

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА ТЕОРІЯ КОДУВАННЯ

УДК 004.03:620.97

B. S. Busygin, S. L. Nikulin, K. L. Sergieieva, O. V. Korobko

PRINCIPAL APPROACHES TO CREATING GEOINFORMATION SYSTEM OF RENEWABLE ENERGY SOURCES IN UKRAINE

Dnipro University of Technology, Dnipro

Анотація. Запропоновано підходи до створення геоінформаційної технології (ГІС-технології) й інструментів геоінформаційної системи виявлення й аналізу ділянок відновлюваних джерел енергії (ВДЕ) в Україні. Сформульовано основні складові ГІС-технології. Вони базуються на методах, моделях та технологіях комплексного аналізу статистичних та аерокосмічних даних з метою створення інструментів спеціалізованої геоінформаційної системи з урахуванням WEB-орієнтованого підходу. Представлені компоненти ГІС-технології призначені для розробки нових програмних й аналітичних інструментів на основі геоінформаційних систем та технологій, а також інших інструментів просторового моделювання в задачах оцінки перспективності територій України з точки зору доцільності розміщення об'єктів ВДЕ.

Ключові слова: геоінформаційна система, геоінформаційна технологія, відновлювальні джерела енергії, інтелектуальний аналіз даних.

Анотация. Предложены подходы к созданию геоинформационной технологии (ГИС-технологии) и инструментов геоинформационной системы выявления и анализа участков возобновляемых источников энергии (ВИЭ) в Украине. Сформулированы основные составляющие ГИС-технологии. Они базируются на методах, моделях и технологиях комплексного анализа статистических и аэрокосмических данных с целью создания инструментов специализированной геоинформационной системы с учетом WEB-ориентированного подхода. Представленные компоненты ГИС-технологии предназначены для разработки новых программных и аналитических инструментов на основе геоинформационных систем и технологий, а также прочих инструментов пространственного моделирования в задачах оценки перспективности территорий Украины с точки зрения целесообразности размещения объектов ВИЭ.

Ключевые слова: геоинформационная система, геоинформационная технология, возобновляемые источники энергии, интеллектуальный анализ данных.

Abstract. The approaches to geoinformation technology (GIS-technology) and geoinformation system tools creation are proposed for detecting and analyzing the renewable energy sources (RES) areas in Ukraine. The main components of GIS-technology are formulated. They are based on methods, models, and technologies of integrated analysis of statistical and aerospace data in order to create on this basis the tools of a specialized geoinformation system, taking into account the WEB-oriented approach. The presented components of GIS-technology are intended to create new software and analytical tools based on geoinformation systems and technologies, as well as other spatial modeling tools in the tasks of assessing the prospects of Ukrainian territories in terms of the feasibility of RES power facilities placing.

Key words: geoinformation system, geoinformation technology, renewable energy sources, data mining.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-4-11>.

Introduction

Recently, the demand for energy resources has increased significantly. It is accompanied by climate changes due to carbon combustion products' emission into the atmosphere, which requires strengthening of Ukraine's energy policy regarding renewable energy technologies (RES).

As of today, Ukraine has adopted a number of laws that stress the priority of alternative energy in the structure of the country's energy complex and aim at developing alternative energy. Among them are:

- The Law of Ukraine "On Alternative Energy Sources" No. 555-IV concerning the legal, economic, ecological and organizational basis for using alternative energy sources, and assisting expansion of their use in fuel and energy complex;
- The Law of Ukraine "On Alternative Fuels" No. 1391-XIV concerning the legal, social, economic, environmental, and organizational basis for production (extraction) and use of alternative fuels;
- The Law of Ukraine "On Energy Saving" No. 74/94-BP concerning the legal, economic, social and ecological basis for energy saving in enterprises, associations, and organizations located on the territory of Ukraine, as well as for its citizens;
- The Law of Ukraine "On Heat Supply" No. 2633-IV concerning stimulation of heat energy production from alternative energy sources;
- The Law of Ukraine "On Regulation of Urban Development" No. 3038-VI concerning improving the investment opportunities in electricity production from alternative sources.

The key aspect in assessing RES potential capacities is choosing the optimum power systems location. An important role is played by open statistical and spatial data obtained from geological, geophysical, meteorological, and aerospace monitoring, as well as by tools of open-source and commercial geoinformation systems and technologies (GIS-technologies) as a means to support making decisions based on a set of multilevel and heterogeneous data. The use of GIS-technologies will link together industry information resources on various Ukrainian administrative bodies and will contribute to successful implementation of the Energy Strategy of Ukraine until 2035, approved by the Cabinet of Ministers of Ukraine (No. 605-r of August 18, 2017).

Objective

The objective is to present the main approaches to geoinformation technology creation for estimating the potential of renewable energy sources in Ukraine.

Problem Statement

In 2018, Ukraine became a full member of the International Renewable Energy Agency – IRENA. According to IRENA (www.irena.org), the share of renewable energy sources in the total final energy consumption of Ukraine will rise from 3% in 2009 to 13.2%-21.8% in 2030, that would result in savings of USD 175 million per year in 2030 [1].

The annual renewable power generation is conditioned by a combination of factors including RES power plant geographic location (height above sea level, terrain slope, distance to the nearest settlements, land type, etc.), technical capabilities of electrical equipment, and economic potential defined as a total amount of solar and geothermal energy, wind energy, energy of biomass, small rivers, tides, and sea waves, etc. [2-6]. According to the State Agency on Energy Efficiency and Energy Saving of Ukraine (SAEE), Ukraine has a significant technically achievable potential for energy production from RES and alternative fuels, which is more than 98.0 million tons per year (Fig. 1).

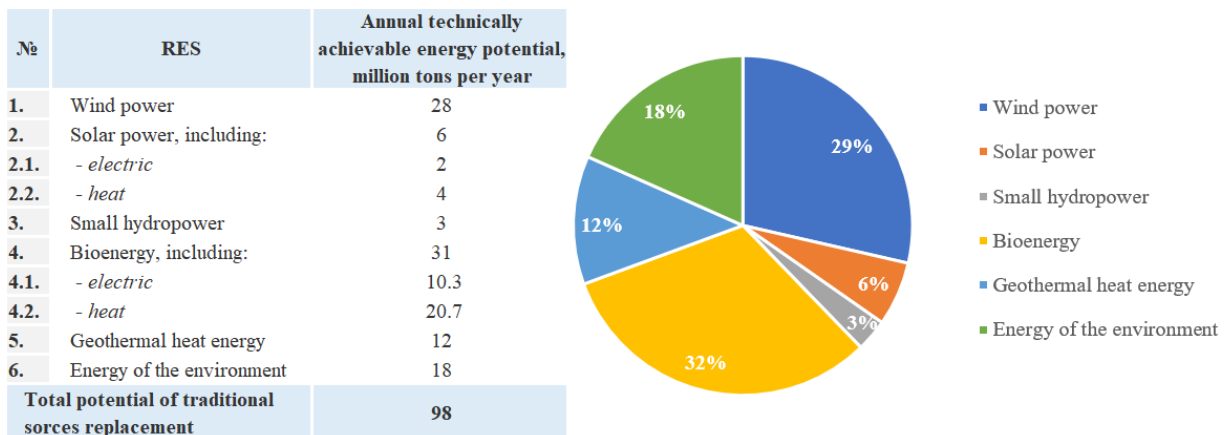


Figure 1 – Potential of renewable energy in Ukraine (saee.gov.ua)

Recent years have seen a rapid development of renewable energy in the world with an annual growth of tens of percent. The total global renewable electricity capacity in the world at the end of 2019 is more than 2.6 TW. The US and China occupy the leading positions with 1.1 TW of installed capacity, India holds third place with 137 GW of installed capacity [5]. By the beginning of 2020, the capacity of renewable energy power systems (not including hydropower) exceeded 1.4 TW. Since 2015, as many renewable power plants have been put into operation as traditional ones. During the recent 15 years, traditional power generation has developed at a rate of 2.0% per year, renewable energy generation has grown by 10-40% annually, while solar energy – by 50% per year. Investments in renewable energy sources abroad are continuously increasing (Fig. 2, Fig. 3) [1].

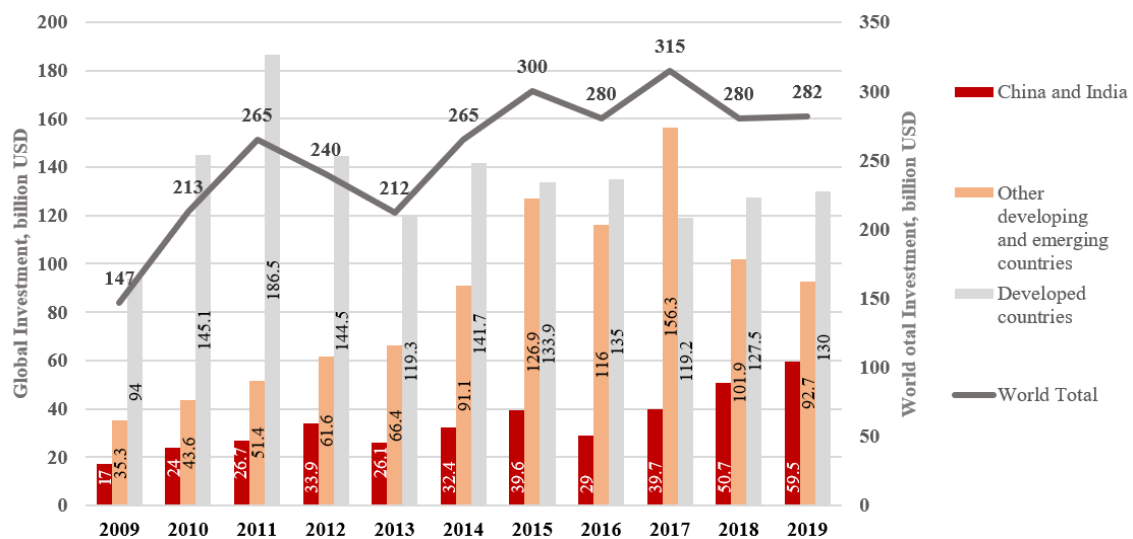


Figure 2 – Global investment in renewable power, billion USD [5]

Nowadays, annual investments in renewable energy sources reach \$ 282 billion. Solar energy is especially actively invested – more than \$ 131 billion per year, and wind energy – about \$ 138 billion [1, 5, 7].

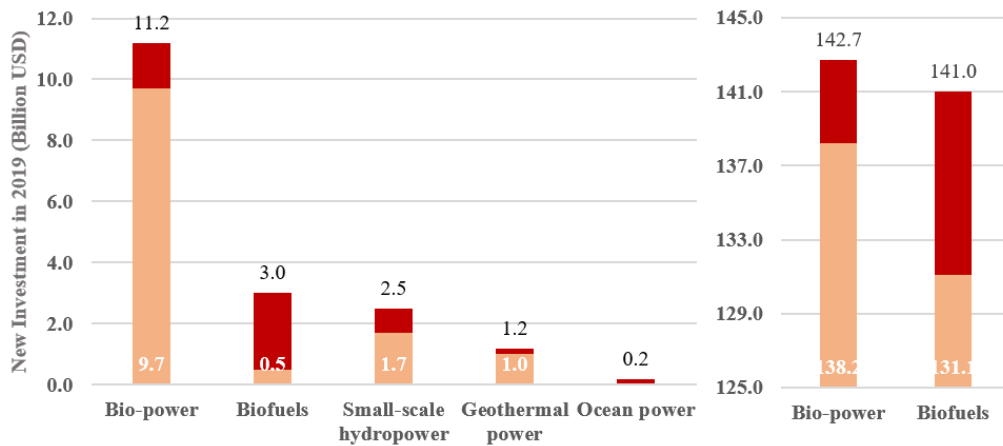


Figure 3 – Global Investment in Renewable Energy by Technology, 2019 [5]

Europe has the most ambitious plans to ensure up to 40%-contribution of RES to the total energy balance by 2040. At the same time, according to experts from the International Renewable Energy Agency, the share of renewable energy in the energy balance of Ukraine should be 13.2-21.8% by 2030 [1].

One of the reasons for accelerating the development of RES is the fact that many technologies of renewable energy generation have sharply fallen in price. At the moment, the development of photovoltaic solar energy converters has led to a reduction of the solar energy cost almost by a factor of 100 in comparison with 1980, and of wind energy cost – by a factor of 10-15.

Ukraine belongs to UNECE (United Nations Economic Commission for Europe) energy region covering 17 countries in South East and Eastern Europe, the Caucasus, and Central Asia [7]. Nowadays Ukraine has large RES potential in this region and is working on attracting foreign investments. According to SAEE, in 2019 about 3.7 billion euros invested in 4500 MW of renewable electricity in Ukraine.

According to UNECE, by the end of 2016, significant (97 MW) non-hydro renewable energy facilities were put into operation in Ukraine including solar power plants with a capacity of 62 MW. During 2018, installed solar power capacity increased by 646 MW [8]. Sizable wind power potential is presented in the UNECE region, with the largest resources in Kazakhstan and Ukraine.

Ukraine sees a rapid increase in the capacity of solar power plants (SPP) (Fig. 4).

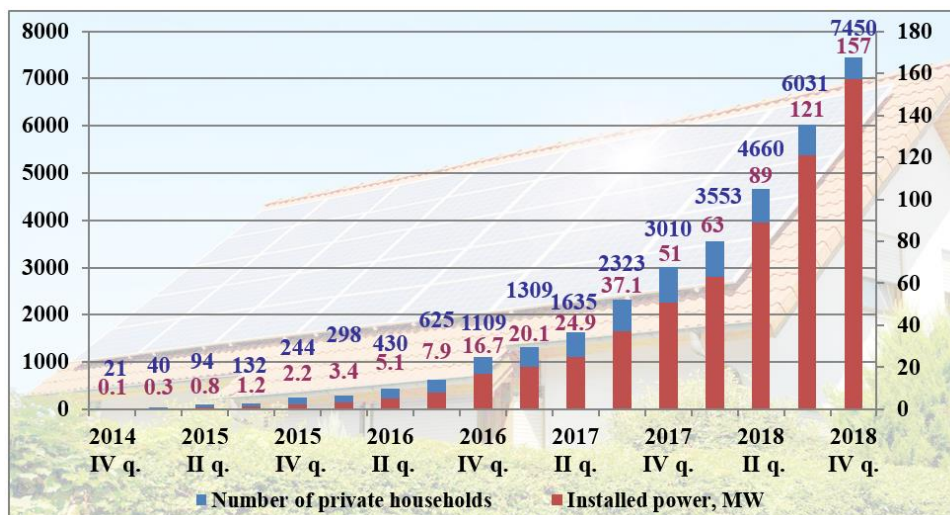


Figure 4 – SES of private households [8]

Several SPPs have been installed and put into operation, the largest of them being the Nikopol Solar Park (Dnipropetrovsk Oblast, capacity – 246 MW), Pokrovske Solar Park (Dnipropetrovsk Oblast, capacity – 240

MW), etc. (Fig. 5). The energy potential of deep heat resources in Ukraine is 20 times higher than the total energy potential of non-renewable fossil fuels such as oil, gas, coal, peat, timber, etc.



Nikopol Solar Park (DTEK), Dnipropetrovsk Oblast
Capacity – 246 MW
Consists of 750 000 operational units



Pokrovske Solar Park (DTEK), Dnipropetrovsk Oblast
Capacity – 240 MW
Consists of 874 000 operational units

Figure 5 – Solar power stations in Ukraine

According to IRENA, in 2030, the share of biomass energy in the Ukrainian RES structure will be 76% (820 PJ/year), sun and wind energy – 7% and 10% respectively. The share of hydropower energy will decrease to 4% by 2030 as compared to 32% in 2009 due to a significant increase in all other RES [1]. The heating sector will consume 73% of renewable energy, the transport sector – 7%, which significantly exceeds the figures from the SAEF forecast [1].

As stated in the *Energy Strategy of Ukraine for the period up to 2035*, by 2025, it is planned to start the Smart Grids implementation plan and create an extensive infrastructure for the development of electric transport. By 2035, RES will be developing dynamically, which will allow to increase their share in the total energy consumption up to 25% [2].

Components of Geoinformation Technology

Nowadays, foreign countries have accumulated successful experience of GIS-technologies use in renewable power generation [9].

The well-known foreign GIS in RES are:

1) Renewable Resources Map and Data, created in the US National RES Laboratory (www.nrel.gov), which is a geoinformation system at the national level. The Internet-version contains dynamic maps, databases, and tools allowing to analyze renewable energy sources in order to determine the most viable technologies for US conditions. The metadata files contain descriptions of data acquisition methods and calculations.

2) The Global Atlas for Renewable Energy created by IRENA (irena.org/globalatlas) provides open access to the datasets and their visualisations on wind and solar resources, as well as tools to find maps of renewable energy resources for locations across the world.

3) Atlas on Renewable Energy of Vermont (www.vtenergydashboard.org/energy-atlas) – regional GIS, developed based on ESRI ArcGIS. It includes detailed information on existing RE facilities in Vermont state (USA), as well as potential areas for building such objects. Data on the following sources are available: solar, wind, hydrothermal, small water streams, biological wastes. The information is displayed in every detail – at the level of individual buildings.

International database (DB) covering territories of various scales (NASA SSE, WRDC, SOLARGIS, METEONORM, etc.) that were developed in the last two decades can be considered as products similar to GIS. Ground-based measurement data, aerospace monitoring, and modeling results (models of general atmospheric circulation and distribution of solar radiation) serve as information basis for them. Several databases include maps of terrain and landscape types [9].

At present, the information WEB resource – UA MAP (www.uamap.org.ua) operates in Ukraine. It accumulates information on renewable energy projects and ensures communication between the initiators of such projects and investors in order to facilitate the attraction of investments in these areas. The UA MAP project was initiated by All-Ukrainian Investment and Sustainable Development Agency, with the support of SAEF, as well as the Danish Energy Agency and the Ukrainian-Danish Energy Center. This resource allows building maps of the existing RES projects' location but does not contain tools for assessing the potential of territories in terms of the possible placement of RES power plants [2].

Another Ukrainian WEB-map of renewable energy projects in Ukrainian Association of renewable energy (uare.com.ua).

Creating geoinformation technology for identifying, analyzing, and mapping prospective renewable energy sites is aimed at solving the problems of assessing the possibilities of RES effective use and can become a serious step in the development of this sector [8-11].

The main components of this technology are (Fig. 6):

1. Collecting and analyzing a complex of terrestrial and aerospace data, forming models for their representation in multidimensional post-relational databases. They should include, in particular, information of the following types: cartographic, thematic with spatial coordinates, attributive text documents, various graphics, reference materials, simulation results, computational and theoretical studies. Particular attention should be paid to the following types of energy sources: wind, solar, biomass, small water streams, hydrothermal, etc. Integrating a complex of spatial terrestrial and aerospace materials by forming the information basis of GIS in the form of thematic databases and knowledge bases [12].

2. Analyzing heterogeneous RES data on different levels. Developing criteria for the selection of sites for RES power plant installation. Identifying areas promising from the point of diverse RES presence taking into account data on existing RES plants and energy systems in a region.

3. Decision-making support based on spatial geological and meteorological data to select the right site for RES plant construction. Creating 2D and 3D maps of resource potential [13-15].

4. Developing software tools for GIS technology, equipped with modern scientific and technical tools of Big Data processing and Data Mining, considering the WEB-based approach. Conducting experimental studies and issuing recommendations on power equipment placement.

GIS technology will provide an opportunity for rapid processing of big heterogeneous data on natural resources and energy balance of the territory, which determines the demand in RES and their contribution to the energy complex. It will also provide data about energy infrastructure objects suitable for integrated use within RES facilities. GIS-technology uses tools of data mining and geoinformation data analysis. They will be used for solving tasks of perspective territories target searching and mapping by means of valid estimates and calculations.

To ensure the possibility of realizing RES geoinformation system in Ukraine, special attention should be given on the one hand, to studying the state of existing geoinformation products in the countries leading in this area, to identifying their basic features, advantages, and disadvantages, and to create on this basis a structure of GIS. On the other hand, the focus should be on mapping the known RES in Ukraine, on analyzing their structure, geographic location, utilization efficiency, economic feasibility, etc. [9].

The second component of research is concerned with the development of software and hardware architecture at all data processing stages, and in particular – the structure and composition of the database.

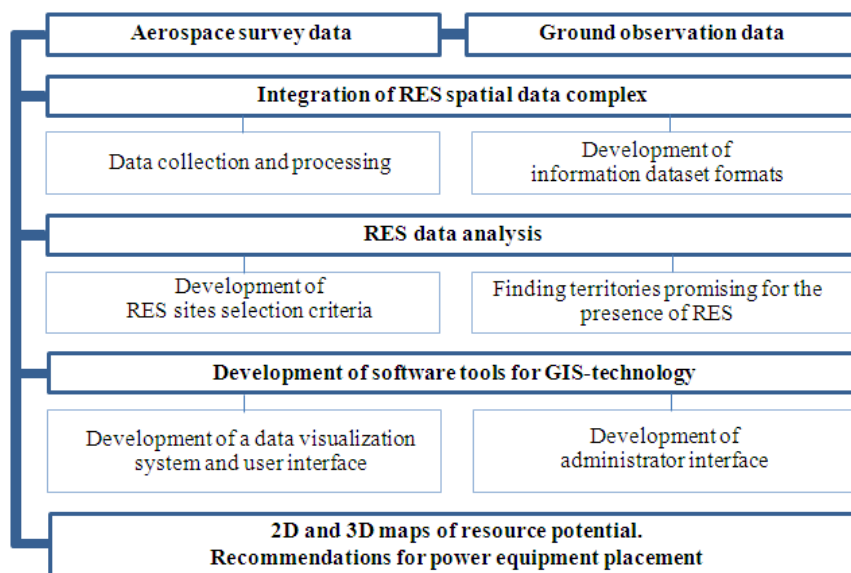


Figure 6 – The main components of the project for RES GIS-technology creation in Ukraine

The third component of the work includes the development and implementation of geoinformation analysis methods for a complex of spatial and attributive data based on the WEB-oriented approach. WEB-component will provide easy access to the project data for their prompt update and analysis [9, 13]. The creation of the WEB-interface will provide decision makers with an opportunity to gain access to databases information content. Besides reports and analytic materials, the results of spatial and attributive data analysis, as well as the maps of

RES spatial distribution will be presented on the WEB-portal with the ability to update them in real-time mode. The cartographic component of the WEB-portal will contain tools of queries and sample creation, comparative analysis, etc. All results will be downloaded by a user from the portal for their further analysis and decision-making.

The fourth component consists of conducting work on the country regions zoning with the purpose to identify areas prospective for the presence of renewable energy sources and suitable for their use in renewable power generation.

Software Tools of RES GIS-Technology

RES GIS software tools comprise two main components: tools for data collecting, storing, processing, and manipulating; and tools for geoinformation analysis and visualizing a set of spatial and attributive data taking into account the WEB-based approach (Fig. 7).

The software implementation of integrated RES data processing can be based on the use of specialized GIS tools such as ESRI ArcGIS, Quantum GIS (QGIS), ERDAS Imagine, ENVI, GIS GRASS, GIS SAGA, etc. The standard GIS toolkit can be supplemented and extended using a software implementation of analysis and modeling methods, for example, using the ERDAS IMAGINE Modeller tool, the Python scripting language in ESRI ArcGIS and QGIS, the GDAL, OGR, Numpy, SciPy, and Matplotlib library tools, etc.

WEB-interface for visualizing maps of perspective territories can be realized on the basis of existing cartographic services, in particular, a WEB-server based on GeoServer, Google Maps JavaScript API. GeoServer uses a Java Virtual Machine and allows publishing results of aerospace survey processing (thematic maps, synthesized image), presented in a raster (e.g., GeoTIFF) and vector (e.g., ESRI Shapefile) formats using different coordinate systems. GeoServer allows connecting to the ESRI ArcGIS ArcMap 10 and specifies the data presentation styles. Google Maps API technology based on HTML and JavaScript will allow visualizing thematic information on the available Google map data, such as vector maps and colorful synthesized images.

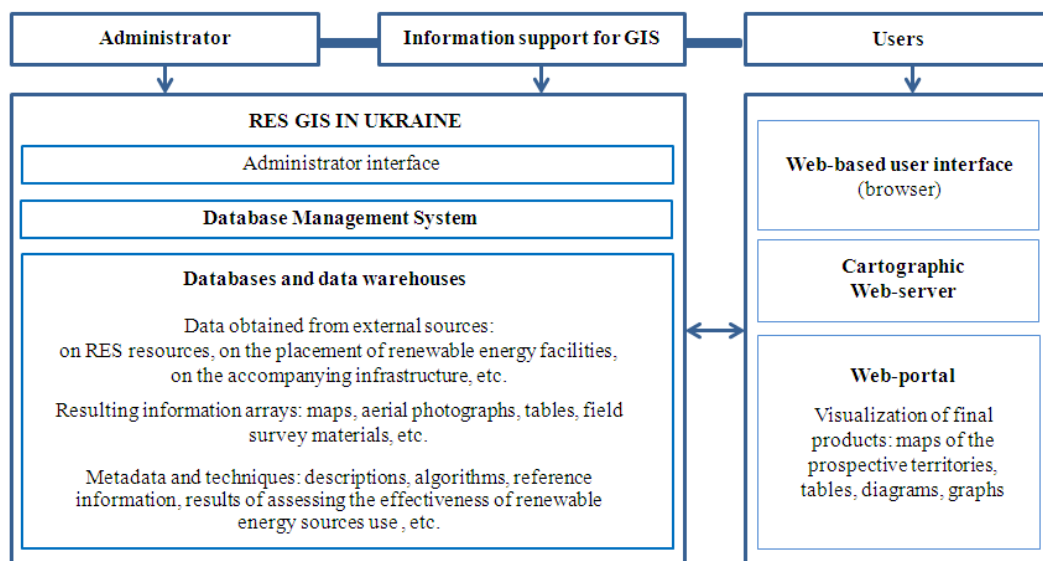


Figure 7 – The basic software components of the project focused on RES GIS-technology creation in Ukraine

The developed geoinformation technology and components of the geoinformation system are realized as part of the specialized RAPID (Recognition, Automated Prediction, Interpretation of Data) GIS, which is a powerful tool for the integrated analysis of spatial data based on Data Mining mathematical apparatus [16].

Conclusions

One of the ways to develop the energy sector of Ukraine is to assess the possibility and effectiveness of renewable energy sources use for regions' energy supply based on datasets covering both natural resources of the territory and its economic characteristics. In this connection, the task of creating specialized technological and software tools for processing, transforming, and analyzing data to estimate RES potential is urgent. The main approaches to geoinformation technology development for estimating RES potential based on the use of a complex of different-level and heterogeneous spatial data from ground-based observations and aerospace surveys aimed at creating 2D and 3D resource potential maps are presented. This will link together the industry information resources for different administrative units of Ukraine.

The creation and implementation of appropriate specialized geoinformation technology and GIS tools will help to expand the network of RES power plants and increase the efficiency of Ukraine's energy supply.

References

- [1] "REMAP 2030 – Renewable Energy Prospects for Ukraine," *Abu Dhabi: IRENA*, 46 p., 2015.
- [2] B. S. Busygin, K. L. Sergieieva, "GIS RES in Ukraine – myth or reality?," in *Proc. the 17th International Conference on Geoinformatics – Theoretical and Applied Aspects*, Kyiv, Ukraine, 2018, pp.1–7. – Access mode: <https://doi.org/10.3997/2214-4609.201801751>.
- [3] S. O. Kudria, "Stan ta perspektyvy rozvytku vidnovliuvanoi enerhetyky v Ukraini," *Visnyk NAN Ukrainy*, 12, pp. 19-26, 2015.
- [4] J. R. Oakleaf, C. M. Kennedy, S. Baruch-Mordo, J.S. Gerber, P.C. West, J.A. Johnson, J. Kiesecker, "Mapping global development potential for renewable energy, fossil fuels, mining and agriculture sectors," *Scientific data*, 6(1), pp. 1–17, 2019.
- [5] "REN21 Global Status Report," *Paris: REN21 Secretariat*, 367 p., 2020. – Access mode: https://www.ren21.net/wp-content/uploads/2019/05/gsr_2020_full_report_en.pdf
- [6] M. M. Tymkiv, D. V. Kasiianchuk, "Analiz vidnovliuvalnykh dzherel enerhii na terytorii Ivano-Frankivskoi oblasti z vykorystanniam heoinformatsiinykh system," *Ekolohichna bezpeka ta zbalansovane resursokorystannia*, 2, pp. 153–159, 2017.
- [7] "REN21 UNECE Renewable Energy Status Report," *Paris: REN21 Secretariat*, 2017. – Access mode: https://unece.org/DAM/energy/se/pp/renew/Renewable_energy_report_2017_web.pdf.
- [8] "Renewables in Ukraine," *KPMG-Ukraine Ltd.*, July 2019, 20 p. – Access mode: <https://assets.kpmg/content/dam/kpmg/ua/pdf/2019/07/Renewables-in-Ukraine-2019.pdf>
- [9] R. Avtar, N. Sahu, A.K. Aggarwal, S. Chakraborty, A. Kharrazi, A.P. Yunus, J. Dou, T.A. Kurniawan, "Exploring Renewable Energy Resources Using Remote Sensing and GIS – A Review," *Resources*, 8, p. 149, 2019. – Access mode: <https://www.mdpi.com/2079-9276/8/3/149>.
- [10] B. P. Sah, P. Wijayatunga, "Geographic Information System-Based Decision Support System for Renewable Energy Development: An Indonesian Case Study," *ADB Sustainable Development Working Paper Series*, 49, 42 p., 2017.
- [11] V. Zatserkovnyi, N. Oberemok, A. Puzyk, "Geoinformation modeling in the problems of renewable energy," *NTU "KhPI" Bulletin: Series "New Solutions in Modern Technologies"*, 1285(9), p. 118-127, 2018.
- [12] L. Matejicek, "Assessment of Energy Sources Using GIS," *Springer*, 2017. – Access mode: <https://www.springer.com/gp/book/9783319526935>.
- [13] Y. Sun, R. Wang, J. Li, J. Liu, "GIS-based multiregional potential evaluation and strategies selection framework for various renewable energy sources: a case study of eastern coastal regions of China," *Energy Science & Engineering*, 3, pp. 123–140, 2017.
- [14] O. Prilepova, Q. Hart, J. Merz, N. Parker, V. Bandaru, B. Jenkins, "Design of a GIS-Based WEB Application for Simulating Biofuel Feedstock Yields," *ISPRS International Journal of Geo-Information*, 3, pp. 929–941, 2014.
- [15] B. S. Busygin, V. V. Vasilev, K. L. Sergieieva, "About development of geoinformation system for renewable energy in Ukraine," *IV Mizhnarodnoi naukovo-praktuchna konferentsiya "Nadrokorystuvannia v Ukraini. Perspektyvy investuvannia"*, Truskavets, pp. 375-380, 2017.
- [16] B. S. Busygin, S. L. Nikulin, "Specialized geoinformation RAPID system: features, structure, tasks," *Geoinformatika*, 1, pp. 22–36, 2016.

Стаття надійшла: 24.05.2021.

Information about the authors

Busygin Borys Serhiiovich – Doctor of Technical Sciences, Professor, Professor of the Department of Information Technologies and Computer Engineering, Dnipro University of Technology.

Nikulin Serhii Leonidovich – Doctor of Geological Sciences, Associate Professor, Professor of the Department of Information Technologies and Computer Engineering, Dnipro University of Technology.

Sergieieva Kateryna Leonidivna – Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Information Technologies and Computer Engineering, Dnipro University of Technology.

Korobko Olga Valeriyivna – Assistant, Head of the Laboratory of the Department of Information Technologies and Computer Engineering, Dnipro University of Technology.

Б. С. Бусигін, С. Л. Нікулін, К. Л. Сергеева, О. В. Коробко

**ОСНОВНІ ПІДХОДИ ДО СТВОРЕННЯ
ГЕОІНФОРМАЦІЙНОЇ СИСТЕМИ ВІДНОВЛЮВАНИХ
ДЖЕРЕЛ ЕНЕРГІЇ В УКРАЇНІ**

Національний технічний університет «Дніпровська політехніка», Дніпро

Б. С. Бусыгин, С. Л. Никулин, Е. Л. Сергеева, О. В. Коробко

**ОСНОВНЫЕ ПОДХОДЫ К СОЗДАНИЮ
ГЕОИНФОРМАЦИОННОЙ СИСТЕМЫ
ВОЗОБНОВЛЯЕМЫХ ИСТОЧНИКОВ ЭНЕРГИИ В
УКРАИНЕ**

Национальный технический университет «Днепро́вская политехника», Днепр

УДК 004.42

В. П. Карашецький, В. І. Яркун

АВТОМАТИЗОВАНА СИСТЕМА МОНІТОРИНГУ ВСТУПНОЇ КАМПАНІЇ ЗАКЛАДУ ВИЩОЇ ОСВІТИ

Національний лісотехнічний університет України, Львів

Анотація. У пропонованій статті розглянуто актуальні завдання моніторингу та синтаксичного аналізу даних при розробленні автоматизованих систем, що забезпечують ефективний моніторинг вступної кампанії до закладу вищої освіти (ЗВО). Розроблено методику і програмне забезпечення автоматизованої системи моніторингу вступної кампанії ЗВО. Розроблене програмне забезпечення представляє веб-сайт. Дана система призначена для періодичного відстежування вступної кампанії ЗВО шляхом синтаксичного аналізу сторінок інформаційної системи "Конкурс" (<https://www.vstup.info/>) і використання зручного графічного інтерфейсу для перегляду даних абітурієнтів та їх швидкого ранжування за різними критеріями.

Ключові слова: вступна кампанія, графічний інтерфейс, ЗВО, моніторинг, синтаксичний аналіз сторінок, Java, Jsoup.

Аннотация. Разработана методика и программное обеспечение автоматизированной системы мониторинга вступительной кампании учреждения высшего образования (УВО). Разработанное программное обеспечение представляет веб-сайт. Данная система предназначена для периодического отслеживания вступительной кампании УВО путем синтаксического анализа страниц информационной системы "Конкурс" (<https://www.vstup.info/>), и использования удобного графического интерфейса для просмотра данных абитуриентов и их быстрого ранжирования по разным критериям.

Ключевые слова: вступительная кампания, графический интерфейс, УВО, мониторинг, синтаксический анализ страниц, Java, Jsoup.

Abstract. The methodology and software of an automated monitoring system for the admission campaign of a higher education institution (HEI) has been developed. The developed software represents the website. This system is designed to periodically monitoring the admission campaign of HEI by parsing the pages of the information system "Competition" (<https://www.vstup.info/>), and using a user-friendly graphical interface to view the data of applicants and quickly rank them according to various criteria.

Key words: admission campaign, graphical interface, HEI, monitoring, parsing of pages, Java, Jsoup.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-12-16>.

Вступ

В ході дослідження розглядаються існуючі моніторингові інформаційні системи які вирішують певні питання вступу до ВНЗ України. Такі системи використовуються багатьма вітчизняними навчальними закладами. Вони спрямовані на спостереження за перебігом вступної кампанії, а результати цих спостережень служать для обґрунтування або прогнозування управлінських рішень.

Актуальність

Останнім часом особливого значення в Україні набуває проблема якості освіти, яка зумовлена зростанням економічного рівня країни. Щорічно в червні, липні проходить вступна кампанія до ЗВО України. Вступ абітурієнтів відбувається на підставі їх рейтингу, що визначається як сума балів отриманих за тестами ЗНО, середнього бала атестата та додаткових балів за визначні досягнення у навчанні. Кожний ЗВО зацікавлений в проведенні постійного моніторингу його вступної кампанії з використанням сучасних засобів інформаційних технологій, які забезпечують оперативну обробку отриманої інформації та її представлення в зручному вигляді. На сьогоднішній день актуальним постає завдання ранжування отриманих даних в результаті моніторингу.

Аналіз досліджень

На сьогоднішній день впроваджена Єдина державна база з питань освіти (ЄДЕБО), доступна за адресою <https://vstup.edbo.gov.ua>, дані якої виступають вихідними параметрами для моніторингу Вступної кампанії.

Інформаційна система "Конкурс" (<https://www.vstup.info/>) відображає інформацію отриману з ЄДЕБО. Система не передбачає фільтрування даних, пошук потрібної інформації здійснюється переходом за посиланнями по сторінках сайту.

Інформаційна система "Вступ. ОСВІТА.UA" (<https://vstup.osvita.ua/>) здійснює інформування громадськості про перебіг вступної кампанії на підставі даних Єдиної державної електронної бази з питань освіти [3]. В даній системі реалізовано пошук лише за ЗВО. Відсутня можливість фільтрування інформації за особистими даними абітурієнтів.

Широкої популярності набув "Сервіс пошуку абітурієнтів" (<https://abit-poisk.org.ua>). Сайт використовує інформацію отриману з ресурсів: ЄДЕБО, ІС Конкурс та Вступ. ОСВІТА.UA. Інтерфейс для здійснення пошуку та фільтрування інформації даного сайту представлений у вигляді командної стрічки, що не завжди зручно некомпетентному користувачу. Основна особливість системи – пошук за ПІБ та роком вступу. На сайті не передбачено можливість фільтрування за базовою освітою та діапазоном конкурсного балу.

В. П. Карашецький, В. І. Яркун, 2021

Ряд публікацій присвячено дослідженню автоматизації процесу формування контингенту студентів (ФКС) у ході поточної вступної кампанії. Наведено процеси, що відбуваються в інформаційно-аналітичній системі. Описано функціонування автоматизованої системи приймальної комісії (ПК) закладу вищої освіти. Представлено зовнішній вигляд основних інтерфейсів автоматизованого робочого місця оператора, його функції і задачі [6]. Досліджується та розробляється автоматизована система супроводу приймальної кампанії ВНЗ, що представляє собою базу даних MS Access, в якій реалізовано підсистему для обміну даними через SOAP інтерфейс [5]. У праці [4] проаналізовано вибір пріоритету, як показника визначеності абітурієнта. Тут подано архітектуру інформаційної системи прогнозування контингенту студентів на основі вибору пріоритету. В роботі [1] розглядаються заходи вдосконалення інформаційного забезпечення вищих навчальних закладів України, проводиться аналіз недоліків вже існуючих та переваг впровадження нових інформаційних систем, які допоможуть якомога краще вирішити проблему управління даними.

Мета

У даній роботі автори намагаються досягнути наступних цілей: розкрити основний зміст моніторингу вступної кампанії як складової забезпечення якості освіти у ЗВО; розробити інформаційну систему моніторингу вступної кампанії ЗВО; здійснити автоматизоване ранжування даних про абітурієнтів за різними критеріями.

Задачі

1. Створити автоматизовану систему моніторингу вступної кампанії до ЗВО, яка спрямована на розроблення суспільно корисного програмного продукту.
2. Здійснювати пошук та фільтрування конкурсних даних абітурієнтів.
3. Організувати зручний інтерфейс для ранжування отриманої інформації.

Розв'язання задач

Заклад вищої освіти, врахувавши стратегію власної роботи, приступає до розроблення автоматизованої інформаційної системи моніторингу вступної кампанії, впровадження якої варто здійснювати у кілька етапів, а саме впроваджувати окремі модулі чи підсистеми поступово покращуючи роботу на певних напрямках своєї діяльності.

В розробленій автоматизованій системі моніторингу вступної кампанії ЗВО дані про вступників отримуються шляхом синтаксичного аналізу (рис.1) сторінок інформаційної системи "Конкурс" з використанням Java-бібліотеки jsoup[2], яка призначена для синтаксичного аналізу (парсингу), отримання і обробки даних, що зберігаються в документах HTML. Jsoup – це бібліотека аналізу даних HTML для Java, орієнтована на гнучкість та простоту використання. Її можна використовувати для витягування окремих даних з HTML-сторінок, а також для змін їх вмісту за допомогою списку дозволених тегів та атрибутів. Jsoup представляє API для парсинга і опрацювання даних, використовуючи методи DOM, CSS та jQuery. Бібліотека реалізує специфікацію WHATWG HTML5 і аналізує HTML в тому ж DOM, що і сучасні браузерери.

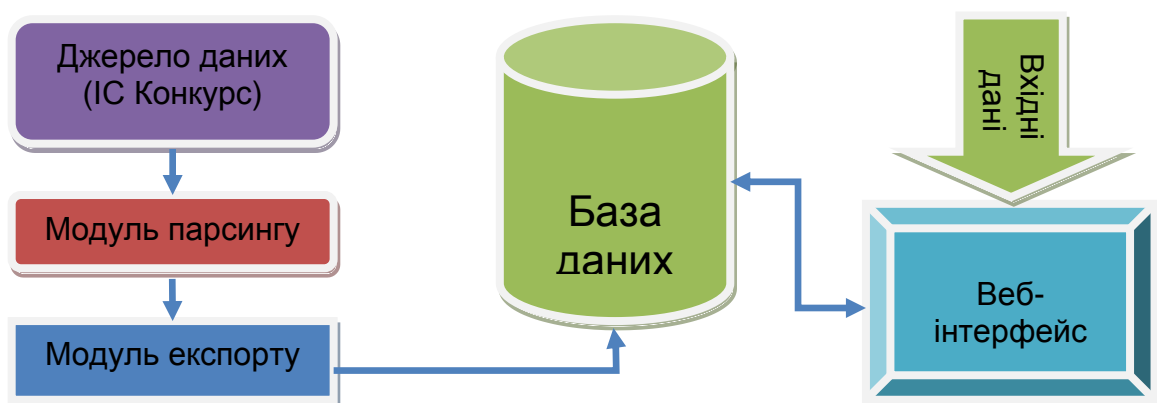


Рисунок 1 – Структурна схема синтаксичного аналізу ІС

Використання бібліотеки jsoup забезпечила наступні функціональні можливості системи моніторингу вступної кампанії закладу вищої освіти:

- 1) Читання та аналіз HTML, використовуючи URL або рядок. Підтримка множинного читання.
- 2) Можливість знаходити і отримувати дані через селектори CSS.
- 3) Маніпулювання HTML-елементами, атрибутами і текстом.

- 4) Можливість очищати користувачський контент HTML сторінки уникаючи атаки XSS.
- 5) Виводити чистий HTML код.
- 6) Можливість опрацювати незамкнуті теги, неявні теги.

Наприклад, для пошуку необхідних HTML елементів <td> та отримання їх контенту, фрагмент програмного коду має наступний вигляд:

```
Document d = Jsoup.connect("http://www.vstup.info/2020/160/i2020i160p713351.html").get( );
Elements tdElements = doc.getElementsByTag("td");
```

Функціонал бібліотеки jsoup дозволив аналізувати дані ІС Конкурс за її URL-адресою. Одержані таким чином дані спочатку зберігаються у динамічному масиві, фільтруються та передаються на запис до бази даних. Це дозволяє збільшити швидкодію подальшої обробки даних.

Головна сторінка системи моніторингу вступної кампанії НЛТУ України представлена на рисунку 2. Система є веб-орієнтованим продуктом та доступна із сайту університету.

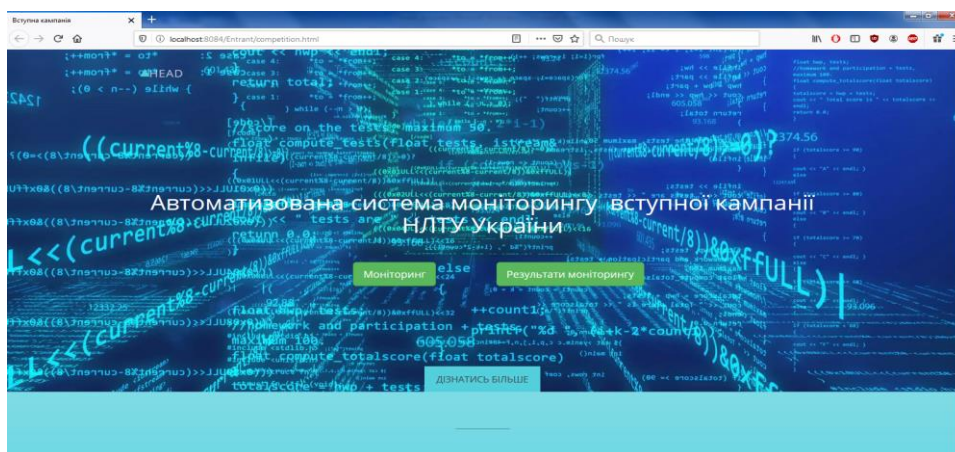


Рисунок 2 – Сторінка автоматизованої системи моніторингу вступної кампанії НЛТУ України

Для забезпечення захисту та цілісності даних реалізовано систему авторизації, яка розділяє права користувачів та адміністратора. Моніторинг вступної кампанії НЛТУ України може виконуватись адміністратором сайту натисненням кнопки “Моніторинг”. Внаслідок цього відбувається запуск скрипта на виконання парсинга, який здійснює синтаксичний аналіз HTML-сторінок. Парсинг можна здійснювати шляхом вибору одного із запропонованих джерел – посилання на сторінку. Наприклад, дані про вступників можуть отримуватися шляхом синтаксичного аналізу сторінок інформаційної системи “Конкурс” з використанням бібліотеки jsoup мови програмування Java.

Після завершення моніторингу з’являється діалогове вікно з повідомленням “Моніторинг завершено” (Рис. 3).

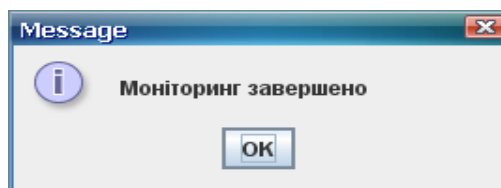


Рисунок 3 – Діалогове вікно з повідомленням про завершення моніторингу

Щоб переглянути результати моніторингу потрібно натиснути кнопку “Результати моніторингу”. Після цього з’являється окреме вікно з графічним інтерфейсом (Рис. 4), в якому користувачі можуть оглянути дані про вступників та виконати ранжування абітурієнтів за різними критеріями (ступенем освіти, спеціальністю, формою навчання, базовою освітою, кількістю балів, статусом заяви та поданням оригіналів документів).

№ п/п	П.І.П.	Ступінь	Спеціальн.	Форма	Базова осв.	Пріоритет	К-сть балів	Статус	Квота	Оригінали	Деталізація
1	Лукачев А. Б.	122	д	МС	—	—	363.68	Рекоменд.	—	—	Українськ.
2	Марків В. П.	122	д	МС	—	—	359.0	Рекоменд.	—	—	Українськ.
3	Чайківський Б.	122	д	МС	—	—	348.82	Рекоменд.	—	+	Українськ.
4	Рудник Е. М.	122	д	МС	—	—	346.08	Рекоменд.	—	—	Українськ.
5	Сивенюк О. Б.	122	д	МС	—	—	342.99	Рекоменд.	—	—	Українськ.
6	Юрчишин Б.	122	д	МС	—	—	336.75	Рекоменд.	—	—	Українськ.
7	Цуррик В. О.	122	д	МС	—	—	336.7	Рекоменд.	—	+	Українськ.
8	Мазурин Н. Б.	122	д	МС	—	—	334.91	Рекоменд.	—	—	Українськ.
9	Тарнавський Б.	122	д	МС	—	—	334.26	Рекоменд.	—	+	Українськ.
10	Жиняль Т. Б.	122	д	МС	—	—	334.2	Рекоменд.	—	—	Українськ.
11	Романюк Б.	122	д	МС	—	—	333.58	Рекоменд.	—	+	Українськ.
12	Куцько Д. Ю.	122	д	МС	—	—	333.32	Рекоменд.	—	—	Українськ.
13	Ободницький Б.	122	д	МС	—	—	331.82	Рекоменд.	—	—	Українськ.
14	Майор О. Б.	122	д	МС	—	—	331.18	Рекоменд.	—	—	Українськ.
15	Хлинич А. Б.	122	д	МС	—	—	300.4	Допущено	Квота №1	+	Українськ.
16	Королюк В. Б.	122	д	МС	—	—	287.99	Допущено	Квота №1	+	Українськ.
17	Плох Д. Ю.	122	д	МС	—	—	326.5	Допущено	—	—	Українськ.
18	Павлюк О. М.	122	д	МС	—	—	323.35	Допущено	—	—	Українськ.
19	Пеньків Д. Б.	122	д	МС	—	—	322.19	Допущено	—	—	Українськ.
20	Понуровський Б.	122	д	МС	—	—	321.94	Допущено	—	—	Українськ.
21	Коваленко Б.	122	д	МС	—	—	319.7	Допущено	—	—	Українськ.
22	Селенко А. І.	122	д	МС	—	—	317.67	Допущено	—	—	Українськ.

Рисунок 4 – Вікно результатів моніторингу вступної кампанії НЛТУ України

Пошук або фільтрування конкурсних даних вступників може здійснюватися у даному вікні відповідно за окремими критеріями чи комплексно. Вибірка за окремим критерієм виконується шляхом встановлення значення критерію у випадіаючому списку чи внесення потрібного значення у поле введення. При комплексному фільтрі додатково необхідно встановити прапорці для вибраних критеріїв та натиснути кнопку "Пошук".

Розроблена система забезпечує виконання наступного функціоналу:

- автоматизований парсинг сторінок інформаційної системи "Конкурс";
- автоматизоване ведення отриманих даних до бази даних;
- здійснення вибірок за багатьма критеріями;
- здійснення пошуку за будь-яким відомим атрибутом та їх набором;
- здійснення вибірки абітурієнтів за діапазоном значень конкурсного балу;
- здійснення ранжування абітурієнтів за різними критеріями;
- автоматизований аналіз та прогнозування попиту на спеціальності.

Висновки

Отже, здійснивши аналіз наявних інформаційних систем, що вирішують актуальні питання вступу до ВНЗ України, вивчивши питання автоматизації моніторингу вступної кампанії і використовуючи сучасні підходи до створення систем такого типу – розроблено методіку та програмне забезпечення автоматизованої системи моніторингу вступної кампанії ЗВО. Система надає можливість періодичного відстеження вступної кампанії ЗВО з метою перегляду конкурсних даних абітурієнтів, їх швидкого ранжування за різними критеріями та автоматизований аналіз попиту на спеціальності. При створенні власних систем моніторингу ЗВО повинен враховувати досвід розроблення та успіх впровадження прототипних систем, орієнтованість на об'єкт дослідження та очікуваний результат. Впровадження системи моніторингу забезпечить підвищення ефективності роботи університету під час вступної кампанії.

Список літератури

- [1] Л. О. Білик, "Вдосконалення інформаційного забезпечення управління вищих навчальних закладів освіти України," *Вісник студентського наукового товариства ДонНУ імені Василя Стуса*, 2013, 1.5: 219–223.
- [2] Библиотека jsoup [Електронний ресурс]. Режим доступу: <http://developer.alexanderklimov.ru/android/library/jsoup.php>. Дата звернення 10.04.2021.
- [3] Вступ. ОСВІТА.УА. [Електронний ресурс]. Режим доступу: <https://vstup.osvita.ua/>. Дата звернення 15.04.2021.
- [4] П. Жежнич, А. Шілінг, "Побудова системи прогнозування контингенту студентів ЗВО на основі вибору пріоритет," *Інноваційні комп'ютерні технології у вищій школі: Матеріали 10-ї науково-практичної конференції*. м. Львів, 21-23 листопада 2018 року, Львів: Видавництво Львівської політехніки, 2018. – 192 с.
- [5] Т. І. Капленко, Н. В. Кас'янок, "Розробка автоматизованої системи супроводу приймальної кампанії ВНЗ," *Сучасні проблеми і досягнення в галузі радіотехніки, телекомунікацій та інформаційних технологій: тези доповідей VIII Міжнародної науково-практичної конференції, 21–23 вересня 2016р., м. Запоріжжя*. – Запоріжжя: ЗНТУ, 2016. – 145–146 с.

- [6] А. М. Мороз, Н. О. Похлебіна, В. А. Хобін, "Інформаційно-аналітична система приймальної комісії ОНАХТ як основа автоматизованого управління формування контингенту студентів," *Automation of technological and business processes*. 12, 4 (Груд. 2020), 36–42.

Стаття надійшла: 12.02.2021.

References

- [1] L. O. Bilyk, "Vdoskonalennia informatsiinoho zabezpechennia upravlinnia vyshehkykh navchalnykh zakla-div osvity Ukrainy," *Visnyk studentskoho naukovoho tovarystva DonNU imeni Vasylia Stusa*, 2013, 1.5: 219–223. – [in Ukrainian].
- [2] Byblyoteka jsoup [Online]. Available: <http://developer.alexanderklimov.ru/android/library/jsoup.php>. [Accessed: 10.04.2021]. – [in Russian].
- [3] Vstup.OSVITA.UA. [Online]. Available: <https://vstup.osvita.ua>. [Accessed: 15.04.2021]. – [in Ukrainian].
- [4] P. Zhezhnych, A. Shilinh, "Pobudova systemy prohnozuvannia kontynhentu studentiv ZVO na osnovi vyboru priorytet," *Innovatsiini kompiuterni tekhnolohii u vyshchii shkoli: Materialy 10-yi naukovo-praktychnoi konferentsii. m. Lviv, 21-23 lystopada 2018 roku*, Lviv: Vydavnytstvo Lvivskoi politekhniky, 2018. – 192 s. – [in Ukrainian].
- [5] T. I. Kapliienko, N. V. Kasianok, "Rozrobka avtomatyzovanoi systemy suprovodu pryimalnoi kampanii VNZ," *Suchasni problemy i dosiahnennia v haluzi radiotekhniki, telekomunikatsii ta informatsiinykh tekhnolohii: tezy dopovidei VIII Mizhnarodnoi naukovo-praktychnoi konferentsii, 21–23 veresnia 2016r., m.Zaporizhzhia*. – Zaporizhzhia: ZNTU, 2016. – 145–146 s. – [in Ukrainian].
- [6] А. М. Мороз, Н. О. Похлебіна, В. А. Хобін, "Інформаційно-аналітична система приймальної комісії ОНАХТ як основа автоматизованого управління формування контингенту студентів," *Automation of technological and business processes*. 12, 4 (Груд. 2020), 36–42. – [in Ukrainian].

Відомості про авторів

Карашецький Володимир Петрович – кандидат технічних наук, доцент, доцент кафедри інформаційних технологій.

Яркун Володимир Ігорович – старший викладач кафедри інформаційних технологій, керівник центру веб-комунікацій.

В. П. Карашецкий, В. И. Яркун

АВТОМАТИЗИРОВАННАЯ СИСТЕМА МОНИТОРИНГА ВСТУПИТЕЛЬНОЙ КАМПАНИИ УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ

Национальный лесотехнический университет Украины, Львов

V. P. Karashetsky, V. I. Yarkun

AN AUTOMATED SYSTEM FOR MONITORING THE ADMISSION CAMPAIGN OF A HIGHER EDUCATION INSTITUTION

Ukrainian National Forestry University, Lviv

УДК 004.9+616.24

V. A. Kovenko, I. V. Bogach, M. V. Baraban

ITEM-BASED COLLABORATIVE FILTERING BASED ON NLP TECHNIQUES

Vinnytsia National Technical University, Vinnytsia

Abstract. The benchmark approach to content-based recommendation systems is exposed in this article. The usage of Word2Vec embeddings made by Google is unleashed. The opportunity of using additional business logic is considered.

Keywords: NLP techniques, Word2Vec, CountVectorizer, cosine similarity, embeddings, content based system, content based recommendation, user based recommendation.

Анотація. В даній роботі представлено підхід к рекомендаційним системам на основі каталогу. Показано використання Word2Vec еMBEDDING, представлених Google. Описана можливість використання додаткової бізнес логіки.

Ключевые слова: NLP підходи, Word2Vec, CountVectorizer, косинусна схожість, еMBEDDING, система основана на каталозі, рекомендації основані на каталозі, рекомендації основані на користувачах.

Анотація. У цій роботі представлено алгоритм для вирішення проблеми рекомендацій на основі каталогу. Показано використання Word2Vec еMBEDDING, представлених Google.

Ключові слова: NLP підходи, Word2Vec, CountVectorizer, косинусна схожість, еMBEDDING, система основана на каталозі, рекомендації основані на каталозі, рекомендації основані на користувачах.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-17-22>.

Introduction

Understanding users' preferences and proposing them the most relevant products is essential for every commercial business which involves the process of interacting with users. As nowadays the web infrastructure is developing rapidly, lot's of commercial activities move to the space of the internet. Thus, the demand for recommendation systems arises. Recommendation system is an engine which goal is to recommend relevant items to users. Many world famous companies like Netflix, Amazon, YouTube, etc, use them to attract more people to their websites and increase their income. The recommendation systems can be divided into two groups: content based [1] and user based [2]. Content based recommendation systems focus on the content, its taxonomy and metadata for making predictions, while the user based ones require user interactions like clicks or ratings the user left for items. Nevertheless user based recommendation systems are much more powerful than content based, they require lot's of computational power that can afford working with big data. On the opposite, when building a content based recommendation system, the one is interested only in the catalog of items, and as a rule the number of items is always smaller than the number of users in the system. Content based recommendation system is a nice start for a small company that just appeared on the market of web products. With the development of the sphere of Natural Language Processing (NLP), new opportunities for content-based recommendation systems appeared. The new approach to recommendation systems is proposed in this article, the problem is stated as measuring the similarities between items' metadata and is addressed as an NLP task. The system uses a hybrid algorithm based on counting words in a statement and Word2Vec model provided by Google. The possibility of using an additional business logic is considered. Finally, the results are viewed with respect to Movielens dataset.

Data preparation

Constructed system was validated using the famous Movielens [1] dataset, which contains catalogues of movies, their metadata and intersections of users with a catalog. The columns that were used are the following: title, movie_id and genres (Fig. 1).

As the constructed algorithm is based on content metadata – it was decided to enrich Movielens dataset with IMDB database to get more relative columns.

	movieid	title	genres
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
1	2	Jumanji (1995)	Adventure Children Fantasy
2	3	Grumpier Old Men (1995)	Comedy Romance
3	4	Waiting to Exhale (1995)	Comedy Drama Romance
4	5	Father of the Bride Part II (1995)	Comedy

Figure 1 – Original Movielens data

As genres columns were already in our dataset, this field was enriched with more data from IMDB, other columns were just added (Fig. 2).

	title	year	titleType	genres	characters	directors	writers	cast
0	Toy Story	1995	movie	[Children, Fantasy, Adventure, Animation, Comedy]	["Mr. Potato Head", "Slinky Dog", "Buzz Lighty...]	[John Lasseter]	[John Lasseter, Pete Docter, Andrew Stanton, J...]	[Joel Cohen, Tom Hanks, Tim Allen, Don Rickles...]
1	Jumanji	1995	movie	[Fantasy, Children, Adventure, Family, Comedy]	["Sarah Whittle", "Alan Parrish", "Van Pelt", ...]	[Joe Johnston]	[Jonathan Hensleigh, Greg Taylor, Jim Strain, ...]	[Scott Kroopf, Robin Williams, Kirsten Dunst, ...]
2	Grumpier Old Men	1995	movie	[Romance, Comedy]	["John Gustafson", "Maria Sophia Coletta Raget...]	[Howard Deutch]	[Mark Steven Johnson]	[Tak Fujimoto, Walter Matthau, Jack Lemmon, An...]
3	Waiting to Exhale	1995	movie	Drama, Romance, Comedy]	["Bernadine Harris", "Robin Stokes", "Savannah...]	[Forest Whitaker]	[Terry McMillan, Ronald Bass]	[Kenneth 'Babyface' Edmonds, Whitney Houston, ...]
4	Father of the Bride Part II	1995	movie	[Comedy, Family, Romance]	["George Banks", "Annie Banks-MacKenzie", "Nin...]	[Charles Shyer]	[Albert Hackett, Frances Goodrich, Nancy Meyer...]	[Elliot Davis, Steve Martin, Diane Keaton, Mar...]

Figure 2 – Data after enriching with IMDB

Data transformation

The MovieLens data is made of strings which describe items' metadata, but for an algorithm to work the transformation of relative columns to the matrix of numbers is needed. For this purpose a hybrid transformation, made of CountVectorizer [3] and Word2Vec [4] model, is used. CountVectorizer is a technique of counting words in a sentence or text corpus. Each word is then represented by its frequency of appearance in a sentence (Fig. 3).

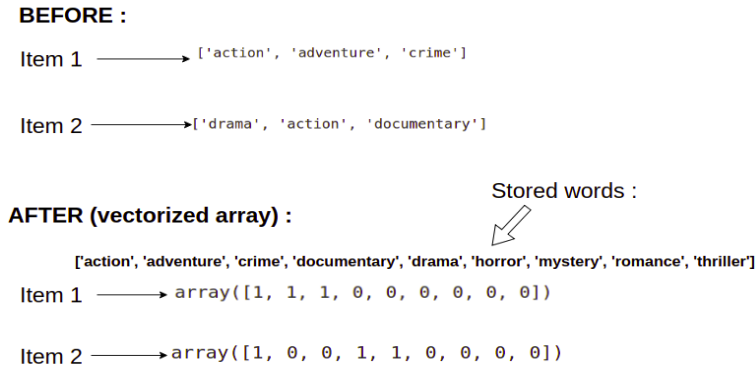


Figure 3 – The usage of CountVectorizer algorithm

That gives a possibility to represent word sequences as vectors of numbers. CountVectorizer is used only on columns that don't have a strong semantic context (directors, writers, cast, characters, etc.) For the fields with semantic context the other algorithm, known as Word2Vec is used. Word2Vec is a powerful machine learning algorithm that is used in the NLP sphere for semantic text classification, finding similarity between words, etc. Word2Vec learns the embedding space of words in which similar words (the ones that tend to appear in the same context) are closer to each other. Thus, the learned Word2Vec embeddings can be used to represent our columns in a numerical way and also save a semantic context of them (Fig. 4).

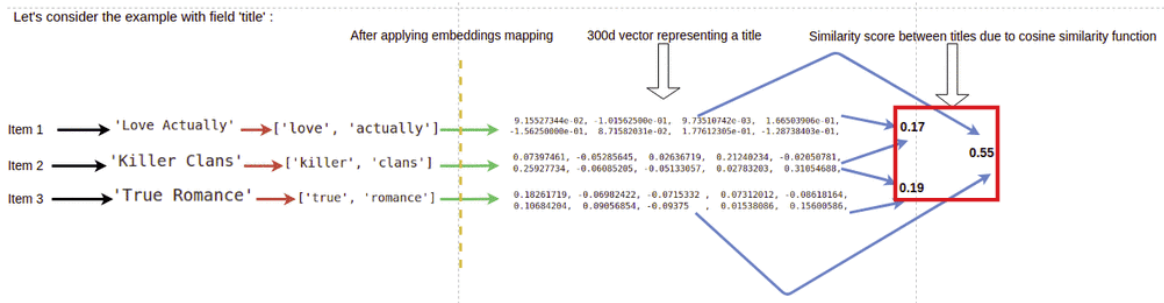


Figure 4 – Example of usage of Word2Vec model

For the purpose of representing not just a single word with Word2Vec, but the whole sentence – the embeddings of sentence's words are averaged. All the highlighted transformations give a possibility to represent content as nested matrices of number related to its metadata, thus the opportunity of using mathematical methods for finding similarity between content can be applied.

Algorithm

The algorithm is based on a cosine similarity, that is a mere choice for NLP tasks (1).

$$k(x, y) = \frac{xy^T}{\|x\| \|y\|} \tag{1}$$

The cosine similarity is computed between each column of each item in the dataset, thus for each item we have a matrix of similarities with others by a particular column. Because of the fact that one data column regarding an item can be much more important for the final recommendation than the other one, the additional business logic is added. The additional set of columns' weights that can be configured manually was added. Each attribute/column matrix is multiplied by a related weight parameter that gives an opportunity to decrease/increase contribution of it to the final similarity calculation. This makes the overall system more flexible and extendable towards new logic. Finally, the similarity matrices by columns are averaged to produce a final similarity matrix for an item (Fig. 5). Then, top N recommendations can be retrieved using a similarity score.

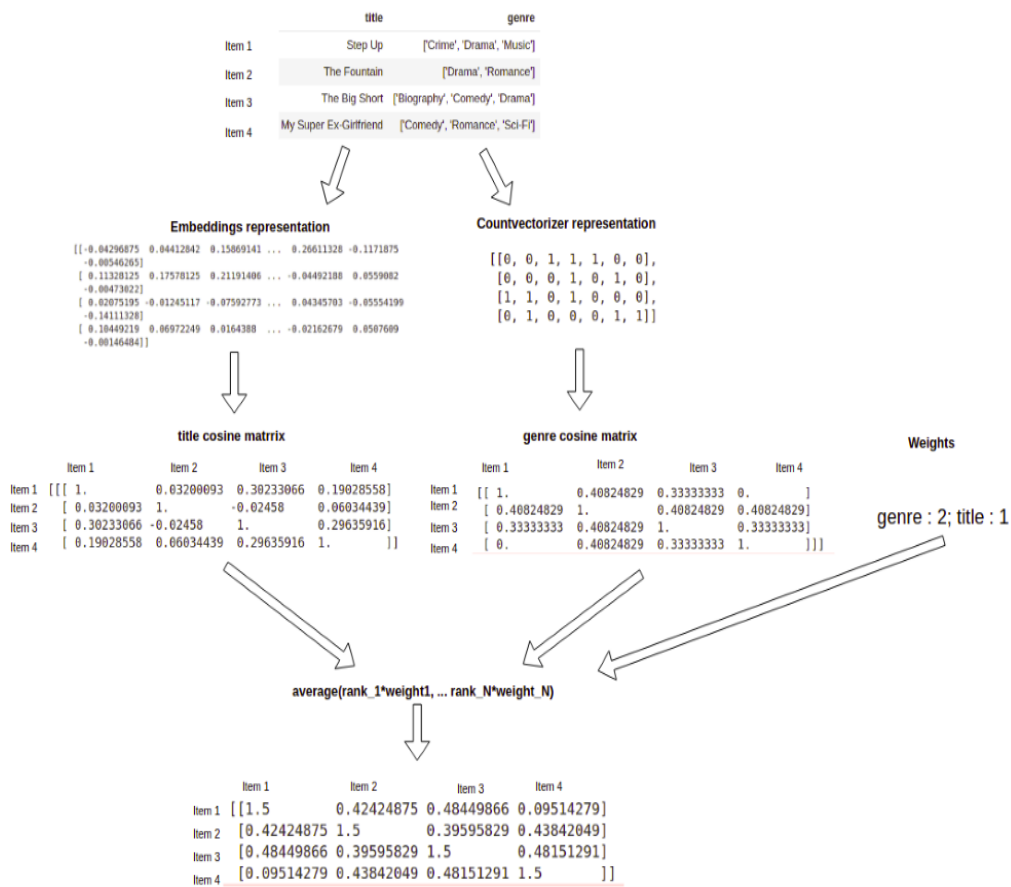


Figure 5 – Example of similarity calculation on two columns of data

Experiments

Experiments were conducted with respect to different weights sets. For each set of weights two experiments were done. For each experiment the item was randomly chosen and the top-4 related recommendations by similarity score were shown. From the first experiment (Fig. 6) it's obvious that given and predicted movies have similar titles by semantics and overlapping genres, as the highest weights were set to genre and title columns.

With different weights, the other picture appears. From the second experiment (Fig. 7), it's obvious that items have the biggest similarity by such fields as: characters, cast, directors, which is not a surprise if to look at weights.

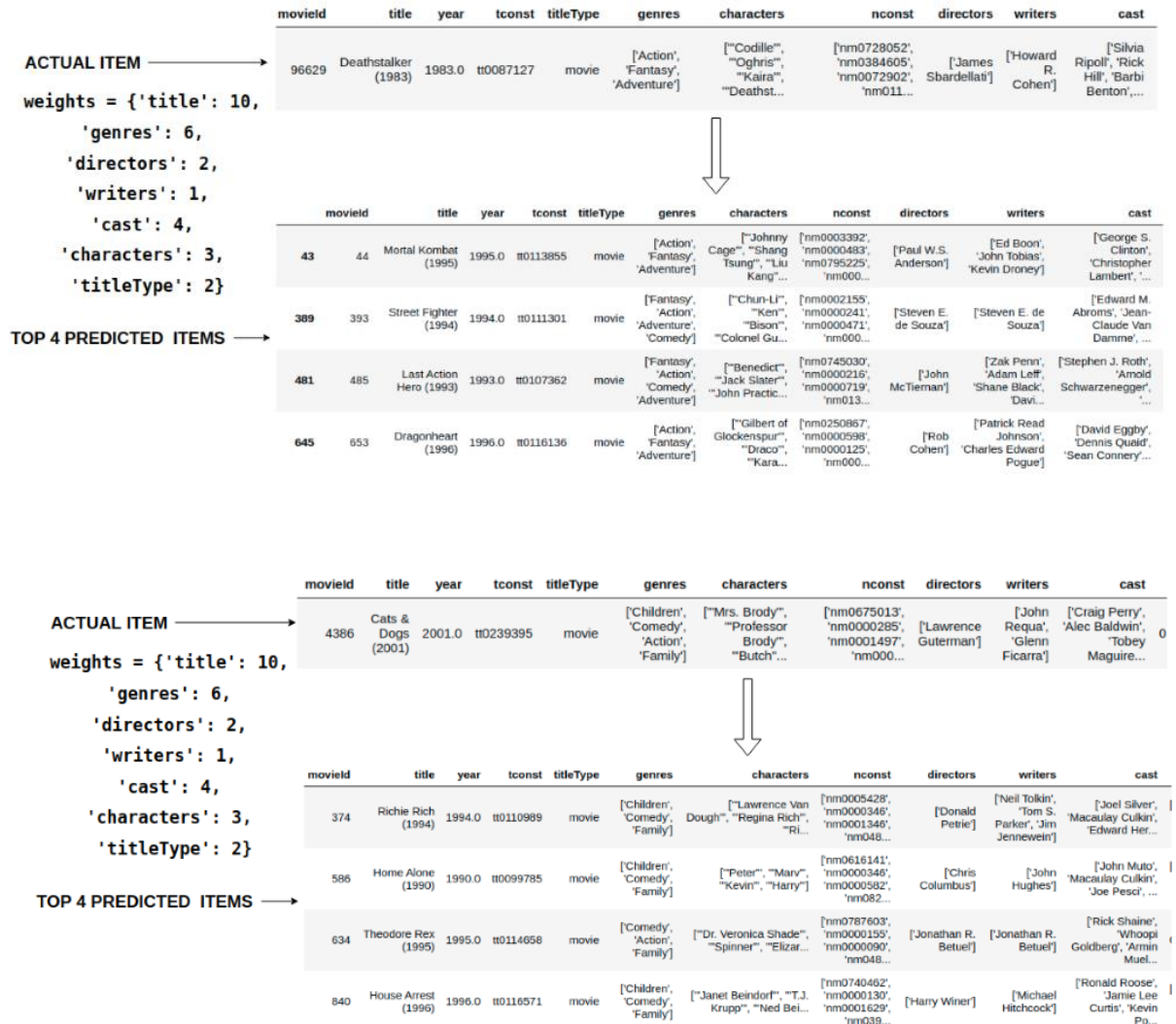
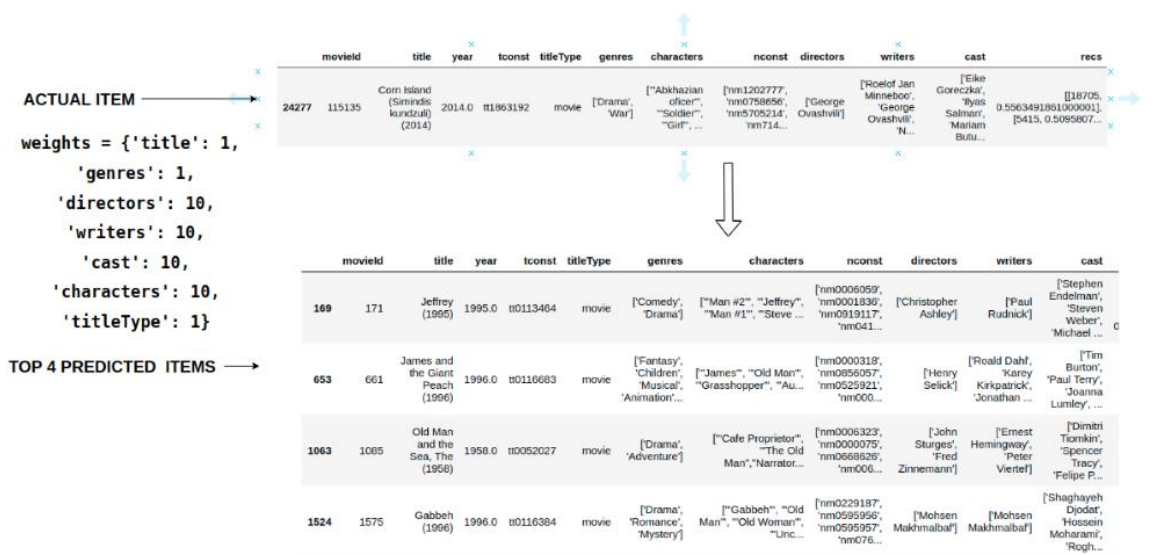


Figure 6 – The first experiment with different weights sets



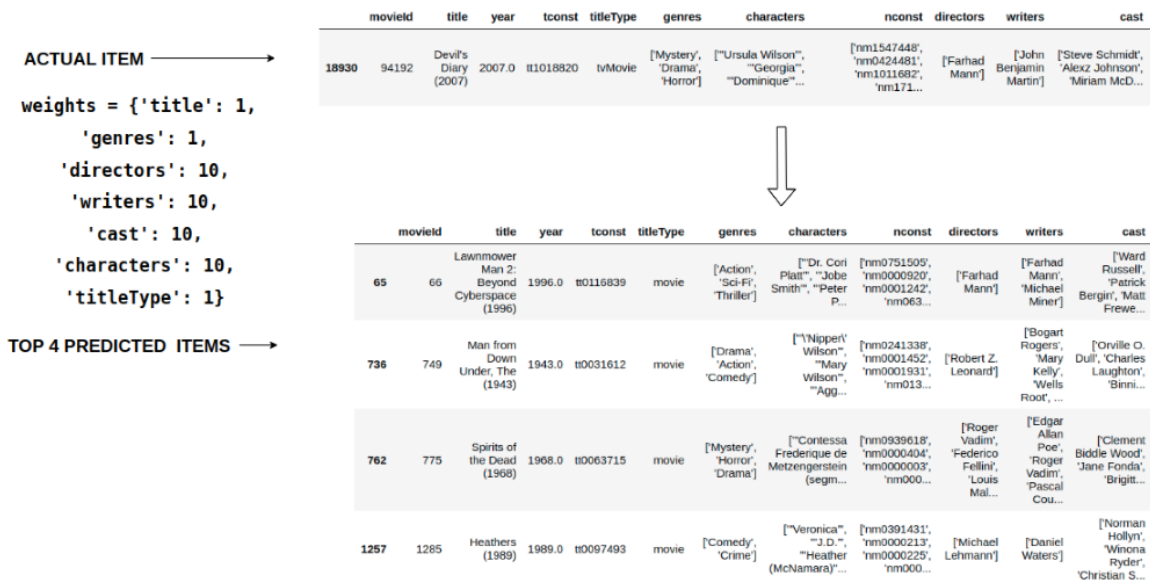


Figure 7 – The second experiment with different weights sets

Summary and further work

Nevertheless, the proposed system is not based on user activity, it has lots of advantages. It's flexible, not power consuming, easy to extend and flexible. Addressing the issue of recommendation systems as an NLP task, gives a lift to usage of novel NLP techniques like BERT [5]. On the opposite side, it's much less powerful than a user based system and is heavily dependent on the quality of catalog metadata. The particular system can also be used in ensemble with a user based recommendation system to construct a hybrid system. To sum up, the proposed algorithm can be used as an alternative to user based system and is an adequate choice for companies which just started their activity on the web market.

References

- [1] Content-Based Recommendation System [Electronic resource]. URL: https://www.researchgate.net/publication/236895069_Content-Based_Recommendation_Systems – Title from the screen.
- [2] Recommendation Systems : User-based Collaborative Filtering using N Nearest Neighbors. [Electronic resource]. URL: <https://medium.com/sfu-big-data/recommendation-systems-user-based-collaborative-filtering-using-n-nearest-neighbors-bf7361dc24e0> – Title from the screen.
- [3] 10+ Examples for Using CountVectorizer [Electronic resource]. URL: <https://kavita-ganesan.com/how-to-use-countvectorizer/#.XIF2wHUzav4> – Title from the screen.
- [4] Distributed Representations of Words and Phrases and their Compositionality [Electronic resource]. URL: <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf> – Title from the screen.
- [5] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Electronic resource]. URL: <https://arxiv.org/abs/1810.04805> – Title from the screen.

Стаття надійшла: 29.08.2021.

Information about the authors

Kovenko Volodymyr Andriyovych – student of the Department r of Automation and Intelligent Information Technologies.

Bogach Ilona Vitalievna – Ph.D., Associate Professor, Associate Professor of the Department of Automation and Intelligent Information Technologies.

Baraban Maria Volodymyrivna – Ph.D., Associate Professo of the Department r of Automation and Intelligent Information Technologies.

В. А. Ковенко, І. В. Богач, М. В. Барабан

**КОЛЛАБОРАТИВНАЯ ФИЛЬТРАЦИЯ НА БАЗЕ
КАТАЛОГА С ИСПОЛЬЗОВАНИЕМ NLP ТЕХНОЛОГИИ**

Винницкий национальный технический университет, Винница

В. А. Ковенко, І. В. Богач, М. В. Барабан

**КОЛЛАБОРАТИВНА ФІЛЬТРАЦІЯ НА ОСНОВІ КАТОЛОГУ
З ВИКОРИСТАННЯМ NLP ТЕХНОЛОГІЇ**

Вінницький національний технічний університет, Вінниця

УДК 004.9: 004.4

Л. В. Крилик, А. А. Яровий, І. В. Безсмертна

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ ІНФОРМАЦІЙНОЇ СИСТЕМИ АВТОМАТИЗАЦІЇ ПОДАВАННЯ ЗАЯВОК ДО КОМУНАЛЬНИХ ПІДПРИЄМСТВ

Вінницький національний технічний університет, Вінниця

Анотація. В роботі розглянуто проблематику частини комунальної галузі, яка відповідає за взаємодію комунального підприємства та осіб, що воно обслуговує. Після проведеного аналізу доведено доцільність діджиталізації та реалізації WEB-орієнтованого програмного продукту, що задовольнив би усі потреби, значно спростив та прискорив процес взаємодії, потребував би мінімальних зусиль на навчання, а також був доступний користувачам різних комп'ютерних та мобільних пристроїв за наявності інтернету. На базі цього, з метою автоматизації та удосконалення процесів отримання актуальної інформації та проектування прозорої системи, було описано особливості реалізації інформаційної системи автоматизації подання заявок до комунальних підприємств, її архітектуру, структуру та програмну реалізацію такими інструментами як NodeJS, ReactJS, GraphQL, MongoDB. Як результат розроблено інформаційну систему автоматизації подання заявок до комунальних підприємств, що повністю задовольняє базові потреби користувачів послуг комунальних підприємств та спрощує процес отримання послуг та інформації.

Ключові слова: інформаційна система, WEB-технології, методології проектування програмного забезпечення, автоматизація подання заявок, комунальні підприємства, моніторинг.

Аннотація. В работе рассмотрена проблематика части коммунальной отрасли, которая отвечает за взаимодействие коммунального предприятия и лиц, которых оно обслуживает. После проведенного анализа доказана целесообразность диджитализации и реализации WEB-ориентированного программного продукта, который удовлетворил бы все потребности, значительно упростил и ускорил процесс взаимодействия, требовал минимальных усилий на обучение, а также был доступен пользователям различных компьютерных и мобильных устройств при наличии интернета. На базе этого, с целью автоматизации и усовершенствования процессов получения актуальной информации и проектирование прозрачной системы, было описано особенности реализации информационной системы автоматизации подачи заявок в коммунальные предприятия, ее архитектуру, структуру и программную реализацию такими инструментами как NodeJS, ReactJS, GraphQL, MongoDB. В результате разработана информационная система автоматизации подачи заявок в коммунальные предприятия, которая полностью удовлетворяет базовые потребности пользователей услуг коммунальных предприятий и упрощает процесс получения услуг и информации.

Ключевые слова: информационная система, WEB-технологии, методологии проектирования программного обеспечения, автоматизация подачи заявок, коммунальные предприятия, мониторинг.

Abstract. The work considers the issue of the part of the utility industry that is responsible for the interaction of the utility company and the people it serves. After the analysis, the feasibility of digitalization and implementation of a WEB-oriented software product was proved, which would satisfy all the needs and significantly simplify and accelerate the interaction process, required minimal training efforts, and was also available to users of various computer and mobile devices with the presence of the Internet. Based on this, in order to automate and improve the processes of obtaining relevant information and designing a transparent system, the features of the implementation of information technology for automating applications to utilities, its architecture, structure and program implementation of tools such as NodeJS, ReactJS, GraphQL, MongoDB were described. As a result, an information system for automating the submission of applications to utilities has been developed, which fully satisfies the basic needs of users of utility services and simplifies the process of obtaining services and information.

Keywords: information system, WEB-technologies, software design methodologies, automation of application submission, utilities, monitoring.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-23-29>.

Вступ

Ми живемо в добу, де панують цифрові технології, у час, коли усі процеси підлягають автоматизації для подальшого їх спрощення, у час, коли більшість користувачів користуються веб-платформами з мобільних пристроїв, оскільки це зручно і доступно у будь-який момент. Галузь комунальних підприємств залишається однією з невід'ємних частин щоденного життя пересічних громадян і у теперішній реалізації не є сучасним рішенням, тому і потребує автоматизації, що дозволило б спростити процеси взаємодії та обміну інформації, прискорити їх та структурувати. Проте, сучасна реалізація потребує сучасних рішень з використанням перевірених технологій обраних під потреби та задачі системи, що вона вирішує. Досягнення поставленої цілі неможливо і без правильних підходів вибору архітектури та користувацького інтерфейсу, адже саме сукупність усіх влучно підібраних інструментів є запорукою створення повноцінної інформаційної технології.

Актуальність

У сучасному світі фактор часу є надзвичайно важливим, тому жодна людина не хоче марнувати його в довгезелних чергах для простого отримання інформації, і тим більше повторювати цю операцію ще декілька разів або витримувати довгезелне очікування на гарячій лінії, яке може не принести жодного результату. Адже це все забирає багато часу, нервів та сил, а також псує настрої і подекуди може навіть нашкодити здоров'ю. Саме тому, у більшості прогресивних країн установи намагаються спростити та автоматизувати такі процеси, щоб їх громадяни були не прив'язані до місця і не марнували так багато часу, а наприклад сидючи вдома чи перебуваючи на роботі за лічені хвилини могли отримати чітку та актуальну інформацію щодо роботи комунального підприємства [1].

Оскільки держава розуміє необхідність діджиталізації комунальної галузі, в результаті низки реформ було створено законопроект, що підтверджує актуальність такого рішення [2].

Варто зазначити, що в ході проведених досліджень, прямих аналогів не знайдено, проте є ряд систем, що вирішують схожі проблеми [3]:

Гаряча лінія – система, що функціонує вже багато років, де можна отримати інформацію (або надати її) про певні поломки. Проте цей спосіб організації має ряд таких недоліків: довге очікування, що подекуди змушує відмовитись від використання; ускладнений, а подекуди неможливий доступ до інформації; неактуальна інформація; відсутність прозорості; присутній значний людський фактор.

Система «Прозорий офіс» – система, що поєднує в собі багато можливостей, але для взаємодії, саме з комунальною галуззю є відносно складною, орієнтованою на більш глобальний рівень та не завжди відповідає специфіці саме комунальної галузі [3].

Ситуаційний центр – система, що в цілодобовому режимі відстежує всі аварії на лініях. Але такий центр орієнтований на більш глобальний рівень та не включає безліч речей, які стосуються комунальних підприємств таких як «ЖЕК» або «ОСББ».

Є також досить зручні веб-системи, але вони орієнтовані у переважній більшості на оплату послуг і розподіл бюджетів. Оскільки мають іншу спеціалізацію, тому не були детально розглянуті у даному дослідженні.

Ряд недоліків у схожих системах: непрозорість та відсутність функціоналу, що б спростив та автоматизував процес подачі заявок та отримання інформації. Все вищевказане стало передумовою для розробки системи автоматизації подання заявок до комунальних підприємств.

Мета

Метою дослідження є розробка інформаційної системи автоматизації подання заявок до комунальних підприємств, огляд структури, інструментів реалізації їх переваг та недоліків, як наслідок виділення особливостей організації цієї системи.

Задачі

1. Дослідити особливості системних вимог та архітектури інформаційної системи автоматизації подання заявок до комунальних підприємств.

2. Розробити структурну організацію та програмну реалізацію інформаційної системи автоматизації подання заявок до комунальних підприємств.

Розв'язання задач

Системні вимоги та архітектура інформаційної системи автоматизації подання заявок до комунальних підприємств

В основу розробки запропонованої архітектури системи автоматизації покладено такі особливості:

- система має забезпечувати взаємодію користувачів та комунальних підприємств;
- наявність модуля інформування, що спростить процедуру отримання інформації для користувачів послуг комунальних підприємств;
- введення модуля локалізації, що надасть змогу користувачу бачити актуальну інформацію, що стосується саме його району проживання або локалізації;
- забезпечення захищеного доступу, тому важливим елементом системи є модуль входу та реєстрації на етапі якого мають бути розділені звичайні користувачі та користувачі-адміністратори. Також на цьому етапі має відбуватись певна фільтрація для уникнення відсотку неправдивої інформації, що може лише змарнувати час на перевірку змісту заявки та самого випадку;
- система має забезпечувати можливості розподілення ролей, оскільки система має однаково задовольняти потреби користувачів; робітників, що працюють в бригадах та користувачів-адміністраторів, що контролюють процес робіт та заявок;
- користувачі з різним рівнем підготовки можуть повноцінно користуватись системою, час на навчання є мінімальним.

Ключовим елементом інформаційної системи автоматизації подання заявок до комунальних підприємств є підсистема, що реалізує саме модуль автоматизації процесу подання заявок, можливість створення і подання заявки в один клік, що значно спрощує цю процедуру та заощаджує час. Також цей модуль включає відображення актуальної інформації про них та реалізації взаємодій між користувачами, користувачами та комунальним підприємством.

Найбільш вдалим користувацьким інтерфейсом з точки зору UX-дизайну для відображення прозорого стану заявок, взаємодії та робіт, які проводились чи ні, оцінки проведених робіт та відгуків, на базі яких можливе формування заохочень до продуктивнішої роботи бригад, що виконують роботи, було обрано артефакт «дошка» з таких гнучких сучасних методологій, як Agile, Scrum, Kanban [4], що ідеально задовольняє потреби системи.

Потреби, які задовольняє реалізація модуля взаємодії із заявками у вигляді дошки:

- користувач з легкістю може орієнтуватись у статусах заявок, оскільки межі колонок є чіткими, а статуси інтуїтивно зрозумілими;

- користувач з привілеями може змінювати статус заявок через інтерфейс максимально швидко та інформативно, що буде відображено в реальному часі та не потребує знань бази даних;

- робочі бригади можуть здійснювати моніторинг кількості робіт та їх статуси, що дає змогу структурувати роботу та зробити її більш продуктивною, а також побачити повний «score» роботи.

Ця дошка повинна включати в себе такі речі як: колонки з чітко визначеними статусами, де кожна колонка демонструє актуальний статус задачі, картки задач з інформацією про них, можливість переходу з одного статусу в інший, виділення відповідальної людини шляхом призначення на неї певної картки задачі [5].

Основними функціональними можливостями, що повинні бути реалізовані в цій системі, є:

- можливість віддаленого доступу через мережу інтернет;
- кросплатформність та кросбраузерність;
- наявність можливості кешування даних для забезпечення швидкодії відображення інформації;
- можливість реєстрації та підтримка різних видів користувачів відокремлених певними можливостями;
- можливість створити та подати заявку;
- відображення списку усіх заявок та надання актуальної інформації;
- відображення актуального статусу робіт по заявці;
- підтримка реєстрації з «розподіленням ролей», що буде надавати певній ролі дещо розширені можливості взаємодії з системою;
- можливість зміни статусу заявки для ролі користувач-адміністратор або користувач-робітник;
- можливість залишити відгук про пророблену роботу;
- підтримка оновлень у реальному часі.

Нижче наведений приклад діаграми активності на верхньому рівні деталізації роботи системи:

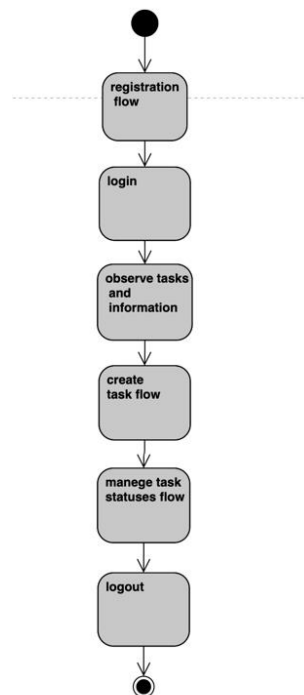


Рисунок 1 – Діаграми активності на верхньому рівні деталізації роботи системи

Структурна організація інформаційної системи автоматизації подання заявок до комунальних підприємств

У системі потрібно реалізувати основні модулі, які б забезпечували виконання основних вимог до системи, а саме:

- взаємодію між об'єктами системи;
- взаємодію між системою та користувачем;
- виконання функцій кожного об'єкту;
- відображення елементів на екрані.

Основними модулями програми є такі:

- модуль реєстрації;
- модуль відображення інформації;
- модуль створення заявок;
- модуль взаємодії з заявкою як сутністю;
- модуль кастомізації особистих даних – профайл.

Користувачами цієї системи є: комунальні підприємства, громадяни (власники приміщень або спостерігачі), ОСББ, виконавці робіт (бригади робітників), органи контролю виконаних робіт та оператори. Діаграму прецедентів системи представлено на рис. 2.

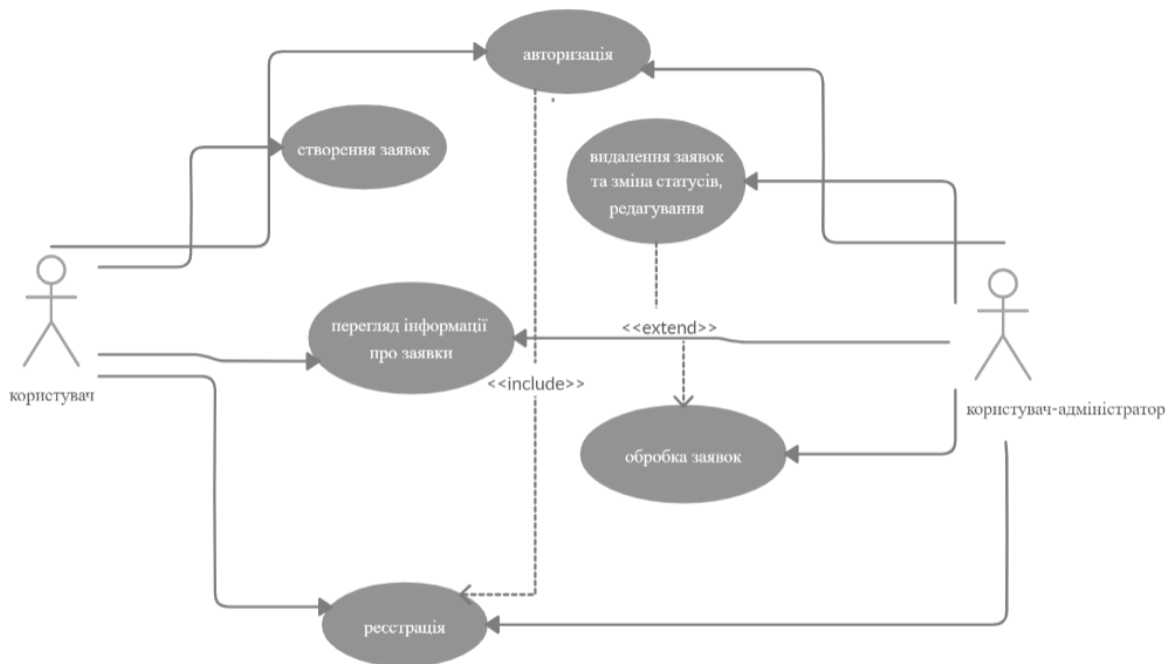


Рисунок 2 – Діаграма прецедентів системи

Деталізовану структурну схему ключового модуля взаємодії з заявками до комунальних підприємств подано на рис. 3.

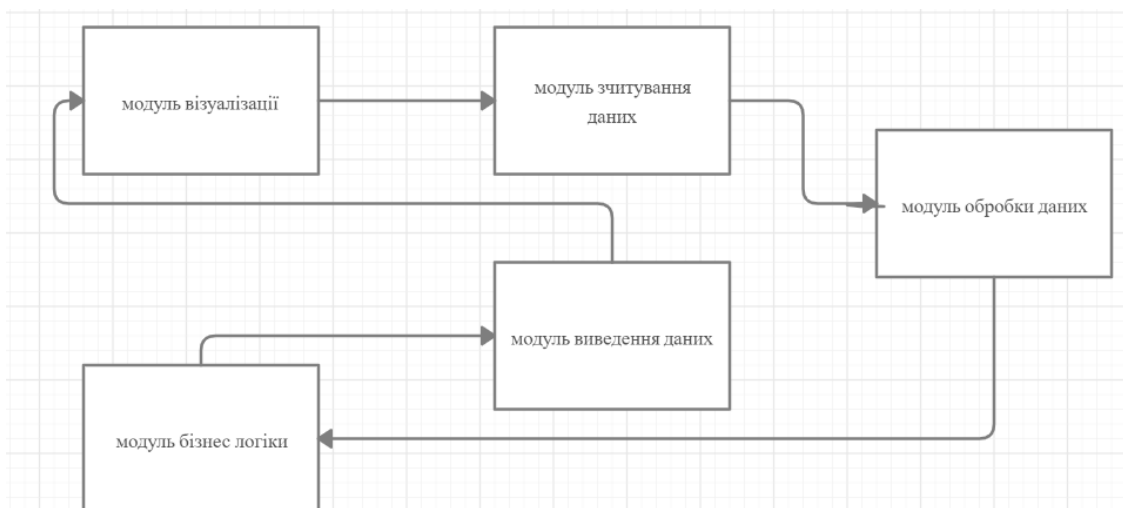


Рисунок 3 – Структурна схема програмного модуля автоматизації подання заявок до комунальних підприємств

Модуль візуалізації – це модуль користувацького інтерфейсу, який є візуальним представленням системи, напряму пов’язаний з модулями зчитування та виведення інформації. Головними функціями

якого є: відображення, можливість введення інформації та надання зв'язків, що будуть в подальшому використовувати інші модулі.

Модуль зчитування даних – модуль, що відповідає за зчитування введених даних, структурує їх, адаптує під певний формат та передає їх на подальшу обробку.

Модуль обробки даних – модуль, що безпосередньо відповідає за обробку отриманих даних та проводить всі відповідні маніпуляції по даним до того, як вмикається апарат бізнес логіки. Так званий про-шарок, що відповідає за з'єднання клієнта і серверу та БД.

Модуль виведення даних – модуль, що відповідає за збір даних та формування їх для подальшого візуального представлення у модулі візуалізації.

Модуль бізнес-логіки – модуль, що вміщає в себе всю бізнес-логіку, а саме функції, що вирішують основні задачі, наприклад: створення заявки, також модуль включає взаємодію між об'єктами системи та їх поведінку, що включає в себе модуль авторизації, модуль створення та подання заявки, модуль допоміжної бізнес-логіки.

Програмна реалізація інформаційної системи автоматизації подання заявок до комунальних підприємств

Критерії вибору програмних засобів:

- можливість створення шаблонів коду;
- наявність об'єктно-орієнтованого проектування;
- сучасний підхід;
- високий рівень оптимізації;
- наявність функціоналу для зберігання локального стану елементів;
- можливість розробки, як фронт так і бек частини сайту з використанням лише однієї мови;
- підтримка бібліотек для роботи з базою даних;
- можливість кешування даних;
- підтримка реалізації кросплатформного веб-додатку з акцентом на користувачів мобільних пристроїв.

Одним із найбільш вдалих варіантів, що надає можливість реалізувати всі вище описані вимоги є JavaScript, але для кращої реалізації доцільно буде використовувати фреймворк. Після проведеного аналізу варіантів реалізації front-end частини відповідно до вимог системи обрано ReactJS [5, 6], який має високий рівень оптимізації та покриває усі потреби.

Для серверної частини єдиним можливим варіантом, що покрив би усі вказані потреби є NodeJS. NodeJS – серверна платформа, що призначена для створення масштабованих розподілених серверних додатків, яка використовує подієво-орієнтовану архітектуру та неблокуючу асинхронну взаємодію [7]. Головна ідея Node.js: використання неблокуючого подієво-орієнтованого введення/виведення, щоб залишатися «легким» та ефективним під час роботи з додатками, що обробляють великі обсяги даних в реальному часі і функціонують на розподілених пристроях.

Щодо реалізації зв'язку клієнт-серверу, за результатами дослідження системи подачі заявок до комунальних підприємств можна стверджувати: при використанні Rest API ми маємо архітектуру з великою кількістю «endpoints», що призводить до величезної кількості запитів у різні місця, проте найбільшим мінусом такої архітектури є складність побудови запитів, а також «over fetching». А це означає, що з виконанням кожного запиту майже завжди тягнеться безліч «непотрібних» даних, що сповільнює роботу системи. Крім того, кешування і офлайн організація є досить складними у реалізації з використанням Rest API. Стосовно GraphQL, використання цієї технології дозволяє вирішити проблему зайвих даних, оскільки має структуру, яка дозволяє в одному запиті отримати усі потрібні запити роблячи запит на один endpoint. Така поведінка забезпечується при організації GraphQL серверу, що має одну точку входу, проте об'єднує множини інформації. Також GraphQL використовується у поєднанні з Apollo як на клієнті так і на сервері, що в свою чергу надає ряд переваг: чітка типізація, кешування, оптимістичний користувацький інтерфейс, що дозволяє оновлювати інтерфейс в процесі відправлення запиту, також таке поєднання зменшує навантаження на мобільному пристрої, що суттєво прискорює роботу системи [8, 9].

Отже, оскільки система подання заявок та їх моніторинг передбачає такі процеси як: створення заявок, корегування даних, відображення інформації про заявки, обробка заявок, коментарі, профайли користувачів, розподіл ролей, а також модуль живого спілкування – це означає, що в системі буде велике навантаження і велика кількість запитів [1]. Алгоритм кешування допомагає вирішити певні проблеми. Apollo кеш передбачає кешування деяких запитів з можливістю оновлення лише тих полів, які змінюються за допомогою функцій запису до кешу, а зчитування даних з кешу допоможе уникнути великої кількості однакових запитів, що в свою чергу зменшує навантаження і додає можливість відображення певної інформації в офлайн режимі. Також така оптимізація позитивно впливає на користувачів з мобільних пристроїв. Функція оптимістичного інтерфейсу добре підходить для реалізації чату, оскільки допомагає відображати повідомлення вже в процесі відправлення.

Отже, реалізація системи потребує впровадження GraphQL на зміну Rest API.

Оскільки програмний модуль буде розроблено у вигляді веб-сайту на технології ReactJS для front-end частини та NodeJS для back-end, то база даних має бути максимально простою та через суміжні технології максимально підходити. Оскільки для запитів було обрано застосування GraphQL, через використання GraphQL база даних має підтримувати зберігання записів в форматі json для більшої зручності.

Основними вимогами до бази даних є:

- здатність до масштабованості;
- можливість розширення;
- швидкість розробки;
- можливість зберігання даних в форматі json;
- гнучкість і легка зміна схем;
- сумісність з GraphQL та Apollo.

Після проведеного аналізу архітектурних потреб системи для реалізації бази даних обрано MongoDB, що має ряд переваг: відсутність схеми, оскільки бізнес-модель ще не до кінця визначена і з великою ймовірністю проект до виходу на ринок зазнає багато змін, в тому числі на рівні організації даних [10]. MongoDB саме те, що потрібно, адже на відміну від SQL в MongoDB немає необхідності створювати таблиці, змінювати їх схеми, створювати міграції, перейматися типами даних, багатий функціонал агрегації. Дані можна отримувати за допомогою MongoDB набагато простіше і зручніше, а саме через map-reduce: групування за складними умовами, переформатування документів миттєво, отримання випадкових документів, сортування.

Також дані можна або перемістити, тому що відсутність схеми це і має на увазі, або просто продублювати в потрібне місце. Тобто, можливість мати одне і те ж поле з одними і тими ж даними, але в різних колекціях. Або два поля в одній колекції, а плюс до них ще одне поле, яке є композицією перших двох [11].

Висновки

Встановлено, що розробка інформаційної системи автоматизації подання заявок до комунальних підприємств є доцільною та актуальною задачею.

1. Сформовано основні потреби та особливості реалізації системи у вигляді Web-технології, що значно спрощує процес взаємодії між комунальним підприємством та особою, яку воно обслуговує, що в свою чергу, значно заощаджує час, структурує інформацію та саму роботу комунальних бригад. Такий підхід надає змогу усім користувачам мати швидкий доступ до платформи за наявності мережі з мобільних пристроїв, планшетів, ноутбуків, персональних комп'ютерів тощо.

2. Здійснено аналіз вимог до інструментальних засобів, яких потребує реалізація системи, та як наслідок реалізовано її за допомогою мови програмування JS, на базі двох бібліотек: ReactJs, NodeJS для клієнтської та серверної частини відповідно. Для запитів обрано мову запитів GraphQL, що забезпечує вищий рівень оптимізації запитів, ніж RestAPI та спроектовано базу даних MongoDB.

Як результат, розроблено інформаційну систему автоматизації подання заявок до комунальних підприємств.

В подальших дослідженнях планується дослідити доцільність застосування методів штучного інтелекту, для подолання «людського фактору» в обробці задач для зменшення часу витраченого оператором на роботу, а також для збільшення швидкодії роботи самої системи. Також планується реалізувати аналіз потреб користувачів та локалізацію інформаційної системи під певний регіон.

Список літератури

- [1] І. В. Безсмертна, Л. В. Крилик, "Перспективи впровадження автоматизованої системи подання заявок та їх відстеження," *Матеріали XLIX науково-технічної конференції факультету інформаційних технологій та комп'ютерної інженерії*, с. 919 – 920, 2020. [Електронний ресурс]. Режим доступу: https://conferences.vntu.edu.ua/public/files/1/vntu_2020_netpub.pdf. Дата звернення: Черв. 25, 2021.
- [2] Закон про житлово-комунальні послуги – Законодавство України, 2021. [Електронний ресурс]. Режим доступу: <https://zakon.rada.gov.ua/laws/show/1875-15>. Дата звернення: Черв. 25, 2021.
- [3] Прозорий офіс, 2021. [Електронний ресурс]. Режим доступу: <https://transparent.vmr.gov.ua/default.aspx>. Дата звернення: Черв. 25, 2021.
- [4] Kanban, 2021. [Online]. Available: <https://kanbanize.com/kanban-resources/getting-started/what-is-kanban-board>. Accessed on: June 25, 2021.
- [5] ReactJs, 2021. [Online]. Available: <https://ru.reactjs.org/>. Accessed on: June 25, 2021.
- [6] Порівняльна характеристика фреймворків для створення користувацького інтерфейсу, 2021. [Електронний ресурс]. Режим доступу: <https://mkdev.me/posts/sravnenie-javascript-freymvorkov-vue-js-react-i-angular-2019>. Дата звернення: Черв. 25, 2021.
- [7] NodeJs, 2021. [Online]. Available: <https://nodejs.org/ru/docs/>. Accessed on: June 25, 2021.
- [8] GraphQL, 2021. [Online]. Available: <https://graphql.org/learn/>. Accessed on: June 25, 2021.

- [9] Порівняння Rest API та GraphQL, 2021. [Електронний ресурс]. Режим доступу: <https://www.howtographql.com/basics/1-graphql-is-the-better-rest/>. Дата звернення: Черв. 25, 2021.
- [10] MongoDB, 2021. [Online]. Available: <https://www.mongodb.com/>. Accessed on: June 25, 2021.
- [11] Переваги MongoDB, 2021. [Електронний ресурс]. Режим доступу: <https://echo.lviv.ua/dev/9693>. Дата звернення: Черв. 25, 2021.

Стаття надійшла: 18.08.2021.

References

- [1] V. Bezsmertna, L. V. Krylyk, "Perspektyvy vprovadzhennia avtomatyzovanoi systemy podannia zaiavok ta yikh vidstezhennia," *Materialy XLIX naukovo-tekhnichnoi konferentsii fakultetu informatsiinykh tekhnolohii ta kompiuternoї inzhenerii*, s. 919 – 920, 2020. [Elektronnyi resurs]. Rezhym dostupu: https://conferences.vntu.edu.ua/public/files/1/vntu_2020_netpub.pdf. Data zvernennia: Cherv. 25, 2021. – [in Ukrainian].
- [2] Law on housing and communal services – Legislation of Ukraine, 2021. [Online]. Available: <https://zakon.rada.gov.ua/laws/show/1875-15>. Accessed on: June. 25, 2021. – [in Ukrainian].
- [3] Transparent office, 2021. [Online]. Available: <https://transparent.vmr.gov.ua/default.aspx>. Accessed on: June. 25, 2021. – [in Ukrainian].
- [4] Kanban, 2021. [Online]. Available: <https://kanbanize.com/kanban-resources/getting-started/what-is-kanban-board>. Accessed on: June 25, 2021.
- [5] ReactJs, 2021. [Online]. Available: <https://ru.reactjs.org/>. Accessed on: June 25, 2021.
- [6] Comparative characteristics of frameworks for creating a user interface, 2021. [Online]. Available: <https://mkdev.me/posts/sravnenie-javascript-freymvorkov-vue-js-react-i-angular-2019>. Accessed on: June. 25, 2021. – [in Ukrainian].
- [7] NodetJs, 2021. [Online]. Available: <https://nodejs.org/ru/docs/>. Accessed on: June 25, 2021.
- [8] GraphQL, 2021. [Online]. Available: <https://graphql.org/learn/>. Accessed on: June 25, 2021.
- [9] Comparison of Rest API and GraphQL, 2021. [Online]. Available: <https://www.howtographql.com/basics/1-graphql-is-the-better-rest/>. Accessed on: June. 25, 2021. – [in Ukrainian].
- [10] MongoDB, 2021. [Online]. Available: <https://www.mongodb.com/>. Accessed on: June 25, 2021.
- [11] Advantages of MongoDB, 2021. [Online]. Available: <https://echo.lviv.ua/dev/9693>. Accessed on: June. 25, 2021. – [in Ukrainian].

Відомості про авторів

Крилик Людмила Вікторівна – кандидат технічних наук, доцент, доцент кафедри комп'ютерних наук.

Яровий Андрій Анатолійович – доктор технічних наук, професор, завідувач кафедри комп'ютерних наук.

Безсмертна Інна Василівна – студентка групи 2КН-20м, факультет інформаційних технологій та комп'ютерної інженерії.

Л. В. Крылик, А. А. Яровой, И. В. Бессмертная

ОСОБЕННОСТИ РЕАЛИЗАЦИИ ИНФОРМАЦИОННОЙ СИСТЕМЫ АВТОМАТИЗАЦИИ ПОДАЧИ ЗАЯВОК В КОММУНАЛЬНЫЕ ПРЕДПРИЯТИЯ

Винницкий национальный технический университет, Винница

L. Krylik, A. Yarovy, I. Bezsmertna

PECULIARITIES OF IMPLEMENTATION OF INFORMATION SYSTEM OF AUTOMATION SUBMISSION OF APPLICATIONS TO UTILITY ENTERPRISES

Vinnitsia National Technical University, Vinnitsia

УДК 004.67(075.8)

Т. Б. Мартинюк, О. В. Войцеховська

ЕФЕКТИВНІСТЬ ОДИНИЧНОГО КОДУВАННЯ ДАНИХ

Вінницький національний технічний університет, Вінниця

Анотація. В роботі проведено аналіз реалізаційних властивостей одиничних кодів як альтернативних до класичних двійкових. Показана можливість формування логіко-часового коду, який є симбіозом двох відомих одиничних кодів: одиничного нормального та одиничного позиційного (маркувального) коду. Це дозволяє забезпечити прискорений принцип запису інформації та меншу енергоємність при її збереженні. Наведено аналітичні вирази для розрахунку апаратних та часових витрат при конкретній реалізації одиничного кодування даних на зсувному регістрі та двійковому лічильнику з дешифратором. Саме відсутність дешифрації сформованих даних потребує збільшення кількості зв'язків у схемах в $\lceil N \log_2 N \rceil$ разів. Широке застосування програмованих логічних ІС дозволяє усунути цю проблему при компактному розміщенні пристроїв з одиничним кодуванням у мікросхемі ПЛІС. Це дає можливість використовувати одиничні коди як альтернативу двійковим кодам при передачі даних, а також у деяких типах пристроїв керування при кодуванні їх станів і у запам'ятовуючих пристроях при адресації їх вмісту. Як приклад, показано ефективність використання одиничних кодів для кодування станів мікропрограмних автоматів, що досягається відносною простотою комбінаційних схем та відсутністю необхідності декодування кодових комбінацій станів автомата.

Ключові слова: одиничний код, ефективність кодування, кодування станів, мікропрограмний автомат.

Аннотация. В работе проведен анализ реализационных свойств единичных кодов как альтернативных классическим двоичным. Показана возможность формирования логико-временного кода, который является симбиозом двух известных единичных кодов: единичного нормального и единичного позиционного (маркировочного) кода. Это позволяет обеспечить ускоренный принцип записи информации и меньшую энергоёмкость при ее сохранении. Приведены аналитические выражения для расчета аппаратных и временных затрат при конкретной реализации единичного кодирования данных на сдвиговом регистре и двоичном счетчике с дешифратором. Именно отсутствие дешифрации сформированных данных вызывает увеличение количества связей в схемах в $\lceil N \log_2 N \rceil$ раз. Широкое применение программируемых логических ИС позволяет устранить эту проблему при компактном размещении устройств с единичным кодированием в микросхеме ПЛИС. Это дает возможность использовать единичные коды в качестве альтернативы двоичным кодам при передаче данных, а также в некоторых типах устройств управления при кодировании их состояний и в запоминающих устройствах при адресации их содержания. В качестве примера показана эффективность использования единичных кодов для кодирования состояний микропрограммных автоматов, что достигается относительной простотой комбинационных схем и отсутствием необходимости декодирования кодовых комбинаций состояний автомата.

Ключевые слова: единичный код, эффективность кодирования, кодирование состояний, микропрограммный автомат.

Abstract. The paper analyzes the implementation properties of unit codes as alternatives to classical binary ones. The possibility of forming a logical-temporal code, which is a symbiosis of two known unit codes: a unit normal and a unit positional (marking) code, is shown. This allows you to provide accelerated principle of recording information and less energy-intensive while storing it. Analytical expressions for calculation of hardware and time expenses at concrete realization of unit data encoding on the shift register and the binary counter with the decoder are resulted. It is the lack of decryption of the generated data that requires increasing the number of connections in the schemes in $\lceil N \log_2 N \rceil$ times. The widespread use of FPGAs eliminates this problem with the compact placement of unit-coded devices in the FPGA chip. This makes it possible to use unit codes as an alternative to binary codes in data transmission, as well as in some types of control devices when encoding their states and in storage devices when addressing their contents. As an example, the efficiency of using unit codes for encoding the states of microprogram automata is shown, which is achieved by the low complexity of combinational circuits and the absence of the need to decode the code combinations of automaton states.

Key words: unit code, coding efficiency, state coding, microprogram automaton.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-30-36>.

Вступ

Принципи машинного кодування числової інформації відіграють значну роль у забезпеченні важливих характеристик комп'ютерних засобів, а саме, швидкодію, надійність, контролездатність, вартість, тощо [1–3]. Ефективність будь-якої системи числення можна оцінити, враховуючи такі показники, як складність алгоритмів перетворення при прийомі даних та видачі результатів, складність подання чисел, складність та швидкодія пристроїв для виконання арифметичних та інших типових операцій над числами [4, 5].

Метою даної роботи є визначення ефективності одного з варіантів альтернативного кодування числової інформації, а саме одиничного кодування з врахуванням апаратних та часових витрат.

Постановка задачі

У роботах [6, 7] детально розглянуто два різновиди одиничного кодування для логіко-часової обробки десяткової інформації: одиничний нормальний та одиничний позиційний (маркувальний) код. В результаті дослідження обох кодів по базовим критеріям, а саме за швидкодією, енергоємністю та заводостійкістю було запропоновано новий код, що є їх симбіозом. Він отримав назву логіко-часового коду (ЛЧК), оскільки був призначений для використання у специфічному однорідному середовищі [6–8]. Враховуючи, що принцип запису даних з використанням одиничного нормального коду є більш швидкісним, а з використанням позиційного коду – найменш енергоємним, то ЛЧК виявився зорієнтованим на максимально можливу схемотехнічну швидкодію при записі інформації та мінімально можливу потужність при її збереженні. Порівняння за цими показниками виконувалось серед двох наведених одиничних кодів [6–8].

Особливість двох різновидів одиничних кодів також детально розглянуто у роботах [9, 10] їх функціональні можливості та класифікацію за базовими ознаками [11] наведено у роботах [12, 13].

Оскільки одиничні коди було запропоновано для подання десяткової інформації [6–9], то доречно ці коди представити у вигляді матриці кодування D_r^n розмірністю $r \times n$, де r – основа системи числення, а n – кількість двійкових розрядів. У цій матриці рядки представляють коди десяткових цифр $a_i \in \{0, \dots, 9\}$, а стовбці відповідають двійковим розрядам $j \in \{0, \dots, 9\}$ кожного коду цифри a_i [14].

Таким чином, матриця кодування для одиничного нормального коду має вигляд:

$$D_{10}^{10} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \quad (1)$$

а для одиничного позиційного коду вигляд:

$$D_{10}^{10} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (2)$$

Як видно з наведених матриць кодування D_{10}^{10} (1) і (2), для представлення кожної десяткової цифри a_i використовується максимальна кількість двійкових розрядів $n = 10$.

Особливості реалізації одиничних кодів

Апаратна реалізація будь-якої недвійкової системи числення, зокрема десяткової, виконується за умови використання схем з багатозначним поданням інформації [15].

Одним з прикладів таких багатозначних функціональних елементів, зорієнтованих на одиничне кодування числової інформації, є багатофункціональні модулі [6–8, 16], які не замінюють двійкові елементи, а розширюють область апаратної реалізації специфічного одиничного кодування числової інформації.

Разом з тим, за своїми функціональними можливостями та принципами схмотехнічної побудови багатофункціональні оптоелектронні модулі [6–8, 16] схожі з реверсивними зсувними регістрами [5]. Але для кількісного оцінювання доцільності вибору одиничної системи кодування розглянемо апаратні витрати на прикладі лічильників з врахуванням необхідності дешифрації двійково-кодуваної інформації на їх виходах.

Для лічильника з двійковим кодуванням та повним дешифратором апаратні витрати, що еквівалентні кількості входів вентилів, визначаються таким чином [17]:

$$P \approx S \lceil \log_2 N \rceil + q \lceil \log_2 N \rceil \cdot N = (S + q \cdot N) \lceil \log_2 N \rceil, \quad (3)$$

де N – кількість станів лічильника; $\lceil \log_2 N \rceil$ – кількість розрядів лічильника; S – умовна кількість елементів на один розряд лічильника, який містить тригерний каскад з входом лічби; q – коефіцієнт пропорційності між кількістю розрядів лічильника та середніми апаратними витратами на один вихід дешифратора з урахуванням кількості логічних та додаткових схем.

Для зсувного регістра з одиничним кодуванням апаратні витрати дорівнюють:

$$P' \approx p \cdot N, \quad (4)$$

де p – апаратні витрати на один розряд регістра, який містить тригерний каскад з роздільними входами.

Аналіз наведених співвідношень свідчить, що за умови

$$N > N_{кр} \approx 2^{p/q}, \quad (5)$$

використовувати одиничне кодування вигідніше, ніж двійкове, причому зі зростаючою ефективністю при подальшому збільшенні N [1, 3 – 5, 17].

При аналізі часових характеристик необхідно врахувати, що час перемикання зсувного регістра з одиничним кодуванням дорівнює часу перемикання тригерного каскаду з роздільними входами t_p , тобто

$$T = t_p. \quad (6)$$

Сумарний час перемикання двійкового лічильника з груповим перенесенням та вихідним лінійним дешифратором можна визначити таким чином [1, 4, 5, 17]:

$$T_{\Sigma} = t_n + t_g, \quad (7)$$

а при багатокаскадній реалізації як:

$$T_{\Sigma} = t_n + (2k + 1)t'_g, \quad (8)$$

де t_n – час перемикання тригерного каскаду з входом лічби; t_g , t'_g – час спрацювання вентиля відповідно з підсиленням вихідного сигналу і з урахуванням міжкаскадного підсилення сигналів; k – кількість каскадів для реалізації кон'юнкцій n змінних.

Отже, швидкодія зсувного регістра з одиничним кодуванням вище, ніж у лічильника з двійковим кодуванням та вихідним дешифратором [17]. Після того, як з'ясувалось, що зсувні регістри з одиничним кодуванням інформації мають переваги перед двійковими лічильниками за апаратними та часовими витратами, виникає питання щодо задач, де використання одиничного кодування є ефективнішим, ніж двійкове.

Враховуючи, що аналіз одиничних кодів показав їх завадостійкість та контролездатність [12, 13], то їх можна використовувати як альтернативу двійковим кодам при передачі даних з урахуванням схематичних можливостей, а також у деяких типах пристроїв керування при кодуванні їх станів і у запам'ятовуючих пристроях при адресації їх вмісту. Ефективність використання одиничних кодів у наведених прикладах пов'язана, в першу чергу, з відсутністю необхідності застосування дешифраторів. Але, разом з тим, виникає необхідність у збільшенні кількості комутаційних зв'язків у $N/\lceil \log_2 N \rceil$ разів. Широке застосування на програмованих логічних ІС (ПЛІС) дозволяє усунути цю проблему при компактному розміщенні не тільки конкретних пристроїв, але й систем і комплексів [18].

Можливості одиничного кодування станів мікропрограмних апаратів

Розглянемо вплив способу кодування станів мікропрограмного автомата (МПА) і, зокрема, одиничного кодування, на складність реалізації його комбінаційної схеми, а також на залежність її від коефіціє-

нта k_p розгалуження мікропрограми, що інтерпретується [19, 20]. Пропонується застосування одиничного кодування для альтернативних R-автоматів [21], в яких як елементи пам'яті використовуються двійкові зсувні регістри на відміну від СТ- і Т-автоматів, запам'ятовувальна частина яких реалізується відповідно на двійкових лічильниках і паралельних регістрах [22].

Відомо, що від обраного способу кодування станів МПА залежить об'єм пам'яті, стійкість автомата та його швидкодія [19, 20]. Етап кодування впливає на складність логічного перетворювача МПА, а вибір коду для розміщення внутрішніх станів МПА дозволяє збільшити надійність автомата [19, 20].

Якщо використати для кодування внутрішніх станів МПА одиничний позиційний (маркувальний) код з метою усунення наступної дешифрації кодових комбінацій, то це приведе до збільшення кількості елементів пам'яті МПА з одночасним спрощенням його логічного перетворювача [23, 24]. Разом з тим, значне збільшення у $N/\lceil \log_2 N \rceil$ разів кількості розрядів R-автомата з одиничним кодуванням у порівнянні з R-автоматом з будь-яким неодноразрядним кодуванням станів досягається шляхом «нарощування» розрядності зсувного регістра. А це не складає труднощів при реалізації такого зсувного регістра з його зв'язками на ПЛІС [18].

Крім того, одиничний спосіб кодування внутрішніх станів МПА дозволяє мінімізувати кількість елементарних кон'юнкцій в ДНФ для функцій переходу МПА [24, 25], що приводить до зменшення складності синтезу комбінаційних схем [23]. Так, величини π і ρ , які визначають кількість елементарних кон'юнкцій (до мінімізації) у запису ДНФ функцій переходу для R- та Т-автомата відповідно, мають такий вигляд [26]:

$$\pi = \sum_{j=1}^r L_j + \sum_{i \in G} \mu(a_i) \cdot K(a_i), \quad (9)$$

$$\rho = \sum_{i \in H} \sigma(a_i) \cdot K(a_i), \quad (10)$$

де L_j – довжина j -го лінійного ланцюга; r – кількість лінійних ланцюгів; G – множина номерів входів лінійних ланцюгів; $K(a_i)$ – кількість одиниць у коді стану a_i ; $\mu(a_i)$ – кількість переходів у стан a_i , що не входять до лінійних ланцюгів; H – множина номерів станів автомата; $\sigma(a_i)$ – кількість переходів у стан a_i .

Значення першого доданку у формулі (9) не змінюється, оскільки воно формується на етапі абстрактного синтезу, тому зменшити можна тільки значення другого доданку за рахунок мінімальної кількості одиниць у кодах станів МПА. Саме це досягається за рахунок використання одиничного позиційного (маркувального) коду вигляду (2).

Доведено, що сумарна складність комбінаційних схем МПА в цьому випадку майже удвічі менша в порівнянні з відомими R-автоматами при реальній складності мікропрограм ($N > 5$), а також через відсутність будь-яких обмежень за типами автоматних графів стосовно R-автоматів [23–25].

Висновки

1. Існування таких специфічних кодів, як одиничні коди, які є альтернативними до класичних двійкових кодів, підтверджується їх ефективним застосуванням для вирішення конкретних практичних задач. Це стосується, наприклад, особливостей кодування станів керуючих автоматів.

2. До основних переваг використання одиничного кодування станів мікропрограмних автоматів можна віднести простоту синтезу комбінаційних схем, їх низьку складність та відсутність необхідності декодування (перетворення) кодових комбінацій станів автомата. Це пов'язано, в першу чергу, з усуненням необхідності дешифрації при використанні одиничних кодів для лічби та кодування керуючої інформації.

3. Розрахунок апаратних та часових витрат при реалізації одиничного кодування цифрових даних показав їх виграти при використанні зсувних регістрів у порівнянні з двійковими лічильниками з деширатором.

Список літератури

- [1] К. Г. Самофалов, В. И. Корнейчук, В. П. Тарасенко, *Электронные цифровые вычислительные машины: учебник для вузов*. Киев, Украина: Вища школа, 1983, 455 с.

- [2] М. П. Матвієнко, *Комп'ютерна логіка: підручник*. Київ, Україна: Видавництво Ліра-К, 2017, 324 с.
- [3] М. Ф. Бондаренко, Н. В. Білоус, А. Г. Руткас, *Комп'ютерна дискретна математика: підручник*. Харків, Україна: «Компанія СМІТ», 2004, 480 с.
- [4] М. П. Матвієнко, В. П. Розен, *Комп'ютерна схемотехніка: навчальний посібник*. Київ, Україна: Видавництво Ліра-К, 2016, 192 с.
- [5] Е. П. Угрюмов, *Цифровая схемотехника: учеб. пособие для вузов*. СПб., Россия: БХВ-Петербург, 2010, 816 с.
- [6] В. П. Кожемяко, *Оптоэлектронные логико-временные информационно-вычислительные среды*. Тбилиси, Грузия: Мецниереба, 1984, 360 с.
- [7] С. В. Свечников, В. П. Кожемяко, Л. И. Тимченко, *Квазиимпульсно-потенциальные оптоэлектронные элементы и устройства логико-временного типа*. Киев, Украина: Наукова думка, 1987, 256 с.
- [8] В. П. Кожемяко, О. Г. Натрошвили, Т. Б. Мартынюк, Л. Ш. Имнаишвили, *Оптоэлектронная схемотехника: учеб. пособие*. Киев: УМК ВО, 1988, 276 с.
- [9] Т. Б. Мартинюк, О. М. Тарасова, М. М. Аль-Хіярі, "Особливості логіко-часового зображення числової інформації," *Вісник Вінницького політехнічного інституту*, № 1, с. 72-76, 2000.
- [10] Т. Б. Мартинюк, М. М. Аль-Хіярі, С. А. Василецький, "Функційна повнота логічно-часового принципу зображення інформації," *Вісник Вінницького політехнічного інституту*, № 2, с. 48-52, 2000.
- [11] Э. Берлекэмп, *Алгебраическая теория кодирования: пер. с англ.*, М., Россия: Мир, 1971, 480 с.
- [12] Т. Б. Мартинюк, Мохамед Салем Нассер, В. В. Власійчук, О. М. Наконечний, "Аналіз можливостей одиничного кодування числової інформації," *Оптико-електронні інформаційно-енергетичні технології*, № 2 (10), с. 39-44, 2005.
- [13] В. П. Кожемяко, Т. Б. Мартинюк, В. В. Дмитрук, В. В. Власійчук, "Класифікація одиничних кодів," *Оптико-електронні інформаційно-енергетичні технології*. № 1 (11), с. 36-42, 2006.
- [14] З. Л. Рабинович, В. А. Раманаускас, *Типовые операции в вычислительных машинах*. Киев, Украина: Техніка, 1980, 264 с.
- [15] К. Г. Самофалов, В. И. Корнейчук, А. М. Романкевич, В. П. Тарасенко, *Цифровые многозначные элементы и структуры: учеб. пособие*. Киев, Украина: Вища школа, 1974, 168 с.
- [16] В. П. Кожемяко, Т. Б. Мартинюк, О. І. Суприган, Д. І. Клімкіна, *Квантові перетворювачі на оптоелектронних логіко-часових середовищах для око-процесорної обробки зображень: монографія*. Вінниця, Україна: УНІВЕРСУМ-Вінниця, 2007, 126 с.
- [17] З. Л. Рабинович, *Основы теории элементарных структур ЭВМ*. М., Россия: Радио и связь. – 1982, 280 с.
- [18] Р. И. Грушвицкий, А. Х. Мурсаев, Е. П. Угрюмов, *Проектирование систем на микросхемах программируемой логики*. СПб., Россия: БХВ-Петербург, 2002, 608 с.
- [19] В. І. Жабін, І. А. Жуков, І. А. Клименко, В. В. Ткаченко, *Прикладна теорія цифрових автоматів*. К., Україна: Книжкове вид-во НАУ, 2007, 364 с.
- [20] В. П. Кожемяко, Т. Б. Мартынюк, К. В. Кожемяко, "Синтез устройства управления на R-автомате," *Управляющие системы и машины*, № 1/2, с. 22–25, 1995.
- [21] Ю. А. Бузунов, И. Г. Буренков, Н. Н. Шипилов, "Микропрограммные автоматы на параллельно-последовательных структурах," *Управляющие системы и машины*, № 2, с. 26–29, 1982.
- [22] Т. Б. Мартинюк, К. В. Кожемяко, А. В. Кожемяко, "До оцінки складності комбінаційних схем R-автоматів," *Вісник Вінницького політехнічного інституту*, № 1, с. 31–34, 1997.
- [23] Т. Б. Мартинюк, А. В. Кожемяко, Н. В. Фофанова, "Два варіанти синтезу мікропрограмних R-автоматів," *Вісник Вінницького політехнічного інституту*, № 4, с. 47–53, 1998.
- [24] Т. Б. Мартынюк, "Особенности синтеза микропрограммных R-автоматов," *Управляющие системы и машины*, № 3, с. 22-26, 1998.
- [25] В. М. Кирпичников, В. А. Складаров, "Синтез микропрограммных автоматов по граф-схемам алгоритмов с малым числом условных вершин," *Управляющие системы и машины*, № 1, с. 77–83, 1978.

Стаття надійшла: 18.08.2021.

References

- [1] K. G. Samofalov, V. I. Kornejchuk, V. P. Tarasenko, *Jelektronnye cifrovye vychislitel'nye mashiny: uchebnyk dlja vuzov*. Kiev, Ukraina: Vishha shkola, 1983, 455 s. – [in Russian].
- [2] M. P. Matviienko, *Kompiuterna lohika: pidruchnyk*. Kyiv, Ukraina: Vydavnytstvo Lira-K, 2017, 324 s. – [in Ukrainian].
- [3] M. F. Bondarenko, N. V. Bilous, A. H. Rutkas, *Kompiuterna dyskretna matematyka: pidruchnyk*. Kharkiv, Ukraina: «Kompaniia SMIT», 2004, 480 s. – [in Ukrainian].

- [4] M. P. Matviienko, V. P. Rozen, *Kompiuterna skhemotekhnika: navchalnyi posibnyk*. Kyiv, Ukraina: Vydavnytstvo Lira-K, 2016, 192 s. – [in Ukrainian].
- [5] E. P. Ugrjumov, *Cifrovaja shemotekhnika: ucheb. posobie dlja vuzov*. SPb., Rossija: BHV-Peterburg, 2010, 816 s. – [in Russian].
- [6] V. P. Kozhemjako, *Optojelektronnye logiko-vremennye informacionno-vychislitel'nye sredy*. Tbilisi, Gruzija: Mecniereba, 1984, 360 c. – [in Russian].
- [7] S. V. Svechnikov, V. P. Kozhemjako, L. I. Timchenko, *Kvaziimpul'sno-potencial'nye optojelektronnye jelementy i ustrojstva logiko-vremennogo tipa*. Kiev, Ukraina: Naukova dumka, 1987, 256 s. – [in Russian].
- [8] V. P. Kozhemjako, O. G. Natroshvili, T. B. Martynjuk, L. Sh. Imnaishvili, *Optojelektronnaja shemotekhnika: ucheb. posobie*. Kiev: UMK VO, 1988, 276 s. – [in Russian].
- [9] T. B. Martyniuk, O. M. Tarasova, M. M. Al-Khiari, "Osoblyvosti lohiko-chasovoho zobrazhennia chyslovoi informatsii," *Visnyk Vinnytskoho politekhnichnoho instytutu*, № 1, s. 72-76, 2000. – [in Ukrainian].
- [10] T. B. Martyniuk, M. M. Al-Khiari, S. A. Vasyletskyi, "Funktsiina povnota lohichno-chasovoho pryntsypu zobrazhennia informatsii," *Visnyk Vinnytskoho politekhnichnoho instytutu*, № 2, s. 48-52, 2000. – [in Ukrainian].
- [11] Je. Berlekjemp, *Algebraicheskaia teorija kodirovanija: per. s angl.*, M., Rossija: Mir, 1971, 480 s.
- [12] T. B. Martyniuk, Mokhamed Salem Nasser, V. V. Vlasichuk, O. M. Nakonechnyi, "Analiz mozhlyvosti odynychnoho koduvannia chyslovoi informatsii," *Optyko-elektronni informatsiino-enerhetychni tekhnologii*, № 2 (10), s. 39-44, 2005. – [in Ukrainian].
- [13] V. P. Kozhemiako, T. B. Martyniuk, V. V. Dmytruk, V. V. Vlasichuk, "Klasyfikatsiia odynychnykh kodiv," *Optyko-elektronni informatsiino-enerhetychni tekhnologii*. № 1 (11), s. 36-42, 2006. – [in Ukrainian].
- [14] Z. L. Rabinovich, V. A. Ramanauskas, *Tipovye operacii v vychislitel'nyh mashinah*. Kiev, Ukraina: Tehnika, 1980, 264 s. – [in Russian].
- [15] K. G. Samofalov, V. I. Kornejchuk, A. M. Romankevich, V. P. Tarasenko, *Cifrovye mnogoznachnye jelementy i struktury: ucheb. Posobie*. Kiev, Ukraina: Vishha shkola, 1974, 168 s. – [in Russian].
- [16] V. P. Kozhemiako, T. B. Martyniuk, O. I. Supryhan, D. I. Klimkina, *Kvantovi peretvoriuvachi na optoelektronnykh lohiko-chasovykh seredovyschchakh dlia okoprotsesornoj obrobky zobrazhen: monohrafiia*. Vinnytsia, Ukraina: UNIVERSUM-Vinnytsia, 2007, 126 c. – [in Ukrainian].
- [17] Z. L. Rabinovich, *Osnovy teorii jelementnykh struktur JeVM*. M., Rossija: Radio i svjaz'. – 1982, 280 s. – [in Russian].
- [18] R. I. Grushvickij, A. X. Mursaev, E. P. Ugrjumov, *Proektirovanie sistem na mikroshemah programmiruemoj logiki*. SPb., Rossija: BHV-Peterburg, 2002, 608 s. – [in Russian].
- [19] V. I. Zhabin, I. A. Zhukov, I. A. Klymenko, V. V. Tkachenko, *Prykladna teoriia tsyfrovyykh avtomativ*. K., Ukraina: Knyzhkove vyd-vo NAU, 2007, 364 s. Ju. G. Karpov, *Teoriia avtomatov*. SPb., Rossija: Piter, 2003, 208 s. – [in Ukrainian].
- [20] V. P. Kozhemjako, T. B. Martynjuk, K. V. Kozhemjako, "Sintez ustrojstva upravlenija na R-avtomate," *Upravljajushhie sistemy i mashiny*, № 1/2, s. 22–25, 1995. – [in Russian].
- [21] Ju. A. Buzunov, I. G. Burenkov, N. N. Shipilov, "Mikroprogrammnye avtomaty na parallel'no-posledovatel'nykh strukturah," *Upravljajushhie sistemy i mashiny*, № 2, s. 26–29, 1982. – [in Russian].
- [22] T. B. Martyniuk, K. V. Kozhemiako, A. V. Kozhemiako, "Do otsinky skladnosti kombinatsiinykh skhem R-avtomativ," *Visnyk Vinnytskoho politekhnichnoho instytutu*, № 1, s. 31–34, 1997. – [in Russian].
- [23] T. B. Martyniuk, A. V. Kozhemiako, N. V. Fofanova, "Dva varianty syntezy mikroprogramnykh R-avtomativ," *Visnyk Vinnytskoho politekhnichnoho instytutu*, № 4, s. 47–53, 1998. – [in Ukrainian].
- [24] T. B. Martynjuk, "Osobennosti sinteza mikroprogrammnykh R-avtomatov," *Upravljajushhie sistemy i mashiny*, № 3, s. 22-26, 1998. – [in Ukrainian].
- [25] V. M. Kirpichnikov, V. A. Skljarov, "Sintez mikroprogrammnykh avtomatov po graf-sheham algoritmov s malym chislom uslovnykh vershin," *Upravljajushhie sistemy i mashiny*, № 1, s. 77–83, 1978. – [in Russian].

Відомості про авторів

Мартинюк Тетяна Борисівна – доктор технічних наук, професор, професор кафедри обчислювальної техніки.

Войцеховська Олена Валеріївна – кандидат технічних наук, доцент, доцент кафедри обчислювальної техніки.

Т. Б. Мартынюк, Е. В. Войцеховская

**ЭФФЕКТИВНОСТЬ ЕДИНИЧНОГО КОДИРОВАНИЯ
ДАНЫХ**

Винницкий национальный технический университет, Винница

T. B. Martyniuk, O. V. Voytsekhovska

EFFICIENCY OF UNIT DATA ENCODING

Vinnytsya National Technical University, Vinnytsia

УДК: 004.312.466:510.644.4

О. О. Семенова, А. О. Семенов, О. О. Войцеховська

ЗАСТОСУВАННЯ НЕЧІТКОГО КОНТРОЛЕРА У ПРОЦЕДУРІ ВЕРТИКАЛЬНОГО ХЕНДОВЕРУ

Вінницький національний технічний університет, Вінниця

Анотація. Через швидкий розвиток технологій мобільного зв'язку виникає ситуація, коли на одній території функціонують одразу декілька різних мереж мобільного зв'язку. У такому випадку, основною проблемою для забезпечення необхідної якості обслуговування користувачів є реалізація оптимального алгоритму швидкого перемикавання між різними мережами, тобто здійснення операції вертикального хендOVER. У роботі запропоновано використовувати нечіткий контролер у інтелектуальному алгоритмі багатокритеріального вертикального хендOVER. Розроблений нечіткий контролер має шість вхідних лінгвістичних змінних та одну вихідну. Подано базу правил, що складається з 64 правил. Проведено моделювання роботи розробленого нечіткого контролера у програмі Matlab.

Ключові слова: гетерогенна мережа, хендOVER, нечіткий контролер.

Аннотация. Быстрое развитие технологий мобильной связи приводит к возникновению ситуации, когда на одной территории функционируют сразу несколько различных сетей мобильной связи. В таком случае, основной проблемой для обеспечения требуемого качества обслуживания пользователей является реализация оптимального алгоритма быстрого переключения между различными сетями, то есть выполнение операции вертикального хэндовера. В работе предложено использовать нечёткий контроллер в интеллектуальном алгоритме многокритериального вертикального хэндовера. У разработанного нечёткого контроллера шесть входных лингвистических переменных и одну выходная. Представленная база правил состоит из 64 правил. Проведено моделирование работы разработанного нечёткого контроллера в программе Matlab.

Ключевые слова: гетерогенная сеть, хэндовер, нечёткий контроллер.

Abstract. Due to the rapid development of mobile technologies, there is a situation when several different mobile networks operate on one area. In this case, the main problem for ensuring the required quality of service for subscribers is implementation of an optimal algorithm for fast switching between different network, i.e. performing a vertical handover operation. The paper proposes to utilize a fuzzy controller in the intelligent algorithm of multicriterial vertical handover. We propose to use a fuzzy controller having six input and one output linguistic variables. The rule base consisting of sixty-four rules is presented. Simulation of the developed fuzzy controller in Matlab program was performed.

Keywords: heterogeneous network, handover, fuzzy controller.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-37-44>.

Вступ

Наразі на Україні актуальним є питання забезпечення якості мобільного зв'язку, що спричинено розвитком ринку телекомунікацій. При цьому, по всій території розгортаються мережі четвертого покоління, в найближчі роки планується розгортання мереж п'ятого покоління. Як відомо, мобільні мережі наступних поколінь є гетерогенними мережами, де різні технології радіодоступу об'єднуються у межах єдиної інфраструктури з метою збільшення покриття і пропускної здатності. Зокрема, у гетерогенних мережах одночасно передаються різні типи трафіку (групові виклики з голосовим або відео трафіком, а також трафіком завантаження файлів) з мобільних станцій, що можуть рухатися з високою швидкістю. Тому, при реалізації гетерогенної мережі найбільшими викликами є ефективне керування радіоресурсами і забезпечення прозорого безшовного пересування абонентів. При цьому, ключовим механізмом, що дозволяє абоненту безшовно пересуватися в гетерогенній мережі є механізм вертикального хендOVER [1].

Актуальність

ХендOVER – це процес передачі управління з'єднанням від однієї базової станції (або точки доступу) до іншої. ХендOVER між базовими станціями однієї мережі називається горизонтальним, хендOVER між окремими безпроводними мережами називається вертикальним [2]. Як відомо, найбільш поширений алгоритм горизонтального хендOVER оснований на значенні потужності приймального сигналу RSS, тобто хендOVER може відбутися у випадку, коли рівень RSS на базової станції стає нижчим за пороговий. В той же час, у гетерогенних мережах неефективно здійснювати операцію хендOVER лише на основі одного критерію, оскільки при передачі мультисервісного трафіку необхідно враховувати параметри якості обслуговування, а також пріоритети користувачів. Таким чином, розроблення алгоритму вертикального хендOVER, який би враховував максимальну кількість параметрів, є актуальною науково-практичною задачею.

Мета

Метою даної роботи є розроблення пристрою для прийняття рішення стосовно здійснення операції вертикального хендOVERу.

Задачі

1. Визначити вхідні критерії для прийняття рішення стосовно здійснення вертикального хендOVERу.

2. Обґрунтувати структурну схему пристрою, що видає рішення стосовно здійснення хендвера.

3. Про моделювати роботу пристрою.

Реалізація хендвера

Операція вертикального хендвера ініціюється за наступних умов: мобільна станція вийшла за межі зони покриття мережі; погіршилися показники якості обслуговування; змінилися пріоритети користувача.

В процедурі вертикального хендвера виділяють три етапи [3]. Перший етап, ініціалізація, передбачає визначення доступних безпроводних мереж. Також на цьому етапі проводиться збір даних, які потрібні для прийняття рішення. Це можуть бути як параметри абонента, так і параметри мережі або параметри каналу зв'язку чи мобільного телефону. Другий етап, прийняття рішення, передбачає визначення мережі, до якої буде здійснено перемикавання обслуговування мобільної станції. На третьому етапі, здійснення хендвера, відбувається передача сеансу зв'язку на обрану мережу. Отже, на етапі прийняття рішення необхідно визначити найкращу мережу з-поміж доступних, розрахувати конкретний момент часу для переключення, а також прийняти рішення – здійснювати операцію хендвера чи ні.

Існуючі алгоритми етапу прийняття рішення поділяють за способом оброблення вхідних параметрів на два типи – математичні та інтелектуальні, причому інтелектуальні алгоритми здатні обробити більшу кількість вхідних параметрів [4].

Інтелектуальні алгоритми засновано на використанні нечітких контролерів, штучних нейронних мереж, генетичних алгоритмів. Зокрема, алгоритми на базі нечітких контролерів здатні обробляти результати неточних вимірювань, тому вони добре підходять для реалізації у мережах мобільного зв'язку, які розгортають у великих містах, де сигнали зазнають втрат через вплив будівель та машин [5]. Алгоритми на базі штучних нейронних мереж при виборі найкращої мережі дозволяють врахувати такі параметри як налаштування користувача та можливості стільникового телефону [6]. Алгоритми на базі генетичних алгоритмів оцінюють також ймовірність успішного хендвера, що, у свою чергу, призводить до підвищення продуктивності мережі [7].

Таким чином, покращити ефективність та якість системи мобільного зв'язку можна шляхом застосування у процедурі вертикального хендвера інтелектуальних алгоритмів. Оскільки використання алгоритму вертикального хендвера на основі нечіткої логіки дозволяє оптимізувати параметри мережі [8], у даній роботі пропонується розробити нечіткий контролер для застосування при реалізації операції вертикального хендвера (рис. 1).

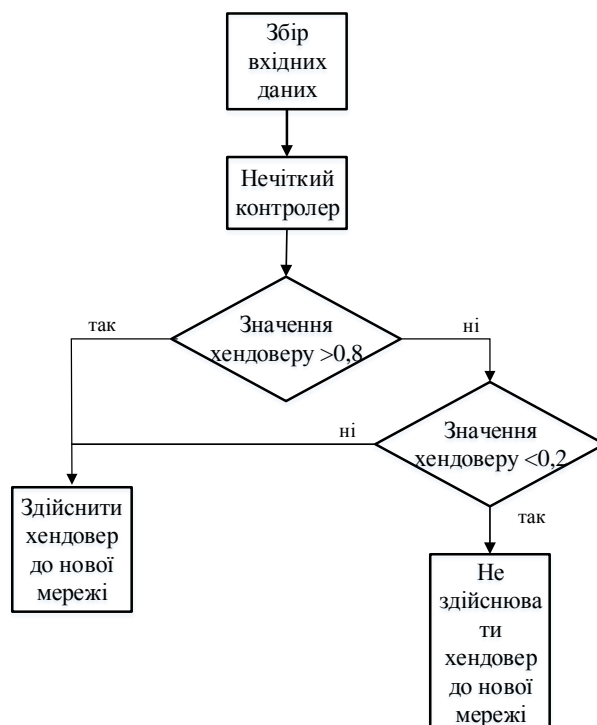


Рисунок 1 – Алгоритм хендверу на основі нечіткої логіки

Нечіткий контролер

У якості вхідних величин нечіткого контролера пропонується використати такі параметри стільникової мережі: рівень потужності сигналу I , доступна смуга частот B , швидкість передачі даних R , джитер

J , відстань між мобільною станцією та базовою станцією D , швидкість руху мобільного абонента S . Функції належності для вхідних та вихідних величин нечіткого контролера подано на рис.2–8.

Вихідною величиною нечіткого контролера є рішення стосовно доцільності здійснення хендвера H , що приймає значення від 0 до 1. Одиниця означає, що необхідно виконати передачу обслуговування мобільного абонента на нову стільникову мережу, а поява на виході нечіткого контролера нуля свідчить про повну недоцільність здійснення передачі обслуговування мобільного абонента на нову мережу. Проміжні значення відображають часткову доцільність здійснення хендвера.

Окрім того, можливо застосовувати такі додаткові параметри як безпечність зв'язку, вартість, енергоспоживання, пріоритет абонента.

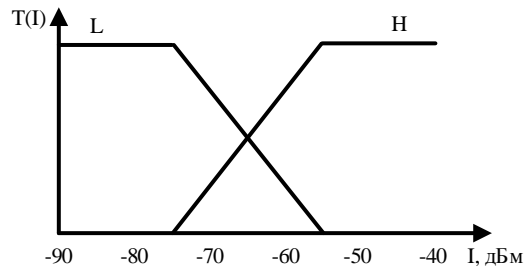


Рисунок 2 – Функції належності для рівня потужності сигналу

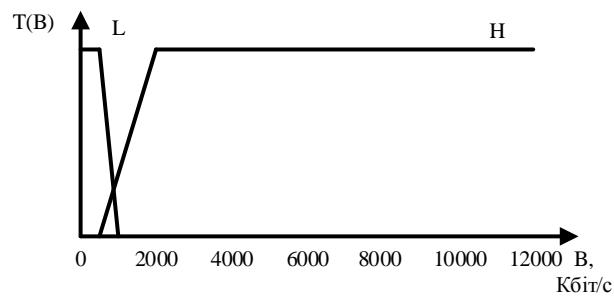


Рисунок 3 – Функції належності для доступної смуги частот

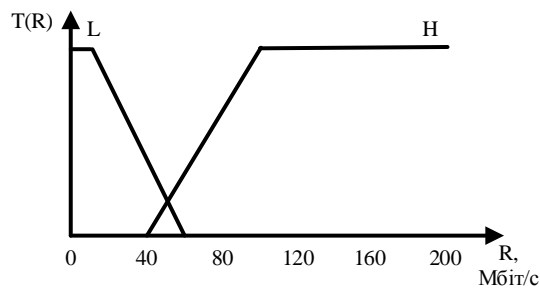


Рисунок 4 – Функції належності для швидкості передачі даних

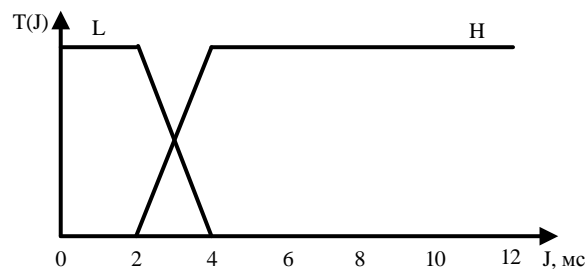


Рисунок 5 – Функції належності для джитера

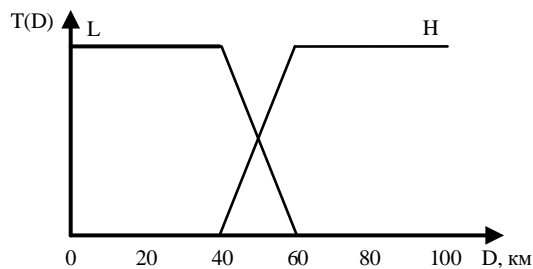


Рисунок 6 – Функції належності для відстані

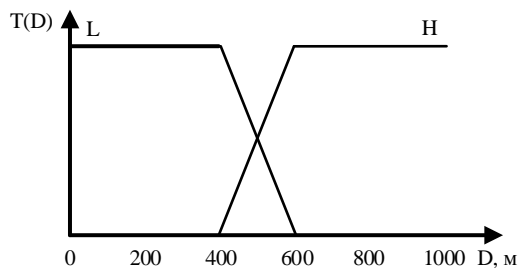


Рисунок 7 – Функції належності для швидкості

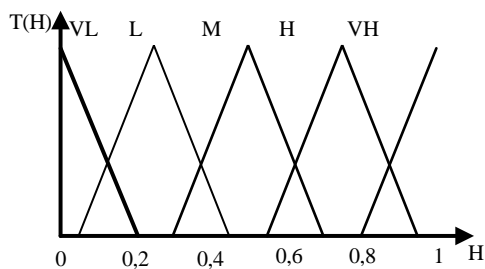


Рисунок 8 – Функції належності для здійснення хендвера

Оскільки база правил функціонування розробленого нечіткого контролера є досить великою та містить 64 правила, в табл.1 подано лише частину з них.

Таблиця 1 – База правил

Номер правила	I	B	R	J	D	S	H
1	L	L	L	L	L	L	VL
2	L	L	L	L	L	H	VL
3	L	L	L	L	H	L	VL
4	L	L	L	L	H	H	L
5	L	L	L	H	L	L	VL
6	L	L	L	H	L	H	VL
7	L	L	L	H	H	L	L
8	L	L	L	H	H	H	L
9	L	L	H	L	L	L	M
10	L	L	H	L	L	H	M
11	L	L	H	L	H	L	L
12	L	L	H	L	H	H	M
...
62	H	H	H	H	L	H	VH
63	H	H	H	H	H	L	VH
64	H	H	H	H	H	H	H

Моделювання

Роботу розробленого нечіткого контролера було промодельовано у програмі Matlab. На рис. 9 подано вигляд нечіткого контролера, база правил представлена на рис.10.

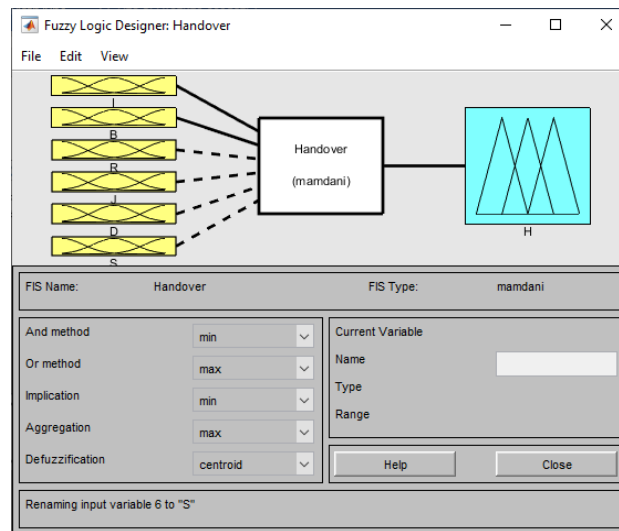


Рисунок 9 – Нечіткий контролер в програмі Matlab

У випадку, показаному на рис.11, були застосовані такі вхідні дані: рівень потужності сигналу складав -71 дБм, доступна смуга частот – 100 кбїт/с, швидкість передачі даних – 43 Мбїт/с, джитер – $8,6$ мс, відстань між мобільною станцією та базовою станцією – 570 м, швидкість руху мобільного абонента – 20 км/год. У результаті моделювання одержуємо значення рішення стосовно здійснення хендвера – $0,396$.

У випадку, показаному на рис.12, були застосовані такі вхідні дані: рівень потужності сигналу складав -72 дБм, доступна смуга частот – 8100 кбїт/с, швидкість передачі даних – 121 Мбїт/с, джитер – $8,83$ мс, відстань між мобільною станцією та базовою станцією – 240 м, швидкість руху мобільного абонента – 85 км/год. У результаті моделювання одержуємо значення рішення стосовно здійснення хендвера – $0,552$.

У випадку, показаному на рис.13, були застосовані такі вхідні дані: рівень потужності сигналу складав -45 дБм, доступна смуга частот – 3000 кбїт/с, швидкість передачі даних – 160 Мбїт/с, джитер – $4,5$ мс, відстань між мобільною станцією та базовою станцією – 350 м, швидкість руху мобільного абонента – 110 км/год. У результаті моделювання одержуємо значення рішення стосовно здійснення хендвера – $0,937$, тобто хендвер буде здійснено.

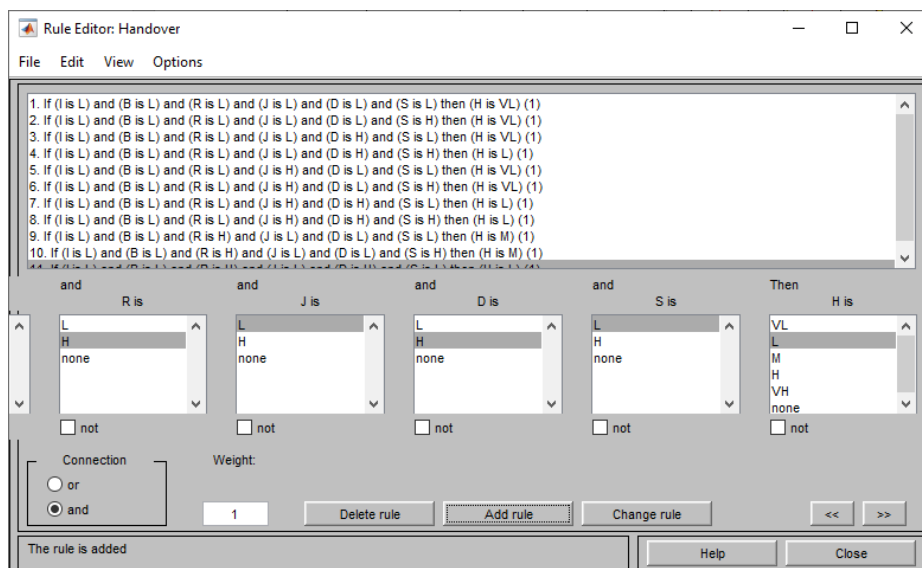


Рисунок 10 – База правил



Рисунок 11 – Результат моделювання



Рисунок 12 – Результат моделювання



Рисунок 13 – Результат моделювання

Висновки

У роботі запропоновано схему прийняття рішення стосовно здійснення інтелектуального вертикального хендвера на базі нечіткого контролера, який враховує шість вхідних параметрів. Правильність роботи пристрою підтверджується проведенням моделюванням у середовищі Matlab. Завдяки оптимізації механізму вертикального хендвера користувачке обладнання може здійснити правильний вибір мережі стільникового зв'язку.

Список літератури

- [1] A. M. Aibinu, A. J. Onumanyi, A. P. Adedigba, M. Ipinyomi, T. A. Folorunso, M. J. E. Salami, "Development of hybrid artificial intelligent based handover decision algorithm," *Engineering Science and Technology an International Journal*, V. 20(2), pp. 381-390, February. 2017. doi: 10.1016/j.jestch.2017.01.005.
- [2] А. Р. Масюк, І. Б. Стрихалюк, М. В. Брич, І. О. Кагало, Г. В. Бешлей, "Алгоритм інтелектуального вертикального хендверу в гетерогенній мобільній мережі на основі хмарних обчислень," *Вісн. Нац. ун-ту «Львів. Політехніка»*, № 874, с. 110-121, 2017.
- [3] Ionut Bosoanca, Anca Vargatu, "An Overview of Vertical Handoff Decision Algorithms in NGWNs and a new Scheme for Providing Optimized Performance in Heterogeneous Wireless Networks," *Informatica Economica*, V. 15(1), pp. 5-21, 2011.
- [4] А. Д. Гришаева, В. Я. Воропаева, "Выбор параметров и разработка критерия оптимизации для процедуры вертикального хэндвера," *Збірник наукових праць XIII Міжнародної науково-технічної конференції аспірантів і студентів. Автоматизація технологічних об'єктів та процесів. Пошук молодих*, Донецьк, 2013, с. 39-41.
- [5] M. T. Islam et al., "Vertical handover decision using fuzzy logic in a heterogeneous environment," in *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, Dhaka, 2013, pp. 1-3. doi: 10.1109/ICIEV.2013.6572621.
- [6] Mahmood Adnan, Hushairi Zen, Al-Khalid Othman, "Vertical Handover Decision Processes for Fourth Generation Heterogeneous Wireless Networks," *Asian Journal of Applied Sciences*, V. 1, Is. 5, pp. 229-235, December. 2013.
- [7] M. Saravanan, A. Prithiviraj, "Genetic based approach to Optimize the Vertical Handover performance among Hetrogenous Network," in *Proceedings of the International Conference on Intelligent Computing Systems (ICICS 2017)*, Salem, India, 2017, pp. 80-84.
- [8] Sukhmandeep Kaur, "Optimization Of User Behaviour Based Handover Using Fuzzy Logic," *International Journal of Scientific & Technology Research*, V. 8, Is. 10, pp.2623-2629, October. 2019.

Стаття надійшла: 11.03.2021.

References

- [1] M. Aibinu, A. J. Onumanyi, A. P. Adedigba, M. Ipinyomi, T. A. Folorunso, M. J. E. Salami, "Development of hybrid artificial intelligent based handover decision algorithm," *Engineering Science and Technology an International Journal*, V. 20(2), pp. 381-390, February. 2017. doi: 10.1016/j.jestch.2017.01.005.
- [1] R. Masiuk, I. B. Strykhalyuk, M. V. Brych, I. O. Kahalo, H. V. Beshley, "Intellectual vertical handover algorithm in heterogeneous mobile network based on cloud technology," *Visnyk of Lviv Polytechnic National University*, № 874, pp. 110-121, 2017. – [in Ukrainian].
- [2] Ionut Bosoanca, Anca Vargatu, "An Overview of Vertical Handoff Decision Algorithms in NGWNs and a new Scheme for Providing Optimized Performance in Heterogeneous Wireless Networks," *Informatica Economica*, V. 15(1), pp. 5-21, 2011.
- [3] A. D. Hryshaeva, V. Ya. Voropayeva, "Selection of parameters and development of an optimization criterion for the vertical handover procedure," in *Proceedings of the XIII International scientific-technical conference of post-graduate students and students. Automation of technological objects and processes. Search for young people*. Donetsk, 2013, pp. 39-41.
- [4] M. T. Islam et al., "Vertical handover decision using fuzzy logic in a heterogeneous environment," in *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, Dhaka, 2013, pp. 1-3. doi: 10.1109/ICIEV.2013.6572621. – [in Russian].
- [5] Mahmood Adnan, Hushairi Zen, Al-Khalid Othman, "Vertical Handover Decision Processes for Fourth Generation Heterogeneous Wireless Networks," *Asian Journal of Applied Sciences*, V. 1, Is. 5, pp. 229-235, December. 2013.
- [6] M. Saravanan, A. Prithiviraj, "Genetic based approach to Optimize the Vertical Handover performance among Hetrogenous Network," in *Proceedings of the International Conference on Intelligent Computing Systems (ICICS 2017)*, Salem, India, 2017, pp. 80-84.

- [7] Sukhmandeep Kaur, "Optimization Of User Behaviour Based Handover Using Fuzzy Logic," *International Journal of Scientific & Technology Research*, V. 8, Is. 10, pp.2623-2629, October. 2019.

Відомості про авторів

Семенова Олена Олександрівна – кандидат технічних наук, доцент, доцент кафедри телекомунікаційних систем та телебачення.

Семенов Андрій Олександрович – доктор технічних наук, професор, професор кафедри радіотехніки.

Войцеховська Ольга Олександрівна – аспірантка, асистент кафедри системного аналізу та інформаційних технологій.

Е. А. Семёнова, А. А. Семёнов, О. А. Войцеховская

**ПРИМЕНЕНИЕ НЕЧЁТКОГО КОНТРОЛЛЕРА В
ПРОЦЕДУРЕ ВЕРТИКАЛЬНОГО ХЭНДОВЕРА**

Винницкий национальный технический университет, Винница

O. O. Semenova, A. O. Semenov, O. O. Voitsekhovska

**USING A FUZZY CONTROLLER FOR THE VERTICAL
HANDOVER PROCEDURE**

Vinnitsia National Technical University, Vinnitsia

УДК 004:78.08

О. В. Сілагін, Д. О. Зелінська, В. А. Гірдвайніс

ОНТОЛОГІЧНЕ МОДЕЛЮВАННЯ БАЗИ ЗНАТЬ З МУЗИЧНОГО НАДЖАНРУ «МЕТАЛ»

Вінницький національний технічний університет, Вінниця

Анотація. У роботі проведено онтологічне моделювання предметної області визначного культурного і музичного феномену, а саме сучасного музичного наджанру «метал». Створено термінологічний словник з даної предметної області із використанням поняття терміносистеми. Проаналізовано можливості середовища розробки онтологій Protégé в контексті заявленої предметної області. Вибрано базовий принцип моделювання онтології у вигляді семантичної мережі, яку можна поступово розширяти, поглиблюючи знання про предметну область. Онтологічну модель бази знань реалізовано засобами середовища Protégé, з використанням знань про відношення функціональності, транзитивності, рефлексивності для структуризації інформації та дотримання принципів її повноти, достовірності та несуперечливості. Вибрано критерій оцінювання коректності онтологічної моделі бази знань. Проведено тестування розробленої онтологічної бази знань та підтверджено високий рівень її коректності в процесі пошуку інформації. Вказано можливі напрямки подальшого розширення онтологічної моделі.

Ключові слова: онтологічна модель, база знань, метал, жанр, Protégé, семантика.

Анотация. В работе проведено онтологическое моделирование предметной области выдающегося культурного и музыкального феномена, а именно современного музыкального супержанра «метал». Создан терминологический словарь по данной предметной области с использованием понятия терминосистемы. Проанализированы возможности среды разработки онтологий Protégé в контексте заявленной предметной области. Выбран базовый принцип моделирования онтологии в виде семантической сети, которую можно постепенно расширять, углубляя знания о предметной области. Онтологическую модель базы знаний реализовано средствами среды Protégé, с использованием знаний об отношении функциональности, транзитивности, рефлексивности для структурирования информации и соблюдения принципов ее полноты, достоверности и непротиворечивости. Выбран критерий оценки корректности онтологической модели базы знаний. Проведено тестирование разработанной онтологической базы знаний и подтвержден высокий уровень ее корректности в процессе поиска информации. Указаны возможные направления дальнейшего расширения онтологической модели.

Ключевые слова: онтологическая модель, база знаний, метал, жанр, Protégé, семантика.

Abstract. The paper presents an ontological modeling of the subject area of a prominent cultural and musical phenomenon, namely the modern musical supergenre "metal". A terminological dictionary on this subject area was created using the concept of terminological system. The possibilities of the Protégé ontology development environment in the context of the stated subject area were analyzed. The basic principle of ontology modeling in the form of a semantic network is chosen, which can be gradually expanded, deepening the knowledge about the subject area. The ontological model of the knowledge base is implemented by means of the Protégé environment, using knowledge of the relationship of functionality, transitivity, reflexivity to structure information and adhere to the principles of its completeness, reliability and consistency. The criterion for assessing the correctness of the ontological model of the knowledge base was chosen. Testing of the developed ontological knowledge base was carried out and the high level of its correctness in the process of information retrieval was confirmed. Possible directions of further expansion of the ontological model are indicated.

Keywords: ontology model, knowledge base, metal, genre, Protégé, semantics.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-45-50>.

Вступ

Поняття онтології знань набуло важливого практичного значення останнім часом у сферах штучного інтелекту, управління знаннями та розробки складних спеціалізованих інформаційних ресурсів. Для фахівців з найрізноманітніших сфер нагальною потребою стає структуризація знань у межах предметної області їх дослідження. Всеосяжна та детальна формалізація знань має на увазі побудову концептуальної схеми предметної області. Така схема будується на основі двох базових термінів, а саме набору понять та інформації про дані поняття [1].

Побудови детальних карт набору понять на сьогоднішній день успішно реалізуються більшістю автоматизованих систем, проте набагато складнішим завданням є правильна подача усієї можливої інформації про поняття. Значимими блоками інформації є властивості об'єктів, відношення між ними, обмеження усіх відомих видів (від обмежень на тип даних і до обмежень на взаємодію певних об'єктів між собою), аксіоми тощо. Редактор онтологій Protégé [2] є одним з найпотужніших засобів формалізації знань, що може реалізовувати більшість взаємодій об'єктів. Окрім того, Protégé є системою з можливістю розширення функціональності, а це означає, що вона постійно перебуває у процесі розвитку та доповнюється зусиллями наукової спільноти.

Актуальність

Тематика предметної області створення онтології, а саме музичне мистецтво, що належить до різновиду «метал» у базовій класифікації музичних жанрів [3], є доволі актуальною на сьогоднішній день як культурний феномен XX і XXI століття і як базис окремої філософської течії, що знайшла багато прихильників по усьому світу. Розроблена онтологічна модель бази знань даного напрямку може бути впроваджена на тематичних веб-ресурсах і значно полегшувати семантичний пошук інформації в межах предметної області. Разом із тим, дана тематика ідеально підходить для проведення дослідження можливостей редактору онтологій Protégé, проектування або встановлення додаткових плагінів у майбутньому в разі необхідності та вирішення питання доцільності розширення функціональності середовища.

Мета

Метою даного дослідження є підвищення коректності семантичного пошуку інформації зацікавленими користувачами, що планується досягти за рахунок створення нової онтологічної моделі бази знань про музичний наджанр «метал».

Задачі

Для досягнення цієї мети і створення нової онтологічної моделі бази знань потрібно вирішити наступні задачі:

1. Дослідити предметну область для створення онтологічної моделі бази знань і відповідно до неї скласти мінімальний словник термінів і сформулювати терміносистему.
2. Створити базову графічну модель у вигляді фрагменту семантичної мережі.
3. Реалізувати онтологічну модель бази знань засобами Protégé, зафіксувавши ключові концепти і терми, після чого закодувати їх.
4. Провести тестування коректності пошуку інформації в реалізованій засобами Protégé онтологічній базі знань та оцінити коректність пошуку за обраним критерієм.

Створення мінімального термінологічного словника для предметної області музичного наджанру «метал»

Початковим етапом для створення онтології предметної області метал-музики є побудова так званої терміносистеми. Терміносистему потрібно розглядати як систематизовану сукупність термінів певної предметної області, що створюється експертами на основі теорії та володіє властивостями системності, повноти, несуперечливості, відносної стабільності, відкритості та динамічності [4]. Важливим є знання того факту, що при наявності ряду теорій в одній предметній області можуть мати місце одразу декілька терміносистем. У подальшому розвитку онтології потрібно розвивати кожен з них окремо, проте у будь-якому випадку основа повинна бути єдиною.

Таким чином, на даному етапі потрібно створити структуризовану мінімальну таблицю з основними дефініціями майбутньої онтології. Необхідно не лише врахувати основні поняття і відокремити їх від другорядних, а і обрати найбільш яскраві приклади певних термінів у вигляді класів та створити їх ієрархію. Ієрархія буде доповнюватись по мірі поглиблення і розширення онтології, проте її базис повинен мати такий вигляд, щоб дати достатнє уявлення про структуру предметної області користувачеві або розробнику, яким би нескінченно малим не був його набір знань у даному напрямку. Таблиця 1 ілюструє саме такий базовий набір знань та уявлень для глибшого розуміння предметної області з музичного наджанру «метал». Терміни подані одночасно українською та англійською мовами, так як дана предметна область значною мірою інтернаціоналізована.

Таблиця 1 – Мінімальний термінологічний словник для уявлення про наповнення онтології метал-музики

Термін	Дефініція	Приклади властивостей відношення	Приклади властивостей даних	Тип даних	Приклади індивідів
Band (гурт)	Колектив людей, об'єднаних якоюсь спільною метою або роботою, які так чи інакше пов'язані з музикою.	Have_motherland (мати країну походження), have_number_of_musicians (мати кількість музикантів у складі), have_rate (мати рейтинг).	Genre (жанр), frontman (фронтмен/лідер), age (вік/кількість років активної діяльності).	String	Gehenna, Immortal, Type O Negative, Olshano, Lucifugum, Drudkh, Electric Wizard, Deeds of Flesh.
Genre (жанр)	Сукупність стилізованих музичних характеристик, що дозволяє об'єднати виконавців наджанру «метал» у менші класифікаційні групи.	Have_origin (мати походження), have_relative_genres (мати споріднені жанри), have_founder (мати засновника, виконавця або гурт).	Age (приблизний вік існування), heyday (роки розквіту), derivative forms (похідні жанри), origins (походження).	String	Black Metal, Death Metal, Doom Metal, Avangard Metal, Gothic Metal, Folk Metal, Gothic Metal, Power Metal, Thrash Metal, Heavy Metal.

Country of origin (країна походження)	Країна, де було засновано гурт або з якої походять більшість членів гурту.	Have_capital (мати столицю), have_cultural_origins (мати культурні основи для виникнення даного жанру).	Region (частина світу, до якої належить дана країна).	String	Norway, Sweden, Ukraine, Poland, USA, France, Germany.
Label (лейбл звукозапису)	Бренд, створений компаніями, що займаються виробництвом, поширенням і просуванням аудіо- та іноді відеозаписів на носіях різних відповідних форматів.	Have_motherland (країна лейблу), have_founder (мати засновника).	Specialties (жанри спеціалізації), status (статус активності), contacts (контакти для зв'язку).	String	Propaganda, Supernal Music, Nuclear Blast, Drakkar Productions, Grieghallen Studio, Шум Деревьев Records, Hammermark Art.
Status (статус гурту)	Стан гурту, його позиція відносно даного моменту часу.	Be_unknown (бути невідомим).	Present_status (теперішній статус), past_status (минулий статус).	Може бути string або bool у спрощеному варіанті.	Active, Unactive, Changed name, Split up.

Створення базової графічної моделі у вигляді фрагменту семантичної мережі

Після виявлення ключових концептів доцільним є побудування ієрархії предметної області, в корені якої буде знаходитись терміносистема, а у вузлах – номенклатура. У якості моделі представлення знань номенклатури використовується семантична мережа знаків-фреймів.

Семантична мережа S будується як об'єднання знаків-фреймів Φ_i : $S = U_i \Phi_j$. Будемо вважати, що загалом сформовано множину термінів Term. Це означає, що потужність множини $\Phi = \{\Phi_i\}$ більша потужності множини Term [5]. Графічне зображення фрагменту семантичної мережі S для побудови онтології метал-музики можна побачити на рисунку 1.

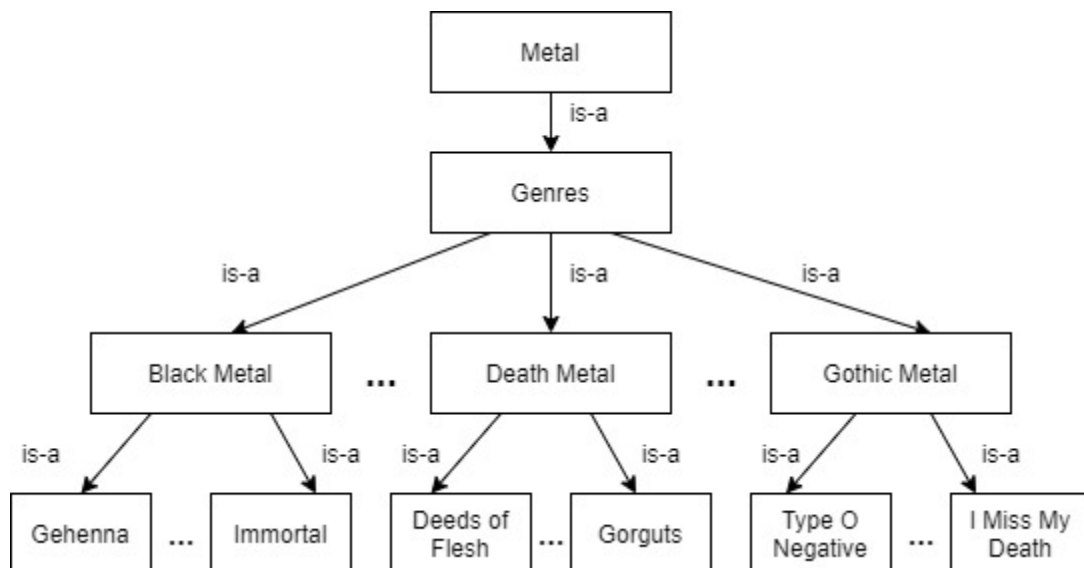


Рисунок 1 – Фрагмент семантичної мережі для побудови онтології метал-музики

Надалі проведемо кодування базису онтології метал-музики. Розпочинати потрібно із налаштування ієрархії класів та підкласів, потім надати їм властивості відношень та властивості даних (визначити усі

необхідні типи даних зі списку або створити нові за потреби), а після цього перейти до створення окремих індивідів у межах кожного класу.

Реалізація онтологічної моделі бази знань засобами Protégé

У Protégé наявні дуже масштабні можливості для роботи з відношеннями, а надання їм властивостей транзитивності, рефлексивності, функціональності та інших дозволяє значно спростити проблеми, що часто постають у межах предметної області. Розглянемо конкретні приклади щоб довести практичне значення встановлення подібних властивостей.

Гурт, що умовно належить до індивідів класу жанру Black Metal (хоча за жанровою класифікацією також належить і до класів Post Metal, Avangard Metal), має назву Olshanoє з 2019 року, хоча до цього мав назву Isa. Проте користувач може не володіти даною інформацією і сприймати відповідно дані індивіди як два абсолютно різних гурти. Як відомо, інформація повинна володіти властивістю повноти, тому важливим буде повідомити користувача про даний факт. Здійснити це можна засобами встановлення відношення еквівалентності. У Protégé даний факт можна реалізувати засобами функції «Same Individual As».

Також слід звернути увагу на функцію «Disjoint With» – її необхідно застосовувати навпаки при забороні використання суперечливих понять або термінів. Даний оператор не можна застосовувати як розмежування жанрів, адже різниця між ними досить нечітка, і один гурт може поєднувати декілька з них. А при визначенні статусу того чи іншого гурту «Disjoint With» є нагальною потребою для забезпечення вимоги несуперечливості – один і той самий гурт не може бути одночасно активним і неактивним. Проте варто врахувати, що він може містити одночасно статус «Active» та «Changed_name» як у попередньому прикладі, тобто бути активним, але змінити назву, і ставити функцію «Disjoint With» між цими двома категоріями не тільки непотрібно, а й категорично заборонено.

При розробці онтології метал-музики варто врахувати реалізацію пошуку індивідів за певними критеріями. Для цього існує Class Expression Editor – зручний інструмент для написання виразів у межах певного класу. Наприклад, користувач хоче знайти музиканта у межах жанру Black Metal, який при цьому мав би вік 38 років. Тому у межах даного жанру потрібно скласти наступну умову: «Musician and have_age value 38». Далі здійснюється пошук усіх індивідів-музикантів даного жанру, яким 38 років. Варто зазначити, що перед цим у всіх індивідів були прописані дані про їх вік і жанр, у якому вони працюють, притому тип даних віку був позначений як «int». При наявності іншого типу даних система видала б помилку і не працювала б з цілочисельним значенням «38».

Проте найпотужнішим засобом для створення запитів на пошук інформації у Protégé є вбудований конструктор мови запитів SPARQL [6]. Основними видами таких запитів є SELECT запит, CONSTRUCT запит, ASK запит та DESCRIBE запит. Найбільш поширеним з даних різновидів є SELECT запит, результати виведення запиту подаються у зручному для користувача табличному вигляді. Коректність виведення запитів є однією з ознак правильності побудови онтології.

Окрім того, графічним відображенням онтології є інструмент Ontograp, що виводить усю семантику предметної області в найбільш зрозумілому та простому поданні.

На рисунку 2 зображено декотрі моменти з процесу створення онтології музичного наджанру «метал» у вигляді колажу, а на рисунку 3 показано власне онтологічний граф в результаті побудови базису онтології.

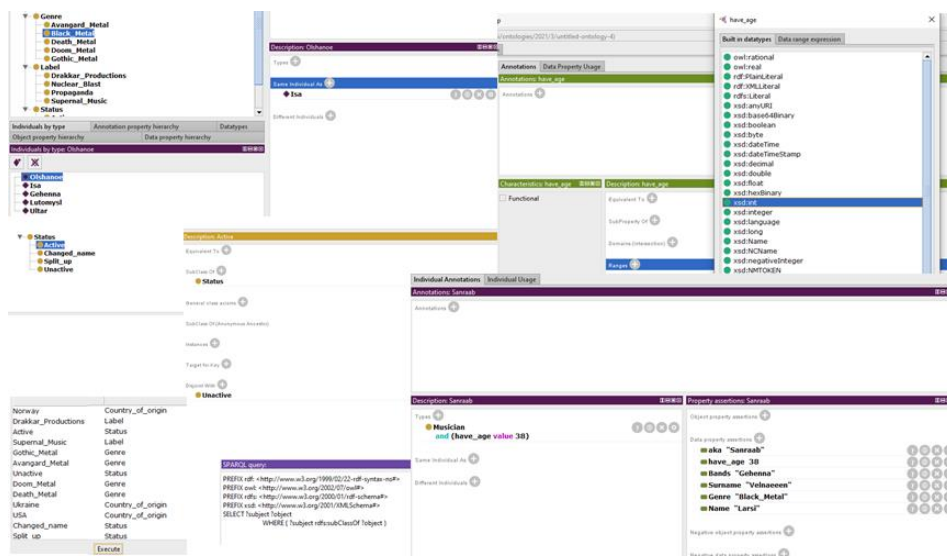


Рисунок 2 – Основні моменти кодування онтології наджанру «метал»

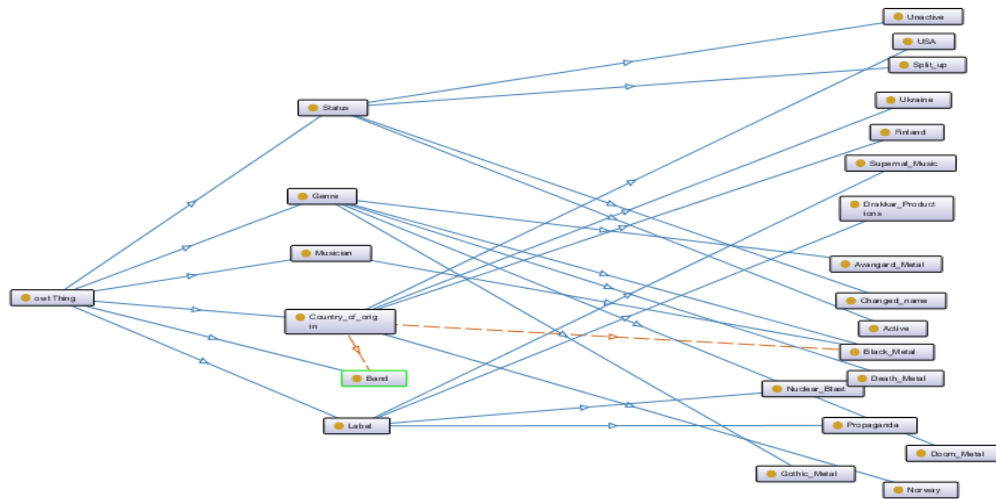


Рисунок 3 – Базовий граф онтології наджанру «метал»

Тестування та оцінювання коректності семантичного пошуку інформації в онтологічній базі знань музичного наджанру «метал»

Гомез-Перез запропонував наступне визначення оцінки онтології у контексті технологій сумісного використання знань: «для складання технічного думки про онтологію, асоційованого з нею програмного середовища і документації, щодо критеріїв. Критеріями можуть виступати вимоги до специфікації, питання компетенції, і / або реальний світ» [7].

Серед сучасних метрик не так багато таких, які враховують усі задані критерії, тому було вирішено скористатись метрикою Single Usability Metrics (SUM), що розроблена аналітиком Джефом Соро [8].

Дана метрика визначається за формулою $SUM = \frac{SM_{comp} + SM_{sat} + SM_{time} + SM_{er}}{4}$. У контексті розробки онтологій показник *Completion* є булевим значенням і позначає досягнення мети здобуття інформації користувачем, *Satisfaction* показує рівень задоволення від структуризації інформації у даній онтології, *Errors* показує виникнення певних суперечливостей або інших помилок при отриманні інформації та *Times* показує середню кількість запитів для отримання тієї чи іншої інформації.

Для визначення коректності пошуку інформації у реалізованій онтологічній базі знань, вона була надана 10 користувачам, кожен з яких реалізував на ній по 10 запитів.

Результати опитування за даною метрикою представлені у таблиці 2.

Таблиця 2 – Результати опитування щодо коректності розробленої онтології

User	Satisfaction	Completion	Errors	Times	Average
1	5	1	0	5	90%
2	5	1	0	7	85,5%
3	5	1	2	3	87,5%
4	5	0	1	4	62,5%
5	4	1	0	4	88,75%
6	4	0	1	3	63,75%
7	5	1	0	3	100%
8	4	1	0	11	76,75%
9	5	1	0	3	100%
10	5	1	1	5	83,75%

Середній показник за метрикою SUM по всім користувачам відповідно становитиме 83,85%, що є досить непоганим показником для онтологічних баз знань. В той же час проведено за аналогічною методикою тестування бази даних наджанру «метал» сайту «Encyclopaedia Metallum», де використовувалась класична реляційна модель організації баз даних, показала набагато нижчі результати. Так середній показник за метрикою SUM по 10 користувачам відповідно становив 75,32%.

Висновки

1. У статті запропоновано нова онтологічна модель організації бази знань (онтологія) предметної області: музичний наджанр «метал».

2. Інструментальними засобами середовища Protégé онтологія предметної області музичний наджанр «метал» реалізована у вигляді бази знань.

3. Проведено тестування реалізованої бази знань та оцінювання коректності семантичного пошуку за метрикою SUM.

4. Мета дослідження – підвищення коректності семантичного пошуку інформації досягається за рахунок застосування в реалізації бази знань нової онтологічної моделі (онтології) про музичний наджанр «метал». У порівнянні з класичною, реляційною моделлю організації баз даних застосування нової онтології дозволяє підняти коректність пошуку на 8%.

Список літератури

- [1] А. О. Никоненко, "Огляд баз знань онтологічного типу," *Штучний інтелект*, № 4, с. 208-219. 2009.
- [2] Protégé. [Online]. Available: <https://protege.stanford.edu/>. Accessed on: May 18, 2021.
- [3] Home – Encyclopaedia Metallum: The Metal Archives. [Online]. Available: <https://www.metal-archives.com/>. Accessed on: May 21, 2021.
- [4] Принципи побудови банків даних. [Електронний ресурс]. Режим доступу: https://web.posibnyky.vntu.edu.ua/fitki/10savchuk_organizaciya_bazdanih_znan/gl_14.html. Дата звернення: May 17, 2021.
- [5] SPARQL. [Online]. Available: <https://znaio.com.ua/SPARQL>. Accessed on: May 21, 2021.
- [6] A. Gomez-Perez, N. Juristo, J. Pazos; in N.J. Mars, editor, "Evaluation and assessment of knowledge sharing technology," *Towards Very Large Knowledge Bases*, Amsterdam: IOS Press, 1995, p. 289-296.
- [7] Расчет SUM. Подход от Джефа Соро. [Електронний ресурс]. Режим доступу: <https://v-shliachkov.medium.com/расчет-sum-подход-от-джефа-соро-c1ecf796f1b2>. Дата звернення: May 25, 2021.

Стаття надійшла: 13.07.2021.

References

- [1] A. O. Nykonenko, "Ohliad baz znan ontolohichnoho typu," *Shtuchnyi intelekt*, № 4, s. 208-219. 2009.
- [2] Protégé. [Online]. Available: <https://protege.stanford.edu/>. Accessed on: May 18, 2021.
- [3] Home – Encyclopaedia Metallum: The Metal Archives. [Online]. Available: <https://www.metal-archives.com/>. Accessed on: May 21, 2021.
- [4] Pryntsy py pobudovy bankiv danykh. [Elektronnyi resurs]. Rezhym dostupu: https://web.posibnyky.vntu.edu.ua/fitki/10savchuk_organizaciya_bazdanih_znan/gl_14.html. Data zvernennia: May 17, 2021.
- [5] SPARQL. [Online]. Available: <https://znaio.com.ua/SPARQL>. Accessed on: May 21, 2021.
- [6] A. Gomez-Perez, N. Juristo, J. Pazos; in N.J. Mars, editor, "Evaluation and assessment of knowledge sharing technology," *Towards Very Large Knowledge Bases*, Amsterdam: IOS Press, 1995, p. 289-296.
- [7] Raschet SUM. Podkhod ot Dzhefa Soro. [Online]. Available: <https://v-shliachkov.medium.com/raschet-sum-podkhod-ot-dzhefa-soro-c1ecf796f1b2>. Accessed on: May 25, 2021.

Відомості про авторів

Сілагін Олексій Віталійович – кандидат технічних наук, доцент кафедри комп'ютерних наук.

Зелінська Дарія Олегівна – магістрантка групи ІКН-20м, кафедра комп'ютерних наук.

Гірдвайніс Владислав Аудрісович – магістрант групи ІКН-20м, кафедра комп'ютерних наук.

О. В. Силагин, Д. О. Зелинская, В. А. Гирдвайнис

ОНТОЛОГИЧЕСКОЕ МОДЕЛИРОВАНИЕ БАЗЫ ЗНАНИЙ МУЗЫКАЛЬНОГО НАДЖАНРА «МЕТАЛ»

Винницкий национальный технический университет, Винница

O. V. Silagin, D. O. Zelinska, V. A. Girdvainis

ONTOLOGICAL MODELING OF THE KNOWLEDGE BASE OF THE MUSICAL SUPERGENRE "METAL"

Vinnitsia National Technical University, Vinnitsia

УДК 004.92

О. Н. Романюк, О. О. Дудник, Д. А. Озерчук

ПІДВИЩЕННЯ ПРОДУКТИВНОСТІ ЗАФАРБОВУВАННЯ ПРИ ВИКОРИСТАННІ ДЛЯ ВИЗНАЧЕННЯ ІНТЕНСИВНОСТЕЙ КОЛЬОРУ ПОВЕРХНІ ДРУГОГО ПОРЯДКУ

Вінницький національний технічний університет, м. Вінниця

Анотація. Підвищення реалістичності відтворення графічних сцен передбачає не тільки збільшення рівня деталізації поверхонь об'єктів реального світу але й використання більш складних моделей освітлення. Це гостро ставить питання про підвищення продуктивності графічних систем, особливо при формуванні динамічних зображень у реальному часі та в інтерактивному режимі, коли передбачається, що траєкторії руху об'єктів не задано заздалегідь, а визначаються діями користувача в процесі взаємодії із системою.

До високопродуктивних методів зафарбовування відносять метод зафарбовування з використанням поверхні другого порядку для визначення інтенсивностей кольорів. Цей метод передбачає розрахунок нормалізованих векторів не для всіх, а тільки для декількох точок поверхні.

Отримано нові аналітичні залежності для визначення інтенсивності кольору точки рядка rasterизації через інтенсивності кольорів сусідніх точок. Отримані залежності не використовують довготривалі мікрооперації множення. Виведено формулу для визначення інтенсивностей кольорів в кінцевих точках цифрових сегментів, на які розбито рядок rasterизації. Наведено структурні схеми для визначення інтенсивностей кольорів. Отримано порівняльні оцінки підвищення продуктивності.

Результати досліджень можуть бути використані у високопродуктивних системах тривимірної графіки.

Ключові слова: зафарбовування, rasterизація, поверхні другого порядку, інтенсивність кольору, тривимірне моделювання, метод Фонга, метод Гуро.

Аннотация. Повышение реалистичности воссоздания графических сцен предусматривает не только увеличение уровня детализации поверхностей объектов реального мира, но и использование более сложных моделей освещения. Это остро ставит вопрос о повышении продуктивности графических систем, особенно при формировании динамических изображений в реальном времени и в интерактивном режиме, когда предполагается, что траектории движения объектов не заданы заранее, а определяются действиями пользователя в процессе взаимодействия с системой.

К высокопроизводительным методам закраски относят метод закраски с использованием для определения интенсивностей цвета поверхности второго порядка. Этот метод предусматривает определение нормализованных векторов не для всех, а только для некоторых точек поверхности.

Получены новые аналитические зависимости для определения интенсивности цвета в точке строки rasterизации через интенсивность цвета соседних точек. Полученные зависимости не используют длинные микрооперации. Выведено формулу для определения интенсивности цвета в конечных точках цифровых сегментов, на которые разбита строка rasterизации. Предоставлено структурные схемы для определения интенсивности цвета. Получены сравнительные оценки повышения производительности.

Полученные результаты могут быть использованы в высокопроизводительных системах трехмерной графики.

Ключевые слова: закраска, rasterизация, поверхности второго порядка, интенсивность цвета, трехмерное моделирование, метод Фонга, метод Гуро.

Abstract. Increasing the realism of the reproduction of graphic scenes involves not only increasing the level of detail of the surfaces of real-world objects, but also the use of more complex lighting models. This raises the question of improving the performance of graphics systems, especially in the formation of dynamic images in real time and interactively, when it is assumed that the trajectory of objects is not set in advance, but determined by user actions in interaction with the system.

High-performance shading methods include the method of shading with the use of second-order surface to determine color intensities. This method involves the calculation of normalized vectors not for all, but only for a few points on the surface.

New analytical dependencies for calculating the color intensity of a point on a rasterization row using color intensity of neighboring points are obtained. The obtained dependencies do not use long-term multiplication microoperations. A formula for computing color intensities at the endpoints of digital segments into which the rasterization string is divided is derived. Structural diagrams for determining color intensities are given. Comparative estimates of productivity increase are obtained.

Obtained results can be used in high-performance three-dimensional graphics systems.

Key words: shading, rasterization, surface of the second order, color intensity, 3D modeling, Phong shading, Gourand shading. DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-51-59>.

Вступ

Етап зафарбовування [1-6] – це етап кінцевої візуалізації, на якому згідно з даними, отриманими на етапі геометричних перетворень, формуються пікселі зображення, для яких визначаються екранні координати та інтенсивності кольору. Цей етап вважається найбільш трудомістким у графічному конвеєрі, оскільки пов'язаний із попіксельними діями, складними обчисленнями та передбачає врахування багатьох параметрів. Рендеринг у загальному обсязі обчислень із формування тривимірних сцен складає 60-80 % [2], тому в значній мірі визначає продуктивність графічних систем.

Підвищення реалістичності відтворення графічних сцен передбачає не тільки збільшення рівня деталізації поверхонь для коректної апроксимації об'єктів реального світу але й використання більш складних моделей освітлення. Це гостро ставить питання про підвищення продуктивності графічних систем, особливо при формуванні динамічних зображень у реальному часі та в інтерактивному режимі, коли передбачається, що траєкторії руху об'єктів не задано заздалегідь, а визначаються діями

користувача в процесі взаємодії із системою. Недостатня продуктивність графічних систем є також завадою до моделювання у сценах фізичних процесів і збільшення кількості динамічних об'єктів.

Оскільки традиційні методи, засоби та підходи не задовольняють вимогам по продуктивності для багатьох галузей застосування тривимірної комп'ютерної графіки, то існує важлива науково-прикладна задача – розробка теоретичних основ високопродуктивного рендерингу тривимірних графічних зображень.

У роботі розглянуто питання використання поверхні другого прядку для високопродуктивного зафарбовування поверхонь тривимірних графічних сцен.

Аналіз методів зафарбовування

Методи, які використовуються для моделювання освітлення, оперують із розсіяним і відбитим світлом [2, 5]. Відбите світло у своєму складі має дві компоненти: дифузну [2, 7] та спекулярну [2, 5, 7]. Розсіяне (фонове) світло – це світло, яке відбивається від навколишнього середовища. Точне моделювання даної складової світла вимагає великих обчислювальних витрат, а тому воно при використанні локальних моделей освітлення апроксимується [8].

Для врахування фонового освітлення можна припустити, що всі поверхні повністю освітлюються розсіяним світлом I_a . При цьому фоновий вклад у дифузне відбиття в будь-якій точці поверхні буде дорівнювати [2, 5, 7] $I_{am} = I_a \cdot k_a$, де k_a – коефіцієнт відбиття розсіяного світла ($0 \leq k_a \leq 1$).

Дифузне відбиття характерно для матових і шорстких поверхонь із хаотичними нерівностями, розміри яких співставні з довжиною хвилі або перевищують її [2, 5, 7]. Дифузне відбиття визначає видимість навколишніх тіл, оскільки кожна точка освітленої поверхні випромінює відбиті промені в усі сторони.

При наявності в сцені точкового джерела світла [1, 2] інтенсивність дифузного відбиття пропорційна косинусу кута між нормаллю до поверхні й напрямком на джерело світла \vec{L} . У цьому випадку для обчислення інтенсивності дифузного відбиття застосовують закон косинусів Ламберта [2, 7]

$$I_d = I_0 \cdot k_d \cdot \cos \varphi,$$

де I_0 – інтенсивність джерела світла, $k_d \in [0,1]$ – коефіцієнт дифузного відбиття, φ – кут між вектором \vec{L} і нормаллю \vec{N} до поверхні (рис. 1).

На рис. 1 відображено також вектор \vec{V} до спостерігача, серединний вектор

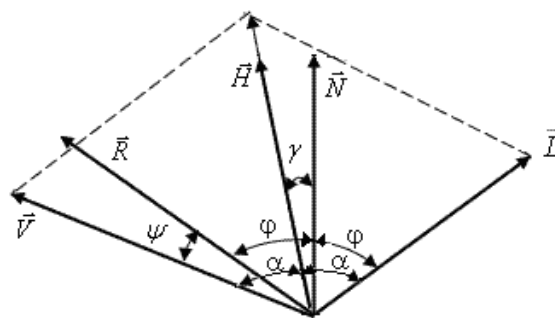


Рисунок 1 – Вектори нормалей до точки поверхні

Історично першим методом зафарбовування був метод однотонного зафарбовування [2], згідно з яким для кожного плоского трикутника визначався вектор нормалі, а на його основі – колір. Складові трикутники об'єкта заповнювалися одним кольором без його градації, що призводило до різкої зміни інтенсивності кольору на межах трикутників. Однотонне зафарбовування вимагає найменших обчислювальних затрат. Воно хоча й має низьку якість, але внесено до функцій Direct3D і підтримується тривимірними акселераторами.

До найпоширеніших методів зафарбовування відносять метод Гуро [1, 2, 4, 9], який забезпечує прийнятний компроміс між швидкодією формування тривимірних зображень та їх якістю. Процес зафарбовування має такі стадії [2]: а) розраховують вектори нормалей до кожної грані; б) шляхом усереднення нормалей усіх граней, яким належить вершина, розраховують нормалі у вершинах трикутника (багатокутника); с) визначають інтенсивності кольору у вершинах багатокутника, використовуючи значення нормалей; д) зафарбовують ділянку, обмежену багатокутником, шляхом

лінійної інтерполяції інтенсивностей кольорів вздовж ребер, а потім і між ребрами вздовж кожного рядка rasterизації. Останнім часом затінення за Гуро часто використовують як проміжну стадію з інтерактивного формування 3D-зображення, покладаючи побудову повноцінної сцени на етап фінального рендерингу.

До недоліків методу Гуро можна віднести [2]: а) метод використовує для визначення інтенсивностей кольору лінійну інтерполяцію, в той час як дифузна та спекулярна складові кольору мають нелінійний характер зміни; б) не враховується локальна кривизна поверхні, оскільки вектори нормалей визначаються тільки для вершин трикутника; в) відблиски відтворюються тільки в разі, якщо вершини трикутників знаходяться в їх зоні (відблиск, який не має спільних точок із вершинами трикутників, або розташований усередині трикутника, не буде сформовано); г) на межах двох трикутників проявляються смуги Маха [2, 5], які пов'язані з літеральним гальмуванням на сітківці ока; д) має місце зміна інтенсивності кольору зображення від кадру до кадру, що виражається в миготінні, особливо відблисків, оскільки при формуванні динамічних зображень змінюється структура та положення вузлів триангуляційної мережі; е) наявність артефакту типу "зірка", який полягає у тому, що відблиск, який повинен мати форму еліпса, має форму зірки. Це пояснюється тим, що ділянки відблиску формуються в різних трикутниках і проявляється ефект смуг Маха; ж) метод не враховує перспективу об'єкта.

При зафарбовуванні за методом Фонга інтенсивність кольору точок визначають за формулою [1, 2, 5, 10]:

$$I = I_a k_a + I_l (k_d \cos \psi + k_s \cos^n \lambda) \quad (1)$$

де I_a, I_l – інтенсивності відповідно розсіяного і направленої джерел світла, k_a, k_d, k_s – коефіцієнти розсіяного, дифузійного і дзеркального світла, ψ – кут між напрямком світла і вектором нормалі, λ – кут між відбитим напрямком світла і спостерігачем, n – коефіцієнт яскравості поверхні, $\cos^n \lambda$ – дистрибутивна функція (BRDF) [8], яка відповідає за оптичні характеристики поверхні.

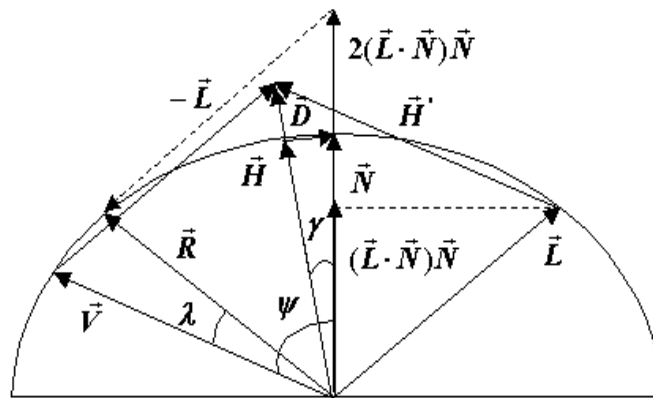


Рисунок 2 – Визначення кутів для моделей освітлення Фонга і Бліна

У більшості випадків при реалізації зафарбовування за Фонгом використовуються моделі освітлення Бліна та Фонга [2, 5, 7, 10].

При використанні моделі освітлення Фонга (рис.2) у формулі (1) $\cos \lambda = \vec{R} \cdot \vec{V} = 2(\vec{L} \cdot \vec{V})(\vec{N} \cdot \vec{V}) - \vec{V} \cdot \vec{L}$. Для моделі освітлення Бліна $\cos \lambda$ заміняють на $\cos \gamma = \vec{H} \cdot \vec{L}$, де $\vec{H} = (\vec{L} + \vec{V}) / |\vec{L} + \vec{V}|$.

Модель освітлення Фонга вважається точнішою, однак потребує більшого обсягу обчислень. Продуктивність зафарбовування при використанні моделі Бліна значно вища порівняно з моделлю Фонга. Це пояснюється тим, що у випадку, коли джерело світла та спостерігач розташовані на нескінченній відстані від об'єкта (найпоширеніший випадок у комп'ютерній графіці), значення вектора \vec{H} для моделі освітлення Бліна розраховується один раз для кожного кадру зображення. При використанні ж моделі освітлення Фонга для кожної точки поверхні знаходять вектор \vec{R} . У подальшому будемо використовувати модель освітлення Бліна як таку, що має меншу обчислювальну складність.

При зафарбовуванні 3D-об'єктів за методом Фонга визначають нормовані вектори до поверхні об'єкта, джерела світла й спостерігача, а також допоміжні вектори залежно від вибору моделі освітлення. Нормалізація вектора [2, 5, 11, 12] потребує виконання трьох операцій ділення, трьох операцій множення, двох операцій додавання та операцію знаходження квадратного кореня, виконання яких достатньо трудомістке. У зв'язку з цим актуальним питанням є зменшення при зафарбовуванні кількості використаних нормалізованих векторів.

Один з таких методів зафарбовування передбачає використання для визначення інтенсивності кольору поверхні другого порядку, для якої нормалізовані вектори визначаються тільки в деяких точках. Інтенсивність кольору в точці на поверхні другого порядку задано поліномом [13]

$$I(x, y) = A \cdot x^2 + B \cdot y^2 + C \cdot x \cdot y + D \cdot x + E \cdot y + F.$$

Із рівняння видно, що необхідно визначити шість невідомих A, B, C, D, E, F . Це передбачає формування й розв'язання системи із шести рівнянь. Оскільки для трикутника задаються інтенсивності кольору тільки в трьох його вершинах, то найпростіше до визначити інтенсивності кольору в середніх точках на ребрах трикутника, що дасть можливість скласти систему із шести рівнянь.

Знаходження невідомих A, B, C, D, E, F наведено в [13]. У подальшому метод, який використовує для визначення інтенсивності кольору наведену формулу, будемо називати «прямим».

Мета статті – підвищення продуктивності зафарбовування при використанні для визначення інтенсивностей кольору поверхні другого порядку за рахунок використання нових аналітичних залежностей.

Модифікований метод зафарбовування

Розглянемо рядок пастеризації (рис. 3), паралельний осі абсцис.

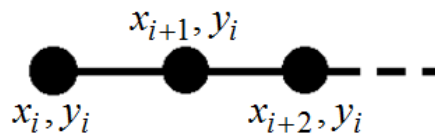


Рисунок 3 – Рядок rasterизації, паралельний осі абсцис

Для поточної точки (x_i, y_i) інтенсивність кольору визначаємо за формулою

$$I(x_i, y_i) = Ax_i^2 + By_i^2 + Cx_i y_i + Dx_i + Ey_i + F. \quad (2)$$

Виконаємо крокове переміщення по осі x і знайдемо інтенсивність кольору в точці (x_{i+1}, y_i)

$$\begin{aligned} I(x_{i+1}, y_i) &= A(x_i + 1)^2 + By_i^2 + C(x_i + 1)y_i + D(x_i + 1) + Ey_i + F = \\ &= Ax_i^2 + 2Ax_i + A + By_i^2 + Cx_i y_i + Cy_i + Dx_i + D + Ey_i + F = \\ &= I(x_i, y_i) + 2Ax_i + A + Cy_i + D. \end{aligned} \quad (3)$$

Для рядка rasterизації значення Cy_i не змінюється, тому його доцільно обчислювати тільки один раз і в подальшому використовувати як операнд.

При зміщенні на дві точки (x_{i+2}, y_i) , інтенсивність кольору дорівнює

$$\begin{aligned} I(x_{i+2}, y_i) &= A(x_i + 2)^2 + By_i^2 + C(x_i + 2)y_i + D(x_i + 2) + Ey_i + F = \\ &= Ax_i^2 + 4Ax_i + 4A + By_i^2 + Cx_i y_i + 2Cy_i + Dx_i + 2D + Ey_i + F = \\ &= I(x_i, y_i) + 4Ax_i + 4A + 2Cy_i + 2D. \end{aligned} \quad (4)$$

З формул (3) і (4) видно, що інтенсивність кольору в поточній точці можна визначити через інтенсивності кольору в двох попередніх точках:

$$\begin{aligned}
 I(x_{i+2}, y_i) &= I(x_i, y_i) + 4Ax_i + 4A + 2Cy_i + 2D = \\
 &= 2[I(x_i, y_i) + 2Ax_i + A + Cy_i + D] - I(x_i, y_i) + 2A = \\
 &= 2I(x_{i+1}, y_i) - I(x_i, y_i) + 2A.
 \end{aligned} \tag{5}$$

Розрахунок за формулою (5) для визначення інтенсивності кольору в точці (x_{i+2}, y_i) значно простіший порівняно з формулою (2), оскільки передбачається використання тільки двох мікрооперацій зсуву та двох мікрооперацій додавання. При цьому не використовується «довга» операція множення.

На рис. 4 наведено структурну схему для визначення $I(x_{i+2}, y_i)$.

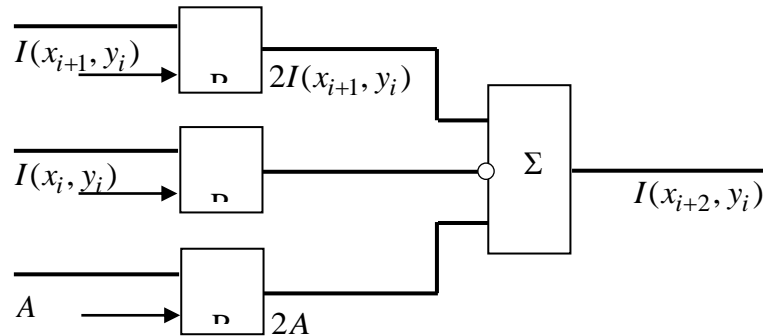


Рисунок 4 – Структурна схема блоку для визначення $I(x_{i+2}, y_i)$

Знайдемо значення інтенсивності кольору в точці (x_{i+1}, y_i) через інтенсивності кольорів в двох найближчих сусідніх точках. Для цього використаємо формулу (5).

$$I(x_{i+1}, y_i) = \frac{I(x_i, y_i) + I(x_{i+2}, y_i) - 2A}{2}. \tag{6}$$

Порівняння формул (5) і (6) показує, що їхня обчислювальна складність однакова.

При зміщенні на три точки, інтенсивність кольору в точці (x_{i+3}, y_i) дорівнює

$$\begin{aligned}
 I(x_{i+3}, y_i) &= A(x_i + 3)^2 + By_i^2 + C(x_i + 3)y_i + D(x_i + 3) + Ey_i + F = \\
 &= Ax_i^2 + 6Ax_i + 7A + By_i^2 + Cx_iy_i + 3Cy_i + Dx_i + 3D + Ey_i + F = \\
 &= I(x_i, y_i) + 6Ax_i + 9A + 3Cy_i + 3D.
 \end{aligned}$$

На жаль, в останній формулі необхідно використовувати операції множення, що, безумовно, впливає на продуктивність рендерингу.

Інтенсивність кольору в точці можна виразити через інтенсивність кольору в трьох попередніх точках згідно виразу

$$\begin{aligned}
 I(x_{i+3}, y_i) &= I(x_i, y_i) + 6Ax_i + 9A + 3Cy_i + 3D = \\
 &= [I(x_i, y_i) + 4Ax_i + 4A + 2Cy_i + 2D] + \\
 &+ [I(x_i, y_i) + 2Ax_i + A + Cy_i + D] - I(x_i, y_i) + 4A = \\
 &= I(x_{i+2}, y_i) + I(x_{i+1}, y_i) - I(x_i, y_i) + 4A.
 \end{aligned} \tag{7}$$

Визначимо інтенсивність кольору $I(x_{i+3}, y_i)$ виключно через значення інтенсивностей кольорів. Для цього знайдемо значення $4A$ через інтенсивність кольору в трьох сусідніх точках, використавши формулу (5).

$$4A = 2 \cdot 2A = 2[I(x_{i+2}, y_i) - 2I(x_{i+1}, y_i) + I(x_i, y_i)] = \\ = 2I(x_{i+2}, y_i) - 4I(x_{i+1}, y_i) + 2I(x_i, y_i).$$

Підставивши отримане значення у формулу (7) знайдемо, що

$$I(x_{i+3}, y_i) = I(x_{i+2}, y_i) + I(x_{i+1}, y_i) - I(x_i, y_i) + \\ + 2I(x_{i+2}, y_i) - 4I(x_{i+1}, y_i) + 2I(x_i, y_i) = \\ = 3I(x_{i+2}, y_i) - 3I(x_{i+1}, y_i) + I(x_i, y_i). \quad (8)$$

Множення на 3 можна замінити на операції додавання та зсуву. Наприклад, $3I(x_{i+2}, y_i) = 2I(x_{i+2}, y_i) + I(x_{i+2}, y_i)$.

$$\text{Тоді } I(x_{i+3}, y_i) = 2I(x_{i+2}, y_i) + I(x_{i+2}, y_i) - 2I(x_{i+1}, y_i) - I(x_{i+1}, y_i) + I(x_i, y_i).$$

У наведеній формулі використовуються виключно операції додавання та зсуву. На рис. 5 наведено структурну схему блоку для визначення інтенсивності кольору $I(x_{i+3}, y_i)$

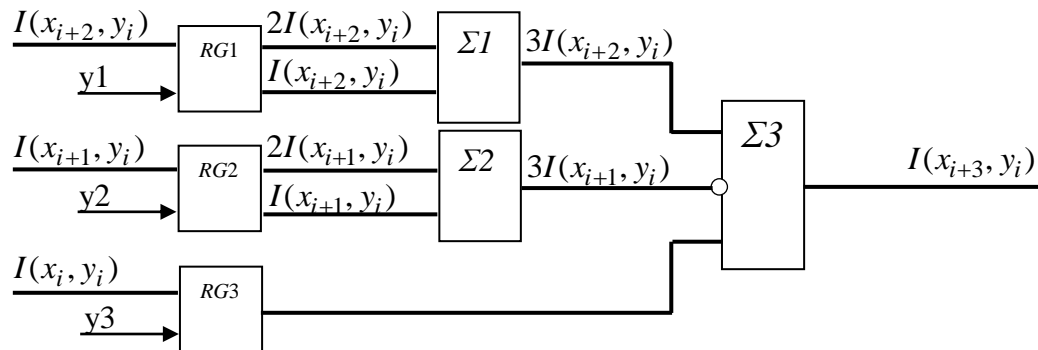


Рисунок 5 – Структурна схема блоку для визначення $I(x_{i+3}, y_i)$

У роботі [2] запропоновано розбивати рядок растеризація на цифрові сегменти довжиною 2^k і визначати інтенсивності кольору в його кінцевих точках. Значення інтенсивності кольору в проміжних точках сегменту у подальшому визначають з використанням кодової лінійної інтерполяції.

Розглянемо визначення інтенсивності кольору в кінцевих точках цифрового сегменту довжиною 2^k

$$I(x_{i+2^k}, y_i) = A(x_i + 2^k)^2 + By_i^2 + C(x_i + 2^k)y_i + D(x_i + 2^k) + Ey_i + F = \\ = Ax_i^2 + 2 \cdot 2^k Ax_i + (2^k)^2 A + By_i^2 + Cx_i y_i + 2^k Cy_i + Dx_i + 2^k D + Ey_i + F = \\ = I(x_i, y_i) + 2^{k+1} Ax_i + 2^{2k} A + 2^k Cy_i + 2^k D.$$

Наведена формула має меншу обчислювальну складність порівняно з використанням формули (2) для визначення інтенсивностей кольору кінцевих точок. Дійсно, в цьому випадку використовується десять операцій множення та п'ять операцій додавання. При використанні отриманої формули для визначення $I(x_{i+2^k}, y_i)$ використовується тільки дві операції множення, чотири операції додавання та $(k+1) + 2k + k + k = 5k + 1$ операцій зсуву.

Порівняємо ресурсні затрати отриманих методів. Як критерій можна використати кількість операцій (множення, додавання/віднімання, зсув). Операція множення є однією з найбільш ресурснозатратних. Множення на числа типу 2^n замінюють зсувом на n розрядів.

Приведемо формулу (2) до вигляду

$$I(x_i, y_i) = x_i(Ax_i + Cy_i + D) + y_i(By_i + E) + F. \quad (9)$$

Тоді для обчислення інтенсивності кольору в точці потрібно буде виконати п'ять операцій множення та п'ять операцій додавання. Використовуючи «прямий» метод для знаходження інтенсивності кольору в n точках потрібно виконати $5n$ операцій множення та $5n$ операцій додавання.

Незалежно від методу, інтенсивність кольору в першій точці рядка растреризації завжди доведеться знаходити «прямим» методом.

Проаналізуємо знаходження інтенсивності кольору з використанням формули (3). Обчислення інтенсивності кольору в кожній точці, починаючи з другої, потребує однієї операції множення, чотирьох операцій додавання та однієї операції зсуву. Тоді для знаходження інтенсивності кольору в n точках потрібно виконати $5+(n-1)=n+4$ операцій множення, $5+4(n-1)=4n+1$ операцій додавання, $(n-1)$ операцій зсуву.

Проаналізуємо обчислення інтенсивності кольору, використовуючи формулу (5). Для кожної точки, починаючи з третьої, знаходження інтенсивності кольору потребує двох операцій зсуву, однієї операції додавання та однієї операції віднімання. Вважаючи, що для знаходження інтенсивності кольору в другій точці рядка растреризації використано попередній метод, для знаходження інтенсивності кольору в n точках потрібно виконати $5+1=6$ операцій множення, $5+4+(n-2)=n+7$ операцій додавання та $(n-2)$ операцій віднімання (оскільки віднімання реалізується через додавання, то потрібно виконати $(n+7)+(n-2)=2n+5$ операцій додавання), $1+2(n-2)=2n-3$ операцій зсуву.

Якщо обчислювати інтенсивність кольору в точці, використовуючи формулу (7), то для кожної точки, починаючи з четвертої, потрібно виконати дві операції додавання та одну операцію віднімання (тобто три операції додавання), а також дві операції зсуву. Якщо інтенсивність кольору в першій точці рядка растреризації знаходити з формули (9), у другій точці – з формули (3), у третій точці – з формули (5), то для обчислення інтенсивності кольору в n точках потрібно виконати 6 операцій множення, $5+4+2+3(n-3)=3n+2$ операцій додавання та $1+2+2(n-3)=2n-3$ операцій зсуву.

Формула (8) у неоптимізованому вигляді містить дві операції множення, операцію додавання та операцію віднімання (дві операції додавання). У випадку, коли інтенсивність кольору в перших трьох точках знаходимо аналогічно до попереднього методу, для знаходження інтенсивності кольору в n точках потрібно виконати $5+1+2(n-3)=2n$ операцій множення, $5+4+2+2(n-3)=2n+5$ операцій додавання та 3 операції зсуву.

Оптимізована формула (8) містить дві операції додавання, дві операції віднімання (всього – чотири операції додавання), дві операції зсуву. У такому випадку, для визначення інтенсивності кольору в n точках потрібно виконати 6 операцій множення, $5+4+2+4(n-3)=4n-1$ операцій додавання та $1+2+2(n-3)=2n-3$ операцій зсуву.

При використанні процесора Intel Xeon Phi [14] операція множення потребує 8 тактів ЦП, операції додавання, віднімання та зсуву – по одному такту ЦП. Щоб отримати загальну кількість тактів ЦП для кожного методу було прийнято $n=16$ та домножено отримані значення на «вартість» операції. Результати вимірювань показано в табл. 1.

Таблиця 1 – Кількість тактів ЦП, потрібних для обчислення значень інтенсивності кольору в 16 точках

Тип формули	Множення, тактів	Додавання, тактів	Зсув, тактів	Всього
Формула (2)	1280	80	0	1360
Формула (9)	640	80	0	720
Формула (3)	160	65	15	240
Формула (5)	48	37	29	114
Формула (7)	48	50	29	127
Формула (8)	256	37	3	296
Оптимізована формула (8)	48	63	29	140

При використанні формули (3) досягається підвищення продуктивності на 67% порівняно з формулою (9), при використанні формули (5) – на 84%.

Висновки

Отримано нові аналітичні залежності для визначення інтенсивності кольору точки рядка растреризації через інтенсивності кольору сусідніх точок. Формули не використовують довготривалі мікрооперації множення. Виведено формулу для визначення інтенсивностей кольору в кінцевих точках цифрових сегментів, на які розбито рядок растреризації. Наведено структурні схеми для визначення

інтенсивностей кольору. Отримано порівняльні оцінки підвищення продуктивності. Результати проведених досліджень можуть бути використані у високопродуктивних системах тривимірної графіки.

Список літератури

- [1] О. Н. Романюк, *Комп'ютерна графіка: навчальний посібник*. Вінниця, Україна: ВДТУ, 2001. 130 с.
- [2] О. Н. Романюк, та А. В. Чорний, *Високопродуктивні методи та засоби зафарбовування тривимірних графічних об'єктів. Монографія*. Вінниця, Україна: УНІВЕСУМ-Вінниця, 2006. 190 с.
- [3] О. Н. Романюк, та О. О. Дудник, "Підвищення реалістичності зафарбовування тривимірних графічних об'єктів," *Вісник ХНТУ*, № 3, с. 269-272, 2016.
- [4] А. Н. Романюк, и Т. А. Замковой, "Алгоритмы рендеринга," *Компьютеры+программы*, № 6, с. 54-57, 1998.
- [5] Д. Херн, и М. П. Бейкер, *Компьютерная графика и стандарт OpenGL*, 3-е издание. Москва, Россия: Издательский дом «Вильямс», 2005, 1168 с.
- [6] О. Н. Романюк, "Метод прискороного зафарбовування тривимірних поверхонь з урахуванням їх локальної кривизни," *Вісник Східноукраїнського національного університету*, № 12, с. 166-172, 2008.
- [7] О. Н. Романюк, "Новий підхід до визначення спекулярної складової кольору," *Оптико-електронні інформаційно-енергетичні технології*, № 2, с. 85-92, 2004.
- [8] О. Н. Романюк, "Класифікація дистрибутивних функцій відбивної здатності поверхні," *Наукові праці Донецького національного технічного університету. Серія «Інформатика, кібернетика і обчислювальна техніка»*, Випуск 9 (132), с. 145-151, 2008.
- [9] О. Н. Романюк, "Новий підхід до підвищення реалістичності зафарбовування тривимірних об'єктів за методом Гуро," *Інформаційні технології та комп'ютерна інженерія*, № 2, с. 106-109, 2005.
- [10] О. Н. Романюк, та А. В. Чорний, "Новий підхід до реалізації процедури зафарбовування за методом Фонга," *Вісник Херсонського державного технічного університету*, Вип. 22, с. 154-160, 2003.
- [11] О. Н. Романюк, та А. А. Шаманський, "Метод зафарбовування тривимірних графічних об'єктів без нормалізації векторів нормалей," *Інформаційні технології та комп'ютерна інженерія*, № 2 (6), с. 111-115, 2006.
- [12] О. Н. Романюк, "Використання квадратичної інтерполяції для зафарбовування тривимірних графічних об'єктів," *Реєстрація, зберігання і обробка даних*, т. 8, № 4, с. 31-37, 2006.
- [13] O. A. Romanyuk, and A. Nast, "A method for accelerated computation of color intensities for shading of three-dimensional graphics objects", in *Współczesne problemy informatyki. Algorytmy i modelowanie*. Legnica, Polska: Wydawnictwo Wyższej Szkoły Menedżerskiej, pp. 213-227, 2007.
- [14] A. Fog, *Instruction tables: Lists of instruction latencies, throughputs and micro-operation breakdowns for Intel, AMD and VIA CPUs*. Kongens Lyngby, Denmark: Technical University of Denmark, 2021.

Стаття надійшла: 30.08.2021.

References

- [1] O. N. Romaniuk, *Kompiuterna hrafika: navchalnyi posibnyk*. Vinnytsia, Ukraina: VDTU, 2001. 130s.
- [2] O. N. Romaniuk, ta A. V. Chorny, *Vysokoproduktyvni metody ta zasoby zafarbovuvannia tryvymirnykh hrafichnykh obiektiv*. Monohrafiia. Vinnytsia, Ukraina: UNIVESUM-Vinnytsia, 2006. 190 s.
- [3] O. N. Romaniuk, ta O. O. Dudnyk, "Pidvyshchennia realistychnosti zafarbovuvannia tryvymirnykh hrafichnykh obiektiv," *Visnyk KhNTU*, № 3, s. 269 272, 2016.
- [4] A. N. Romanjuk, i T. A. Zamkovej, "Algoritmy renderinga," *Computers programs*, № 6, pp. 54–57, 1998.
- [5] D. Hern, i M. P. Bejker, *Komp'juternaja grafika i standart OpenGL*, 3-e izdanie. Moskva, Rossija: Izdatel'skij dom «Vil'jams», 2005, 1168 s.
- [6] O. N. Romaniuk, "Metod pryskorenoho zafarbovuvannia tryvymirnykh poverkhon z urakhuvanniam yikh lokalnoi kryvyzny," *VISNIK of the East Ukrainian National University*, № 12, pp. 166–172, 2008.
- [7] O. N. Romaniuk, "Novyi pidkhid do vyznachennia spekuliarnoi skladovoi koloru," *Optoelectronic information-power technologies*, № 2, pp. 85–92, 2004.
- [8] O. N. Romaniuk, "Klasyfikatsiia dystributyvnykh funktsii vidbyvnoi zdatnosti poverkhni," *Naukovi pratsi Donetskoho natsionalnogo tekhnichnogo universytetu. Seriiia «Informatyka, kibernetyka i obchysliuvalna tekhnika»*, Vypusk 9 (132), pp. 145–151, 2008.

- [9] O. N. Romaniuk, "Novyi pidkhdid do pidvyshchennia realistychnosti zafarbovuvannia tryvymirnykh ob'ektiv za metodom Huro," *Information Technologies and Computer Engineering*, № 2, pp. 106–109, 2005.
- [10] O. N. Romaniuk, ta A. V. Chornyi, "Novyi pidkhdid do realizatsii protsedury zafarbovuvannia za metodom Fonha," *Visnyk Khersonskoho derzhavnoho tekhnichnoho universytetu*, Vyp. 22, pp. 154–160, 2003.
- [11] O. N. Romaniuk, ta A. A. Shamanskyi, "Metod zafarbovuvannia tryvymirnykh hrafichnykh ob'ektiv bez normalizatsii vektoriv normaliei," *Information Technologies and Computer Engineering*, № 2 (6), pp. 111–115, 2006.
- [12] O. N. Romaniuk, "Vykorystannia kvadratychnoi interpoliatsii dlia zafarbovuvannia tryvymirnykh hrafichnykh ob'ektiv," *Reiestratsiia, zberihannia i obrobka danykh*, t. 8, № 4, pp. 31–37, 2006.
- [13] O. A. Romanyuk, and A. Hast, "A method for accelerated computation of color intensities for shading of three-dimensional graphics objects", in *Współczesne problemy informatyki. Algorytmy i modelowanie*. Legnica, Polska: Wydawnictwo Wyższej Szkoły Menedżerskiej, pp. 213-227, 2007.
- [14] A. Fog, *Instruction tables: Lists of instruction latencies, throughputs and micro-operation breakdowns for Intel, AMD and VIA CPUs*. Kongens Lyngby, Denmark: Technical University of Denmark, 2021.

Відомості про авторів

Романюк Олександр Никифорович – доктор технічних наук, професор, завідувач кафедри програмного забезпечення.

Дудник Олександр Олександрович – кандидат технічних наук, доцент, доцент кафедри програмного забезпечення.

Озерчук Дмитро Анатолійович – магістрант кафедри програмного забезпечення.

A. N. Romanyuk, A. A. Dudnik, D. A. Ozerchuk

ПОВЫШЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ ЗАКРАСКИ ПРИ ИСПОЛЬЗОВАНИИ ДЛЯ ОПРЕДЕЛЕНИЯ ИНТЕНСИВНОСТИ ЦВЕТА ПОВЕРХНОСТИ ВТОРОГО ПОРЯДКА

Винницкий национальный технический университет, г. Винница

O. N. Romanyuk, O. O. Dudnik, D. A. Ozerchuk

IMPROVING THE PERFORMANCE OF SHADING WHILE USING THE SECOND ORDER SURFACES TO DETERMINE COLOR INTENSITIES

Vinnitsia National Technical University, Vinnitsia

КОМП'ЮТЕРНІ СИСТЕМИ І КОМПОНЕНТИ

УДК 681.325.5

О. Д. Азаров, О. І. Черняк, В. В. Туйчев

ВЕКТОРНИЙ МЕТОД ЛОКАЛІЗАЦІЇ ПОМИЛОК
ПІДВИЩЕНОЇ ЕФЕКТИВНОСТІ

Вінницький національний технічний університет, Вінниця

Анотація. У статті розглянуто проблеми, що виникають під час передавання і зберігання інформації, а саме пошкодження даних під впливом зовнішніх завад. Обґрунтовано актуальність даної теми. Проведено аналіз існуючих підходів до побудови завадостійких кодів, а саме: кодів Ріда-Соломона, кодів Хеммінга, векторних кодів. Запропоновано власний метод побудови дерева згортки при векторному кодуванні, що дозволяє локалізувати подвійні помилки та відновити пошкоджені ділянки даних, за допомогою одного контрольного розряду на слово даних. Запропоноване дерево згортки надає можливість простого розширення розрядності даних. Таким чином, описаний метод достовірного передавання і зберігання інформації має мінімальну надлишковість даних та за рахунок регулярної будови дерева – підвищену ефективність.

Ключові слова: коди Хеммінга, коди Ріда-Соломона, векторні коди, метод фруктових садів, надлишковість, передача інформації.

Анотация. В статье рассмотрены проблемы, возникающие при передаче и хранении информации, а именно повреждения данных под влиянием внешних помех. Обоснована актуальность данной темы. Проведен анализ существующих подходов к построению помехоустойчивых кодов, а именно: кодов Рида-Соломона, кодов Хемминга, векторных кодов. Предложено собственный метод построения дерева свертки при векторном кодировании, который позволяет локализовать двойные ошибки и восстановить поврежденные участки данных, с помощью одного контрольного разряда на слово данных. Предложенное дерево свертки предоставляет возможность простого расширения разрядности данных. Таким образом, описанный метод достоверной передачи информации и хранения имеет минимальную избыточность данных и за счет регулярной строения дерева - повышенную эффективность.

Ключевые слова: коды Хэмминга, коды Рида-Соломона, векторные коды, метод фруктового сада, избыточность, передача информации.

Abstract. The article considers the problems that arise during the transmission and storage of information, namely data corruption under the influence of external interference. The relevance of this topic is substantiated. An analysis of existing approaches to the construction of noise-tolerant codes, namely: Reed-Solomon codes, Hamming codes, vector codes. Our own method of constructing a convolution tree with vector coding is proposed, which allows to localize double errors and recover damaged parts of the data, using one control bit per data word. The proposed convolution tree allows you to easily extend the bit size of the data. Thus, the described method of reliable transmission and storage of information has a minimal redundancy of data and due to the regular structure of the tree - increased efficiency.

Key words: Hamming codes, Reed-Solomon codes, vector codes, orchard method, redundancy, information transfer.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-60-67>.

Вступ

Розширення інформаційної галузі потребує підвищення достовірності при передаванні та зберіганні великих обсягів даних. Всі існуючі підходи до вирішення вказаних задач базуються на введенні у дані надлишковості для зберігання інформації з метою подальшого аналізу появи помилок. Відомо ряд публікацій, що стосуються використання надлишковості для підвищення ефективності обробки інформації [1-11]. До важливих критеріїв вирішення проблем достовірного передавання і зберігання великих обсягів даних, відносять високу ефективність виявлення помилок при незначній надлишковості. Розробка методу, що задовільнить ці критерії є актуальною задачею.

Традиційними способами виявлення та виправлення помилок є коди Ріда-Соломона, коди Хеммінга та інші [12-14]. Всі вони призводять до великої надлишковості, оскільки для локалізації помилок аналізують інформацію про код кожного окремого слова. Збільшення надлишковості даних призводить до зменшення швидкості їх передавання і обробки. Тому перспективним напрямком досліджень для підвищення достовірності передавання і зберігання інформації є локалізація множинних помилок за рахунок аналізу кодів групи сусідніх слів з метою зменшення надлишковості даних.

Одним з ефективних підходів до підвищення достовірності є методи локалізації помилок в двійкових даних та їх подальшого виправлення за допомогою векторних кодів [15-17]. Такі методи передбачають додавання одного контрольного біта до кожного двійкового слова і використання певних розрядів декількох сусідніх слів даних для його обчислення. Відомі публікації, де вказано, що використання векторного методу дозволяє виявляти та виправляти помилки в двійкових даних. Проте, даний підхід потребує подальшого дослідження та обґрунтування. У даній статті авторами запропоновано метод векторного кодування, який дозволяє локалізувати одинарні і подвійні помилки при передаванні і зберіганні великих обсягів інформації за допомогою лише одного контрольного розряду. Запропонований метод використовує вектори з регулярною будовою, що надає можливість простого нарощування розрядності кодових слів.

Актуальність

Однією з найважливіших науково-технічних проблем в сучасному світі є створення автоматизованих систем управління та контролю для виконання різних завдань. У процесі автоматизованого управління та контролю відбувається інтенсивний обмін значними обсягами інформації між окремими частинами системи. При цьому обсяг запам'ятовуваної інформації, а також швидкість її обробки та передачі постійно ростуть. Таким чином, розширення сфери впровадження автоматизованих систем управління потребує вирішення задачі досягнення високої достовірності передавання і зберігання даних при незначній їх надлишковості. Саме тому тема даного дослідження є актуальною.

Мета

Метою статті є розробка методу формування контрольних бітів для зменшення надлишковості даних при їх передаванні та зберіганні.

Задачі

1. Проведення існуючих підходів до побудови кодів з виправленням помилок;
2. Розробка методу векторного кодування з можливістю розширення розрядності.

Аналіз існуючих підходів до побудови кодів з виправленням помилок

Одним з найвідоміших прикладів контролюючих кодів – є коди Хеммінга [14]. Побудова кодів Хеммінга базується на перевірці на парність числа одинарних символів: до послідовності додається такий елемент, щоб число одинарних символів у створеній послідовності було парним (1).

$$r_1 = i_1 \oplus i_2 \oplus \dots \oplus i_k \quad (1)$$

Далі виконується складання по модулю 2

$$S = i_1 \oplus i_2 \oplus \dots \oplus i_n \oplus r_1 \quad (2)$$

Якщо в результаті (2), $S = 0$, то помилки немає, якщо $S = 1$, то одинарна помилка. Такий код називається $(k+1, k)$ або $(n, n-1)$. Перше число – кількість елементів в послідовності, а друге – кількість інформаційних бітів. Для кожного числа перевірючих символів $r = 3, 4, 5, \dots$ існує класичний код Хеммінга з маркуванням (3)

$$(n, k) = (2^r - 1, 2^r - 1 - r), \quad \text{тобто } - (7,4), (15,11), (31,26) \quad (3)$$

При інших значеннях k – в результаті виходить усічений код, наприклад при $k=5$, кодом Хеммінга буде $(9,5)$, який являється усіченим від класичного $(15,11)$. Для кожного класичного коду Хеммінга існують ненульові синдроми і відповідні їм конфігурації помилок, по яким в результаті можна їх локалізувати та виправити. Слід зауважити, що залишаються непоміченими помилки, що виникають одночасно в двох, чотирьох, і т. д. – в парному кількості розрядів. Передбачається, що подвійні, а тим більше множинні помилки малоімовірні.

Проте, існує багато різних задач, в ході яких трапляються не лише одинарні помилки. Накладання помилок зменшує ймовірність знайти та локалізувати помилку. Противагою даному методу є код Ріда-Соломона, який дозволяє локалізувати та виправляти не лише одинарні помилки [15]. Коди Ріда-Соломона (n,k) визначені на r -бітових символах при всіх n і k , для яких виконується умова (4)

$$0 < k < n < 2^r + 2, \quad (4)$$

де k – число інформаційних символів, n – число кодових символів в блоці. Для більшості (n,k) -кодів Ріда-Соломона (далі РС) справджується рівність

$$(k, n) = (2^r - 1, 2^r - 1 - 2 * t), \quad (5)$$

де t – кількість помилкових символів, які може виправити код, а $n-k = 2t$ – число контрольних символів.

Код РС має найбільшу мінімальну відстань можливу для лінійного коду $d(\min) = n-k+1$. Той факт, що $2t$ послідовних степенів α -корні породжуючого многочлена $g(x)$ або, що спектр містить $2t$ послідовних нульових компонентів, є важливою властивістю коду, що дозволяє виправляти t помилок. Не зважаючи на те, що даний алгоритм є надзвичайно потужним інструментом, він може мати надлишковість в 25% від вхідних даних [13].

Найбільш простими способами побудови кодів, локалізуючих і коригуючих помилки передачі або зберігання двійкової інформації, є коди, що базуються на виявленні кратності конкретних бітів даних.

Скоттом Е. та Гетшелем Д. [17] запропоновано векторний код, що дає можливість локалізувати і виправити помилки за допомогою лише одного контрольного розряду на слово. Метод дозволяє ефективно виявляти і виправляти помилку теоретично будь-якої кратності в межах блоку певної кількості слів. Як стверджують автори, така модель помилок найбільш вірогідна при передаванні, записуванні і зчитуванні потокової інформації. Для контролю вони запропонували так званий, метод фруктового саду, концепція якого базується на візуальному враженні, що складається в людини, при проходженні яблуневого саду з регулярно розміщеними рядами дерев. На рисунку 1 надана схема локалізації одинарної помилки за допомогою даного методу при використанні двох векторів згортки за модулем 2.

Всі вектори згортки в один контрольний розряд назвемо деревом згортки. З рисунку видно, що одинарна помилка призведе до порушення парності у двох контрольних розрядах K1 і K2, які назвемо контрольним слідом.

	1			1		2			2	
	1			1	2				2	
	1			■				2		
	1		2	1		2				
	1	2		1	2					
	K1			K2						

Рисунок 1 – Схема локалізації помилки дво-векторним кодом

Відстань між бітами контрольного сліду дозволяє однозначно встановити місце знаходження помилки. Отже, дво-векторний код дозволяє локалізувати і виправити одинарну помилку в межах блоку слів, кількість яких визначається довжиною дерева згортки.

У роботі [16] запропоновано використання три-векторного дерева згортки для локалізації одинарних та виявлення багатократних помилок, як це представлено на рисунку 2.

	2				1					3
		2			1					3
			2		1				3	
				2	1		3			
					2	1	3			
					K					

Рисунок 2 – Дерево згортки для симетричного три-векторного коду

На відміну від дерева згортки на рисунку 1, дана версія дозволяє локалізувати більше одинарних помилок. Таке дерево дозволяє легко визначити співвідношення для обчислення контрольних розрядів. Нехай елементи нульового ($k=0$) розряду є контрольними, тоді для масиву m -розрядних двійкових слів A вони обчислюються додаванням по модулю два за виразом (6).

$$a_i^0 = \left(\sum_{k=1}^{m-1} (a_i^k \oplus a_{i-k}^k \oplus a_{i+k}^k) \right)_{\text{mod} 2}, \quad (6)$$

де a_i^k – k -й розряд i -го слова.

При декодуванні потоку m -розрядних двійкових слів-кодів перевіряється зберігання парності для всіх контрольних бітів, тобто виконується перевірка правильності виразу (7).

$$a_i^0 \oplus \left(\sum_{k=1}^{m-1} (a_i^k \oplus a_{i-k}^k \oplus a_{i+k}^k) \right)_{\text{mod} 2} = 0 \quad (7)$$

Помилковий біт локалізується за допомогою двох розташованих підряд контрольних розрядів, для яких при декодуванні виявлено порушення парності. Якщо під час декодування потоку m -розрядних слів виявлено порушення парності у контрольних розрядах a_i та a_{i+t} , де $t < 2m$, то при $i < m$ ідентифікується

помилка у слові з номером i , а при $i \geq m$ ідентифікується помилка у слові з номером $(i+t)$. В обох випадках помилковий розряд матиме номер t .

Однак, у випадку виникнення подвійної помилки у даному блоці можна навести приклади, коли один і той самий контрольний біт є елементом двох контрольних слідів. Подвійне порушення парності такого розряду не буде поміченим.

Метод векторного кодування для локалізації подвійних помилок з можливістю розширення розрядності

Для виправлення подвійних помилок в словах з розрядністю p 'ять, (включаючи контрольний) у [17] запропонована форма дерева згортки, зображена на рисунку 3, що відповідає виразу (8).

$$a_i^0 = \sum_{j=0}^n (a_i^j \oplus a_{i+j+1}^j) \oplus a_{i+16}^0 \oplus a_{i+10}^1 \oplus a_{i+14}^2 \oplus a_{i+13}^3 \oplus a_{i+12}^4. \quad (8)$$

№ розряду	№ слова																
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
4	1					2							3				
3	1				2									3			
2	1			2											3		
1	1		2								3						
0	К	2															3

Рисунок 3 – Дерево згортки три-векторного коду

Така можливість базується на тому, що на кожному розрядному зрізі у згортці беруть участь три розряди, по одному з кожного вектора. Причому, відстані першим та другим, другим та третім, першим та третім розрядами є унікальними для даного зрізу і на других розрядних зрізах в дереві більше не повторюються. Тому навіть якщо відбудеться накладання біта контрольного сліду одного дерева та іншого, що приведе до неможливості визначення двох відстаней, то все одно залишиться третя унікальна відстань між бітами контрольного сліду, яка дозволить однозначно визначити помилковий розряд. Таким чином, при використанні векторного коду для локалізації n -кратних помилок потрібно будувати дерево згортки, яке має $(n+1)$ векторів, причому повинні бути унікальними всі можливі відстані між векторами на кожному розрядному зрізі D .

Недоліком відомого дерева згортки [17], представленого на рисунку 2, є відсутність простої і зрозумілої закономірності побудови третього вектора для різних розрядностей. Це ускладнює нарощування розрядності слів при використанні векторного кодування для трьох і більшої кількості векторів.

Розглянемо умови яким повинно відповідати дерево згортки для локалізації і виправлення двох помилок. При наявності трьох векторів в формуванні одного контрольного розряду беруть участь три інформаційних розряди однієї ваги з різних слів. Контрольні розряди, в формуванні яких бере участь інформаційний розряд a_i^j назовемо контрольним слідом інформаційного розряду.

Нехай $\{a_f^k, a_g^k, a_h^k\} \subset D_i$, тоді $\{a_{f+b}^k, a_{g+b}^k, a_{h+b}^k\} \subset D_{i+b}$. Знайдемо контрольний слід розряду a_b^k . Нехай $b = f + x = g + y = h + z$, тоді правдивою є рівність (9)

$$\begin{aligned} \{a_b^k, a_{g+x}^k, a_{h+x}^k\} &\subset D_{i+x}; \\ \{a_{f+y}^k, a_b^k, a_{h+y}^k\} &\subset D_{i+y}; \\ \{a_{f+z}^k, a_{g+z}^k, a_b^k\} &\subset D_{i+z}. \end{aligned} \quad (9)$$

Визначимо умову накладання контрольних слідів від двох розрядів в двох точках.

Нехай $\{a_k^i, a_p^i, a_m^i\} \subset D_r$, нехай далі $k > p > m$ і $k-p=b=m$, тобто в дереві на рівнях i та j існують дві пари рівновіддалених вершин. Тоді для довільного a_d^i ($d = k + x = p + y$) отримаємо контрольний слід (10), при чому $k=d-x$, $p=d-y$ і, отже, $k-p=y-x=h$, тобто контрольний слід a_d^i включає два контрольних розряди, що знаходяться на відстані h один від одного.

$$\begin{aligned} \{a_d^i, a_{p+x}^i\} &\subset D_{r+x}; \\ \{a_{k+y}^i, a_d^i\} &\subset D_{r+y}, \end{aligned} \quad (10)$$

Аналогічно для довільного a_c^j ($c = b + z = m + \tau$) отримаємо контрольний слід (11)

$$\begin{aligned} \{a_c^j, a_{m+z}^j\} &\subset D_{r+z}; \\ \{a_{b+\tau}^j, a_c^j\} &\subset D_{r+\tau}; \end{aligned} \quad (11)$$

при чому $b=c-z$, $m=c-\tau$ і, отже, $b-m=\tau-z=h$, тобто контрольний слід a_d^i включає два контрольних розряди, які знаходяться на відстані h один від одного. В тому випадку, коли $\tau = y$, то $z = \tau - h = x$, тобто два контрольних розряди контрольних слідів a_d^i і a_c^j співпадають. Отже, умова $k-p=b-m$ є достатньою, для того щоб виникло співпадіння двох різних контрольних слідів. Необхідність даної умови доводиться від зворотного. Дійсно, нехай контрольні сліди розрядів a_b^k, a_c^p мають спільні контрольні розряди d_j і d_{j+t} . Це означає, що як a_b^k так і a_c^p є елементами дерев згортки з вершинами в слові j в слові $j+t$.

$$\begin{aligned} \{a_b^k, a_c^p\} &\subset D_j; \\ \{a_b^k, a_c^p\} &\subset D_{j+t}. \end{aligned} \quad (12)$$

Оскільки формула згортки для всіх розрядів одна, то із (12) слідує (13)

$$\{a_{b+t}^k, a_{c+t}^p, a_b^k, a_c^p\} \subset D_{j+t}; \quad (13)$$

тобто дерево містить два розряди на рівні k з відстанню між ними t і два розряди на рівні p з відстанню між ними t .

Таким чином, наявність двох пар рівновіддалених вершин в дереві є необхідною і достатньою умовою співпадіння контрольних слідів різних розрядів в двох точках. Отже, для того щоб не було накладання контрольних слідів в двох точках, дерева згортки не повинні включати в себе рівновіддалені розряди однієї ваги. Оскільки, при три-векторному дереві контрольний слід складається з трьох контрольних розрядів, то при подвійній помилці накладання контрольних слідів більше чим в двох точках неможливе.

У три-векторному дереві на кожному рівні знаходиться по три розряди a_i, a_l, a_k , де i, l, k – номери слів такі, що $i < l < k$. Нехай $l-i=x$, $k-l=y$, $k-i=z$. Для j -го рівня повинні виконуватись умови

$$\begin{aligned} x_j &\neq y_j; \\ x_j &\neq z_j; \\ y_j &\neq z_j. \end{aligned} \quad (14)$$

Кількість слів, що беруть участь в одному дереві згортки позначимо як R . При чому, $z_{\max}=R$. Для зменшення R потрібно щоб множини $\{x\}$, $\{y\}$, $\{z\}$ містили натуральні числа, що заповнюють без проміжків відрізок числової вісі від одиниці до R . Надамо більш строгі формулювання умови для дерева згортки з мінімальним R .

Для будь-якого n визначити $\{x_1, \dots, x_n\}$, $\{y_1, \dots, y_n\}$, $\{z_1, \dots, z_n\}$ такі, що x, y, z – натуральні; $\{x\} \cap \{y\} = \emptyset$, $\{y\} \cap \{z\} = \emptyset$, $\{x\} \cap \{z\} = \emptyset$, причому, $\{\{x\} \cap \{y\} \cap \{z\}\} = \{1, 2, 3, \dots, 3n+1\}$.

У даній роботі авторами пропонується конфігурація дерева згортки, яка при всіх позитивних властивостях відомого рішення для будь-якої розрядності надає просту і зрозумілу закономірність побудови дерева з трьох векторів, представлену виразом для визначення контрольного розряду $K_i = a_i^p$:

$$V1_i = \left(\sum_{j=1}^{n-1} (a_j^j) \right)_{\text{mod}2};$$

$$V2_i = \left(\sum_{j=0}^{n-1} (a_{i+j+1}^j) \right)_{\text{mod}2} ;$$

$$V3_i = \left(\sum_{j=0}^{(n-1)/2} (a_{i+2n+2-j}^{n-1-2j}) \right)_{\text{mod}2} \oplus \left(\sum_{j=0}^{(n-3)/2} (a_{i+3n+2-j}^{n-2-2j}) \right)_{\text{mod}2} ;$$

$$a_i^0 = V1_i \oplus V2_i \oplus V3_i.$$

На рисунку 4 зображено приклад запропонованого авторами дерева згортки для виправлення подвійних помилок, де К – контрольний розряд.

В запропонованому варіанті дерева згортки прослідковується проста закономірність його будови, що дозволить застосовувати даний метод для контролю слів даних будь-якої розрядності, а також при необхідності нарощувати розрядність пристрою контролю.

№ розряду	№ слова																
	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17
4	1					2						3					
3	1				2												3
2	1			2							3						
1	1		2														3
0	К	2								3							

Рисунок 4 – Запропонований варіант дерева згортки для три-векторного коду,

Висновки

У статті подано теоретичні аспекти модифікованого алгоритму векторного коду. Було описано існуючі методи кодування даних, проведено аналіз їх переваг та недоліків. Наведено аналітичні вирази для опису побудови дерева згортки три-векторного коду, яке дозволяє виправляти подвійні помилки за допомогою одного контрольного розряду на слово. На основі проведеного аналізу запропоновано власний метод, який дозволяє будувати векторні коди з можливістю розширення розрядності даних за рахунок регулярної будови кодового дерева згортки, що і є підвищенням ефективності.

Список літератури

- [1] А. Д. Азаров, А. И. Черняк, "Полнофункциональная побитовая обработка результатов аналого-цифрового преобразования," на III міжнародній наук.-практ. конф. *Методи та засоби кодування, захисту й ущільнення інформації*, Вінниця, с. 208–209. 2011.
- [2] Olexiy D. Azarov, Olexander G. Murashchenko, Olexander I. Chernyak, Andrzej Smolarz, Gulzhan Kashaganova, "Method of glitch reduction in DAC with weight redundancy," in *16th Conference on Optical Fibers and Their Applications*, Proc. SPIE 9816, 98161T, Lublin and Naleczow, Poland, 2015; doi: 10.1117/12.2229045; <http://dx.doi.org/10.1117/12.2229045>.
- [3] О. Д. Азаров, О. І. Черняк, "Метод побудови швидкодіючих фібоначчєвих лічильників," *Проблеми інформатизації та управління*, №2 (46), с. 5–8. 2014.
- [4] О. Д. Азаров, О. І. Черняк, "Визначення довжини перенесення при додаванні в системах числення з адитивними та мультиплікативними співвідношеннями між вагами розрядів," *Наукові праці Донецького національного технічного університету. Серія: Обчислювальна техніка та автоматизація*, № 74, с. 401–407. 2004. ISSN 1996-1588.
- [5] О. Д. Азаров, О. І. Черняк, "Структурна організація побітового множення і ділення кодів золоті пропорції," *Проблеми інформатизації та управління*, №3(21), с. 5–13. 2007. ISSN 2073-4751.
- [6] О. Д. Азаров, О. І. Черняк, "Розрядність пристроїв порозрядного додавання в АМ-системах числення," *Наукові праці Вінницького національного технічного університету* № 4, с. 1–9. 2010. [Електронний ресурс]. Режим доступу: <http://praci.vntu.edu.ua/index.php/praci/article/view/233>. Дата звертання: Лис. 2020.
- [7] О. Д. Азаров, О. І. Черняк, "Структурна організація побітового додавання і віднімання кодів золоті 1-пропорції з урахуванням знаків," *Інформаційні технології та комп'ютерна інженерія*, № 3(22), с. 13–16. 2011. ISSN 1999-9941.

- [8] О. Д. Азаров, О. І. Черняк, "Аналіз витрат обладнання пристроїв побітової арифметики у системі числення золоті 1-пропорції," *Проблеми інформатизації та управління*, № 2(38), с. 5–9. 2012. ISSN 2073-4751.
- [9] О. Д. Азаров, О. І. Черняк, *Повнофункціональна побітова потокова арифметика зі зменшеними витратами обладнання: монографія*. Вінниця, Україна: ВНТУ, 2013.
- [10] О. Д. Азаров, О. І. Черняк, "Обмеження адитивних співвідношень при порозрядній потоковій обробці в АМ-системах числення," *Інформаційні технології та комп'ютерна інженерія*, № 3(31), с. 67–71. 2014. ISSN 1999-9941.
- [11] О. Д. Азаров, О. І. Черняк, О. Г. Муращенко, "Порозрядне додавання в АМ-системах числення на основі адитивних перетворень," *Проблеми інформатизації та управління*, № 1(45), с. 14–21. 2014. ISSN 2073-4751.
- [12] Richard W. Hamming *The Art of Doing Science and Engineering*, Australia: GORDON AND BREACH SCIENCE PUBLISHERS, 2005.
- [13] Stephen B. Wicker, Vijay K. Bhargava *Reed-Solomon Codes and Their Applications*. Wiley-IEEE Press, 1999. ISBN: 978-0-7803-5391-6.
- [14] А. И. Королев, *Коды и устройства помехоустойчивого кодирования информации*. Минск, 2002
- [15] А. И. Черняк, А. П. Стахов, В. П. Марценюк, В. И. Пилипчак, О. А. Пленсак, "Устройство кодирования по векторному методу," *МКИД G 06 F 11/10 №1451700 A1*, 15.01.89.
- [16] О. П. Шафеева, "Векторные коды для локализации ошибок в двоичных данных," *Омский научный вестник*, №3 (32). 2005.
- [17] Э. Скотт, Д. Гетшель, "Исправление многобитовых ошибок при помощи одного контрольного бита на слово," *Электроника*, № 9, с. 40–47. 1981.

Стаття надійшла: 09.08.2021.

References

- [1] O. D. Azarov, O. I. Chernyak, "Polnofunktsionalnaya pobitovaya obrabotka rezultatov analogo-tsifrovogo preobrazovaniya," in *3-th international scientific-practical Conference on Methods and Means of Encoding, Protection and Compression of Information* (ММЕРСІ 2011), Vinnytsya, Ukraine, April 20-22, 2011, s. 208–209.
- [2] Olexiy D. Azarov, Olexander G. Murashchenko, Olexander I. Chernyak, Andrzej Smolarz, Gulzhan Kashaganova, "Method of glitch reduction in DAC with weight redundancy," in *16th Conference on Optical Fibers and Their Applications*, Proc. SPIE 9816, 98161T, Lublin and Naleczow, Poland, 2015; doi: 10.1117/12.2229045; <http://dx.doi.org/10.1117/12.2229045>.
- [3] O. D. Azarov, O. I. Chernyak, O. G. Murashchenko, "The construction method for high-speed Fibonacci counters," *Informatics and control problems*, №2(46), pp. 5–8. 2014.
- [4] O. D. Azarov, O. I. Chernyak, "Vyznachennya dovezhyny perenesennya pri dodavanni v systemah chislennya z adydyvnymy i multiplikativnymy spivvidnoshennyamy mizh vagamy rozryadiv," *Scientific papers of Donetsk National Technical University. Series: Informatics, Cybernetics and Computer Science*, № 74, pp. 401–407. 2004. ISSN 1996-1588.
- [5] O. D. Azarov, O. I. Chernyak, "Structurna organizatsiya pobitovogo mnojennya i dilenyarjlsd kodiv zolotoi proporsii," *Informatics and control problems*, №3(21), pp. 5–13. 2007. ISSN 2073-4751.
- [6] O. D. Azarov, O. I. Chernyak, "Rozryadnist prystroiv porozryadnogo dodavannya v AM-systemah chislennya," *Scientific Works of Vinnytsia National Technical University* № 4, pp. 1–9. 2010. [Online resource]. Available: <http://praci.vntu.edu.ua/index.php/praci/article/view/233>. Access on: Nov. 2020.
- [7] O. D. Azarov, O. I. Chernyak, "The structural organization of bit-serial adding and subtraction for golden 1-ratio codes with signs," *Information technology and computer engineering*, № 3(22), pp. 13–16. 2011. ISSN 1999-9941.
- [8] O. D. Azarov, O. I. Chernyak, "Analiz vytrat obladnannya prystroiv pobitovoi aryfmetryky u systemi chyslennya zolotoi 1-proporsii," *Informatics and control problems*, №2(38), pp. 5–9. 2012. ISSN 2073-4751.
- [9] O. D. Azarov, O. I. Chernyak, *The full-function pipe-line bit-serial arithmetic with reduced hardware expenses: monography*. Vinnytsia, Ukraine: VNTU, 2013.
- [10] O. D. Azarov, O. I. Chernyak, "The limitations of additive relationships at order-sequential pipeline processing in AM numerical systems," *Information technology and computer engineering*, №3(31), pp. 67–71. 2014. ISSN 1999-9941.
- [11] O. D. Azarov, O. I. Chernyak, O. G. Murashchenko, "Porozryadne dodavannya v AM-systemah chyslennya na osnovi adydyvnyh spivvidnoshen," *Informatics and control problems*, №1(45), pp. 14–21. 2014. ISSN 2073-4751.
- [12] Richard W. Hamming *The Art of Doing Science and Engineering*, Australia: GORDON AND BREACH SCIENCE PUBLISHERS, 2005.
- [13] Stephen B. Wicker, Vijay K. Bhargava *Reed-Solomon Codes and Their Applications*. Wiley-IEEE

Press, 1999. ISBN: 978-0-7803-5391-6.

- [14] A. I. Korolev *Kody i ustroystva pomehoustoychevogo kodirovaniya informatsii*. Minsk, 2002.
- [15] O. I. Chernyak, O. P. Stakhov, V. P. Martsenyuk, V. I. Pilipchak, O. A. Plevsak, "Ustroystvo kodirovaniya po vektornomu metodu," *МКИ4 G 06 F 11/10 №1451700 AI*, 15.01.89.
- [16] O. P. Shafeeva "Vektornye kody dlya lokalizatsii oshibok v dvoichnyh danyh," *Ovskiy nauchniy vestnik*, №3 (32). 2005.
- [17] E. Scott, D. Getshel, "Correcting multi-bit errors with one check bit per word," *Electronics*, № 9, pp. 40–47. 1981.

Відомості про авторів

Азаров Олексій Дмитрович – доктор технічних наук, професор, завідувач кафедри обчислювальної техніки.

Черняк Олександр Іванович – кандидат технічних наук, доцент кафедри обчислювальної техніки.

Туйчев Владислав Володимирович – аспірант кафедри обчислювальної.

А. Д. Азаров, А. И. Черняк, В. В. Туйчев

ВЕКТОРНЫЙ МЕТОД ЛОКАЛИЗАЦИИ ОШИБОК ПОВЫШЕННОЙ ЭФФЕКТИВНОСТИ

Винницкий национальный технический университет, г. Винница

O. D. Azarov, O. I. Chernyak, V. V. Tuychev

VECTOR METHOD OF ERRORS LOCALIZATION WITH HIGHER EFFICIENCY

Vinnitsia National Technical University, Vinnitsia

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ОБЧИСЛЮВАЛЬНІ МЕТОДИ

УДК 004.4, 004.6

Т. А. Чупілко, Ю. В. Ульяновська, М. Ф. Мормуль, А. Е. Лагода

**PYTHON ДЛЯ ОБРОБКИ ДАНИХ І МОДЕЛЮВАННЯ
ФІНАНСОВО-ЕКОНОМІЧНИХ ПОКАЗНИКІВ**

Університет митної справи та фінансів, Дніпро

Анотація. У статті розглянуто аспекти ефективної обробки даних. Значна увага приділяється проблемам, що виникають при моделюванні і прогнозуванні даних та роль досліджень для прийняття рішень. Визначаються етапи роботи з даними та особливості, що є притаманними кожному етапу. Особливе місце у роботі займає опис можливостей програмної обробки даних з використанням мови Python, яка набуває все більшої популярності завдяки простоті, гнучкості, відкритому коду, зручності роботи з даними у різних форматах, а також багатьом розробленим пакетам, які сприяють швидкій та ефективній обробці інформації. Розглядаються NumPy, Pandas, які надають структури даних і функції, що дозволяють зробити роботу зі структурованими даними простою і швидкою, найпопулярніший інструмент для візуалізації даних Matplotlib, пакети для різних обчислювальних задач SciPy, Statsmodels, а також пакет, орієнтований на машинне навчання Scikit-learn. Наводиться приклад використання Python для задач митної сфери. Авторами створено програму для розрахунку, в якій використовуються вище зазначені пакети. Будуються регресійні моделі для аналізу поповнення державного бюджету України надходженнями від митних органів за рахунок ввізного та вивізного мита. Проводиться аналіз моделей на основі економетричних методів моделювання та розраховуються прогнозні оцінки надходжень.

Ключові слова: Python, обробка даних, моделювання, прогнозування, регресійна модель.

Аннотация. В статье рассмотрены аспекты эффективной обработки данных. Значительное внимание уделяется проблемам, возникающим при моделировании и прогнозировании данных, а также роль исследований для принятия решений. Определяются этапы работы с данными и особенности, которые присущи каждому этапу. Особое место в работе занимает описание возможностей программной обработки данных с использованием языка Python, которая приобретает все большую популярность благодаря простоте, гибкости, открытому коду, удобству работы с данными в различных форматах, а также многим разработанным пакетам, которые способствуют быстрой и эффективной обработке информации. Рассматриваются NumPy, Pandas, которые создают структуры данных и функции, позволяющие сделать работу со структурированными данными простой и быстрой, самый популярный инструмент для визуализации данных Matplotlib, пакеты для различных вычислительных задач SciPy, Statsmodels, а также пакет, ориентированный на машинное обучение Scikit-learn. Приводится пример использования Python для задач таможенной сферы. Авторами создана программа для расчета, в которой используются вышеуказанные пакеты. Строятся регрессионные модели для анализа пополнения государственного бюджета Украины поступлениями от таможенных органов за счет ввозной и вывозной пошлины. Проводится анализ моделей на основе эконометрических методов моделирования и рассчитываются прогнозные оценки поступлений.

Ключевые слова: Python, обработка данных, моделирование, прогнозирование, регрессионная модель.

Abstract. The article considers aspects of efficient data processing. Much attention is paid to the problems that arise in modeling and forecasting data and the role of research in decision-making. The stages of work with data and features that are inherent in each stage are determined. A special place in the work is described by the possibilities of software data processing using Python, which is becoming increasingly popular due to simplicity, flexibility, open source, ease of working with data in various formats, as well as many developed packages that facilitate fast and efficient information processing. NumPy, Pandas, which provide data structures and functions that make working with structured data simple and fast, the most popular tool for data visualization Matplotlib, packages for various computational tasks SciPy, Statsmodels, as well as a package focused on machine learning Scikit-learn. An example of using Python for customs tasks is given. The authors have created a program for calculation, which uses the above packages. Regression models are being built to analyze the replenishment of the state budget of Ukraine with revenues from customs authorities at the expense of import and export duties. The analysis of models on the basis of econometric methods of modeling is carried out and forecast estimates of receipts are calculated.

Key words: Python, data processing, modeling, forecasting, regression model.

DOI: <https://doi.org/10.31649/1999-9941-2021-51-2-68-77>.

Вступ

В сучасному світі компанії, підприємства, установи мають справу з великим обсягом даних. Останнім часом все більшої популярності набувають технології, що дозволяють працювати з великими масивами інформації. Підхід до обробки даних залежить, насамперед, від їх типу, мети використання, можливостей підприємства чи установи в організації збору, систематизації, аналізу даних, які часто бувають обмеженими. Наразі невелика кількість підприємств і установ мають змогу інвестувати значні кошти в розвиток цих технологій для власних потреб. Великі компанії, безумовно, мають можливості для розвитку бізнес-аналітики. Для більшості технік роботи з даними обсяг даних може бути і великим, і малим. Наскільки великі дані потрібні для результату, визначається інтересами компанії. Не завжди компанії мають справу із генерованими даними великих обсягів. Найчастіше є певна база даних компанії, яку і потрібно використовувати.

До основних технік аналізу відносяться кластерний та факторний аналіз, моделювання і прогнозування на основі економетричних та оптимізаційних методів, визначення викидів, штучний інтелект, сітьові графи, машинне навчання. Деякі з цих методів розвивалися давно, інші з'явилися нещодавно. При правильному виборі різні техніки і технології є однаково ефективними для оцінювання ситуації в різних сферах діяльності та прийняттю відповідних управлінських рішень. Тому важливим є розуміння, яка те-

хніка для розв'язання якої проблеми підходить, як ці техніки працюють, і як їх застосовувати для моделювання показників.

В різних технологіях для вивчення даних використовуються схожі базові математичні інструменти: і у стандартних пакетах обробки даних та відповідних бібліотеках і модулях у популярній об'єктно-орієнтованій мові Python, і в широко відомому і розповсюдженому пакеті MS Office, і в інших програмних продуктах на кшталт мови R, що є поширеною для статистичного аналізу даних.

В означеній темі роботи автори використовували роботи іноземних авторів [1], [2], де розглядається використання Python та програмних пакетів саме для аналізу даних. В опублікованих матеріалах [3] – [5] порушені питання, пов'язані з проблематикою цієї роботи. Є певні роботи українських та іноземних вчених щодо міркувань про доступ до великих даних, імплементацію їх у статистику, реальну користь їх і т. ін. Незважаючи на значний науковий доробок щодо загальних питань про дані, дослідження із застосуванням інструментарію ефективної обробки даних із застосуванням мов програмування відсутні, і, зокрема, в митній сфері.

Актуальність

Моделювання охоплює різні галузі, різноманітні показники. Особливо важливим є моделювання фінансово-економічних показників, яке дає змогу прогнозувати їх значення при наявності певної тенденції, дати оцінку прогнозу точкову та інтервальну, що визначається довірчим інтервалом. Спеціальні критерії дозволяють оцінити якість побудованої моделі. Тож, аналітика даних сприяє покращенню процесу прийняття рішень у будь-якій сфері діяльності. Але, разом з тим, потрібні інструменти для високоефективної та швидкої обробки даних.

Останнім часом мова програмування Python та велика кількість бібліотек з відкритим кодом, які динамічно поповнюються, є дуже популярним і потужним інструментом, що дозволяє ефективно обробляти дані, моделювати і прогнозувати показники, використовуючи сумісну можливість написання коду та готові рішення.

Технології опрацювання даних визначаються характером даних та метою дослідження.

Але є спільні проблеми при використанні різних технологічних інструментів – відбір і підготовка даних, а також фахове опрацювання результатів.

Щоб скористатися навіть абсолютно автоматизованою системою обробки даних, їх потрібно правильно відібрати, відсортувати, нормалізувати, обрати метод аналізу, і після обробки даних програмою провести саме аналіз, інтерпретацію, прогнозування і т. ін.

На етапі підготовки даних виникають певні проблеми, пов'язані з різними форматами даних, отриманих із різних джерел, з обмеженим доступом до даних, обумовленим цінністю даних та конфіденційністю і суворою регламентованістю. Дані можуть мати різні одиниці виміру, різні рівні агрегування.

Покращення якості даних, розуміння, як дані взаємодіють між собою, оцінювання розподілів і приведення до певного формату неможливе без знання фундаментальних основ відповідного математичного апарату.

Сучасні пакети підтримують багато популярних методів моделювання і оцінювання моделей. Застосування потребує вивчення мови і освоєння пакетів, в яких для використання певного методу чи функції потрібно задати різноманітні параметри для виконання.

Ще одна проблема пов'язана з доступом до даних. Офіційні статистичні дані, які є основою для моделювання і прогнозування фінансово-економічних показників в масштабах країни, зокрема в митній сфері, є досить обмеженими і, в основному, агрегованими. Часто наводяться на офіційних сайтах або в статистичних збірниках дані, що не є нормалізованими, або їх кількість недостатня для побудови адекватних моделей. Тож, задачі моделювання обмежуються тією статистикою, що є офіційно доступною.

Процес моделювання залежить від якості даних і, не менше, від професіоналізму аналітика.

Мета

Метою роботи є аналіз ефективних інструментів для обробки та моделювання даних; застосування Python та бібліотек для аналізу, моделювання і прогнозування фінансово-економічних показників, на прикладі офіційних даних митних надходжень до державного бюджету України по окремим видам, а саме, вивізному та ввізному миту.

Задачі

1. Проаналізувати проблеми у роботі з даними та аспекти застосування інструментів для ефективної обробки даних.
2. Застосувати найбільш придатні інструменти для конкретної задачі моделювання фінансово-економічних показників (на прикладі митних надходжень до бюджету від окремих показників).

Розв'язання задач

Інструменти для ефективної та швидкої обробки даних

Для обробки даних доцільно використовувати технології, найбільш прийнятні для певного типу і задач, що розв'язує дослідник.

У деяких задачах достатньо використання зручного, зрозумілого і доступного інструменту, як MS Excel, що має широкі можливості і пакет аналізу, хоча і дещо обмежений, але придатний для отримання результатів у першому наближенні для уявлення про характер даних. За допомогою електронних таблиць неможливо в програмному режимі запустити виробничу модель, наприклад, штучного інтелекту, але за їх допомогою можна проаналізувати характер даних, змодельовати і спрогнозувати результат. Цей результат можна отримати на основі класичних підходів теорії ймовірностей та математичної статистики щодо нормування даних, кореляційного та регресійного аналізу, оцінювання прогнозних точкових та інтервальних значень, а також за допомогою процедур для визначення оптимальних розв'язків лінійних та нелінійних задач оптимізації.

Для використання автоматизації обробки даних в програмному режимі потрібні знання тієї чи іншої мови програмування. Але для розуміння сутності аналізу, що використовується в технологіях обробки і різних прикладних пакетах таких, як Statistika, SPSS і т. ін., не обов'язково знати, як пишеться код. Ці потужні інструменти включають різноманітний аналіз, в тому числі, регресійний, факторний, кластерний, побудову моделей за допомогою нейронних мереж і багато іншого, а також дають можливість отримати графічне відображення результатів, якщо дозволяє розмірність і постановка задачі.

У процесі роботи з даними можна виділити декілька етапів.

1. Призначення цілі дослідження. При цьому готується проектне завдання і оцінюються мета дослідження і вартість роботи.
2. Збір і підготовка даних, так званий «розвідувальний аналіз». Певні складнощі виникають уже на цьому етапі. Дані можуть бути розрізненими, в різних форматах і потребують нормалізації і приведення до однородності. Матриці можуть бути не повністю заповненими, виродженими. Обов'язково потрібно підібрати алгоритм для заповнення порожнеч. Нерідко в даних бувають значні відхилення, тобто, викиди, які потрібно усунути, тобто очистити дані. Інакше ніякі методи моделювання не приведуть до адекватної моделі. Таким чином, процес підготовки даних є дуже кропітким і рутинним, майже «ручним», потребує інтелектуального підходу і розуміння цілі дослідження.
3. Аналіз і моделювання даних, тобто вибір моделі та оцінювання її параметрів, в машинному навчанні – «тренування моделі». На цьому етапі потрібно зрозуміти, як дані пов'язані між собою, оцінити розподіли даних, визначити і усунути викиди, а також перевірити наявність мультиколінеарності в системі та таких негативних явищ, як гетероскедастичність та автокореляція, які потребують додаткових перетворень змінних та особливих методів. Для цього використовуються певні статистичні методи і просте моделювання. Постають питання: чи пов'язані між собою досліджувані фактори і показник, чи є мультиколінеарність в системі даних, чи можна зменшити кількість змінних і тим самим спростити модель, яку форму залежності обрати для моделювання, яким способом звести модель до лінійної форми і т. ін. На цьому етапі потрібні знання предметної області, а також математики, теорії ймовірностей і математичної статистики. Тільки після указаних досліджень і перетворень даних можна скористатися готовими рішеннями – пакетами програм. Сам процес моделювання – «тренування моделі» – означає побудову різних моделей на одному наборі даних, випадково відібраних із загальної сукупності. Кількість даних можна варіювати з допомогою параметрів, що задаються для обраного методу. Можна обрати найкращою модель за певними критеріями, наприклад, метод найменших квадратів, або метод, заснований на дереві рішень (знову ж таки, з різними параметрами, що можна варіювати), або метод абсолютних відхилень і т.п. Можна тренувати набір даних декілька разів, змінюючи параметри, і таким чином, досягти найкращого результату. Тож, побудова моделі – ітераційний процес і потребує навичок дослідника.
4. Перевірка адекватності моделі і значимості факторів моделі. Після отримання найкращого результату (наприклад, порівнюються сума квадратів відхилень і обирається набір параметрів, що дає найменше із усіх) оцінювання якості моделі відбувається за статистичними критеріями. Якщо якість незадовільна, то модель потрібно «перенавчити».
5. Застосування моделі до незнайомих даних – так званий «тренувальний сет» обирається із тієї ж вибірки – «прогностичне моделювання», тобто визначається прогноз.

Описаний підхід застосовується для задач моделювання і прогнозування при машинному навчанні (Machine Learning). Python, наприклад, має свою бібліотеку Scikit-learn з різноманітними алгоритмами.

Машинне навчання наразі є дуже популярною і перспективною технологією серед аналітиків (data-scientists). Ринок машинного навчання швидко зростає. З 2016 року його обсяг подолав позначку в \$ 1

млрд, а до 2025 року, судячи з прогнозів, він може збільшитися до \$ 39,98 млрд. 60% компаній в світі вже використовують машинне навчання.

Серед завдань, які можуть вирішуватися засобами машинного навчання, можна зазначити задачі моделювання і прогнозування показників в залежності від одного або декількох факторів або оптимізаційні задачі. Використовуються як традиційні методи економетричного аналізу, включаючи однофакторні, багатофакторні моделі на основі методу найменших квадратів, так і нетрадиційні, типу, дерева рішень з великою кількістю встановлюваних параметрів, що дають гнучкість моделювання параметрів моделей.

Набувають популярності так звані «нейронні мережі». При моделюванні використовуються поняття ризику, кількісні ознаки якого обчислюються у відповідності до числових характеристик дискретних та неперервних випадкових величин.

За останні десять років Python перетворився в одну із найважливіших мов програмування, застосовуваних у науці про дані, в машинному навчанні та розробці програмного забезпечення загального призначення в академічних установах і промисловості. Поліпшені бібліотеки для Python сприяли тому, що він став серйозним конкурентом в рішенні задач створення додатків обробки даних.

У багатьох сучасних середовищах застосовується загальний набір успадкованих бібліотек, написаних на FORTRAN і C, що містять реалізації алгоритмів лінійної алгебри, оптимізації, інтегрування та ін. Тому численні компанії використовують Python як «клей» для об'єднання написаних за багато років програм.

Пакети Python для роботи з даними

NumPy, скорочення від «Numerical Python», – основний пакет для виконання наукових розрахунків на Python. Поверх NumPy побудовано інші бібліотеки.

Основні можливості пакету: швидко і ефективно можна створювати об'єкти багатовимірних масивів ndarray; має функції для виконання обчислень з елементами одного масиву або математичних операцій з декількома масивами; надає засоби для читання і запису на диски наборів даних, представлених у вигляді масивів; використовує операції лінійної алгебри, перетворення Фур'є і генератор випадкових чисел; має засоби для інтеграції з кодом, що написаний на C, C++ або Fortran.

NumPy значно прискорює роботу з масивами. Як засіб зберігання і маніпуляції даними, масиви NumPy значно ефективніші за вбудовані в Python структури даних.

Багато засобів обчислень, орієнтовані на Python, або використовують масиви NumPy в якості основної структури даних, або якимось іншим способом організують інтеграцію з NumPy.

Pandas надає структури даних і функції, що дозволяють зробити роботу зі структурованими даними простою і швидкою. Завдяки цій бібліотеці Python перетворився в потужне і продуктивне середовище аналізу даних. Основні об'єкти *pandas* – це *DataFrame* – двовимірна таблиця, в якій рядки і стовпці мають мітки, і *Series* – об'єкт одновимірного масиву з мітками.

У бібліотеці *pandas* поєднуються висока продуктивність засобів роботи з масивами, притаманна NumPy, і гнучкі можливості маніпулювання даними, властиві електронним таблицям і реляційним базам даних (наприклад, на основі SQL). Оскільки маніпулювання даними, їх підготовка і очищення грають дуже велику роль в аналізі даних, *pandas* є одним з основних інструментів.

Основні можливості бібліотеки: має розвинені засоби індексування, що дозволяють просто змінювати форму наборів даних, формувати зрізи, виконувати агрегування і вибирати підмножини; структури даних з позначеними осями підтримують автоматичне або явне вирівнювання даних, що виключає появу типових помилок при роботі з невіривняні даними і даними з різних джерел, які порізно індексовані; має вбудовану функціональність часових рядів; одні і ті ж структури даних здатні підтримувати як часові ряди, так і дані інших видів; арифметичні операції, чкі виконуються з об'єктами, як з числовими даними; має гнучку обробку відсутніх даних (дозаповнення); інтеграція даних; підтримка з'єднання і інших реляційних операцій, наявних в популярних базах даних (наприклад, на основі SQL).

Багато засобів, присутні в *pandas*, або є частиною мови R, або надаються додатковими пакетами.

Сама назва *pandas* утворена від *panel data* (панельні дані), що застосовуються в економетриці для позначення багатовимірних структурованих наборів даних, так і від фрази Python data analysis.

Matplotlib – найпопулярніший в Python інструмент для створення графіків і інших способів візуалізації двовимірних даних, підходить для створення графіків, придатних для публікації. Хоча є можливості візуалізації в інших пакетах, *matplotlib* використовується найчастіше і тому добре інтегрована з іншими частинами екосистеми.

SciPy – набір пакетів, призначених для вирішення різних стандартних обчислювальних задач. Деякі з них: *scipy.integrate* – підпрограми чисельного інтегрування і розв'язання диференціальних рівнянь; *scipy.linalg* – підпрограми лінійної алгебри і розкладання матриць, доповнюють ті, що включені в *numpy.linalg*; *scipy.optimize* – алгоритми оптимізації функцій (знаходження екстремумів) і пошуку коренів; *scipy.signal* – засоби обробки сигналів; *scipy.sparse* – алгоритми роботи з розрідженими матрицями і розв'язання розріджених систем лінійних рівнянь; *scipy.special* – обгортка навколо

SPECFUN, написаної на Fortran-бібліотеці, що містить реалізації багатьох стандартних математичних функцій, в тому числі гамма-функції; scipy.stats – стандартні безперервні і дискретні розподіли ймовірностей (функції щільності ймовірності, формування вибірки, функції безперервного розподілу ймовірності), різні статистичні критерії і додаткові описові статистики.

Scikit-learn є основним інструментарієм програмістів для машинного навчання на Python. У ньому є підмодулі для наступних моделей: класифікація: метод опорних векторів, метод найближчих сусідів, випадкові ліси, логістична регресія і т. ін.; регресія: Lasso, гребнева регресія і т. ін.; кластеризація: метод k середніх, спектральна кластеризація і т. ін.; зниження розмірності: метод головних компонент, відбір ознак, матрична факторизація і т. ін.; вибір моделі: пошук на сітці, перехресний контроль, метрики; попередня обробка: виділення ознак, нормування.

Scikit-learn орієнтований головним чином на прогнозування і передбачення.

Statsmodels – пакет статистичного аналізу. У порівнянні із Scikit-learn, пакет Statsmodels містить алгоритми класичної (перш за все частотної) статистики та економетрики. У нього входять наступні підмодулі: регресійні моделі: лінійна регресія, узагальнені лінійні моделі, лінійні моделі зі змішаними ефектами і т. ін.; дисперсійний аналіз (ANOVA); аналіз часових рядів: AR, ARMA, ARIMA, VAR і інші моделі; непараметричні методи: ядерна оцінка щільності, ядерна регресія; візуалізація результатів статистичного моделювання.

Пакет statsmodels орієнтований більшою мірою на статистичне виведення, він дає оцінки невизначеності і p-значення параметрів. Використовується разом з NumPy і Pandas.

В Python є бібліотеки для зручного і швидкого зчитування даних в форматах електронних таблиць, баз даних, csv та ін.

Приклад використання Python для моделювання митних надходжень до державного бюджету України

Для цієї роботи були використані дані офіційної статистики [6].

Зазначимо, що інформація у відкритому доступі є дуже обмеженою.

Для задачі є консолідовані дані, які включають повний обсяг надходжень від митних органів до державного бюджету України, а також надходження від ввізного та вивізного мита, наведені в табл. 1.

Таблиця 1 – вихідні дані для моделювання надходжень до державного бюджету України від митних органів всього та за окремими видами.

Рік	Надходження до державного бюджету України, У, грн.	Надходження до державного бюджету України ввізного мита X1, грн.	Надходження до державного бюджету України вивізного мита, X2, грн.
2013	1,40036E+11	1,2550E+10	2,4900E+08
2014	3,57084E+11	1,3056E+10	1,7900E+08
2015	5,34694E+11	1,7422E+10	2,4500E+08
2016	6,16219E+11	2,0004E+10	3,7000E+08
2017	6,98405E+11	2,2257E+10	6,4300E+08
2018	8,33615E+11	2,3301E+10	5,1600E+08
2019	8,79833E+11	2,2778E+10	2,3000E+08
2020	8,77603E+11	2,1538E+10	2,5700E+08

Проведемо аналіз окремих складових у надходженнях, зокрема ввізного та вивізного мита.

За даними Таблиці 1 оцінимо наявність і тісноту зв'язку між ними та загальними надходженнями від митних органів, вигляд і тип моделі, параметри регресії, адекватність моделі, статистичну значимість параметрів, наявність автокореляції, визначимо прогнозне значення показника (точкову та інтервальну оцінку) і побудуємо довірчі зони регресії. У якості інструмента моделювання використаємо Python та бібліотеки NumPy, Statsmodels, Matplotlib, Xlrd (для зчитування даних із файлу Excel).

Найпоширенішою сучасною методикою моделювання структурованих даних є економетричне моделювання. За допомогою регресійного аналізу оцінюється залежність показника від одного чи декількох факторів. Найкращий результат дає метод найменших квадратів відхилення вихідних даних від змодельованих. Адекватність моделі оцінимо за допомогою критерію Фішера. Оцінювання статистичної значимості параметрів регресій, а також довірчих інтервалів регресій, проведемо на основі критерію Стюдента. Окрім цього, отримаємо іншу статистику по моделі і застосуємо модель для прогнозу.

Нижче наведено лістинги результатів програмного виконання розрахунків на Python та графіки для наочного уявлення про побудовані моделі.

Спочатку визначимо, яким чином надходження до бюджету від митних органів залежать від надходжень за ввізне мито. На рис.1 і рис.2 представлені результати виконання.

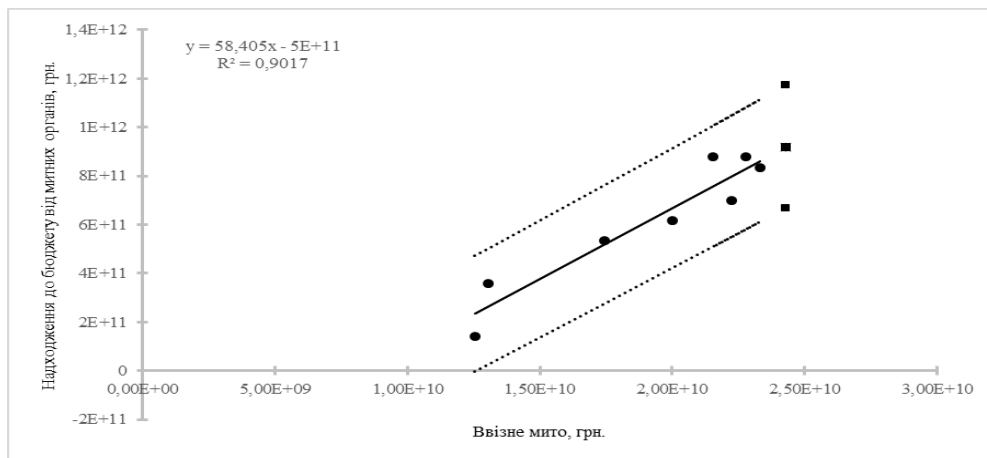


Рисунок 1 – Регресія надходжень до бюджету України від митних органів на ввізне мито, довірчий інтервал прогнозу, довірча зона регресії, побудовані з надійністю 0,95

```

=====
                        OLS Regression Results
=====
Dep. Variable:          y      R-squared:                0.902
Model:                  OLS    Adj. R-squared:           0.885
Method:                  Least Squares      F-statistic:              55.03
Date:                    Tue, 03 Aug 2021    Prob (F-statistic):       0.000309
Time:                    22:34:43          Log-Likelihood:           -211.97
No. Observations:       8                AIC:                      427.9
Df Residuals:           6                BIC:                      428.1
Df Model:                1
Covariance Type:        nonrobust
=====
                        coef    std err          t      P>|t|      [0.025    0.975]
-----
x1                58.4053      7.873        7.418    0.000     39.141    77.670
const             -4.991e+11    1.54e+11    -3.245    0.018    -8.75e+11 -1.23e+11
=====
Omnibus:                1.280    Durbin-Watson:           1.337
Prob(Omnibus):          0.527    Jarque-Bera (JB):        0.653
Skew:                   0.145    Prob(JB):                 0.722
Kurtosis:               1.631    Cond. No.                 9.46e+10
=====

```

Warnings:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
 [2] The condition number is large, 9.46e+10. This might indicate that there are strong multicollinearity or other numerical problems.

Кореляція:

```
[[1.    0.9496]
 [0.9496 1.    ]]
```

Значення фактору для заданого прогнозу показника: [[2.56681055e+10]
 [2.85201172e+10]]

Process finished with exit code 0

Рисунок 2 – Лістинг виконання програми (модель залежності митних надходжень до бюджету від ввізного мита)

Проаналізуємо основні результати моделювання.

Застосовано модель OLS, метод Least squares. Залежна змінна – у. Рівняння регресії виведено на рис. 1. Коефіцієнт кореляції дорівнює 0,95, що свідчить про сильну кореляцію між фактором і показником. Коефіцієнт детермінації скорегований – 0,885: зміна показника обумовлена зміною фактору на 88,5%.

F-статистика свідчить про адекватність моделі: розраховане значення дорівнює 55,03, критичне значення для степенів вільності задачі і рівня значимості 0,05 дорівнює 5,98. Розраховані значення t-статистики 7,42 для нахилу та -3,25 для перетину регресії. Обидва параметри є статистично значимими з довірчою імовірністю 0,975. Критичне значення t-статистики 2,45.

Відповідно, довірчі інтервали параметрів регресії при значимості 0,025: для нахилу: (39,14; 77,67), для перетину: (-8,75E+11; -1,23E+11). Статистика Дарбіна–Уотсона свідчить про відсутність автокореляції в моделі. Коваріаційна матриця вірно специфікована.

Модель може бути використана для прогнозу показника. Визначено прогнозні оцінки (точкові та інтервальні). Коефіцієнт еластичності за середніми показниками за останні чотири роки дорівнює 1,69, що означає, що показник є еластичним по фактору, причому темп зростання наповнення бюджету від митних надходжень уповільнюється, зокрема, за рахунок ввізного мита.

Побудуємо і проаналізуємо регресію митних надходжень до бюджету на вивізні мито. Аналогічно до раніше викладеного, отримаємо результат виконання програми. Для наочного представлення результату маємо діаграму з рівнянням регресії на рис.3.

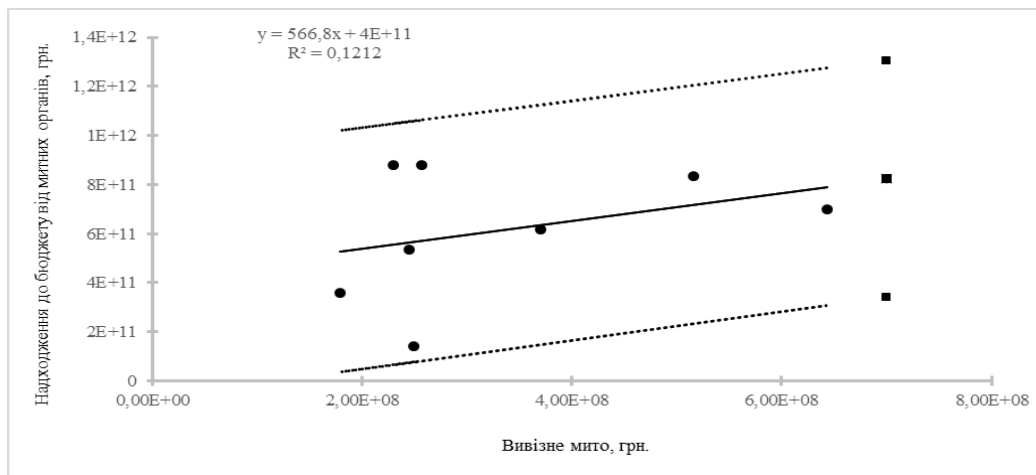


Рисунок 3 – Регресія надходжень до бюджету України від митних органів на вивізні мито, довірчий інтервал прогнозу, довірча зона регресії, побудовані з надійністю 0,95

На рис 4 представлено відповідний лістинг.

Застосовано ту ж модель OLS, що і в попередньому випадку. Залежна змінна – у. Рівняння регресії виведено на рис.5. Коефіцієнт кореляції дорівнює 0,348, що свідчить про дуже слабу кореляцію між фактором і показником. Коефіцієнт детермінації 0,1212 дуже низький, близький до нуля: зміна надходжень від митниці обумовлена зміною надходжень від вивізного мита на 12%. F-статистика свідчить про неадекватність моделі: розраховане значення дорівнює 0,83, критичне значення для степенів вільності задачі і рівня значимості 0,05 дорівнює 5,98. Розраховані значення t-статистики 0,91 для нахилу та 1,86 для перетину регресії. Обидва параметри статистично не значимо відрізняються від нуля з рівнем значимості 0,125. Критичне значення t-статистики 2,45 для степенів вільності моделі і заданого рівня значимості.

Дуже широка довірча зона пояснюється значним розкидом вихідних даних, і, відповідно, широкий довірчий інтервал прогнозу не має практичної цінності. Статистика Дарбіна–Уотсона свідчить про відсутність автокореляції в моделі. Коваріаційна матриця вірно специфікована.

Модель, очевидно, не можна рекомендувати використовувати для прогнозу показника.

Коефіцієнт еластичності за середніми показниками за останні чотири роки дорівнює 0,34, збільшуючись до значення 0,5, тобто показник є нееластичним по фактору, причому темп зростання наповнення бюджету від митних надходжень пришвидчується, за рахунок вивізного мита.

```

Результат розрахунку:

                                OLS Regression Results
=====
Dep. Variable:                  y      R-squared:                  0.121
Model:                          OLS    Adj. R-squared:            -0.025
Method:                        Least Squares  F-statistic:                0.8276
Date:                          Tue, 03 Aug 2021  Prob (F-statistic):        0.398
Time:                          22:29:22    Log-Likelihood:           -220.73
No. Observations:              8      AIC:                       445.5
Df Residuals:                  6      BIC:                       445.6
Df Model:                      1
Covariance Type:              nonrobust
=====
                                coef    std err          t      P>|t|      [0.025    0.975]
-----
x1                566.7962    623.032      0.910     0.398    -957.708    2091.300
const            4.267e+11    2.3e+11     1.856     0.113   -1.36e+11    9.89e+11
=====
Omnibus:                  0.138    Durbin-Watson:            0.410
Prob(Omnibus):           0.933    Jarque-Bera (JB):         0.213
Skew:                   -0.204    Prob(JB):                 0.899
Kurtosis:                2.313    Cond. No.                 8.94e+08
=====

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 8.94e+08. This might indicate that there are
strong multicollinearity or other numerical problems.
Кореляція:
[[1.    0.3482]
 [0.3482 1.    ]]
Значення фактору для заданого прогнозу показника: [[4.44254957e+08]
 [4.93616619e+08]]

Process finished with exit code 0

```

Рисунок 4 – Лістинг виконання програми (модель залежності митних надходжень до бюджету від вивіз-ного мита)

Зробимо загальні висновки щодо двох останніх моделей. Для обох моделей використано метод найменших квадратів, що дає найкращу апроксимацію вихідних даних з найменшою сумою квадратів похибок. Оцінки, отримані цим методом за теоремою Гауса-Маркова є ефективними і незміщеними. Для обох моделей розрахунок відбувався за допомогою створеної авторами програми на Python та з використанням бібліотек NumPy, Statsmodels, Matplotlib, Xlrd. Попередня підготовка даних відбувалася в Excel.

В роботі показано можливість програмної обробки даних. Не вдалося отримати більш повні статистичні дані з офіційного сайту, щоб можна було провести аналіз по великих масивах аналогічних даних. Але проблеми, що виявилися при моделюванні за наведеними агрегованими даними, відтворилися б аналогічно. Окрім того, в лістингах указане застереження, що кількість даних для коректного розрахунку має бути великою. Якщо кількість даних недостатня, то потрібно використовувати відкореговані оцінки, що відомо також з теорії математичної статистики.

Можна зазначити, що при невеликій кількості даних можна було б скористатися статистичним аналізом, що надає Excel, і це було б найефективніше рішення. Але метою роботи було використання можливостей Python для моделювання. Та ж сама програма дала б ефективний розрахунок для великих масивів даних, що не є проблемою в програмуванні безпосередньо. Також можна зазначити, що дані можуть зберігатися у базах даних. В Python є відповідні засоби роботи і з такими даними.

Тож, маємо відповідні економічні висновки, що виявляють великі проблеми в надходженнях від митних органів і, зокрема, від вивізного мита.

Можна будувати різні моделі однофакторні, і багатфакторні лінійні і нелінійні, і таким чином, отримати повну картину відносно того, які процеси відбуваються в галузі, зокрема, в митній, виявити позитивні і негативні явища. На основі результатів моделювання можна отримати науково обгрунтовані прогнози і приймати відповідні управлінські рішення.

Висновки

1. Проведено аналіз проблем, що виникають при обробці даних та ефективних інструментів для моделювання та прогнозування даних.
2. Побудовано моделі для аналізу митних надходжень до державного бюджету України за рахунок ввільного та вивільного мита. Було використано мову програмування Python та відповідні до задач пакети.
3. Зазначено можливість використання результатів моделювання фінансово-економічних показників для прийняття управлінських рішень.

Список літератури

- [1] У. Маккини, *Python и анализ данных*. М., Россия: ДМК Пресс, 2020, 540 с.
- [2] С. Деви, М. Арно, А. Мохамед, *Основы Data Science и BigData. Python и наука о данных*. Петербург, Россия: Питер, 2017, 336 с.
- [3] Т. А. Чупілко, "Актуальні проблеми високоефективної обробки даних. Моделювання показників за допомогою мови програмування Python," у *Актуальні напрями розвитку технічного та виробничого потенціалу національної економіки*. Дніпро, Україна: Пороги, 2021, с. 151–163.
- [4] Т. А. Чупілко, "Базовий інструментарій у сучасних технологіях комп'ютерної бізнес-аналітики," у *Міжнар. наук. конф. Інноваційні технології, моделі управління кібербезпекою ІТМК-2020*, Дніпро, 2020, т. 2, с. 53–54.
- [5] Т. А. Чупілко, "Комп'ютерні технології та економіко-математичні методи в управлінні бізнес-процесами на підприємстві," у *Міжнар. наук. конф. Інноваційні технології, моделі управління кібербезпекою ІТМК-2020*, Дніпро, 2020, Т.1, с. 26–28.
- [6] Міністерство фінансів України. [Електронний ресурс]. Режим доступу: <http://mof.gov.ua>. Дата звернення: 20 серпня, 2021.

Стаття надійшла: 29.08.2021

References

- [1] U. Makkyun, *Python y analiz dannykh*. M., Russia: DMK Press, 2020, 540 p.
- [2] S. Devy, M. Arno, A. Mokhamed, *Osnovy Data Science i BigData. Python y nauka o dannkh*. Peterburg, Russia: Pyter, 2017, 336 p.
- [3] T. A. Chupilko, "Aktualni problemy vysokoeffektyvnoi obrobky danykh. Modeliuvannia pokaznykiv za dopomohoiu movy proqramuvannia Python," u *Aktualni napriamy rozvytku tekhnichnoho ta vyrobnychoho potentsialu natsionalnoi ekonomiky*. Dnipro: Porohy, 2021, pp. 151–163.
- [4] T. A. Chupilko, "Bazovi instrumentarii u suchasnykh tekhnolohiiakh kompiuternoi biznes-analityky," in *Mizhnar. Nauk. Konf. Innovatsiini tekhnolohii, modeli upravlinnia kiberbezpekoiu ITMK-2020*, Dnipro, 2020, t. 2, pp. 53–54.
- [5] T. A. Chupilko, "Kompiuterni tekhnolohii ta ekonomiko-matematychni metody v upravlinni biznes-protsesamy na pidpriemstvi," in *Mizhnar. Nauk. Konf. Innovatsiini tekhnolohii, modeli upravlinnia kiberbezpekoiu ITMK-2020*, Dnipro, 2020, t. 1, pp. 26–28.
- [6] Ministerstvo finansiv Ukrainu. [Online]. Available: <http://mof.gov.ua>. Accessed on: August 20, 2021.

Відомості про авторів

Чупілко Тетяна Анатоліївна – кандидат технічних наук, доцент.

Ульяновська Юлія Вікторівна – кандидат технічних наук, доцент, завідувач кафедри.

Мормуль Микола Федорович – кандидат технічних наук, доцент, доцент.

Лагода Анастасія Едуардівна – студентка.

Т. А. Чупілко, Ю. В. Ульяновская, Н. Ф. Мормуль, А. Э. Лагода

PYTHON ДЛЯ ОБРАБОТКИ ДАННЫХ И МОДЕЛИРОВАНИЯ ФИНАНСОВО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Университет таможенного дела и финансов, Днепр

T. A. Chupilko, Yu. V. Ulianovska, M. F. Mormul, A. E. Lagoda
**PYTHON FOR DATA PROCESSING AND SIMULATION OF
FINANCIAL AND ECONOMIC INDICATORS**

University of Customs and Finance, Dnipro

ДО ВІДОМА АВТОРІВ

Найновіші правила оформлення і подання статей знаходяться на сайті журналу
<http://itce.vntu.edu.ua/>