послуг мобільного зв'язку кількість даних, які підлягають обробці, передачі та

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

зберіганню постійно зростає. Одночасно з цим мобільні передачі даних з високою ємністю, високою швидкістю та високим покриттям стали центром уваги в галузі, яка є актуальною проблемою з точки зору телекомунікацій. Існуюча система бездротового зв'язку все частіше стикається з новими викликами та проблемами в телекомунікаційній мережі, що вказує на необхідність пошуку нових рішень, де одним з таких рішень є застосування зв'язку за допомогою БПЛА.

На відміну від традиційного наземного стільникового зв'язку та мереж супутникового зв'язку основні переваги мереж зв'язку за допомогою БПЛА є наступними: 1) завдяки низькій вартості та можливості швидкого розгортання БПЛА вони підходять для вирішення надзвичайних ситуацій, а також для завдань з обмеженою тривалістю, наприклад, екстрене відновлення зв'язку, коли інфраструктура зв'язку може бути сильно пошкодженою, наприклад, через непередбачені ситуації стихійного лиха, або розвантаження даних, коли базові станції стільникового зв'язку перевантажені тощо. Тому за допомогою БПЛА, оснащених засобами зв'язку можна швидко встановити зв'язок, а також забезпечити покриття зі зв'язком для областей, де не вистачає комунікаційної інфраструктури, наприклад, для інспекцій гірських електростанцій, або для інспекцій на кордоні та надання допомоги при стихійних лихах тощо.

У порівнянні з наземним зв'язком, зв'язок БПЛА має деякі переваги з наявністю кращих сигналів зв'язку та меншої чутливості до наземних будівель. Через висоту базової станції на традиційний бездротовий зв'язок можуть легко впливати перешкоди, як високі гори та високі будівлі, де відповідно якістю зв'язку буде значно погіршена. Тому на основі того, що БПЛА мають перевагу в повітрі вони можуть з легкістю уникати перешкоди та встановлювати надійні лінії зв'язку.

Таким чином, з такими перевагами, як швидке реагування, висока маневреність, низька вартість, гнучке розгортання, хороший канал і висока керованість, бездротовий зв'язок за допомогою БПЛА має широке застосування в безпеці національної оборони, реагуванні на надзвичайні ситуації та ліквідації наслідків катастроф. Це має практичне значення для повного вивчення характеристик БПЛА в бездротовому зв'язку за допомогою БПЛА та покращення якості зв'язку, рівня доступності та енергоефективності шляхом оптимізації траєкторії БПЛА, розподілу ресурсів і схем передачі.

У літературі зв'язок за допомогою БПЛА має три типових сценарії, а саме: ретрансляція за допомогою БПЛА, повсюдне покриття за допомогою БПЛА, а також поширення інформації та збір даних за допомогою БПЛА. Зокрема, очікується, що використання БПЛА як повітряних ретрансляторів стане перспективним рішенням для покращення продуктивності існуючого бездротового зв'язку. У порівнянні з традиційним зв'язком з одним стрибком, технологія ретрансляції може значно покращити покриття мережі, ефективно придушити втрату бездротових каналів для доставки інформації та підвищити надійність бездротового зв'язку.

Телекомунікаційна мережа та телекомунікаційний зв'язок. Телекомунікаційна мережа є системою, яка дозволяє передавати дані та інформацію між різними точками через різні канали зв'язку. Такі мережі можуть використовувати різні типи передавальних середовищ, такі як кабелі (оптоволоконні, мідні), радіохвилі (бездротовий зв'язок) та супутники [11]. Телекомунікаційні мережі здатні забезпечувати передачу інформації між різними користувачами через спеціальні канали зв'язку, які складаються з різних компонентів [11-12]:

- передавачі та приймачі, де пристрої здійснюють кодування та передачу сигналів (наприклад, модеми або маршрутизатори) з наступним отриманням та декодуванням;
- канали зв'язку, де можуть бути фізичні або бездротові середовища, через які передаються сигнали;
- комутаційне обладнання, де пристрої, які управляють потоками даних направляють їх до відповідних адресатів (в якості комутаційного обладнання можуть слугувати комутатори, маршрутизатори та концентратори);
- протоколи зв'язку, що складається з набору правил та стандартів, які забезпечують процес передачі даних для коректності та безпеки, де прикладами можуть бути TCP/IP для інтернету та LTE для мобільного зв'язку;
- центральні вузли, де великі комутаційні центри або дата-центри координують та обробляють значний обсяг трафіку, забезпечуючи його розподіл та зберігання.

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

Телекомунікаційні мережі класифікують за деякими критеріями як за типом переданої інформації, так і за радіусом дії. За типом переданої інформації можуть бути голосові та мультимедійні дані. За радіусом дії телекомунікаційні мережі поділяються на: локальні мережі (LAN), які забезпечують зв'язок в межах обмеженої території, наприклад, всередині будівлі; регіональні або міські мережі (MAN), які з'єднують декілька локальних мереж в межах міста або агломерації; глобальні мережі (WAN), які покривають великі географічні відстані, об'єднуючи міські та локальні мережі, наприклад, інтернет. До основних компонентів МТМ відносяться базові станції, контролери базових станцій, мобільні комутатори та системи підтримки операцій [12].

Ретрансляційне з'єднання. Ретрансляційний зв'язок (ретрансляція) є методом передачі сигналів у телекомунікаційних мережах, при якому сигнал передається від одного передавача до іншого через проміжні вузли або ретранслятори, що забезпечує передачу сигналів на великі відстані або в умовах, де прямий зв'язок між кінцевими точками неможливий або ускладнений [13]. На рис. 1 показано ретранслятор телекомунікаційної мережі.

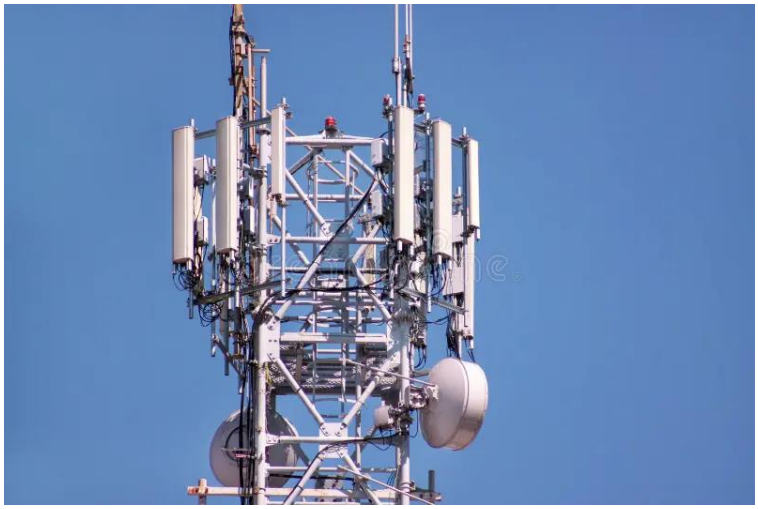


Рис. 1. Ретранслятори телекомунікаційної мережі, базова приймально-передавальна станція, яка містить баштовий передавач та ретранслятор антени бездротового зв'язку
Джерело: сформовано авторами

Розглянемо основні аспекти ретрансляційного зв'язку, який може використовуватися в поєднанні з БПЛА [14, 15].

1) Ретранслятори – пристрої, які приймають сигнал, підсилюють його та передають далі, де пристрої можуть бути встановлені на землі (на вежі або будівлі) або в космосі (супутники). У мобільних мережах ретранслятори використовуються для покращення покриття в міських зонах, у віддалених районах або в місцях з перешкодами (наприклад, гірські райони).

2) Принцип дії ретрансляторів обумовлений прийманням сигналу від джерела або іншого ретранслятора з підсиленням його та передачею наступному ретранслятору або кінцевому приймачу. Цей процес може повторюватися декілька разів, що дозволяє сигналу подолати великі відстані або складні умови.

3) Типи ретрансляційного зв'язку можуть поділятися на:

- наземна ретрансляція, яка використовує наземні станції (вежі, будівлі) для передачі сигналів;
- супутникова ретрансляція, яка використовує супутники на орбіті для прийому і передачі сигналів, що дозволяє забезпечити глобальне покриття та зв'язок у віддалених районах;
- радіорелейний зв'язок, який використовує мікрохвильові радіохвилі для передачі сигналів між наземними станціями на великі відстані.

4) Переваги ретрансляційного зв'язку, що охоплює наступне:

- збільшення (розширення) покриття, що дозволяє забезпечити зв'язок у віддалених або важкодоступних місцях;

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

- підсилення сигналу, що підвищує якість сигналу, знижуючи втрати при передачі на великі відстані;
- гнучкість, яка може використовуватися при різних умовах і для різних видів зв'язку (голосовий зв'язок, передача даних, телевізійне мовлення тощо).

5) Недоліки ретрансляційного зв'язку обумовлені наступним:
- затримки, які можливі в передачі сигналу через проходження через кілька ретрансляторів;
- складність, де встановлення та обслуговування ретрансляторів може бути складним і дорогим;
- вразливість, де ретранслятори можуть бути вразливими до технічних збоїв або навмисних пошкоджень тощо.

Тому ретрансляційний зв'язок є важливим елементом телекомунікаційних систем та МТМ, що забезпечує надійний та ефективний зв'язок на великі відстані, а також за умов складного рельєфу.

Ретрансляційне з'єднання з використанням БПЛА. Після того, як вище було зазначено про телекомунікаційні мережі та ретрансляційне з'єднання далі розглянемо чотири сценарії застосування ретрансляційного з'єднання в поєднанні з БПЛА [14]. Дальність зв'язку БПЛА обумовлена їх функціональними можливостями та характеристиками, де можуть використовуватися як 2,4 ГГц для управління БПЛА, так і 5 ГГц для передачі відео з дальністю зв'язку, який може досягати близько 5-6 км. Проте, ці параметри можуть бути значно кращими, де частота може сягати близько 800-900 МГц для управління БПЛА та 1,3-1,4 ГГц для передачі відео в діапазоні 25-30 км. Розглянемо ситуацію для чотирьох сценаріїв, де відстань між джерелом і пунктом призначення складає 5 км, а один БПЛА може використовуватися в якості проміжного ретранслятора, як показано на рис. 2 (а). Коли ця відстань може бути більшою ніж 5 км, то в такому випадку доцільно використовувати декілька БПЛА в якості проміжних ретрансляторів для обміну інформацією, як показано на рис. 2 (б).

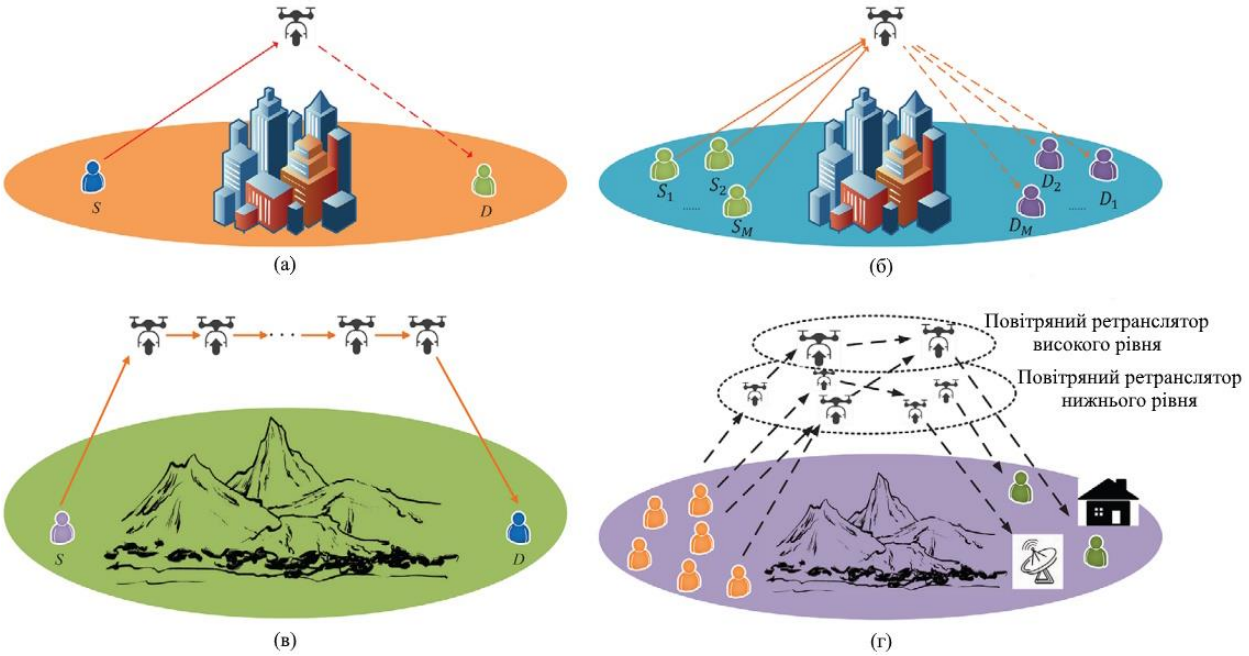


Рис. 2. Сценарії ретрансляції зв'язку з використанням БПЛА, де (а) окремі мережі ретрансляції БПЛА; (б) багатокористувацькі мережі ретрансляції БПЛА; (в) багатоходові мережі ретрансляції БПЛА; (г) Інтернет БПЛА
Джерело: сформовано авторами на основі [15]

Сценарій 1. Мережі ретрансляції одиночних БПЛА. На рис. 2 (а) показано бездротову систему з парою джерело-приймач S і D, де обидва розташовані на земній поверхні й між ними прямий зв'язок відсутній, що обумовлюється деякими умовами, наприклад, надмірна відстань або серйозні перешкоди зв'язку між парою джерело-приймач. Тому в даному випадку БПЛА може

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

використовуватися в якості ретрансляційного вузла для обміну інформацією між S та D. Даний сценарій є найпростішим, так як лише пара вузлів джерела та вузла призначення повинні передавати та обмінюватися інформацією, а наявність однієї одиниці БПЛА може бути достатньо для ефективної передачі інформації. Однак, з точки зору практичності можуть виникнути певні складнощі, які обумовлені наступним:

1) незалежно від того, чи є пара джерело-приймач на земній поверхні стаціонарною чи мобільною, на місці події можуть виникати різні труднощі;

2) пара джерело-отримувач може мати декілька джерел, яким необхідно обмінюватися інформацією, що призводить до двосторонньої схеми ретрансляції;

3) режим ретрансляції, такий як підсилення та передача, або декодування та передача також можуть провокувати різні проблеми для оптимізації ретрансляції.

4) для ретрансляції БПЛА необхідно вибрати найбільш зручне та фіксоване місце, або оптимальну траєкторію польоту БПЛА для здійснення ретрансляції.

Згідно з першим сценарієм проблема оптимізації траєкторії БПЛА є найбільш критичною, де використовуються статичні та мобільні реле БПЛА. Дослідження статичних реле БПЛА зосереджені на підвищенні продуктивності бездротових мереж за допомогою оптимізації розгортання БПЛА. Дослідження мобільних ретрансляторів БПЛА забезпечує створення різноманітних схем планування БПЛА, де можуть розроблятися схеми мобільної ретрансляції для режимів підсилення-передача, або декодування-передача, що призначено для альтернативного доступу до вузла джерела та вузла призначення, де БПЛА може летіти по прямій лінії й кожен цикл польоту може бути розділений на дві фази, як показано на рис. 3.

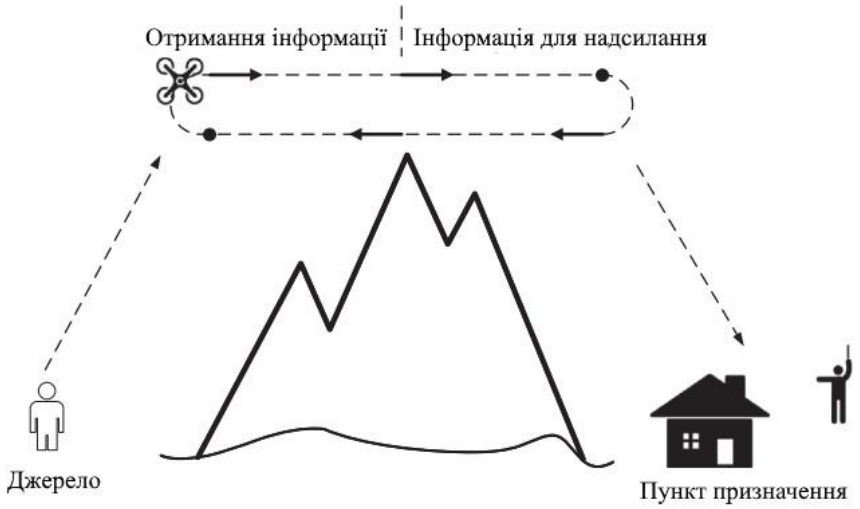


Рис. 3. Проєктування траєкторії ретрансляції БПЛА режимів підсилення-передача, декодування-передача.
Джерело: сформовано авторами на основі [15]

Далі розглянемо декілька етапів сценарію. Перший етап відповідає за отримання, обробку та збереження інформації, де БПЛА повинен максимально швидко переміщуватися до вихідного вузла, а потім повертатися. У випадку, якщо швидкість та тривалість польоту БПЛА будуть мати достатньо великі значення, БПЛА протягом певного часу може затриматися над вихідним вузлом, перш ніж повернутися, щоб скористатися найкращим каналом зв'язку, де може отримувати інформацію та зберігати її в буфері даних без перешкод.

Другий етап відповідає за відправку інформації, де БПЛА повинен якнайшвидше переміщуватися до вузла призначення, а потім повернутися. На даному етапі БПЛА також може затриматися на певний час, якщо швидкість та тривалість польоту будуть мати достатньо великі значення, де здійснюється відправка інформації з буферу даних до вузла призначення без перешкод. Однак у схемі мобільної ретрансляції може бути затримка передачі інформації, яка може зростати зі збільшенням тривалості польоту БПЛА.

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

Сценарій 2. Багатокористувацькі мережі ретрансляції БПЛА. Згідно з рис. 2 (б), якщо один БПЛА використовується для передачі інформації для кількох пар користувачів, розташованих на земній поверхні, позначених як S_1, S_2 … S_m та D_1, D_2 … D_m вважається, що всі пари користувачів неспроможні до прямого зв'язку й можуть бути встановлені зв'язком за допомогою реле БПЛА. На відміну від попереднього сценарію багатокористувацькі мережі ретрансляції БПЛА вимагають додаткового розгляду протоколів множинного доступу, так як БПЛА потрібно обслуговувати кілька пар джерело-приймач у мережі.

Для багатокористувацьких пар ретрансляційних зв'язків у сценарії 2 найбільш складними проблемами є ефективне проєктування протоколу планування множинного доступу та розподіл потужності, коли БПЛА обслуговує декілька користувачів одночасно. Так як в багатокористувацьких мережах ретрансляції БПЛА декілька пар користувачів передають інформацію через загальний БПЛА, вони в свою чергу можуть створювати перешкоди один одному, де необхідно призначити різний ортогональний спектр або часові діапазони для кожної пари користувачів, з метою уникнення перешкод та їх розрізняння. Одночасно з цим ефективне керування живленням для різних користувачів може повною мірою використовувати ресурси живлення та покращити продуктивність передачі багатокористувацької мережі.

Однак через різні стани каналів кожної пари користувачів під час польоту БПЛА не може передавати інформацію в реальному часі, тому можливість буферизації є доволі важливим питанням. Зокрема, протягом усього польоту БПЛА безперервно приймає та декодує інформацію від вихідних вузлів S_1, S_2 зі збереженням в буфері даних, який потім доставляється до відповідних вузлів призначення D_1, D_2 у відповідний час.

Традиційні релейні системи можуть передавати та приймати інформацію в різний час в ортогональних частотних діапазонах, тоді як релейні системи можуть передавати і отримувати інформацію одночасно на одній частоті. Тому порівняно з ретранслятором, реле може значно покращувати спектральну ефективність системи бездротового зв'язку та забезпечує більшу гнучкість для проектування мережі ретрансляції.

Сценарій 3. Багатоходові мережі ретрансляції БПЛА. Згідно з рис. 2 (в) декілька БПЛА можуть використовуватися в якості ретрансляційних вузлів для підтримки зв'язку між парою джерело-приймач S та D, які розташовані на земній поверхні, але прямий зв'язок між ними відсутній. Вся пара джерело-приймач і ретранслятори БПЛА можуть обмінюватися інформацією тільки з лівим та правим найближчими сусідніми вузлами.

Декілька БПЛА потрібні для забезпечення зв'язку, оскільки один БПЛА може не впоратися з ретрансляцією, оскільки вони знаходяться занадто далеко один від одного або середовище надто складне. Ретрансляція за допомогою декількох БПЛА принесе деякі нові проблеми.

1) Перш за все, необхідно знайти відповідну кількість ретрансляторів БПЛА, щоб два сусідні ретранслятори не були занадто далеко один від одного, щоб впливати на якість зв'язку між ними, і не були занадто близько один до одного, щоб створювати надмірні взаємні перешкоди. Це визначається різними факторами, включаючи специфіку місцевості, відстань між і, комунікаційні можливості БПЛА тощо.
2) По-друге, існує проблема спільної оптимізації позицій або траєкторій декількох БПЛА.
3) Багатоходова ретрансляція БПЛА вимагає відповідного дизайну протоколу багатоходової ретрансляції, що призводить до більш складного спільного планування роботи декількох БПЛА. Зокрема, розробка протоколу багатоходової ретрансляції пов'язана з напрямком передачі інформації та режимом ретрансляції.

З метою покращення продуктивності багаточкового зв'язку БПЛА, неминуче потрібно виконати спільну оптимізацію траєкторії декількох БПЛА, щоб повною мірою використати мобільність БПЛА і розробити відповідні протоколи багатоточкового зв'язку.

Для односторонньої передачі інформації схема багатоточкової ретрансляції БПЛА може бути розроблена так, як показано на рис. 2 (в). Траєкторія БПЛА обмежується лише початковим і кінцевим розташуванням і спільно оптимізується відповідно до мети зв'язку. Застосовується стратегія декодування-передача з рівною пропускною здатністю для прийому і передачі інформації. Інформація передається від вузла-джерела до вузла призначення послідовно за допомогою декількох

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

стрибків. Тому затримка пересилання, спричинена передачею інформації через БПЛА становить часові інтервали.

Сценарій 4. Інтернет БПЛА. З розширенням застосування БПЛА виконання завдань покриття зв'язку невеликою кількістю БПЛА може більше не відповідати складним та мінливим фактичним потребам у зв'язку. Для спільного ефективного виконання таких завдань за допомогою БПЛА реалізується мультиплатформний та багатоцільовий зв'язок, який може використовувати Інтернет в поєднанні з БПЛА в якості доповнення до ретрансляційного зв'язку БПЛА. На рис. 2 (г) показано масштабний сценарій ретрансляції, де група БПЛА формує ретрансляційну мережу для надання послуг зв'язку для двосторонніх багатокористувацьких мереж. У цьому сценарії група БПЛА об'єднана в мережу для широкомасштабного ретрансляційного зв'язку, утворюючи багатошарову ретрансляційну площину і забезпечуючи гнучкі і надійні з'єднання.

У цьому сценарії зв'язок між наземними користувачами і БПЛА - це не просто зв'язок «точка-точка», а зв'язок «один-до-багатьох» або навіть «багато-до-багатьох», який може змінюватися з часом. Таким чином, традиційний односторонній або двосторонній ретрансляційний зв'язок між БПЛА може більше не відповідати вимогам сервісу, що призведе до появи нової архітектури, яку називають Інтернетом БПЛА. Для різних умов зв'язку можна розробити різні схеми розгортання Інтернету БПЛА або більш складні мережеві рішення для мобільних БПЛА, які змінюються відповідно до вимог. Розташування ретрансляційних вузлів безпосередньо впливає на якість покриття і зв'язність мережі, а отже, визначає топологію і надійність всієї мережі. Крім того, важливими питаннями є спільна комунікація і оптимізація маршрутизації декількох БПЛА.

У порівнянні з ретрансляційним зв'язком «точка-точка», Інтернет БПЛА має більш жорсткі вимоги до надійності і навантаження на систему ретрансляційного зв'язку БПЛА. У багатокористувацькій зоні розподілу розумне і ефективне розгортання мережі з декількома БПЛА, де механізм оптимізації маршрутизації є важливими засобами для забезпечення ретрансляційного зв'язку для складного Інтернету БПЛА.

Перш за все, необхідно розробити схему безперервного розгортання БПЛА з низьким рівнем складності для ефективного покращення покриття мережі та пропускної здатності ретрансляційної мережі. Масштабний сценарій ретрансляційного зв'язку показаний на рисунку 2 (г). Для досягнення надійного і стабільного ретрансляційного зв'язку ключова технологія полягає в розумному і ефективному позиційному розгортанні БПЛА. Завдяки розгортанню вторинної високорівневої ретрансляційної площини БПЛА на низькорівневій ретрансляційній площині БПЛА, вона забезпечує високу пропускну здатність, багатоканальні паралельні шляхи для ретрансляційного зв'язку, тим самим покращуючи пропускну здатність мережі і надійність ретрансляційної системи.

Крім того, коли мережа ретрансляційного зв'язку БПЛА передає великий обсяг даних у реальному сценарії великомасштабного ретрансляційного зв'язку, може існувати кілька можливих шляхів передачі даних, і одночасна передача даних може проходити через деякі ключові вузли ретрансляції одночасно. Тому неминуче виникають проблеми вибору шляху передачі даних та уникнення колізій даних. Для типових сценаріїв, таких як багатоходова передача і розподіл даних в повітряній мережі ретрансляторів БПЛА, необхідна оптимізація схем маршрутизації для багатошляхових даних і механізмів уникнення завад в Інтернеті БПЛА. Ці технології можуть спільно оптимізувати маршрутизацію і планування ресурсів Інтернету БПЛА, зменшити міжшарову затримку передачі і конфлікт сигналів, тим самим ефективно зменшуючи затримку передачі даних в Інтернеті БПЛА.

Недоліки використання сценаріїв. Одним з суттєвих недоліків всіх чотирьох сценаріїв є оптимізація траєкторії БПЛА. Незважаючи на те, що БПЛА можуть мати унікальні характеристики, такі як мобільність та гнучкість, використовуючи БПЛА для бездротової передачі даних є складнішим в порівнянні з традиційними ретрансляційними комунікаціями. Оптимізація траєкторії є складною проблемою через велику кількість фізичних обмежень та високу мобільність БПЛА. Наприклад, при дослідженні траєкторії БПЛА на предмет максимальної швидкості до вузла призначення необхідно врахувати кілька факторів, таких як інформаційно-причинне обмеження, обмеження на потужність передачі, обмеження на початкове місцезнаходження, обмеження на швидкість та обмеження на кінцеве місцезнаходження.

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

Крім того, розв'язання задачі оптимізації безперервної траєкторії БПЛА також стикається з аналітичними проблемами, оскільки необхідно вирішити незліченну кількість змінних оптимізації (наприклад, положення декількох БПЛА). Також в бездротових мережах, що складаються з БПЛА, оптимізація траєкторії повинна враховувати кореляцію між мобільністю і продуктивністю мережі. При оптимізації траєкторії індекси оптимізації для різних типів БПЛА і різних польотних завдань відрізняються. Типові показники оптимізації включають найшвидший набір висоти, найбільшу тривалість польоту, найвищу енергоефективність і найшвидше прибуття. Тому аналіз відповідної траєкторії за різними індексами оптимізації є складним завданням.

Висновки. На основі проведеного літературного огляду було з'ясовано, що сучасні методи управління мобільними телекомунікаційними мережами спонукають до залучення безпілотних літальних апаратів з метою покращення з'єднання та передачі інформації на великі відстані. На основі наданої характеристики у вигляді чотирьох сценаріїв було з'ясовано, що ретрансляційне з'єднання є невід'ємною частиною бездротового зв'язку для БПЛА.

У сценарії 1 розглядається використання одиночних БПЛА як ретрансляційних вузлів для передачі інформації між джерелом (S) і приймачем (D), коли прямий зв'язок між ними відсутній. БПЛА діє як проміжна ланка, забезпечуючи зв'язок. Хоча цей сценарій є простим, можуть виникати труднощі, пов'язані з: варіаціями в мобільності та стаціонарності пари S-D; наявністю кількох джерел інформації, що потребують ретрансляції; різними режимами ретрансляції (підсилення-передача або декодування-передача); вибором оптимального місця або траєкторії польоту БПЛА для ретрансляції.

У сценарії 2 розглядається використання одного БПЛА для передачі інформації між кількома парами користувачів, де всі користувачі неспроможні до прямого зв'язку. Основні припущення щодо використання Багатокористувацьких мереж ретрансляції БПЛА містять наступне: множинний доступ (що складається з розробки ефективного протоколу планування множинного доступу для уникнення перешкод між користувачами); розподіл потужності (де важливо оптимізувати розподіл потужності для покращення продуктивності передачі в багатокористувацькій мережі); буферизація (де БПЛА повинні мати можливість буферизації даних, оскільки прямий реальний час передачі може бути неможливим через різні стани каналів); спектральна ефективність (де використання релейних систем дозволяє проводити одночасну передачу та прийом інформації на одній частоті, що підвищує спектральну ефективність та гнучкість мережі).

У сценарії 3 розглядається використання кількох БПЛА для ретрансляції інформації між джерелом (S) та приймачем (D) на земній поверхні, де було з'ясовано наступне: кількість ретрансляторів (де необхідно визначити оптимальну кількість БПЛА, щоб сусідні ретранслятори не були надто далеко один від одного, що вплине на якість зв'язку, і не були занадто близько, що створюватиме взаємні перешкоди); оптимізація позиціювання траєкторії (де важливо оптимізувати позиції або траєкторії кількох БПЛА для забезпечення ефективного зв'язку); протокол багатоходової ретрансляції (де необхідно розробити протокол багатоходової ретрансляції, що враховує напрямок передачі інформації та режим ретрансляції); одностороння передача (де для односторонньої передачі інформації необхідно використовувати схему декодування-передача з рівною пропускною здатністю для прийому і передачі інформації, що мінімізує затримки)

У 4 сценарії розглядається використання Інтернету БПЛА для забезпечення масштабного і гнучкого ретрансляційного зв'язку у багатокористувацьких мережах, де було з'ясовано наступне: масштабний зв'язок (де група БПЛА може утворювати мережу для двосторонніх багатокористувацьких зв'язків, забезпечуючи багатошарову ретрансляційну площину); гнучкі з'єднання (де зв'язок між наземними користувачами і БПЛА може змінюватися з часом, що вимагає нових архітектурних рішень); оптимізація розгортання (позиційне розгортання БПЛА безпосередньо впливає на якість покриття і зв'язність мережі, визначаючи топологію і надійність); комунікація та маршрутизація (де спільна оптимізація комунікації і маршрутизації кількох БПЛА є ключовою для надійного зв'язку); ефективне розгортання (де необхідно розробити схеми безперервного розгортання БПЛА з низьким рівнем складності для покращення покриття мережі і пропускної здатності); вимоги до надійності (де Інтернет БПЛА матиме значно високі вимоги до надійності і навантаження на систему).

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

Тому проаналізувавши всі 4 сценарії можна зробити висновок, що кожен з сценаріїв ретрансляції може містити ряд переваг, які можуть покращувати якість з'єднання для обміну інформації. Проте, одним з суттєвих недоліків використання все ж таки залишається оптимізація траєкторії БПЛА, що потребує подальшого дослідження та аналізу.

Список бібліографічного опису

1.Sharma, A., Vanjani, P., Paliwal, N., Basnayaka, C. M. W., Jayakody, D. N. K., Wang, H. C., & Muthuchidambaranathan, P. (2020). Communication and networking technologies for UAVs: A survey. *Journal of Network and Computer Applications*, *168*, 1-18. URL: https://doi.org/10.1016/j.jnca.2020.102739 (дата звернення: 21.07.2024)
2.Mozaffari, M., Saad, W., Bennis, M., Nam, Y. H., & Debbah, M. (2019). A tutorial on UAVs for wireless networks: Applications, challenges, and open problems. *IEEE communications surveys & tutorials*, *21*(3), 2334-2360. URL: https://doi.org/10.1109/COMST.2019.2902862 (дата звернення: 21.07.2024)
3.Nawaz, H., Ali, H. M., & Laghari, A. A. (2021). UAV communication networks issues: A review. *Archives of Computational Methods in Engineering*, *28*(3), 1349-1369. URL: https://doi.org/10.1007/s11831-020-09418-0 (дата звернення: 21.07.2024)
4.Hentati, A. I., & Fourati, L. C. (2020). Comprehensive survey of UAVs communication networks. *Computer Standards & Interfaces*, *72*, 1-33. URL: https://doi.org/10.1016/j.csi.2020.103451 (дата звернення: 21.07.2024)
5.Ghamari, M., Rangel, P., Mehrubeoglu, M., Tewolde, G. S., & Sherratt, R. S. (2022). Unmanned aerial vehicle communications for civil applications: A review. *IEEE Access*, *10*, 102492-102531. URL: https://doi.org/10.1109/ACCESS.2022.3208571 (дата звернення: 21.07.2024)
6.Sun, Y., Wang, T., & Wang, S. (2019). Location optimization and user association for unmanned aerial vehicles assisted mobile networks. *IEEE Transactions on Vehicular Technology*, *68*(10), 10056-10065. URL: https://doi.org/10.1109/TVT.2019.2933560 (дата звернення: 21.07.2024)
7.Guirado, R., Padró, J. C., Zoroa, A., Olivert, J., Bukva, A., & Cavestany, P. (2021). Stratotrans: Unmanned aerial system (uas) 4g communication framework applied on the monitoring of road traffic and linear infrastructure. *Drones*, *5*(1), 1-14. URL: https://doi.org/10.3390/drones5010010 (дата звернення: 21.07.2024)
8.Khan, S. K., Naseem, U., Siraj, H., Razzak, I., & Imran, M. (2021). The role of unmanned aerial vehicles and mmWave in 5G: Recent advances and challenges. *Transactions on Emerging Telecommunications Technologies*, *32*(7), pp. 1-18. URL: https://doi.org/10.1002/ett.4241 (дата звернення: 21.07.2024)
9.Mkiramweni, M. E., Yang, C., Li, J., & Zhang, W. (2019). A survey of game theory in unmanned aerial vehicles communications. *IEEE Communications Surveys & Tutorials*, *21*(4), 3386-3416. URL: https://doi.org/10.1109/COMST.2019.2919613 (дата звернення: 21.07.2024)
10. Mozaffari, M. (2018). Wireless communications and networking with unmanned aerial vehicles: fundamentals, deployment, and optimization, 177-190. URL: http://hdl.handle.net/10919/83921 (дата звернення: 21.07.2024)
11. Ezeigweneme, C. A., Umoh, A. A., Ilojianya, V. I., & Adegbite, A. O. (2024). Telecommunications energy efficiency: optimizing network infrastructure for sustainability. *Computer Science & IT Research Journal*, *5*(1), 26-40. URL: https://doi.org/10.51594/csitrj.v5i1.700 (дата звернення: 21.07.2024)
12. Rakushev, M., Zuiko, V., Pantiushenko, R., Nozdrachov, O., Khomenko, V., & Chornomaz, O. (2022, December). The Technique of the Information and Telecommunication System" Terminal" Use for UAVs Acquired Data. In *2022 IEEE 4th International Conference on Advanced Trends in Information Theory (ATIT)*, 208-212. URL: https://doi.org/10.1109/ATIT58178.2022.10024215 (дата звернення: 21.07.2024)
13. Bashir, M. N., & Yusof, K. M. (2020). A review of relay network on uavs for enhanced connectivity. *Jurnal Teknologi*, *82*(1). 173-183. URL: https://doi.org/10.11113/jt.v82.13183 (дата звернення: 21.07.2024)
14. Euler, S., Maattanen, H. L., Lin, X., Zou, Z., Bergström, M., & Sedin, J. (2019, April). Mobility support for cellular connected unmanned aerial vehicles: Performance and analysis. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-6). URL: https://doi.org/10.1109/WCNC.2019.8885820 (дата звернення: 21.07.2024)
15. Li, B., Zhao, S., Miao, R., & Zhang, R. (2021). A survey on unmanned aerial vehicle relaying networks. *IET communications*, *15*(10), 1262-1272. URL: https://doi.org/10.1049/cmu2.12107 (дата звернення: 21.07.2024)

References

1. Sharma A., Vanjani P., Paliwal N., Basnayaka C. M. W., Jayakody D. N. K., Wang H. C., & Muthuchidambaranathan P. (2020). Communication and networking technologies for UAVs: A survey. *Journal of Network and Computer Applications*, *168*, 1-18. URL: https://doi.org/10.1016/j.jnca.2020.102739 (access date: 21.07.2024)
2. Mozaffari M., Saad W., Bennis M., Nam Y. H., & Debbah M. (2019). A tutorial on UAVs for wireless networks: Applications, challenges, and open problems. *IEEE communications surveys & tutorials*, *21*(3), 2334-2360. URL: https://doi.org/10.1109/COMST.2019.2902862 (access date: 21.07.2024)
3. Nawaz H., Ali H. M., & Laghari A. A. (2021). UAV communication networks issues: A review. *Archives of Computational Methods in Engineering*, *28*(3), 1349-1369. URL: https://doi.org/10.1007/s11831-020-09418-0 (access date: 21.07.2024)
4. Hentati A. I., & Fourati L. C. (2020). Comprehensive survey of UAVs communication networks. *Computer Standards & Interfaces*, *72*, 1-33. URL: https://doi.org/10.1016/j.csi.2020.103451 (access date: 21.07.2024)

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

5. Ghamari M., Rangel, P., Mehrubeoglu M., Tewolde G. S., & Sherratt R. S. (2022). Unmanned aerial vehicle communications for civil applications: A review. *IEEE Access*, *10*, 102492-102531. URL: https://doi.org/10.1109/ACCESS.2022.3208571 (access date: 21.07.2024)
6. Sun Y., Wang T., & Wang S. (2019). Location optimization and user association for unmanned aerial vehicles assisted mobile networks. *IEEE Transactions on Vehicular Technology*, *68*(10), 10056-10065. URL: https://doi.org/10.1109/TVT.2019.2933560 (access date: 21.07.2024)
7. Guirado R., Padró J. C., Zoroa A., Olivert J., Bukva A., & Cavestany P. (2021). Stratotrans: Unmanned aerial system (uas) 4g communication framework applied on the monitoring of road traffic and linear infrastructure. *Drones*, *5*(1), 1-14. URL: https://doi.org/10.3390/drones5010010 (access date: 21.07.2024)
8. Khan S. K., Naseem U., Siraj H., Razzak I., & Imran M. (2021). The role of unmanned aerial vehicles and mmWave in 5G: Recent advances and challenges. *Transactions on Emerging Telecommunications Technologies*, *32*(7), pp. 1-18. URL: https://doi.org/10.1002/ett.4241 (access date: 21.07.2024)
9. Mkiramweni M. E., Yang C., Li J., & Zhang W. (2019). A survey of game theory in unmanned aerial vehicles communications. *IEEE Communications Surveys & Tutorials*, *21*(4), 3386-3416. URL: https://doi.org/10.1109/COMST.2019.2919613 (access date: 21.07.2024)
10. Mozaffari M. (2018). Wireless communications and networking with unmanned aerial vehicles: fundamentals, deployment, and optimization, 177-190. URL: http://hdl.handle.net/10919/83921 (access date: 21.07.2024)
11. Ezeigweneme C. A., Umoh A. A., Ilojianya V. I., & Adegbite A. O. (2024). Telecommunications energy efficiency: optimizing network infrastructure for sustainability. *Computer Science & IT Research Journal*, *5*(1), 26-40. URL: https://doi.org/10.51594/csitrj.v5i1.700 (access date: 21.07.2024)
12. Rakushev M., Zuiko V., Pantiushenko R., Nozdrachov O., Khomenko V., & Chornomaz O. (2022, December). The Technique of the Information and Telecommunication System" Terminal" Use for UAVs Acquired Data. In *2022 IEEE 4th International Conference on Advanced Trends in Information Theory (ATIT)*, 208-212. URL: https://doi.org/10.1109/ATIT58178.2022.10024215 (access date: 21.07.2024)
13. Bashir M. N., & Yusof K. M. (2020). A review of relay network on uavs for enhanced connectivity. *Jurnal Teknologi*, *82*(1). 173-183. URL: https://doi.org/10.11113/jt.v82.13183 (access date: 21.07.2024)
14. Euler S., Maattanen H. L., Lin X., Zou Z., Bergström M., & Sedin J. (2019, April). Mobility support for cellular connected unmanned aerial vehicles: Performance and analysis. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1-6). URL: https://doi.org/10.1109/WCNC.2019.8885820 (access date: 21.07.2024)
15. Li B., Zhao S., Miao R., & Zhang R. (2021). A survey on unmanned aerial vehicle relaying networks. *IET communications*, *15*(10), 1262-1272. URL: https://doi.org/10.1049/cmu2.12107(access date: 21.07.2024)

© Радзівілов Г.Д., Дєгтяр О.А., Безносенко С.Ю., Ткаченко А.Л.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-45
УДК: 355
Радзівілов Григорій Данилович, к.т.н., професор
https://orcid.org/0000-0002-6047-1897
Масесов Микола Олександрович, к.т.н., старш. наук. співр.
https://orcid.org/0000-0003-4537-4295
Дєгтяр Олег Андрійович, провід. наук. співр., докт. наук з держ. управління, професор
https://orcid.org/0000-0001-6413-3580
Військовий інститут телекомунікацій та інформатизації імені Героїв Крут, м. Київ, Україна

СВІТОВІ ТЕНДЕНЦІЇ ЗІ СТВОРЕННЯ ТА РОЗВИТКУ ВІЙСЬКОВОЇ ТЕХНІКИ

Радзівілов Г.Д., Масесов М.О., Дєгтяр О.А. Світові тенденції зі створення та розвитку військової техніки. Стаття аналізує порядок та тенденції створення та розвитку військової техніки у модерний період. Здійснено огляд загальних трендів військової сфери у світі відповідно до мілітарних галузей : озброєння та військова техніка, сухопутні війська, військово-морські сили, військово-повітряні сили, військово-космічні війська, військовий зв'язок тощо. Проаналізовано практики найсучасніших та інженерно передових армій світу щодо реагування на військову цифровізацію та, зокрема, цифровізацію мілітарно-технічного середовища : США (проєкти «ШІ та автономна зброя», «Кібервійна», «Інформаційна війна», «Розробка, тестування, відпрацювання та імплементація використання зброї»), ЄС (Європейська програма оборонних досліджень The European defense research policy, Європейська оборонно-промислова стратегія The European Defence Industrial Strategy). Пристосовно до подій в Україні та неспровокованої повномасштабної збройної агресії рф від 24.02.2022 р., концептуалізовано необхідність запровадження трендів військових технологій, орієнтованих на діджитал-трансформацію в Україні. Це, зокрема, розбудова безпілотних систем (БПЛА), протидія проявам гібридних війн (кібератаки, дезінформація, гібридні та терористичні загрози тощо) та застосування у процесі створення та розвитку військової техніки технологій розширеної (AR) та віртуальної (VR) реальностей. Визначено місце вітчизняного військово-оборонного комплексу в формуванні сучасної військової політики та військово-технічного оснащення в умовах Євроінтеграції та реалізації західноорієнтованого курсу.
Ключові слова: військова техніка, військово-оборонний комплекс, військова цифровізація, мілітарна інформатизація, мілітарно-технічне середовище, кібербезпека, військові технології

Radzivilov H., Masesov M., Diegtiar O. World trends in military equipment creation and development. The article analyzes the order and trends in the creation and development of military equipment in the modern period. An overview of the general trends of the military sphere in the world was carried out in accordance with the military branches: weapons and military equipment, land forces, naval forces, air forces, space forces, military communications, etc. The practices of the most modern and engineering-advanced armies of the world in response to military digitalization and, in particular, the digitalization of the military-technical environment were analyzed: the USA (projects "AI and autonomous weapons", "Cyber war", "Information warfare", "Development, testing and implementation of the use of weapons", EU (The European Defense Research Policy, The European Defense Industrial Strategy). Adaptable to the events in Ukraine and the unprovoked full-scale armed aggression of the Russian Federation from February 24, 2022, the need to introduce military technology trends focused on digital transformation in Ukraine is conceptualized. These include, in particular, raising cyber security standards, the development of unmanned aerial vehicles (UAVs), countering manifestations of hybrid wars (cyber attacks, disinformation, hybrid and terrorist threats, etc.) and the use of advanced technologies in the process of creating and developing military equipment (AR) and virtual (VR) realities. The place of the military-defense complex in the formation of modern military policy and military-technical equipment is determined.
Keywords: military equipment, military-defense complex, military digitalization, military informatization, military-technical environment, cyber security, military technologies

Постановка проблеми у загальному вигляді. Процес створення та розвиток військової техніки доцільно розглядати у якості феномену забезпечення державної стабільності, територіальної цілісності, незалежності, суверенітету та недоторканості.

Військова безпека держави де-факто є гарантією її міжнародного статусу, декларацією захищеності від порушень міжнародного права проти останньої. У Преамбулі Північноатлантичного договору (Вашингтон, округ Колумбія, 04.04.1949 р.) ключовою інтенцією транскордонного військово-оборонного співробітництва держав-членів НАТО визначено насамперед колективну оборону та колективну безпеку. При цьому, тенденція щодо розвитку військових технологій та змагання у озброєнні між членами НАТО, союзниками НАТО (що є не повноправними членами Альянсу) та ін. державами, що не входять до блоку, є міжполітичною аксіомою з часів завершення Другої світової війни.

Глобалізація політичних процесів та зростання геополітичної напруги, змагання між такими державами, як США та рф/Китай та ін. та військові конфлікти зумловлюють необхідність використання у військово-технічній сфері передових технологій. Під останніми можна розглядати

© Радзівілов Г.Д., Масесов М.О., Дєгтяр О.А.

як загальні мілітарно-виробничі тенденції, так і тенденції окремих передових держав щодо мілітарного забезпечення – про обидва фактори говоритимемо далі.

Аналіз останніх публікацій. Проблема огляду світових тенденцій створення та розвитку військової техніки є предметом вітчизняних та зарубіжних наукових обговорень.

Серед представників вітчизняної науки доцільно виділяти напрацювання І. Чепкова, О. Горішної, В. Варакути, П. Нора, Р. Шевченка. У працях останніх розглядаються тренди та тенденції діджиталізації військового виробництва та діяльності армії, проаналізовано структурні особливості гіпотетичного розвитку виготовлення озброєння в Україні пристосовно до вимог модерної мілітарної структури взаємозв'язків між державами.

Зарубіжна наукова парадигма щодо огляду світових тенденцій створення та розвитку військової техніки, в свою чергу, переважно концентрується на огляді досвіду розвинених держав світу в даному контексті. Виділяємо аналітичні роботи К. Л. Гартцля, С. Марковскі, Р. Меттьюза, Х. Арієс, Б, Гайгеріха та ін. Напрацювання вищезазначених авторів переважно спрямовуються на огляд практик та досвіду підвищення інформаційно-діджитального потенціалу складової мілітарного забезпечення розвинених держав світу, як-от США, ЄС та ін.

Метою статті є огляд та комплексний аналіз світових тенденцій зі створення та розвитку військової техніки – як на прикладі передових армій та військово-оборонних комплексів світу (США, ЄС) та Організації Північноатлантичного договору (НАТО), так і на прикладі досвіду України у її євроінтеграційному поступі та протидії неспровокованій російській повномасштабній агресії.

Виклад основного матеріалу дослідження.

Огляд проблеми тенденційних підходів до створення та розвитку військової техніки пропонуємо розпочати із окреслення системних складників мілітарної модернізації за галузями : озброєння та військова техніка, сухопутні війська, військово-морські сили, військово-повітряні сили, військово-космічні війська, військовий зв'язок.

У галузі озброєння та військової техніки пріоритетність дослідницьких напрямів та концентрація мілітарних розробок насамперед спрямовується на інженерне, комп'ютерне та телекомунікаційне забезпечення. До сфери науково-розробницького інтересу НАТО, наприклад, у даній розробницькій парадигмі включено таку підгалузеву методологію, як траєкторно-маршрутне планування руху безпілотних літальних апаратів (БПЛА), використання бездротових ретрансляційних та інформаційних мереж, користуванням Інтернет-мережею для обліку даних щодо дронів, літаків, а також амуніції та спорядження, придатного для застосування в межах допомоги потерпілому військовому [1].

Галузь сухопутних військ щодо аспектів потенційних розробок та тенденцій удосконалення переважно концентрується на покращення військово-технічних показників механізованих і танкових військ. В доктрині НАТО придатними до потенційного подальшого покращення визнаються такі галузі, як підвищення швидкісних здатностей бойових танків (battle tank) та бойових машин піхоти (infantry fighting vehicle), стандартизація діяльності бойових інженерних машин та радарів, що можуть бути використані для типографічної орієнтації та предметної орієнтації на місцевості [1].

Сфера розвитку галузі військово-морських сил за специфікою технічно-військової генерації відповідно стандартів НАТО доцільно розглядати з позиції генерації та розвитку безпілотних систем морського базування та автоматизованого використання у сфері гідрозв'язку, військово-морських досліджень та навігаційно-засобових систем підводного планування з метою поглиблення потенціалу морського флоту та балансування сил не лише на сухопутному напрямку, але у сфері аквамілітарного потенціалу держави [1].

Розвиток та планування удосконалення військово-космічних військ як напрям військово-технічного створення та розвитку, знову-таки, доцільно розглядати крізь призму доктрини НАТО та розробок її членів. Міністерство оборони США, наприклад, визначило військово-технологічний розвиток зброї гіперзвукового призначення власним стратегічним орієнтиром у 2020 р. На виконання даного кластеру США, як член НАТО та одночасно передова армія світу за аспектами матеріально-технічного оснащення та професійного навчання, затвердило Стратегію гіперзвукової зброї 2020 р. (US Hypersonic Weapons Strategy 2020), відповідно до якої нараз нарощує виробництво передової гіперзвукової ракети повітряно-стартового призначення AGM-183A [2]. Наявність альтернативних розробок у країн-суперниць США на міжнародній арені (як-от рф та запуск в

© Радзівілов Г.Д., Масесов М.О., Дєгтяр О.А.

експлуатацію гіперзвукової ракети «Циркон» у 2022 р.) є індикатором успішного американських, а відтак – і загальносвітових напрацювань демократично-мілітарного світу в галузі технологічного оснащення армій.

Військовий зв'язок у якості аспекту передового мілітарно-технологічного забезпечення в рамках доктрини НАТО, подібно до загального кластеру озброєння та військової техніки, концентрується на необхідності вдосконалення насамперед мереж бездротового зв'язку (МБЗ) та бездротових сенсорних мереж (БСМ), за допомогою яких забезпечуються передові рішення для військових щодо детекції та радіолокаційного/радіоелектронного приглушення ворожих літальних апаратів та розвідувальних БПЛА, метою яких є генерація та збір даних щодо військових частин, фортифікацій, мілітарних напрацювань тощо (актуально як в умовах стану війни між державами, так і в умовах мирного часу) [1].

Від загальних концептів створення, розвитку військово-технічного спектру за доктриною НАТО пропонуємо перейти до огляду зазначеної тематики пропорційно партикулярним напрацюванням США та ЄС як передових мілітарно-інноваційних світових одиниць.

Військові інновації в США сегментується за конкретними напрямами, як-от «ШІ та автономна зброя», «Кібервійна», «Інформаційна війна», «Розробка, тестування, відпрацювання та імплементація використання зброї».

Проєктний складник «ШІ та автономна зброя» (AI and Autonomous Weapons), ініційований Міністерством оборони США, першочергово спрямований на використання лазерних технологій в галузі розробок дронів та ракет, дозволяє використовувати методи «автоматичного управління» зазначеними видами озброєнь. Прикладом слугують, зокрема, американські ракети Helfire, відзначені близьким до безпохибкового рівнем високоточності. Використання лазерних технологій, що автоматизує штучний інтелект (ШІ), також пропонується Міністерством оборони США до використання як елемент просування піхотних військ під час наступальних операцій та боїв на місцевості [3].

Кіберрозвиток США у галузі військово-технічного виробництва, в свою чергу, розкривається за ініціативою Міністерства оборони США «Кібервійна». За даним концептом, кіберінфраструктура розглядається як підсистема національної та військової безпеки, а захищеність даних (зокрема, мілітарних та державно-управлінських) є пріоритетом мілітарно-урядової діяльності. Супутники, військове обладнання та технології оборонного виробництва Міністерство оборони США також визначає мілітарних кіберпріоритетом до 2040 р. [4]. Підсегмент «Інформаційна війна», водночас, походить від «Кібервійни» та концентрується на ролі та місці засобів масової інформації (ЗМІ) у допомозі збройним формуванням та просуванні стандартів розвитку військового виробництва та нарощування військового потенціалу шляхом комунікації із владою [3].

Сегмент цифровізації військово-оборонного комплексу США «Розробка, тестування, відпрацювання та імплементація використання зброї» пов'язані із правом командирів на ресурсне вдосконалення власних формувань пропорційно потребам. Міністерство оборони США спільно із USAWC (United States Army War College) у 2020 р. визначило порядок використання наявних боєприпасів стандартною операційною процедурою для кожного підрозділу, спрямованою на збирання розвідувальних даних та керування системами зброї, якщо це необхідно для виконання завдань без загрози фізичної шкоди. При цьому командири розробляють вимоги, а сержанти забезпечують повне впровадження останніх в рядовий склад [3].

В свою чергу, у ЄС політика створення та розвитку військової техніки в сучасному форматі забезпечується за допомогою Європейської програми оборонних досліджень The European defense research policy та Європейської оборонно-промислової стратегії The European Defence Industrial Strategy). Так, відповідно до Європейської програми оборонних досліджень The European defense research policy, у 2016 році Єврокомісія опублікувала План дій у сфері європейської оборони (COM(2016)0950), у якому запропонувала створити EDF та інші заходи для підвищення ефективності витрат держав-членів на спільну оборону та сприяння конкурентоспроможності, де було передбачено витрати на оборонні дослідження (PADR) з бюджетом 90 мільйонів євро на 2017-2019 роки та з бюджетом 500 мільйонів євро на 2019-2020 роки. EDF почав працювати 1 січня 2021 року із загальним узгодженим бюджетом майже 8 мільярдів євро на період 2021-2027 років відповідно Регламенту (ЄС) 2021/697). Пріоритетами виконання Європейської програми оборонних досліджень The European defense research policy визначено цифровізацію військового управління,

© Радзівілов Г.Д., Масесов М.О., Дєгтяр О.А.

автоматизацію літально-апаратного забезпечення та технологічні рішення у сфері радіоелектронної боротьби та дистанційних військових дій [4].

В той же час, ключовими наративами Європейської оборонно-промислової стратегії The European Defence Industrial Strategy, представленої 05.03.2024 р. з метою збільшення витрат на оборону, є покращення спільних закупівель, забезпечення передбачуваності для оборонної промисловості та підвищення оперативної сумісності між збройними силами країн Європи та НАТО на засадах посилення співпраці з українськими оборонно-промисловими підприємствами. Законодавча пропозиція щодо нормативного акту про заснування Європейської програми оборонної промисловості (EDIP) і рамки заходів для забезпечення своєчасної доступності та постачання оборонної продукції є вирішальним кроком у реалізації цієї стратегії, згідно з чим EDIP має мобілізувати 1,5 мільярда євро з бюджету ЄС протягом 2025-2027 років задля продовження підвищення конкурентоспроможності оборонної технологічної та промислової бази ЄС у її прагненні реалізувати власні продемократичні станларти поза межами Союзу в мілітарному прояві [4].

Потрібно зазначити, що окрім США та ЄС як піонерів нововведень у галузі стандартизації військово-оборонної діяльності, доцільно також розглядати вітчизняний досвід відповідності сучасним мілітарним реаліям як своєрідну тенденційну рамку. Так, повномасштабне вторгнення рф до України від 24.02.2022 р. поставило питання щодо необхідності військової, стратегічної та інформацйно-інноваційної протидії ворогові, яку Україна реалізувати у таких мілітарно-розробницьких стартапах, як підвищення стандартів кібербезпеки, розбудова безпілотних систем (БПЛА), протидія проявам гібридних війн та застосування у процесі створення та розвитку військової техніки технологій розширеної (AR) та віртуальної (VR) реальностей [5].

Підвищення стандартів кібербезпеки як механізм військово-мілітарних інновацій в Україні передбачає побудову защиеності систем державних органів, зокрема, оборонного комплексу, від кібернападів ворога (рф) з метою збереження, по-перше, внутрішньомілітарних спроможностей та стабільності та, по-друге, опцій захисту життів громадян в умовах воєнного стану [5].

Розбудова безпілотних систем (БПЛА), в свою чергу, спрямована на розвідувальні та ударні операції по законних військових об'єктах на території рф. Йдеться, зокрема, про SHARK, "Лелека-100", "Фурія", "Валькірія", PD-2 [5].

Протидія проявам гібридних війн ((кібератакам, дезінформації, гібридним та терористичним загрозам) форматується крізь призму діяльності не лише мілітарного, але й субмілітарного (інформаційно-мілітарного) комплексу. Компетенція щодо даного питання в Україні може концептуалізуватися через діяльність Центру протидії дезінформації (ЦПД), що агрегує вітчизняний інформаційний простір на предмет внутрішніх та зовнішніх фейків, що дестабілізують суспільно-інформаційне та військове середовище [5].

Наостанок, сегментарне застосування у процесі створення та розвитку військової техніки технологій розширеної (AR) та віртуальної (VR) реальностей в Україні необхідно розглядати радше як бонус-трекер підвищення мілітарно-військової кваліфікації. За допомогою AR можна проводити навчання, симуляції, а також підвищувати стандарти розвідки та віддаленої навігації на місцевості, в той час як VR допомагає підвищувати професійно-психологічну компетенцію бійців шляхом симуляції бойових навчань, віртуальних тренувань та фактологічного проведення ментальних тренінгів за допомогою комп'ютерних систем та комунікації із фахівцями [5].

Виходячи із зазначеного, можна говорити про комплексний характер створення та розвитку військової техніки як окремого сегменту, що включає в себе не лише галузеві мілітарні, але й субмілітарні тенденції забезпечення військових пропорційно вимогам та критеріям часу.

Висновки і перспективи подальших досліджень. Аналіз світових тенденцій зі створення та розвитку військової техніки дозволив дійти наступних умовиводів.

По-перше, загальносвітовими пріоритетами мілітарного розвитку, згідно доктринально-предметного інтересу НАТО, є сфери озброєння та військової техніки, сухопутних військ, військово-морських сил, військово-повітряних сил, військово-космічних військ, військового зв'язку тощо.

По-друге, піонером у галузі військово-мілітарних розробок та інновацій можна назвати США, де Міністерство оборони США координує та розробляє фахову сегментацію за такими напрямами, як проєкти «ШІ та автономна зброя», «Кібервійна», «Інформаційна війна», «Розробка,

© Радзівілов Г.Д., Масесов М.О., Дєгтяр О.А.

тестування, відпрацювання та імплементація використання зброї», більш детально означені нами вище.

По-третє, і заключне, повномасштабне вторгнення рф до України від 24.02.2022 р. поставило перед вітчизняним військово-оборонним та фахово-мілітарним комплексом завдання інноваційного нарощування озброєнь та мілітарних нововведень. Серед ключових таких новацій виділяємо розбудову безпілотних систем (БПЛА), протидію проявам гібридних війн (кібератаки, дезінформація, гібридні та терористичні загрози тощо) та застосування у процесі створення та розвитку військової техніки технологій розширеної (AR) та віртуальної (VR) реальностей.

Список використаних джерел

1. Писаренко, Т. та ін. Аналіз світових технологічних трендів у військовій сфері. ДНУ «Український інститут науково-технічної експертизи та інформації». Київ, 2021. 110 с.
2. US Congressional Budget Office. U.S. Hypersonic Weapons and Alternatives. USCBO official website. Режим доступу : https://www.cbo.gov/publication/58924
3. Hartzell, C. L. Future Weapons Technology of 2040. Army University Press. 2023. AUP official website. Режим доступу : https://www.armyupress.army.mil/Journals/NCO-Journal/Archives/2023/July/Future-Weapons-Technology-of-2040/
4. European Parliament. Defence industry. EP official website. Режим доступу : https://www.europarl.europa.eu/factsheets/en/sheet/65/defence-industry
5. Кондрат, С. Військові Технології: 6 Трендів 2023-2026. Ukraine Lifehacker. 2023. Офіційний веб-сайт ULH. Режим доступу : https://www.ukraine-lifehacker.com/military-technologies-trends

References

1. Pysarenko, T., et al. (2021) Analysis of global technological trends in the military sphere. Kyiv: DNU "Ukrainian Institute of Scientific and Technical Expertise and Information".
2. Swagel, P. L. (2023). U.S. Hypersonic Weapons and Alternatives. *US Congressional Budget Office.* URL: https://www.cbo.gov/publication/58924
3. Hartzell, C. L. (2023). Future Weapons Technology of 2040. Army University Press. *AUP official website.* URL: https://www.armyupress.army.mil/Journals/NCO-Journal/Archives/2023/July/Future-Weapons-Technology-of-2040/
4. European Parliament. Defense industry. *EP official website.* URL: https://www.europarl.europa.eu/factsheets/en/sheet/65/defense-industry
5. Kondrat, S. (2023). Military Technologies: 6 Trends 2023-2026. Ukraine Lifehacker. *ULH Official Website.* URL: https://www.ukraine-lifehacker.com/military-technologies-trends

© Радзівілов Г.Д., Масесов М.О., Дєгтяр О.А.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-46
УДК 621.396
Селепина Йосип Романович, к.т.н., доцент
https://orcid.org/ 0000-0002-2421-1844
Бондаренко Олександр Олександрович, студент
https://orcid.org/ 0000-0001-9456-6715
Луцький національний технічний університет, Луцьк, Україна

ІНТЕГРАЦІЯ СИСТЕМ ЗВ'ЯЗКУ НА БАЗІ WI-BAS В ГІБРИДНИХ МЕРЕЖАХ

Селепина Й.Р., Бондаренко О.О. Інтеграція систем зв'язку на базі Wi-Bas в гібридних мережах. В роботі проведено аналіз технології Wi-BAS як ключового елемента в гетерогенних безпроводових мережах. Висвітлено основні переваги Wi-BAS, зокрема можливість забезпечення широкосмугового доступу на великих відстанях, що є важливим для віддалених і сільських районів, де використання традиційної інфраструктури малоефективне. Особлива увага приділена інтеграції Wi-BAS з сучасними мобільними технологіями, такими як LTE та 5G, що забезпечує гнучкість і масштабованість мережі. Розглянуто механізми керування частотним спектром і процедури спектрального моніторингу для підвищення ефективності та мінімізації перешкод у роботі безпроводових мереж. Запропоновано алгоритм спектрального сенсингу для забезпечення стабільності роботи мережі в умовах зростаючого навантаження.
Ключові слова: гетерогенні мережі, безпроводовий зв'язок, управління, частотний спектр.

Selepyna Y., Bondarenko O. Integration of communication systems based on Wi-Bas in hybrid networks. The article analyzes Wi-BAS technology as a key element in heterogeneous wireless networks. The main advantages of Wi-BAS are highlighted, particularly the ability to provide broadband access over long distances, which is crucial for remote and rural areas where traditional infrastructure is inefficient. Special attention is given to the integration of Wi-BAS with modern mobile technologies such as LTE and 5G, which ensures network flexibility and scalability. Mechanisms for frequency spectrum management and spectral monitoring procedures are also discussed to improve efficiency and minimize interference in wireless network operations. The authors propose a spectrum sensing algorithm to ensure network stability under increasing load conditions.
Keywords: heterogeneous networks, wireless communication, management, frequency spectrum.

Постановка наукової проблеми. У сучасному світі бездротові технології відіграють ключову роль у розвитку інформаційного суспільства, забезпечуючи швидкий і надійний зв'язок у будь-якій точці планети. Зростаюча потреба в мобільності, збільшення кількості підключених пристроїв, зростання обсягу переданих даних та вимоги до якості обслуговування (QoS) стимулюють розвиток нових технологій і рішень у сфері телекомунікацій. Однією з таких технологій є WiBAS (WiMAX Broadband Access Systems), яка забезпечує широкосмуговий бездротовий доступ до Інтернету на великих відстанях. Проте, у світі, де інфраструктура зв'язку складається з різноманітних технологій, що виконують різні завдання, важливим стає питання інтеграції WiBAS з іншими бездротовими технологіями. Завдяки своїм характеристикам WiBAS може заповнити прогалини в забезпеченні послуг, які є недоступними або економічно недоцільними при використанні традиційних технологій мобільного зв'язку 3G/4G, або з використанням дротової інфраструктури, що часто є високовартісною та технічно і технологічно складною в реалізації у сільських віддалених районах, районах з гірським рельєфом, тощо.

Інтеграція різних бездротових технологій дозволяє створювати гібридні мережі, які поєднують переваги кожної з них і таким чином забезпечують користувачам надійний і безшовний зв'язок, незалежно від їхнього місцезнаходження та умов використання.

Іншим питанням є інтеграція WiBAS з новими високопродуктивними технологіями мобільного зв'язку, зокрема LTE та 5G, оскільки це дозволяє значно підвищити гнучкість і масштабованість мережі. Наприклад, у міських умовах, де є щільна забудова та висока концентрація користувачів, LTE та 5G можуть забезпечити високу швидкість і мінімальні затримки для мобільних пристроїв. У той же час, WiBAS може бути використаний для забезпечення зв'язку на більших відстанях або в умовах, де традиційні мобільні мережі виявляються малоефективними.

Аналіз досліджень. Гетерогенні мережі, які поєднують різні технології доступу, такі як LTE, 5G, Wi-Fi та WiMAX, є важливою складовою сучасної телекомунікаційної інфраструктури. Широкий спектр досліджень і розробок в цьому напрямку визначають, що однією з основних проблем у створенні гетерогенних мереж є інтероперабельність між різними стандартами зв'язку та протоколами. Кожен з елементів гетерогенних мереж потребує складних механізмів керування трафіком для забезпечення безперебійної взаємодії з іншими складовими мережі. Дослідження показують, що інтеграція Wi-BAS із мобільними мережами забезпечує зменшення затримок і підвищення якості обслуговування (QoS) завдяки інтелектуальному управлінню трафіком і

© Селепина Й.Р., Бондаренко О.О.

частотним ресурсом. З цієї точки зору, нова технологія Wi-BAS (WiMAX Broadband Access System) демонструє значний потенціал для розширення можливостей традиційних мереж зв'язку, особливо в умовах гетерогенних мереж [1-3].

Однією з ключових переваг Wi-BAS є можливість забезпечення широкосмугового доступу на великих відстанях, до 50 км від базової станції та підтримує різні топології мереж, включаючи точка-точка (point-to-point) і точка-багатоточок (point-to-multipoint). Це робить технологію ідеальною для покриття сільських та віддалених районів, де інші види зв'язку можуть бути недоступними або занадто дорогими. Wi-BAS може підтримувати швидкість передачі даних до 70 Мбіт/с та забезпечує високу якість обслуговування за рахунок пріоритезації трафіку та управління ресурсами мережі, що дозволяє використовувати її для підтримки високоякісних мультимедійних сервісів, таких як відеоконференції, VoIP, та стрімінг відео в режимі реального часу, що є важливим для бізнесу та критично важливих додатків. Wi-BAS працює в діапазоні частот від 2 до 66 ГГц, що дозволяє гнучко підходити до планування мережі і враховувати місцеві особливості частотного планування, а також уникати завад від інших бездротових систем. Завдяки використанню технологій модуляції та корекції помилок, Wi-BAS демонструє високу стійкість до перешкод, що забезпечує надійний зв'язок навіть у складних умовах. Проте для повноцінної інтеграції з іншими технологіями необхідно вирішити низку проблем, пов'язаних із сумісністю протоколів, управлінням QoS та оптимізацією трафіку [4].

Мета роботи. Метою дослідження є аналіз вимог та проблем створення гетерогенних мереж з використанням технології Wi-BAS, як оптимального рішення для розширення можливостей традиційних мереж зв'язку.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження.

Зважаючи на рівень сьогоднішніх вимог споживачів до операторів мереж зв'язку, для підтримання рівня конкурентоспроможності, очікуваний перехід до ери LTE повинен завершитись в найближчі роки, проте більшість операторів мобільного зв'язку планують підтримувати свою існуючу інфраструктуру GSM/UMTS ще досить довго. Інтеграція різних бездротових технологій у єдину гібридну мережу може значно збільшити витрати на розгортання, налаштування та обслуговування. Це включає вартість обладнання, ліцензії на використання частотного спектру, а також витрати на навчання персоналу і технічну підтримку. Саме тому зараз на користувацькій ділянці домінує плоска операторська мережа Ethernet із підтримкою застарілого трафіку для забезпечення зворотної сумісності. Прогнозується, що на шляху еволюції мобільних мереж до 4G оптимальним є домінування мікро- та піко-стільників [5]. Досвід показує, що малі стільники можуть ефективно підтримувати підвищену ємність мережі, знизити затримки і покращити якість обслуговування в умовах високої щільності користувачів. Використання малих стільників дозволяє операторам мобільного зв'язку розширити покриття та підвищити пропускну здатність без необхідності значного збільшення потужності передавачів або побудови нових макростільників. Завдяки цьому технологія малих стільників стала ключовим елементом у побудові гетерогенних мереж, де різні типи базових станцій працюють разом для забезпечення безперебійного з'єднання та високошвидкісної передачі даних.

Пошук рішень для реалізації зворотного зв'язку малих стільникових зв'язків, виявив переваги технології Wi-BAS, що може забезпечити з'єднання «останньої милі», виконуючи інтенсивне агрегування трафіку [1]. У міських районах, де щільність абонентів мобільного зв'язку дуже висока, з'єднання «останньої милі» можна оптимально встановити за допомогою транзитного зв'язку Wi-BAS (рис. 1). Аналіз функцій Wi-BAS показує, що технологія здатна забезпечити необхідний нам «безшовний» перехід завдяки наступним можливостям:

• агрегація трафіку LTE з кількох районів у містах;

• пріоритизація бездротового трафіку на основі VLAN/біт пріоритету/DSCP із вісьмома класами пріоритету для наскрізного QoS;

• обмін будь-якою сигнальною інформацією (інтерфейси 3GPP R.10x2) між підключеними стільниками LTE;

• низька наскрізна затримка та висока конфіденційність даних користувача відповідно до строгих вимог LTE;

• Підтримка застарілого трафіку (E1, TDM/ATM) для переходу зі змішаних сайтів RAN на LTE.

• Підтримка синхронного Ethernet.

© Селепина Й.Р., Бондаренко О.О.

Вказані можливості додатково можна розширити за допомогою технології транзитного зв'язку E-Band PtP (60 ГГц). Ці технології мають подібні характеристики продуктивності та обидві дуже привабливі з точки зору техніко-економічних перспектив.

Типова система (рис. 1) складається з базової станції, яка обслуговує кілька розкиданих термінальних станцій. Базова станція збирає й обробляє застарілий TDM/ATM і пакетований трафік з усіх обслуговуваних ліній і пересилає його до базової мережі, завжди зберігаючи наскрізну якість обслуговування [1]. Wi-BAS комірка фізично поділена на сектори (зазвичай два або чотири) та обслуговується радіосистемою базової станції (BRS), зовнішньою частиною BS. BRS контролює радіозв'язок між BS і розподіленими термінальними станціями TS і зв'язується із зовнішньою частиною BS. Внутрішня частина BS агрегує трафік від усіх TS і надає мережеві інтерфейси магістралі. MSAD (Multi-Service Access Device), внутрішня частина TS, забезпечує інтерфейс користувача. Усі Wi-BAS елементи мережі для будь-якої кількості систем можна налаштовувати, контролювати та контролювати дистанційно за допомогою розширеної системи керування мережею.

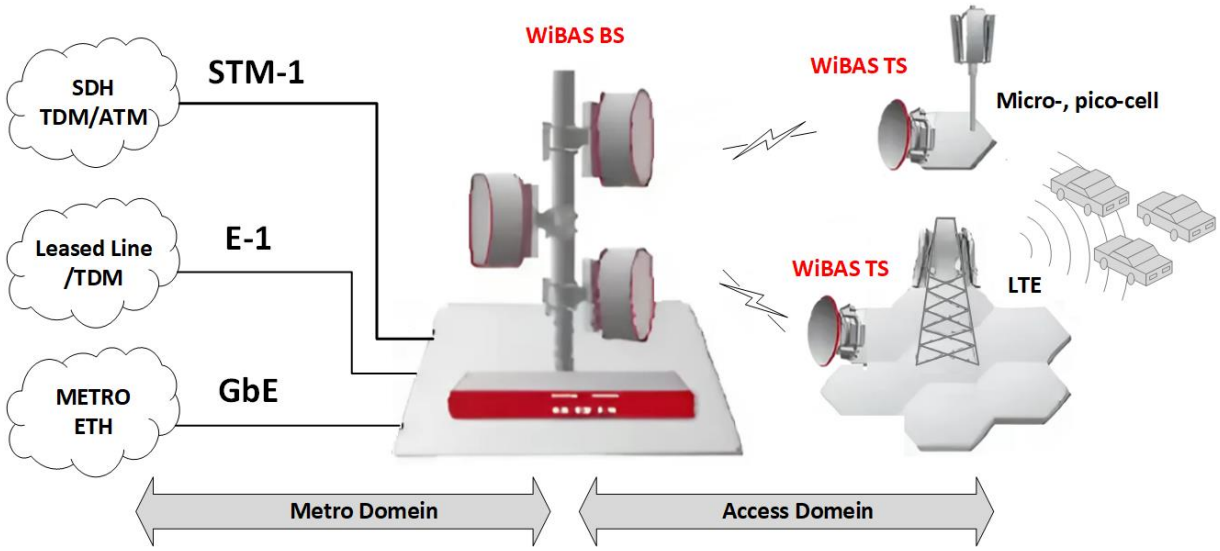


Рис.1. Схема застосування транспортного зв'язку Wi--BAS

Такий економічно-розумний підхід до мережевого планування крім того може подолати обмеження пов'язане з необхідністю прямої видимості між передавачем і приймачем, зокрема властиве для міліметрової хвилі у 5G бездротових мережах, сигнал в яких може інтенсивно відбиватися або поглинатися перешкодами, що призводить до втрат і зниження якості зв'язку.

Детальні параметри системи наведені в таблиці 1.

Табл. 1. Параметри базової станції BS та абонентської станції SS Wi-BAS

Параметри базової станції BS		Параметри абонентської станції SS	
Максимальна кількість вузлів	100	Підсилення антени (дБі)	−1
Максимальна щільність потужності (дБм/підканал)	− 60	Максимальна передача потужності (Вт)	0,5
Ranging Back off Start	2	Модуляція та кодування	64 QAM
Ranging Back off End	4	Розмір буфера	64 Кб
Профіль PHY	Wireless OFDM 20 МГц	Профіль PHY	Wireless OFDMA 20 МГц
Тип профілю PHY	OFDM	Тип профілю PHY	OFDM

Вибравши WiBAS (PtMP) як основну технологію транспортного зв'язку для малих стільників, увесь макрорегіон може покриватися єдиною базовою станцією, стратегічно розміщеною в точці агрегації. У реальних мережах очікуваний діапазон малих комірок становить від 100 м до 300 м, що, з точки зору планування радіозв'язку, забезпечує певну гнучкість щодо вибору доступного місця розміщення базової станції. Перевагами чотирисекторної системи точка-

© Селепина Й.Р., Бондаренко О.О.

багато точок є висока масштабованість (кілька малих комірок на базову станцію), висока надійність (захист на базовій станції), зменшення витрат на експлуатацію та обслуговування. При умові необхідності розширення системи кожен новий термінал для обслуговування мікро-стільника, можна розгорнути без відвідування базової станції.

Так само Wi-BAS розширює охоплення інших широкосмугових технологій, таких як Wi-Fi, WiMAX і xDSL. Система легко інтегрується з існуючою мережевою інфраструктурою та може бути використана для одночасного надання послуг широкосмугового доступу зокрема великим бізнес-клієнтам.

Проте, при розгортанні та гібридних бездротових мереж, зокрема з використанням Wi-BAS як останньої ланки передачі інформації до споживача виникають деякі складності. Оскільки, різні бездротові технології використовують різні частотні діапазони, протоколи зв'язку, методи модуляції і стандарти якості обслуговування (QoS), то це ускладнює інтеграцію різних компонентів у єдину мережу. Забезпечення взаємодії між цими технологіями вимагає спеціальних рішень, таких як шлюзи, конвертери протоколів або адаптивні маршрутизатори.

У зв'язку зі зростанням насиченості радіопростору сигналами різноманітних комплексів і пристроїв передавання інформації, особливо гостро стоїть проблема забезпечення електромагнітної сумісності та перешкодостійкості системи зв'язку [6]. Очевидно, що кожна бездротова технологія використовує певний частотний діапазон, який може перекриватися з іншими технологіями. Це призводить до ризику взаємних перешкод і зниження ефективності роботи мережі. Саме тому, управління частотним спектром стає критично важливим завданням, яке потребує ретельного планування і динамічного розподілу частот.

Управління частотним спектром в мережах Wi-BAS включає в себе кілька важливих процесів та методів, спрямованих на оптимізацію використання частотного спектру для запобігання перешкодам, підвищення продуктивності та забезпечення сумісності з іншими бездротовими системами [7]. Зокрема:

- Wi-BAS працює в широкому діапазоні частот від 2 до 66 ГГц, що включає як ліцензовані, так і неліцензовані діапазони. Використання ліцензованих частот забезпечує захист від перешкод з боку інших користувачів, але вимагає отримання ліцензій від регуляторних органів. Неліцензовані діапазони можуть бути більш доступними, але вони також використовуються іншими технологіями, що збільшує ймовірність взаємних перешкод.

- Wi-BAS підтримує динамічний розподіл частот, що дозволяє адаптувати використання спектру в реальному часі залежно від умов навантаження і стану навколишнього середовища. Система може автоматично перемикатися на менш завантажені частоти або змінювати ширину каналу (наприклад, 3,5 МГц, 7 МГц або 14 МГц) для підвищення пропускної здатності, таким чином допомагає уникати частотних конфліктів і оптимізувати використання доступного спектру.

- Для мінімізації перешкод між різними базовими станціями або іншими бездротовими системами, що працюють в тих же або сусідніх частотних діапазонах, Wi-BAS використовує різні методи управління інтерференцією. Це може включати використання розширених методів модуляції, таких як OFDM (Orthogonal Frequency-Division Multiplexing), а також динамічне регулювання потужності передавачів, щоб знизити вплив перешкод на інші пристрої.

Для забезпечення цих функцій в системі постійно здійснюється моніторинг спектру для постійного відстеження стану частотного діапазону [7-8]. На сьогодні активно розвивається ряд паралельних технологій спектрального моніторингу, пристосованих до систем різного рівня складності. Сюди входить і використання сучасних спектральних аналізаторів, систем моніторингу на основі програмно-конфігурованих радіо (SDR - Software-Defined Radio), або централізованих систем управління спектром, що дозволяють здійснювати аналіз частотного спектру в реальному часі. Таким чином виконується ряд завдань, від ідентифікації джерел перешкод і перевірки відповідності частотних діапазонів, до виявлення перевантажених частотних діапазонів і прийняття рішень щодо оптимізації використання спектру, а також забезпечення глобального огляду стану частотного спектру в межах всієї мережі.

Технічне забезпечення, при цьому, може бути інтегрованим в базові станції або розміщене окремо для постійного моніторингу радіочастотного середовища, включаючи мобільні платформи моніторингу.

Хоча Wi-BAS системи не є когнітивними радіомережами в повному сенсі, вони інтегрують кілька сучасних методів моніторингу спектру, включаючи централізоване управління спектром, динамічний доступ до спектру, спектральний сенсинг і використання спектральних аналізаторів.

© Селепина Й.Р., Бондаренко О.О.

Спектральний сенсинг, при цьому, як процес виявлення наявності або відсутності сигналів у певному частотному діапазоні, є ключовим компонентом когнітивних радіомереж і системи управління спектром. Для реалізації процедури спектрального сенсингу розроблено алгоритм (рис. 2), що включає реалізацію функціональних кроків, кожен з яких відіграє важливу роль у точному і надійному визначенні стану спектру.

Алгоритм спектрального сенсингу складається з кількох кроків, які забезпечують ефективне виявлення сигналів у певному частотному діапазоні. Основними завданнями цього процесу є точне і надійне визначення стану спектра для оптимізації використання частотних ресурсів і мінімізації взаємних перешкод у безпроводових мережах. Залежно від умов та вимог, алгоритм може бути налаштований на використання різних методів виявлення сигналів, що дозволяє адаптуватися до специфічних сценаріїв використання. Ця технологія є важливою для когнітивних радіомереж, де пристрої можуть автоматично перемикатися на менш завантажені частоти для мінімізації перешкод.

Рис.2.Алгоритм реалізації процедури спектрального сенсингу

Хоча переваги впровадження систем моніторингу спектру очевидні, залишається низка проблем, що наразі не мають єдиного рішення. Зокрема, моніторинг великих ділянок радіочастотного спектру вимагає спеціального обладнання, яке може бути громіздким і дороговартісним. З'являються рішення використання БПЛА для здійснення ефективного моніторингу великих територій, що вимагає стратегічного розгортання радіодавачів і створення оптимізованих траєкторій польоту [8]. Останні дослідження спрямовані на розв'язання проблеми проектування траєкторії БПЛА за допомогою штучного інтелекту, однак стандартний дизайн траєкторії та підхід до оптимізації для роїв БПЛА для спільного моніторингу спектру відсутні.

Наступним питанням є врахування обмежень на максимальну затримку, яка допускається при отриманні та аналізі інформаційних даних, з метою зменшення похибок при регулюванні в режимі реального часу. Вирішення цієї проблеми вимагає розробки інноваційних алгоритмів штучного інтелекту та передових технологій обробки сигналів. Крім того, збереження конфіденційності

© Селепина Й.Р., Бондаренко О.О.

накладає додаткові обмеження на типи даних, доступних пристроям збору інформації, ускладнюючи отримання інформації з бездротових мереж. Це підкреслює необхідність розробки нових, високонадійних алгоритмів, призначених для отримання конкретних аналітичних даних із неконфіденційних наборів даних, сприяючи вдосконаленню технології моніторингу спектру для узгодженню роботи та оптимізації гетерогенних безпроводових мереж.

Висновки. Виконано аналіз умов інтеграції технології Wi-BAS як проміжної ланки в гетерогенних безпроводових мережах, що пропонує надійне, високоефективне та комплексне рішення для надання послуг доступу віддаленим користувачам і створення нових джерел доходу для операторів. Показано можливості системи для бездоганної інтеграції з існуючими операторськими мережами завдяки когнітивним функціям керування доступним радіоресурсом, які здатні самостійно виявляти зміни у спектрі, адаптуватися до них, навчатися на основі попереднього досвіду та забезпечувати ефективне використання спектру з мінімальними перешкодами для інших користувачів. Показано, що для забезпечення оптимальної сумісності необхідно впроваджувати ефективні технології та алгоритми спектрального моніторингу. Запропоновано алгоритм реалізації процедури спектрального сенсингу для контролю за частотним діапазоном, що є критичним для підтримки надійності та ефективності безпроводових мереж. Використання цих технологій дозволяє оптимізувати розподіл спектру, виявляти і усувати перешкоди, а також забезпечувати стабільну роботу мережі в умовах постійно зростаючого навантаження на радіочастотний ресурс. Незважаючи на значні переваги, які можуть бути досягнуті завдяки гібридним безпроводовим мережам, їх створення і підтримка вимагають ретельного планування, технологічної компетентності та значних ресурсів. Подолання цих викликів є ключовим завданням для забезпечення стабільної та ефективної роботи таких мереж у різних умовах експлуатації.

Список бібліографічного опису

1.WiBAS-C System. Опис системи. Вид. 3 англ. Intracom S.A. Telecom solutions, Пеанія, Афіни, 2012. 121 с.
2.Бешлей Г., Шкоропад Ю., Бешлей М., Климаш М. Конвергенція гетерогенних безпроводових мереж для майбутніх комунікацій: архітектура, QoS та управління ресурсами. Інформаційні та комунікаційні технології, електронна інженерія. Т. 2, № 2. 2022. С. 20-32. DOI: https://doi.org/10.23939/ictee2022.02.020
3.Атаєро А., Адегоке Е., Алатіше А., Ор'я М. Гетерогенні безпроводові мережі: огляд архітектур взаємодії. Міжнародний журнал інженерії та технологій. Т. 2. 2012. С. 220-232.
4.Лян Г., Ю Х. Алгоритм вибору мережі для гетерогенних безпроводових мереж на основі характеристик послуг і переваг користувачів. Wireless Com Network, Т. 241. 2018. DOI: https://doi.org/10.1186/s13638-018-1264-5
5.Аль-Хававреx М.С., Зрейкат А.І. Аналіз продуктивності мережі WIMAX у різних моделях розповсюдження. Міжнародний журнал комп'ютерних наук і інформаційної безпеки (IJCSIS), Т. 15, № 1, січень 2017.
6.Ума М., Ганапаті П. Ефективний моніторинг трафіку для безпроводових мереж. Міжнародний журнал комп'ютерних додатків 53(9). 2012. С. 51-59. DOI: https://doi.org/10.5120/8453-2255
7.На Ся, Ван Конг, Пінг Хуайчжень, Чжао Чжунцю, Чень Юцін, Ван Пейпей, Ду Хуачжень, Дін Шен, Ю Йонгтан. Оптимізаційні алгоритми в мережах безпроводового моніторингу: огляд. Neurocomputing, Т. 489. 2022. С. 584-598. DOI: https://doi.org/10.1016/j.neucom.2021.12.072
8.Тесті Е., Джорджетті А. Аналітика безпроводових мереж для нової ери патрулювання і моніторингу спектра. IEEE Wireless Communications, 99. 2024. С. 1-7. DOI: https://doi.org/10.1109/MWC.012.2300374

References

1. WiBAS-C System. System description. Ed 3 en. Intracom S.A. Telecom solutions, Peania, Athens, 2012. 121p.
2. Beshley H., Shkoropad Y., Beshley M., Klymash M. Convergence of heterogeneous wireless networks for future communications: architecture, QoS and resource management. Information and communication technologies, electronic engineering. Vol. 2, No. 2. 2022. pp. 20-32. DOI: https://doi.org/10.23939/ictee2022.02.020
3. Atayero A., Adegoke E., Alatishe A., Orya M. Heterogeneous Wireless Networks: A Survey of Interworking Architectures. International Journal of Engineering and Technology. Volume 2. 2012. pp. 220-232.
4. Liang G., Yu H. Network selection algorithm for heterogeneous wireless networks based on service characteristics and user preferences. Wireless Com Network, Vol. 241. 2018. DOI: https://doi.org/10.1186/s13638-018-1264-5
5. AL-Hawawreh M.S., Zreikat A.I. Performance Analysis of a WIMAX Network in Different Propagation Models. International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 1, January 2017.
6.Uma M., Ganapathi P. An Efficient Network Traffic Monitoring for Wireless Networks. International Journal of Computer Applications 53(9). 2012. pp. 51-59. DOI: https://doi.org/10.5120/8453-2255
7. Na Xia, Cong Wang, Huaizhen Peng, Zhongqiu Zhao, Yuqing Chen, Peipei Wang, Huazheng Du, Sheng Ding, Yongtang Yu. Optimization algorithms in wireless monitoring networks: A survey. Neurocomputing, Volume 489. 2022. pp. 584-598. DOI: https://doi.org/10.1016/j.neucom.2021.12.072
8. Testi E., Giorgetti A. Wireless Network Analytics for the New Era of Spectrum Patrolling and Monitoring. IEEE Wireless Communications, 99. 2024. pp. 1-7. DOI: https://doi.org/10.1109/MWC.012.2300374

© Селепина Й.Р., Бондаренко О.О.

DOI: https://doi.org/10.36910/6775-2524-0560-2024-56-47
УДК 621.396.
Soproniuk Ivan[1], PhD student
https://orcid.org/0009-0006-2831-0790
Komar Oleksii [2], PhD, Associate Professor
https://orcid.org/0009-0002-2994-6556
[1]Ukrainian State University of Railway Transport, Kharkiv, Ukraine
[2]National Aviation University, Kyiv, Ukraine

ADAPTIVE APPROACH TO SPECTRUM MONITORING IN COGNITIVE RADIO NETWORKS THROUGH SIGNAL DETECTION OPTIMIZATION

Soproniuk I., Komar O. Adaptive approach to spectrum monitoring in cognitive radio networks through signal detection optimization. The article considers the improvement of the adaptive algorithm of the spectral monitoring method for cognitive radio networks by introducing adaptive wavelet transforms and filters. The use of adaptive Morle and Dobechy wavelet transforms, as well as adaptive Kalman, LMS, and RLS filters is proposed, which allows dynamically changing parameters depending on the conditions of the radio environment. The comparative analysis with traditional methods showed that adaptive methods significantly increase the efficiency of signal detection in conditions of low SNR values, reducing the noise level, improving the accuracy of signal detection and reducing the probability of false alarms. The results of the study confirm the perspective of using adaptive methods to increase the reliability and efficiency of spectral monitoring in real operating conditions.
Keywords: spectral monitoring, cognitive radio networks, adaptive wavelet transforms, adaptive filters, signal processing methods, low SNR, signal detection

Сопронюк І.І., Комар О.М. Адаптивний підхід до спектрального моніторингу в когнітивних радіомережах за рахунок оптимізації детектування сигналів. У статті розглянуто удосконалення адаптивного алгоритму методу спектрального моніторингу для когнітивних радіомереж шляхом впровадження адаптивних вейвлет-перетворень та фільтрів. Запропоновано використання адаптивних вейвлет-перетворень Морле та Добеші, а також адаптивних фільтрів Калмана, LMS та RLS, що дозволяє динамічно змінювати параметри залежно від умов радіосередовища. Проведений порівняльний аналіз з традиційними методами показав, що адаптивні методи значно підвищують ефективність детектування сигналів в умовах низьких значень SNR, зменшуючи рівень шуму, покращуючи точність виявлення сигналів і знижуючи ймовірність хибних тривог. Результати дослідження підтверджують перспективність використання адаптивних методів для підвищення надійності та ефективності спектрального моніторингу в реальних умовах експлуатації.
Ключові слова: спектральний моніторинг, когнітивні радіомережі, вейвлет-перетворення, адаптивні фільтри, методи обробки сигналів, низький SNR, детектування сигналів.

Statement of a scientific problem.

The adaptive algorithm of the spectrum monitoring method (Fig. 1) requires further enhancement through the integration of adaptive wavelet transforms and adaptive filters capable of adjusting parameters in real-time based on the conditions of the radio environment, ensuring stable and efficient operation of the telecommunication system. The necessity for these enhancements is substantiated by the following reasons.

1. Rapidly changing radio environment conditions. Static signal processing methods are inadequate for effectively adapting to these changes, which leads to a reduction in the accuracy of signal detection. Adaptive wavelet transforms, however, can dynamically adjust their parameters to optimize signal analysis, thereby ensuring higher detection accuracy in fluctuating environments.

2. Mitigation of noise and distortion impacts. As the Signal-to-Noise Ratio (SNR) decreases, the influence of noise and distortions on signals significantly increases. Adaptive filters, which can adjust their parameters in real-time, provide more effective noise filtering and minimize distortions, thus enhancing the quality of signal detection. This adaptive capability is critical for maintaining signal integrity under challenging conditions.

3. Improvement in True Positive Rate (TPR) for signal detection. Simulation results have demonstrated that the TPR for signal detection diminishes as SNR decreases. The use of adaptive wavelet transforms allows for better isolation of useful signals from noise, thereby increasing the TPR even under low SNR conditions. This improvement is crucial for reliable communication in noisy environments.

4. Optimization of computational resources. Adaptive methods can optimize the use of computational resources by adjusting the parameters of filters and transforms according to current conditions. This capability reduces processing delays (PD) and enhances system responsiveness, leading to more efficient use of available resources while maintaining high processing speed.

© Soproniuk I.I., Komar O.M.

5. Reduction in False Positive Rate (FPR). Adaptive filters can lower the likelihood of false positives (FPR) by more accurately separating the useful signal from noise. This reduction is particularly important for decreasing the number of false alarms, thereby improving the overall reliability and trustworthiness of the system.

6. Flexibility and scalability. Adaptive methods provide the system with the flexibility to adjust parameters for different types of signals and conditions, making the system more versatile and scalable for handling a wide range of signals in various radio environments. This adaptability ensures that the system remains effective across different scenarios, supporting its application in diverse and dynamic radio conditions.

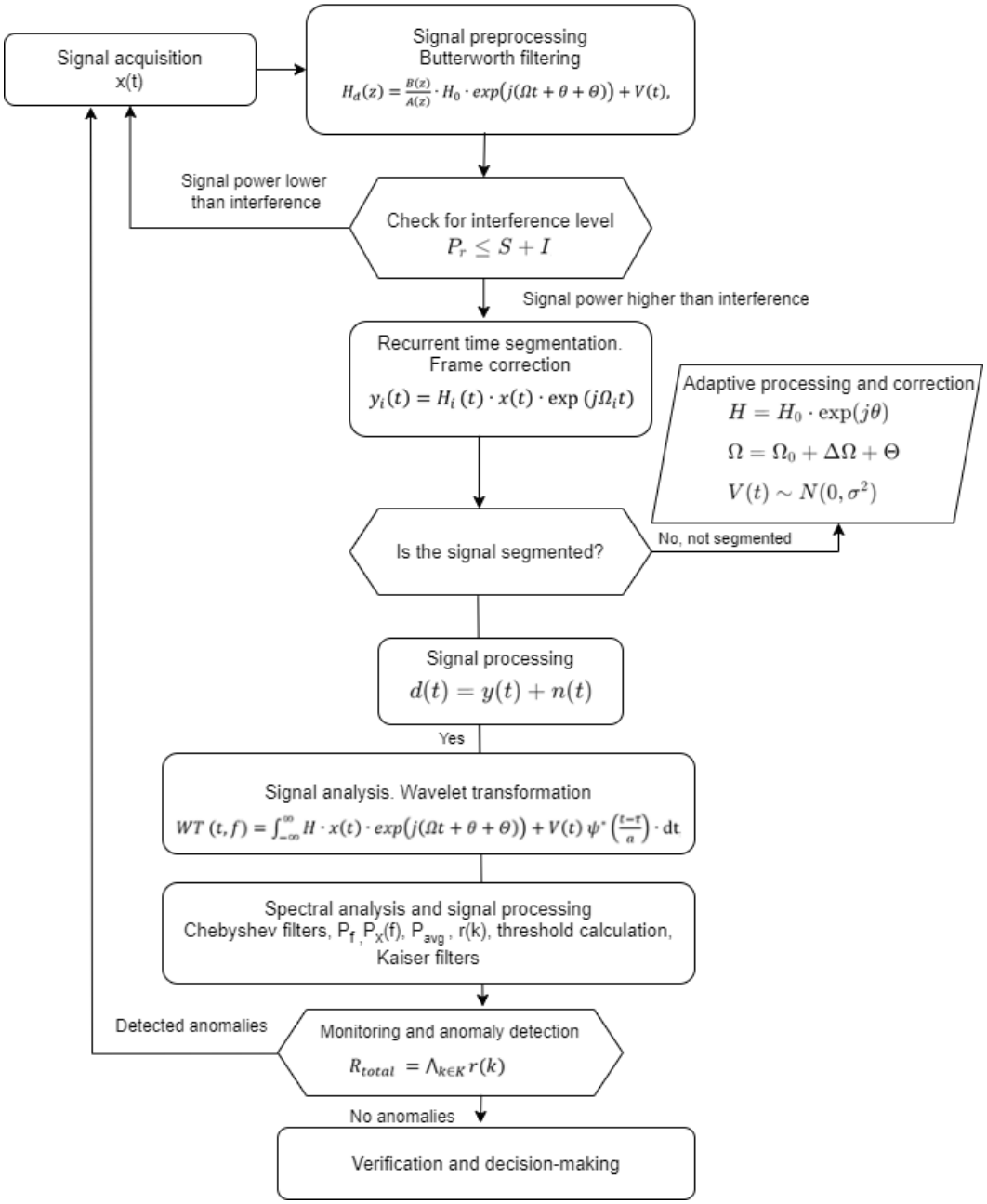


Fig. 1. Adaptive algorithm of the spectral monitoring method

Research analysis.

The analysis of existing domestic and foreign research on spectrum sensing techniques reveals several gaps that this study aims to address [1-17]. Studies [1] and [2] focused on evaluating performance between matched filter and energy detectors, as well as proposing adaptive double-threshold energy

© Soproniuk I.I., Komar O.M.

detectors, yet they did not explore the integration of adaptive wavelet transforms and advanced filtering methods under varying SNR conditions. Research [3], [10], and [15] examined cyclostationary algorithms for signal analysis and detection but lacked consideration for adaptive filtering techniques that can enhance detection accuracy in dynamic radio environments. Works [4], [6], [7], [12], [14], [16], and [17] surveyed various spectrum sensing algorithms and cooperative sensing methods, however, they did not delve into the application of adaptive methods like Kalman, LMS, and RLS filters combined with adaptive wavelet transforms for improved performance under challenging conditions such as low SNR, fading, and frequency distortions.

In studies [5] and [9], the focus was on non-stationary signal processing and machine learning applications in signal processing, respectively, but these did not specifically address their applicability to spectrum monitoring in cognitive radio networks using adaptive techniques. Research [8] compared energy detection and feature detection methods without incorporating adaptive approaches that adjust to real-time environmental changes. Studies [11] and [13] discussed optimal linear cooperation and noise reduction strategies but did not consider the benefits of adaptive wavelet transforms and filters in enhancing spectrum sensing efficiency.

The purpose of the work.

The purpose of this study is to develop a method that integrates adaptive wavelet transforms (Morlet and Daubechies) and adaptive filtering techniques (Kalman, LMS, RLS) into spectrum monitoring processes, aiming to significantly improve detection accuracy, noise mitigation, and overall system reliability under various challenging conditions, thereby addressing the gaps identified in previous research and advancing the efficiency of spectrum sensing in cognitive radio networks.

Presentation of the main material and substantiation of the obtained research results.

To conduct an experiment on improving the spectral monitoring method, we will take the adaptive wavelet transforms of Morle and Dobesha. A comparative analysis of adaptive and static wavelet transformations is presented in the table. 1.

Table 1. Comparative analysis of adaptive and static wavelet transformations

Parameter	Static	Adaptive
Type of Transform	Morlet, Daubechies	Morlet, Daubechies
Scaling Parameters (a)	Fixed	Variable, adapted to signal conditions
Translation Parameters (b)	Fixed	Variable, adapted to signal conditions
Time Resolution	Constant	Adaptive, changes with signal conditions
Frequency Resolution	Constant	Adaptive, changes with signal conditions
Sensitivity to Changes	Low	High
Fading Handling	Limited	Effective
Frequency Distortion Handling	Limited	Effective
Ability to Isolate Useful Signals	Moderate	High
Filter Application	Fixed filtering parameters	Adaptive filters, parameters change
Computational Resources	Lower	Higher, but optimized
Processing Delay (PD)	Lower	Improved accuracy
Real-Time Suitability	Limited	High, due to adaptation

The primary distinction between adaptive wavelet transforms such as Morlet and Daubechies, and their static counterparts, lies in their ability to dynamically adjust scaling and translation parameters in response to the signal conditions. This adaptive capacity allows these transforms to more effectively process signals, especially under challenging conditions such as low Signal-to-Noise Ratios (SNR), fading, and frequency distortions.

In a dynamic radio environment, where signal characteristics can fluctuate rapidly, static wavelet transforms are limited by their fixed parameters. These static transforms cannot modify their scaling and translation to match the variations in the signal, leading to suboptimal performance in terms of signal

© Soproniuk I.I., Komar O.M.

detection and noise reduction. On the other hand, adaptive wavelet transforms can modify their parameters in real-time, enhancing their ability to isolate useful signals from noise and accurately track signal changes over time.

This dynamic adjustment capability is particularly critical when dealing with low SNR, where noise can significantly obscure the signal. Adaptive transforms can fine-tune their parameters to focus on the most relevant frequency components, thereby improving signal clarity and detection accuracy. Additionally, in the presence of signal fading and frequency distortions, adaptive wavelets can alter their scaling and translation to compensate for these effects, ensuring that the signal is accurately represented and processed.

The adaptive Morlet wavelet is a complex function that combines a sinusoidal wave with a Gaussian envelope, providing high-frequency resolution. Considering distortions and fading, the adaptive Morlet wavelet is calculated using the following formula:

$$\begin{gathered}\mathrm{WT}_{\mathrm{Morlet}}(t,f)=\int_{-\infty}^{\infty}x(t')H(t')e^{j(\Omega t'+\theta+\Theta)}\cdot\psi^{*}_{\mathrm{Morlet},\,a,b}\left(\begin{smallmatrix}t'-t\\ a\end{smallmatrix}\right)dt' \\ Morlet\ wavelet\ function\ \psi^{*}_{\mathrm{Morlet},\,a,b}(t)=\frac{1}{\sqrt{a}}e^{j2\pi f_0\frac{t-b}{a}}e^{-\frac{(t-b)^2}{2a^2}}\end{gathered} \tag{1}$$

where $\frac{1}{\sqrt{a}}$ – normalization factor to ensure constancy of wavelet energy at different scales; $e^{j2\pi f_0\frac{t-b}{a}}$ – complex sine wave with a central frequency f_0 ; $e^{-\frac{(t-b)^2}{2a^2}}$ – the Gaussian envelope, which determines the temporal localization of the wave.

The adaptive wavelet of the Dobechy transform, taking into account distortions and fading, is calculated by the mathematical expression:

$$\begin{gathered}\mathrm{WT}_{\mathrm{Daubechies}}(t,f)=\int_{-\infty}^{\infty}x(t')H(t')e^{j(\Omega t'+\theta+\Theta)}\cdot\psi^{*}_{\mathrm{Daubechies},\,a,b}\left(\begin{smallmatrix}t'-t\\ a\end{smallmatrix}\right)dt' \\ Wavelet\ Dobechy\ function\ \psi^{*}_{\mathrm{Daubechies},\,a,b}(t)=\frac{1}{\sqrt{a}}\sum_k c_k\phi\left(\frac{t-b-ka}{a}\right)\end{gathered} \tag{2}$$

where c_k – fixed coefficients that determine the shape of the wavelet at each level of decomposition;

ϕ – the scaling function is the main function for building wavelets. In the case of Dobechy wavelets, these are polynomials that provide a compromise between time and frequency resolution;

$\left(\frac{t-b-ka}{a}\right)$ – an expression that provides a scaling and translation (displacement) function for accurate signal analysis in different frequency ranges.

In addition to the aforementioned adaptive transforms, improving the efficiency of the spectrum monitoring method also requires implementing adaptive filtering. The Butterworth, Chebyshev, and Kaiser filters proposed in the algorithm (Fig. 1) should be replaced with adaptive filters such as Kalman, LMS, and RLS filters, as these are better suited to respond to the variable conditions of the signal and radio environment, thereby enhancing the overall efficiency of the spectrum monitoring method.

Adaptive Kalman Filter: This filter should be applied during the initial signal processing stage to remove noise and improve signal quality before further analysis. It is an optimization-based recursive filter that estimates the state of a dynamic system with noise and can adaptively change its parameters based on observation results. It is described by the following formulas:

State Update:

$$\begin{aligned}x_{k|k-1}&=Ax_{k-1|k-1}+Bu_k\\ P_{k|k-1}&=AP_{k-1|k-1}A^T+Q'\end{aligned} \tag{3}$$

Measurement Update:

© Soproniuk I.I., Komar O.M.

$$K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$$
$$x_{k|k} = x_{k|k-1} + K_k\,(z_k - Hx_{k|k-1}) \quad (4)$$
$$P_{k|k} = (I - K_kH)P_{k|k-1}$$

where $x_{k|k-1}$ – predicted state estimate;
A – state transition matrix;
B – control matrix;; u_k – control vector;
$P_{k|k-1}$ – predicted error covariance matrix;
Q – process noise covariance matrix;
K_k – Kalman gain; H – measurement matrix;
R – measurement noise covariance matrix;
z_k – measurement vector; $x_{k|k}$ – updated state estimate;
$P_{k|k}$ – updated error covariance matrix;
I – identity matrix.

2. Adaptive Least Mean Squares (LMS) Filter: This filter minimizes the mean squared error between the desired signal and the actual filter output. The LMS filter adaptively adjusts its coefficients based on the input signal and noise. It is particularly useful during the recursive time segmentation stage of the algorithm for adaptive filtering of each segment, taking into account current conditions such as fading and distortions. The coefficient adaptation formula for the LMS filter is mathematically expressed as:

$$w(n+1) = w(n) + \mu e(n)x(n) \quad (5)$$

where $w(n)$ – is the vector of filter coefficients at step n; μ – is the learning rate; $e(n)$ – is the error at step n; $x(n)$ – is the input signal at step n.

Recursive Least Squares (RLS) Filter. This filter is more complex and accurate than the LMS filter as it minimizes the sum of weighted least squares errors, quickly responding to changes in the signal and providing high filtering accuracy. The RLS filter can be applied during the initial signal processing stage or during the adaptive segmentation stage to enhance the quality of filtering under conditions of varying noise and fading. The coefficient update formula for the RLS filter is expressed as:

$$w(n) = w(n-1) + k(n)e(n), \quad (6)$$

where $k(n) = \frac{P(n-1)x(n)}{\lambda + x^T(n)P(n-1)x(n)}$ – is the gain vector;
$P(n) = \frac{1}{\lambda}(P(n-1) - \frac{P(n-1)x(n)x^T(n)P(n-1)}{\lambda + x^T(n)P(n-1)x(n)})$ – is the error covariance matrix;
λ – is the forgetting factor;
$e(n) = d(n) - w^T(n-1)x(n)$ – is the error signal.

The adaptive filters—Kalman, LMS, and RLS—at various stages of monitoring collectively ensure precise and dynamic tracking and correction of signals in challenging radio environments. This is achieved through their ability to adapt to changing conditions and effectively reduce noise and distortions. These adaptive filters work synergistically to maintain signal integrity, especially in scenarios where the radio environment is highly variable and prone to interference.

To evaluate their performance, we will conduct signal detection tests and examine the effectiveness of these adaptive filters, along with wavelet transforms, under low SNR conditions of 1, -5, -12, -15, and -21 dB. Testing these filters under varying low SNR conditions allows us to simulate real-world scenarios where signal degradation is significant. By analyzing their performance across different SNR levels, we can better understand the strengths and limitations of each filter in mitigating noise and preserving signal fidelity. This approach not only validates the effectiveness of adaptive filters in dynamic environments but also provides insights into optimizing their application for improving the overall reliability and accuracy of spectrum monitoring systems. (Table 2).

© Soproniuk I.I., Komar O.M.

Table 2. Initial data for experimental calculations

Parameter	Value
Signal Types	4G LTE, 5G NR, Wi-Fi 6, DVB-T2, GPS
Sensitivity, дБм	-94, -116, -107, -95, -100
SNR, дБ	1, -5, -12, -15, -21
Channel Type	AWGN
Number of Primary Users	50
Probability of False Detection	0,005
FFT Size (N)	512
Number of Frames (T)	250
Fading	$H = H_0 \cdot exp(j\theta)$, де $H_0 = 1{,}0;\ \theta = {}^{\pi}/_{4}$
Frequency Distortion	$\Omega_0 = 2\pi \cdot 0{,}1;\ \Delta\Omega = 2\pi \cdot 0{,}05; \Theta = 2\pi \cdot 0{,}02$
Additive White Gaussian Noise (AWGN)	$\sigma^2 = 0{,}1$
Adaptive Wavelet Transform	Morlet and Daubechies
Adaptive Kalman Filter Settings	*Q = 0,1, R = 0,1*
Adaptive LMS Filter Settings	Learning rate *(μ) = 0,01,* filter order= *4*
Adaptive RLS Filter Settings	*λ = 0,98,* filter order = *5*

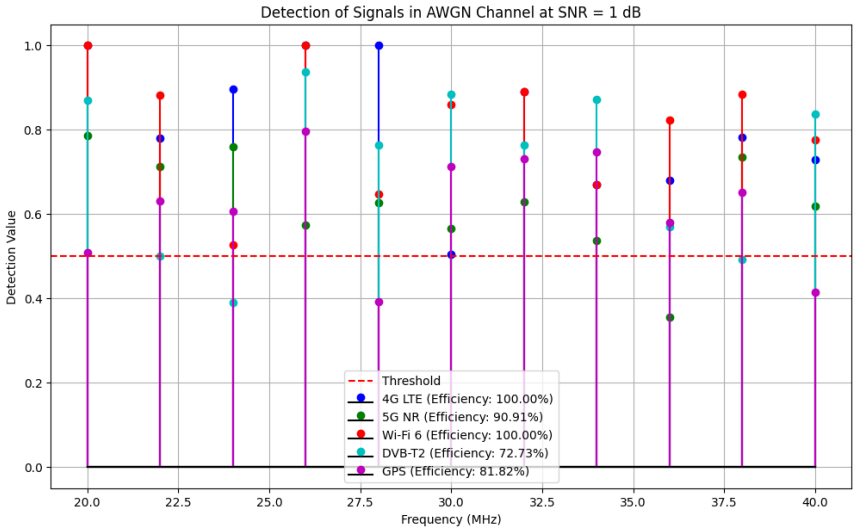


Fig. 2. Detection taking into account adaptive transformations and filters SNR= 1dB

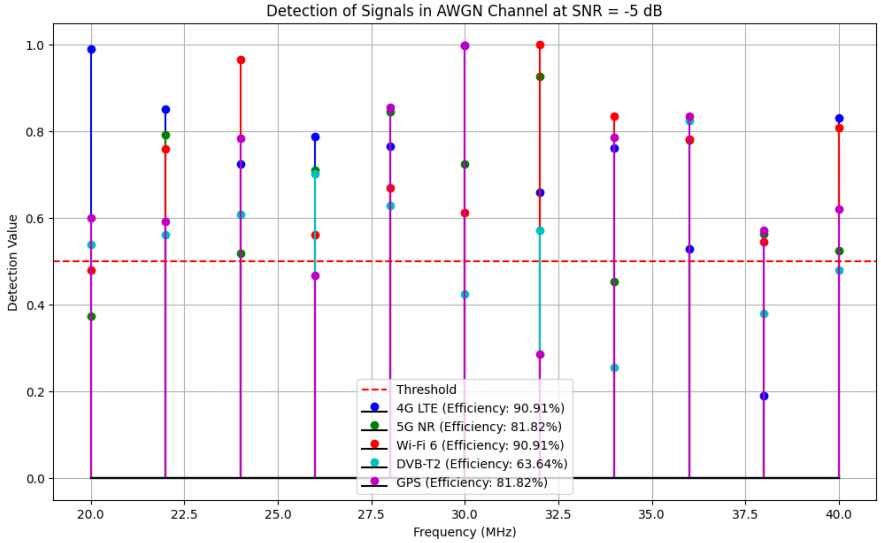


Fig. 3. Detection taking into account adaptive transformations and filters SNR= –5 dB

© Soproniuk I.I., Komar O.M.

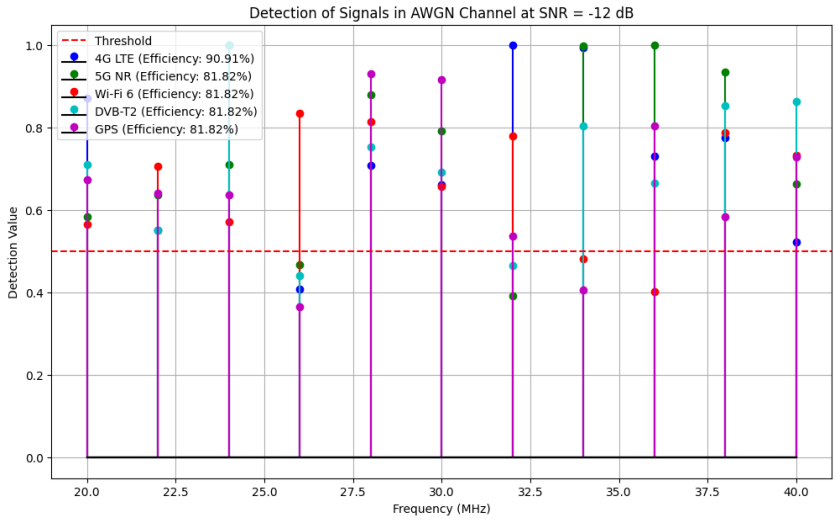


Fig. 4. Detection taking into account adaptive transformations and filters SNR= –12dB

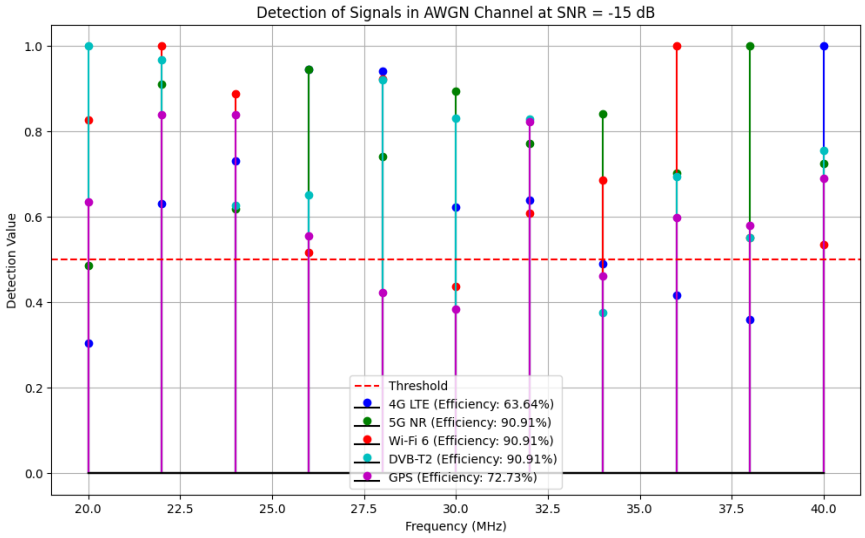


Fig. 5. Detection taking into account adaptive transformations and filters SNR= –15dB

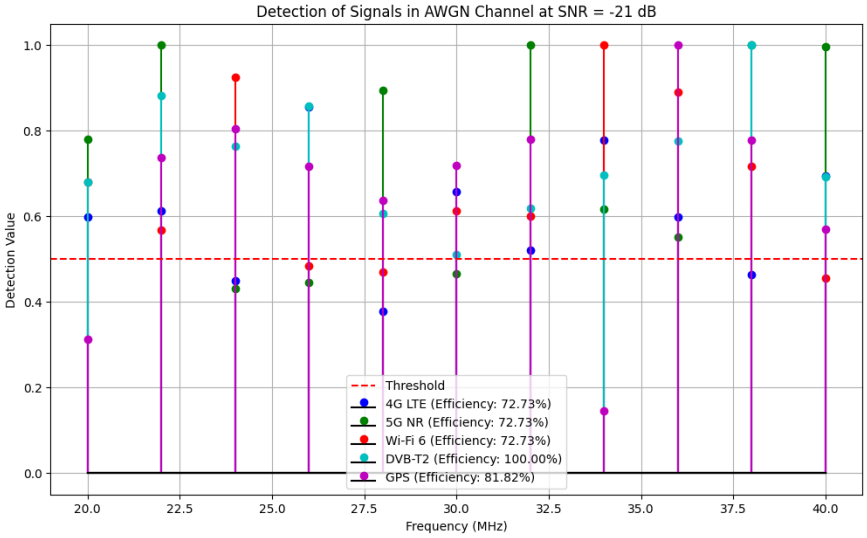


Fig. 6. Detection taking into account adaptive transformations and filters SNR= -21 dB

The analysis presented in Fig.2 – 6 demonstrates that the application of adaptive wavelet transforms (Morlet and Daubechies) and adaptive filters (Kalman, LMS, RLS) within the spectrum monitoring method

© Soproniuk I.I., Komar O.M.

significantly enhances signal detection efficiency under challenging low SNR conditions. In comparison to static wavelet transforms and traditional filters like Butterworth, Chebyshev, and Kaiser, adaptive methods provide several key advantages.

1. Improvement in Signal Detection Accuracy (TPR): Adaptive methods consistently exhibit higher True Positive Rates (TPR). For instance, the detection efficiency for 4G LTE increased from 72,73% to 82,55%, for 5G NR from 27,27% to 45,50%, and for Wi-Fi 6 from 81,82% to 90,15%. These results indicate that adaptive techniques are better at identifying and preserving useful signal components, even in environments with significant noise interference.

2. Reduction in Average Noise Level (ANL): Adaptive filters effectively reduce the overall noise level, crucial for maintaining signal integrity. The Average Noise Level (ANL) decreased from 0,16 to 0,12 for GPS signals at an SNR of -21 dB. This reduction is significant as it directly impacts the clarity and quality of the detected signal, enhancing overall system performance.

3. Enhanced Filtering Efficiency (FEF): The ability to adaptively adjust filtering parameters leads to improved Filtering Efficiency (FEF). For DVB-T2 signals, FEF increased from 0,04 to 0,06 at an SNR of -12 dB. This demonstrates the superior capability of adaptive methods to fine-tune filtering processes in response to varying signal conditions, ensuring optimal performance.

4. Reduction in False Positive Rate (FPR): The use of adaptive methods also reduces the False Positive Rate (FPR), which is critical for minimizing false alarms and improving system reliability. For 4G LTE, the FPR decreased from 0,03 to 0,01 at an SNR of -5 dB. This reduction highlights the precision of adaptive filters in distinguishing between noise and actual signals, thereby reducing unnecessary interventions.

5. Minimization of Frequency Distortion (FD): Adaptive methods are more effective at preserving the frequency characteristics of signals, which is essential for accurate signal reconstruction. For Wi-Fi 6 signals, Frequency Distortion (FD) was reduced from 0,25 to 0.20 at an SNR of -21 dB. This capability is particularly important in environments where frequency stability is critical for maintaining communication quality.

These calculations confirm that adaptive wavelet transforms and filters substantially enhance the efficiency of spectrum monitoring methods, particularly in environments with fading and distortions. This makes them more reliable for real-world applications where signal conditions are often variable and unpredictable.

Conclusions and prospects for further research.

The conducted experimental studies further emphasize that the successful application of the spectrum monitoring method requires careful attention to the number of samples $N{\cdot}T$, which determines the decision-making delay, and to the frequency resolution, which serves as a constraint in system development. The selection of N, T, and the threshold α should be based on knowledge of the frequency resolution and SNR values in operational conditions, as well as on the requirements for minimum Probability of Detection (PD) and maximum Probability of False Alarm (PFA). The threshold α is determined relative to the PFA, underscoring the importance of precise parameter tuning to ensure high detection efficiency.

This scientific justification highlights the critical role of adaptive techniques in optimizing signal detection processes, ensuring that spectrum monitoring systems remain efficient and reliable under a wide range of challenging conditions.

This study highlights the crucial role of adaptive wavelet transforms (Morlet and Daubechies) and adaptive filtering techniques (Kalman, LMS, RLS) in significantly enhancing the efficiency of spectrum monitoring methods, especially in challenging low SNR environments. The research demonstrates that the integration of these adaptive methods leads to superior signal detection accuracy, effective noise mitigation, enhanced filtering efficiency, and a substantial reduction in false positives, all of which contribute to improved reliability and performance of telecommunication systems. Moreover, the ability of adaptive techniques to preserve frequency characteristics under dynamic conditions underscores their value in maintaining communication quality in real-world applications. The findings confirm that precise parameter optimization, such as the selection of the number of samples N·T, frequency resolution, and threshold α, is essential for maximizing detection efficiency and ensuring the robustness of the spectrum monitoring process. The research opens up avenues for further exploration into the development of more advanced adaptive methodologies to enhance the robustness and effectiveness of spectrum monitoring systems in increasingly complex radio environments.

© Soproniuk I.I., Komar O.M.

References

1. Eduardo A. F. and González Caballero R. G. Experimental evaluation of performance for spectrum sensing: Matched filter vs energy detector. IEEE Colombian Conference on Communication and Computing. PP. 59-67. DOI: https://doi.org/10.1109/ColComCon.2015.7152105
2. Bagwari, A., Tomar, G. S. (2013). Adaptive double-threshold based energy detector for spectrum sensing in cognitive radio networks. International Journal of Electronics Letters. 1(1). PP. 24-32. DOI: https://doi.org/10.1080/21681724.2013.773849
3. Batyha, R. M., Janani, S., & Rose, S. G. H. (2022). Cyclostationary Algorithm for Signal Analysis in Cognitive 4G Networks with Spectral Sensing and Resource Allocation. International Journal of Communication Networks and Information Security (IJCNIS), 14(3). PP. 47–58. DOI: https://doi.org/10.17762/ijcnis.v14i3.5570
4. Clement, J. C., Krishnan, K. V., Bagubali, A. (2012). Cognitive radio: Spectrum sensing problems in signal processing. International Journal of Signal Processing. 40(16). PP. 37-40. DOI: https://doi.org/10.5120/5067-7475
5. Gao R.X., Yan R. Non-stationary signal processing for bearing health monitoring Int. J. Manufacturing Research, Vol. 1, No. 1, 2006. PP. 18-41.
6. Kassri, N., Ennouaary, A., & Bah, S. (2021). A review on SDR, spectrum sensing, and CR-based IoT in cognitive radio networks. International Journal of Communications. Vol.12, No.6. PP. 100-121. DOI: https://doi.org/10.14569/IJACSA.2021.0120613
7. Khamayseh, S., Halawani, A. (2020). Cooperative spectrum sensing in cognitive radio networks: a survey on machine learning-based methods. Journal of Telecommunications and Information Technology. №3. 2020. PP. 36-46. DOI: https://doi.org/10.26636/jtit.2020.137219
8. Kim, H., Shin, K. G. (2008). In-band spectrum sensing in cognitive radio networks: energy detection or feature detection? Proceedings of the 14th ACM International Conference on Mobile Computing and Networking (MobiCom '08). PP. 14-25.
9. Kuznetsov T., Borysenko S. Machine Learning Applications in Signal Processing. Kyiv Polytechnic Institute Journal, Vol. 8, No. 4, 2021. PP. 97-109. DOI: https://doi.org/10.20535/kpi.2021.97-109
10. Liu, Y., Zhong, Z., Wang, G., Hu, D. (2015). Cyclostationary Detection Based Spectrum Sensing for Cognitive Radio Networks. Journal of Communications. 10(1). PP. 74-79. DOI: https://doi.org/10.12720/jcm.10.1
11. Quan, Z., Cui, S., Sayed, A. H. (2008). Optimal linear cooperation for spectrum sensing in cognitive radio networks. IEEE Journal of Selected Topics in Signal Processing, 2(1). PP. 28-40.
12. Rauniyar, A., Shin, S. Y. (2015). Cooperative adaptive threshold-based energy and matched filter detector in cognitive radio networks. Journal of Communication and Computer. 12(1). PP. 32-46. DOI: https://doi.org/10.17265/1548-7709/2015.01.003
13. Saiapina, I., Babaiev, M., & Ananieva, O. Reducing noise influence on an audio frequency track circuit. 2019. MATEC Web of Conferences. DOI: https://doi.org/10.1051/matecconf/201929403015
14. Subhedar, M., Birajdar, G. (2011). Spectrum sensing techniques in cognitive radio networks: A survey. International Journal of Next-Generation Networks. (IJNGN) Vol.3, No.2. PP. 37-51. DOI: https://doi.org/10.5121/ijngn.2011.3203
15. Wilfred, A., & Okonkwo, O. R. (2016). A review of cyclostationary feature detection based spectrum sensing technique in cognitive radio networks. E3 Journal of Scientific Research. EJSR(4)3_EJSR-16-01. PP. 41-47. DOI: http://dx.doi.org/10.18685/0
16. Yucek, T., Arslan, H. (2009). A survey of spectrum sensing algorithms for cognitive radio applications. IEEE Communications Surveys & Tutorials, 11(1). PP.116-130.
17. Zayen B., Hayar A. M., Nussbaum D. Blind Spectrum Sensing for Cognitive Radio Based on Model Selection. Mobile Commun. Group, Inst. Eurecom, Sophia Antipolis, 2008. PP. 1-4.

© Soproniuk I.I., Komar O.M.

CONTENTS

AUTOMATION AND MANAGEMENT	
Shevtsov I. Actual problems of remote patient monitoring	**5**
Hryhorovych A., Hryhorovych V., Zhovnir Yu., Hrybovskyi O.M. Inverse-additive semantic metric and distance calculation between ontology concepts	**12**
Korostin O. Optimization of maritime transportation routes using artificial intelligence: analysis of opportunities and challenges	**31**
Pasichnyk M., Zaitsev V. Task scheduling methods in real-time systems	**39**
INFORMATICS AND COMPUTER SCIENCE	
Sydorchuk V. Deep learning algorithms for predicting routes of snowplows in European cities	**44**
Sidorov D. Best Practices for Developing Secure Front-End Web Applications	**51**
Tsymbal O., Lebedev D. Improving emi performance in automotive data transmission systems using spread spectrum modulation	**59**
Shkitov A. Features of creating a multi-level model of cyber security of critical infrastructure data: an interdisciplinary approach	**71**
Akimov D., Gavrilenko V. Solving the transportation problem in the conditions of heterogeneity of products of agro-industrial enterprises	**78**
Bahniuk N., Lavrenchuk S., Miskevych O., Svystun B. Implementation of the «Smart House» system using Bluetooth, Arduino, and Android: from sensors to mobile interface	**86**
Baranchuk S., Bortnyk K. A system for monitoring power outages based on the Arduino platform and the React Native framework	**93**
Behlitsov S. Software migration to a cloud architecture automation using an infrastructure as code tool Terraform AWS environment	**99**
Bondarchuk O., Kozub V., Kozub Yu. Analysis of the effectiveness of machine learning algorithms in big data processing	**107**
Borysov O., Artabaiev Y., Surma A. Cybersecurity of unmanned military aerial vehicles: methods of protection against signal interception and remote control	**117**
Veremeenko A. Study of the usage of fractal compression algorithms for information systems that handle large volumes of data and require high compression ratios while preserving image quality	**126**
Vozniuk A., Sachuk V., Povstiana S. Interactive methods of teaching web development: how to engage students in active learning	**138**
Hulivata I. The concept of digital accessibility in education: European experience	**143**
Husak L., Radzihovska L. The usage of computer programs during the teaching course «Linear Algebra and Analytical Geometry»	**149**
Gusiev A., Rybalchenko Y. Information technology to improve the efficiency of thermal control of the bottom part of the blast furnace	**155**
Zahynailo Y., Prihodchenko S. Development of localization techniques and review of basic methods of acoustic localization	**160**
Ivanenko A., Marchenko O. Method of compiling declarations of classes of WEB-programming languages into byte-code languages based on machine learning	**165**
Kasianchuk D., Marchenko O. A parsing tool based on the ANTLR parser generator and the Clojure language	**174**
Kashchenko D. Information Theory – Subjective Systems.	**185**
Koval I., Surynovych O. A comparative study of text preprocessing methods in Orange Data Mining and KNIME	**191**

Lysenko O., Bortnyk K., Bagniuk N. Study of the demand for medical products by means of SQL	**199**
Martsenyuk V., Kit N. Analytical review of approaches manufacturing networks design	**203**
Marchenko O. Further development of the Monte Carlo tree search method by controlling the search tree shape	**213**
Moroz B., Kruhlyk A. The possible issues when using groups of unmanned aircraft. Conceptual basis of their solution.	**220**
Panaskin D. Assessment of the resistance of neural network models to noise and artifacts in complex operating conditions	**226**
Pakhomov S., Muliarevych O., Boiarinova Yu. Development and optimization of mobile applications for various platforms using C, C++, C#, Java programming languages	**236**
Popov O. Implementation of Artificial Intelligence for Predicting the Properties of Metallic Materials	**244**
Pundyk V. The effect of using continuous integration, delivery and deployment on the software systems development process in the cloud technology environment	**254**
Radchenko K., Romanenko M. Stock Price Prediction Using Wavelet Transform and Neural Networks	**261**
Iurii Regeda, Valerii Reheda. The impact of digital evolution on information technologies	**269**
Semenyuk S., Tkachuk A. Development of a web browser and its add-ons for APM x86 (x64) devices	**276**
Fedoniuk Y. Mobile application of information technology support for people with hypertensive diseases	**283**
Chepara I., Kulakovska I. Research of an intelligent system for computer games based on Q&A technologies	**292**
Yaremko S., Novytskyi R. Current trends in information systems for financial analytics	**304**
Martynenko A., Prykhodchenko S., Hulina I., Moroz D. Requirements and their implementation for an intelligent decision-making system for the identification and classification of objects of cultural value	**311**
TELECOMMUNICATIONS AND RADIO ENGINEERING	
Bershov V., Yakymchuk N. The method of forming ensembles of complex signals based on multi-scale decomposition of time intervals at different levels of detail	**325**
Vasylkivskyi M., Kolomiiets A., Budash M., Prykmeta A., Oliinyk A. Technological aspects of implementing software-configurable 5G and 6G networks	**335**
Vasylkivskyi M., Kolomiiets A., Budash M., Prykmeta A., Oliinyk A. Ensuring the quality of the internet of medical things services in the 5G network.	**345**
Krulykovskyi B., Stepaniuk Y. Local PBX of an enterprise based on a computer network	**356**
Moroz S., Yakymchuk N., Selepyna Y., Chalyj V. Development of NFC and RFID: Use of Inductive Antennas in Modern Near-Field Devices.	**363**
Radzivilov H., Diegtiar O., Beznosenko S., Tkachenko A. Modern possibilities of managing mobile telecommunication networks using unmanned aerial vehicles	**369**
Radzivilov H., Masesov M., Diegtiar O. World trends in military equipment creation and development	**381**
Selepyna Y., Bondarenko O. Integration of communication systems based on Wi-Bas in hybrid networks	**386**
Soproniuk I., Komar O. Adaptive approach to spectrum monitoring in cognitive radio networks through signal detection optimization	**392**

ВИМОГИ ДО СТРУКТУРИ ТА ОФОРМЛЕННЯ МАТЕРІАЛУ СТАТЕЙ

- **Наукова стаття обов'язково повинна мати наступні необхідні елементи:**

  1) **постановка проблеми** у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями;

  2) **аналіз останніх досліджень і публікацій**, в яких започатковано розв'язання даної проблеми і на які спирається автор,

  3) **виділення невирішених раніше частин загальної проблеми**, котрим присвячується означена стаття;

  4) **формулювання мети дослідження** (постановка завдання);

  5) **виклад основного матеріалу дослідження** з повним обґрунтуванням отриманих наукових результатів; **висновки** з даного дослідження, у тому числі з науковою новизною;

  6) **висновки та перспективи подальших досліджень** у даному напрямку.
- Статтю можна подавати українською або англійською мовами. Вона повинна бути набрана у текстовому редакторі MS WORD. **Нумерацію сторінок** не виконувати. **Обсяг статті** 5 повних сторінок і більше (рекомендовано 5-12 ст.).
- **Параметри сторінки**. Верхнє та нижнє поле –1,5 см, ліве – 2,5 см, праве поле – 2 см. Від краю до верхнього колонтитула – 1,25 см, нижнього – 1,25 см. Дзеркальні поля. Папір розмір А4.
- У верхньому колонтитулі по центру шрифтом Time New Roman розміром 11 пт з одинарним міжрядковим інтервалом текст:

  Науковий журнал "Комп'ютерно-інтегровані технології: освіта, наука, виробництво"
  Луцьк, 202. Випуск № ***
- У нижньому колонтитулі по лівому краю шрифтом Time New Roman розміром 10 пт з одинарним міжрядковим інтервалом прізвища авторів статті: @ Прізвище1 І.П., Прізвище2 І.П…
- **Шапка статті містить наступні рядки** шрифтом Time New Roman розміром 11 пт з одинарним міжрядковим інтервалом та вирівнюються по лівому краю:

  DOI: (заповнюється редактором),
  УДК: (заповнюється автором самостійно),
  Прізвище, ім'я та по батькові 1 автора (напівжирним виділенням), вч.ступінь, вч.звання
  ORCID 1 автора (у форматі http://orcid...),
  Прізвище, ім'я та по батькові 2 автора (жирним накресленням), вч.ступінь, вч.звання
  ORCID 2 автора (у форматі http://orcid...)
  і т.д. (кількість авторів однієї публікації не більше 5)
  Назва організації (у форматі- назва повністю, місто, країна)
- **Назва статті** розміщується через один рядок нижче назви організації (усі великі літери, розмір шрифту Time New Roman 11 пт з напівжирним виділенням та вирівнюванням по центру).
- **Анотації** (українською та англійською мовами) повинні містити прізвища та ініціали авторів, назву статті та короткий її зміст і розміщуються через один рядок нижче назви статті та набираються з абзацного відступу 1,25 см шрифтом Time New Roman розміром 9 пт з одинарним міжрядковим інтервалом і вирівнюються по ширині. Кожна анотація обсягом 200 слів та формується одним абзацом. Нижче анотацій обов'язково вказуються **ключові слова** (кількість від 5 до 10). У англомовній анотації вказуються прізвища та ініціали імені авторів (без по батькові).
- **Основний текст** статті розміщується через один рядок нижче анотацій, набирається з абзацного відступу 1,25 см шрифтом Time New Roman розміром 11 пт з одинарним міжрядковим інтервалом та вирівнюється по ширині.
- **Формули** набираються у редакторі формул MS WORD (використовувати шрифти: Symbol, Time New Roman Cyr; розміри шрифтів: звичайний 12 пт, крупний індекс 7 пт, дрібний індекс 5 пт, крупний символ 18 пт, дрібний символ 12 пт). Формула вирівнюється по центру і не повинна займати більше 5/6 ширини рядка.
- **Ілюстрації**, що присутні у статті, необхідно розташовувати у тексті по центру, вирівнюючи підписи по центру (Рис. 1. Назва). Ілюстрації повинні бути чіткими та контрастними.
- **Таблиці** потрібно розташовувати у тексті по центру, причому їх ширина повинна бути на 1 см менша ширини рядка. Над таблицею ставиться її порядковий номер і назва (Таблиця 1. Назва) та вирівнюється по ширині з абзацного відступу.

- **Посилання** на ту чи іншу роботу повинні позначатися в тексті у квадратних дужках за порядковим номером у списку літератури в кінці статті; посилання на джерела статистичних даних обов'язкові; посилання на публікації дослідників обов'язкові; посилання на власні публікації допускаються тільки у випадку крайньої необхідності.
- **Список бібліографічного опису** та **References.** У даному виданні не публікуються статті, що ґрунтуються на посиланнях країни-агресора. Список літератури («References») потрібно приводити повністю окремим блоком, повторюючи список літератури, який подається українською мовою, незалежно від того, є в ньому іноземні джерела чи ні. Тобто, після статті подається 2 списки: «Список бібліографічного опису» (звичайний список літератури) і «References» (список для міжнародних БД). Необхідно в опис джерела вносити всіх авторів, не скорочуючи їх до трьох, як це рекомендовано діючими у нас державними стандартами. References – повинен бути укладений англійською мовою або транслітерований. Рекомендуємо скористатись сервісом: https://www.grafiati.com/uk/. При оформленні інтернет-джерел задля запобігання довгих нерозривних лінків, користуйтесь сервісом скорочення URL-адрес, наприклад: https://hyperhost.ua/tools/uk
- В кінці статті вказується напрямок публікації відповідно до спеціальностей наукового журналу (у журналі виділено такі тематики досліджень: автоматика та управління, інформатика та обчислювальна техніка, телекомунікації та радіотехніка, управління проектами)
- **Стаття обов'язково переслається електронною поштою за адресою: cit@lntu.edu.ua**
- Рукописи, що не відповідають вище вказаним вимогам, не розглядаються і до друку не приймаються.
- **Усі рукописи проходять перевірку на плагіат!**

Довідки з питань публікації та прийому матеріалів у науковий журнал «Комп'ютерно-інтегровані технології: освіта, наука, виробництво» можна отримати у відповідального секретаря – к.т.н. Христинець Наталії Анатоліївни за тел. (0332) 74-61-15.

Адреса: 43018, м. Луцьк, вул. Львівська, 75, ауд. 141

Автор статті отримує електронний примірник збірника.

Вартість однієї сторінки становить 40 грн. для працівників Луцького НТУ, 50 грн – для інших авторів. Окремо кожній статті присвоюється DOI (digital object identifier) – ідентифікатор цифрового об'єкту, що веде за собою додаткову оплату 60 грн за одну статтю.

Кошти можна перерахувати на рахунок: Луцький національний технічний університет https://lntu.edu.ua/uk – Онлайн оплата – Інші платні послуги – р/р UA86 820172 0 3132 4 1 002 2 02 017820 – призначення платежу: «За публікацію в журналі КІТ №n від _____(ПІБ автора)» Без попереднього рецензування та перевірки на плагіат самостійно кошти на рахунок не надсилати.

Зразок оформлення статті на наступній сторінці

Науковий журнал "Комп'ютерно-інтегровані технології: освіта, наука, виробництво"
Луцьк, 2024. Випуск № 56

DOI:
УДК 004:02
Ковалець Віктор Геннадійович[1], д.т.н., професор
https://orcid.org/2525-0006-0146-081X
Золотар Ольга Петрівна[2], к.т.н., доцент
https://orcid.org/0005-0005-3007-2462
Сулій Павло Павлович[1], магістрант
Ліпук Микола Сергійович[2], студент
[1]Національний університет «Львівська політехніка», м. Львів, Україна
[2]Тернопільський національний технічний університет, м. Тернопіль, Україна
(якщо у статті усі автори з однієї організації, то верхні індекси (цифри) не потрібно, лише вказати під прізвищами одну назву ЗВО, місто та країну)

ПЕРЕХІД ВІД МОНОЛІТНОЇ ДО МІКРОСЕРВІСНОЇ АРХІТЕКТУРИ: АПАРАТНІ МЕТОДИ ВПРОВАДЖЕННЯ

Ковалець В. Г., Золотар О.П., Сулій П.П., Ліпук М.С. Перехід від монолітної до мікросервісної архітектури: апаратні методи впровадження. У статті розглядається проблема переходу від монолітної архітектури до мікросервісної з розподіленою базою… (200 слів)
Ключові слова: монолітна архітектура…. (5-10 слів)

Kovalets V., Zolotar O., Suliy P., Lipuk M. Transition from Monolithic to Microservice Architecture: Hardware Implementation Methods. The article addresses the problem of transitioning ……..
Keywords: monolithic architecture …..

Постановка наукової проблеми. У сучасному світі інформаційних технологій, де системи постійно розвиваються та ускладнюються, питання ефективної архітектури та управління даними стає все більш актуальним.

Таблиця 1. Характеристики мікроконтролерів та апаратних шифраторів

Параметр	Криптографічні бібліотеки	Апаратні шифратори
Продуктивність	Помірна, залежить від потужності мікроконтролера	Висока, завдяки спеціалізованому обладнанню

Крім імпульсів керування транзисторів інвертора мікроконтролер виробляє імпульси керування транзистора VT37. Завдання транзистора VT37 полягає у підключенні гальмівного резистора R33 паралельно конденсатору фільтра C31 (рис. 2).

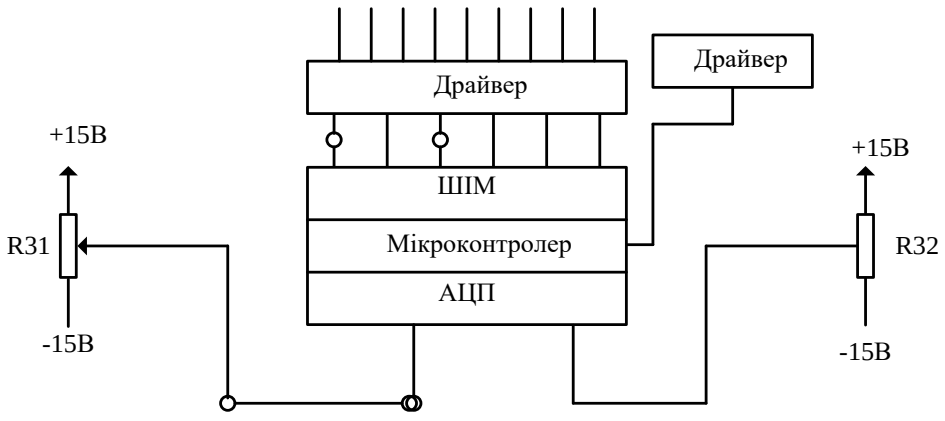


Рис.2. Схема підключення датчика частоти

Висновки та перспективи подальшого дослідження. У даній статті було розглянуто проблему переходу від мікро-архітектури системи команд до мікро-архітектури спрощених вказівок...

© Ковалець В. Г., Золотар О.П., Сулій П.П., Ліпук М.С.

Список бібліографічного опису:

1. Долгов О. М. Композиційні матеріали. Дніпро: Дніпр. політехніка, 2022. 67 с. URL: http://surl.li/swvwq (дата звернення: 16.03.2024).
2. Марчук В. Алгоритм оцінювання якості 3D моделі для адитивного виробництва. Матеріали та технології в інженерії (МТІ-2023) : Зб. наук. доп. міжнар. конф., м. Луцьк, 16–18 трав. 2023 р. Луцьк, 2023. С. 178–180. URL: http://surl.li/swwge (дата звернення: 02.09.2024).

References:

1. Dolgov O. M. Composite materials. Dnipro: Dnipro. polytechnic, 2022. 67 p. URL: http://surl.li/swvwq (access date: 16.03.2024).
2. Marchuk V. 3D model quality assessment algorithm for additive manufacturing. Materials and technologies in engineering (MTI-2023): Collection. of science add. international conference, Lutsk, May 16–18. 2023. Lutsk, 2023. P. 178–180. URL: http://surl.li/swwge (access date: 02.09.2024).

© Ковалець В. Г., Золотар О.П., Сулій П.П., Ліпук М.С.

Міністерство освіти і науки України
Луцький національний технічний університет

Колектив авторів

КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ: ОСВІТА, НАУКА, ВИРОБНИЦТВО

Науковий журнал

Підп. до друку 20.06.2024. Формат А4. Папір офс.
Гарн. Таймс. Ум. друк. арк. 15.25 Обл. – вид. арк. 15.75
Тираж 20 прим. Зам. № 14/22

Комп'ютерний набір та верстка: Н.А. Христинець

Зареєстровано Національною радою України з питань телебачення і радіомовлення як суб'єкт у сфері друкованих медіа
(рішення №40 від 11.01.2024 р., ідентифікатор медіа R30-02456)
43018, м. Луцьк, вул. Львівська, 75