ISSN 2664-9969

UDC 004.623:504.064.36
DOI: 10.15587/2706-5448.2024.302396

Volodymyr Slipchenko,
Liubov Poliahushko,
Olha Krush,
Volodymyr Rudyk

INVESTIGATION OF APPROACHES TO DESIGNING COMPLEX DATABASE STRUCTURES IN SYSTEMS OF INTEGRATED MONITORING OF ENVIRONMENTAL, ECONOMIC, ENERGY AND SOCIAL PARAMETERS OF THE TERRITORY

The object of research is the traditional and universal approach of designing the database structure in systems of integrated monitoring of ecological, economic, energy and social parameters of the territory, which include diverse data from various subject areas. In the course of the study, an analysis was performed based on a set of criteria such as scalability, ease of updating data, absence of empty fields, volume of the database, number of tables and fields, ease and speed of execution of requests for a sample set of indicators of the research object. The comparison of these approaches took place on the example of water resources monitoring, since it has several subsystems and a large number of indicators that are used for assessment. It is established that the proposed universal approach to designing complex database structures made it possible to reduce the volume of the database by 2.25 times due to the absence of empty fields. In particular, in the considered example, the filling factor of the database with the traditional approach is 1.75 times less than with the proposed universal approach. It should be noted that the rate of table filling for the traditional design approach can vary depending on the number of indicator values, while the table filling rate for the universal approach is always close to 100 %. Also, the proposed database design approach makes it possible to speed up data loading and processing. For example, with the same volume of significant information, the minimum speed of sampling the characteristics of one research object is 3.87 times greater in a database developed according to the principles of the universal approach than according to the rules of the traditional approach. The proposed structure of the database is successfully used in the system of complex eco-energy-economic monitoring. The developed structure of the database can serve as an effective basis for the formation of an electronic data bank at the level of the enterprise, region and country.

Keywords: *database design, database structure, complex monitoring, environment, public health.*

Received date: 02.03.2024
Accepted date: 19.04.2024
Published date: 22.04.2024

How to cite
Slipchenko, V., Poliahushko, L., Krush, O., Rudyk, V. (2024). Investigation of approaches to designing complex database structures in systems of integrated monitoring of environmental, economic, energy and social parameters of the territory. Technology Audit and Production Reserves, 2 (2 (76)), 38–43. doi: https://doi.org/10.15587/2706-5448.2024.302396

1. Introduction

One of the priority areas of development of any country in the world is to establish a balance between meeting the modern needs of humanity and protecting the interests of future generations, including their need for a safe and healthy environment. This corresponds to the general concept of sustainable development of the countries of the world, but its implementation is impossible without a comprehensive analysis of the ecological, economic and social parameters of the development of states, in particular Ukraine. Therefore, a system of complex eco-energy-economic monitoring (CEEEM) [1, 2] was developed, which is intended for effective management decision-making to ensure high quality of life and health of the population. The CEEEM system provides the work of various specialists, namely: an ecologist, an economist, an energy expert, a doctor, a lawyer, and a decision-maker. Specialists can use various information to monitor the facility and provide recommendations to improve the situation, including information related to another area. For example, to calculate the environmental tax, an economist uses information on emissions, discharges and disposal of solid household waste, to forecast the incidence of the population, a doctor takes data on discharges into water bodies, emissions of harmful substances into the atmosphere, etc. Therefore, the question of optimizing the structure of such systems and researching different approaches to their design is relevant.

Work [3] describes a database (DB) for diagnosing the state of the river system, [4] describes the DB structure for the placement of solid household waste. The work [5] contains the DB structure for providing geo-informational monitoring of the lands of the nature reserve fund, and the work [6] describes the DB structure of the ecological state of soils. All of them are designed according to the fundamental principle of building a DB, that is, during physical modeling, models optimized for a specific task were created. This approach does not allow replicating the data model to different subject areas, does not provide a simple expansion of the information in the DB, and does not contribute to the use of established solutions.

The approach to creating the most universal DB model for a wide range of tasks within the subject area is used in work [7] for the DB of water objects and services of the water resource quality monitoring system. But for a complex analysis of the territory, which would cover all monitoring subsystems [8], optimization for a specific task would lead to a large number of specialized tables, and their processing to the breakdown of the complex system into subsystems. For example, in [9], the authors provide a description of the structure of a set of databases for work on open rock, which includes separate databases for storing information on air quality, waste management, work with maps, and land use.

Other systems that focus on more than one monitoring subsystem, such as «Open Environment» [10] and RBCA [11], work according to the same principle.

The aim of this research is to compare DB design approaches and develop a unified, flexible DB structure that will allow storing various types of specialized information without the need to restructure the DB structure in the CEEEM system.

2. Materials and Methods

The paper considers two approaches to building an infological model of the subject area, namely: the traditional approach and the approach of creating a universal model.

The traditional approach generates a model with the following features:

– each object of the subject area is an entity of the model, and the number of entities depends on the subject area and is a variable value;

– an entity attribute is a property of a real-world object and only attributes important for the given task are taken into account;

– expansion of the data set implies a change in the composition of entities and their attributes.

Features of the universal model are:

– description of the subject area by hierarchies of real objects, that is, each entity is responsible for a group of objects of the subject area and is their generalization, the set of entities is fixed;

– entity attribute is a common property for a group of real-world objects, only attributes important for the given task are taken into account;

– expanding the data set does not involve changing the composition of entities and their attributes.

A comparison of traditional and universal approaches is carried out on the example of water resource monitoring [8], as it has several subsystems and a large number of indicators that are evaluated.

Hygienic assessment of the safety and quality of drinking water is carried out according to indicators of epidemic safety (microbiological, parasitic), sanitary-chemical (organoleptic, physico-chemical, sanitary-toxicological) and radiation indicators. Epidemic safety is evaluated by 11 indicators, there are also 66 sanitary and chemical indicators for evaluating the safety and quality of water, divided into three tables, and 6 radiation indicators. In addition, water quality is assessed by indicators of specific alpha- and beta-activity of drinking water (2 indicators) and indicators of physiological completeness of the mineral composition of drinking water (9 indicators) [12].

3. Results and Discussion

The traditional approach generates entities that correspond to each group of indicators with a complete list of all indicators in the form of attributes of the corresponding entities [13]. That is, in our case, there will be 7 entities (tables) to store the values of indicators, 2 entities (tables) to store information about the research object and research results. There are 9 entities in total.

As a result, let's get 98 attributes responsible for water quality and divided into 7 tables, 2 attributes for describing the research object, 3 attributes for describing a specific study and 9 special key attributes for relationships between entities. In total, there will be a maximum of 107 attributes, most of which will be empty, since not all objects are measured by all indicators.

The model looks as shown in Fig. 1.

To build a universal model, it is necessary to move away from specifics and rise to the upper level of abstraction. For this, a set of general entities and their characteristics is distinguished. The entity for the description of the water supply facility will be similar, and the entity for the description of indicators will be one universal, instead of 7 specific ones, with such characteristics as name, units of measurement, type of indicator, standards. An example of a universal model is presented in Fig. 2.

As a result, let's get 10 attributes responsible for water quality and divided into 3 tables, 2 attributes for describing the research object and 4 special key attributes for relationships between entities. There will be a total of 16 attributes that contain only the necessary data. That is, if not all indicators are measured at the facility, only the results of the measurements will be stored. The results of the comparison of design approaches are shown in the Table 1.

The values of statistical indicators for two variants of the DB structure of the water resource monitoring subsystem are presented in the Table 2, and in Fig. 3 shows a graph built on their basis. Both databases contain information on the results of the study of 20 objects in the city of Kyiv. The results of the study were taken from the website of Kyivvodokanal PrJSC [14]. The study contains only 8 of 40 sanitary-chemical indicators of safety and quality of drinking water and 1 indicator of the physiological adequacy of the mineral composition of drinking water, namely, total alkalinity.

As can be seen from Fig. 3, with the same number of research objects, the traditional approach generates more entities and attributes, which significantly affect the volume of the database on disk. Although the universal approach generates significantly more records in the tables, the total volume of data is smaller due to the absence of empty fields. Thus, the volume of the database obtained with the traditional design approach (DB1) is 2.25 times larger than the similar database obtained with the application of the universal design approach (DB2). This fact explains the filling factor of the database, which is always less than 100 % with the traditional approach (in the considered example, the filling factor of DB1 is 1.75 times less than that of DB2).

ISSN 2664-9969

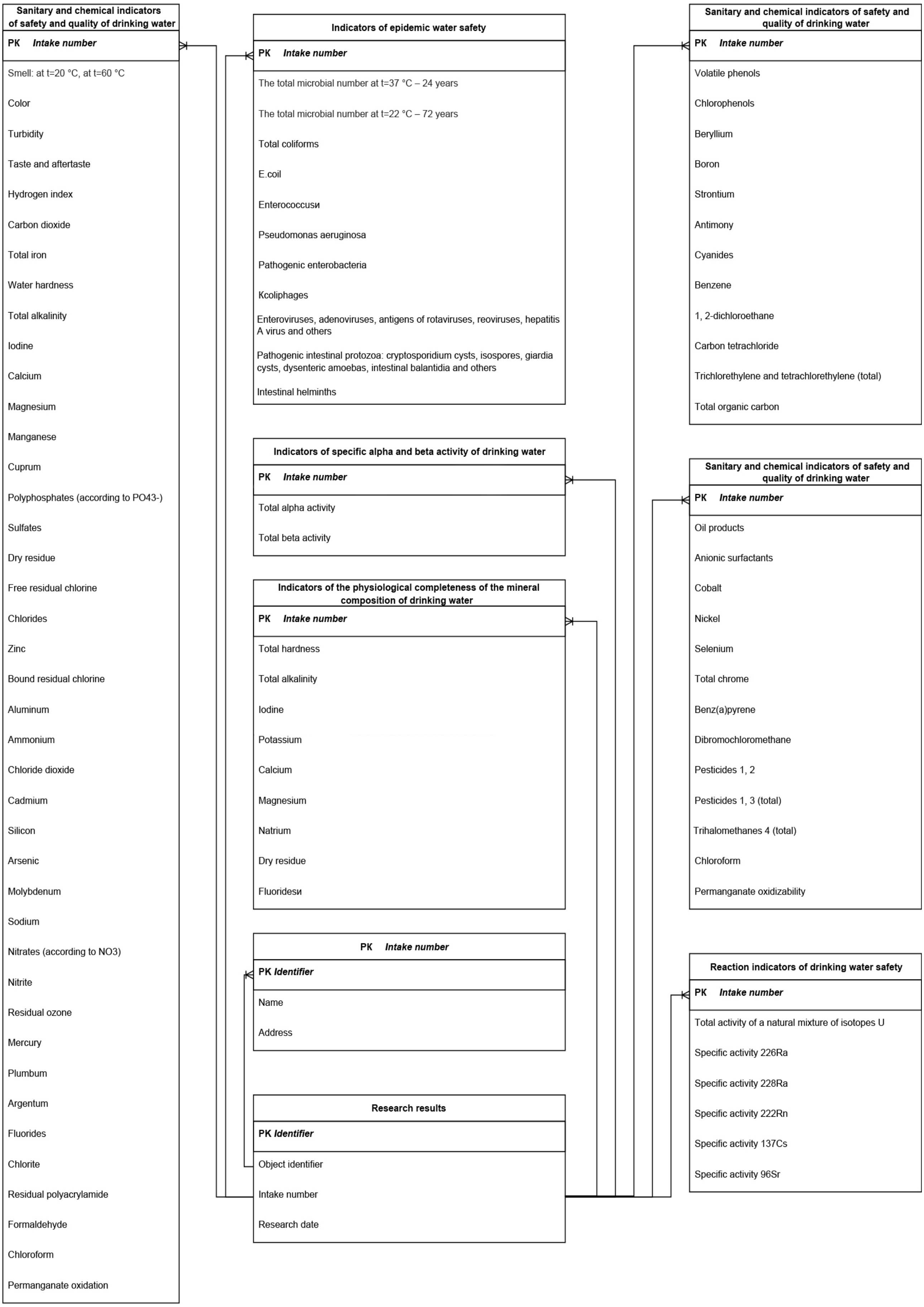

Fig. 1. An example of a traditional model for describing water quality at a water supply facility

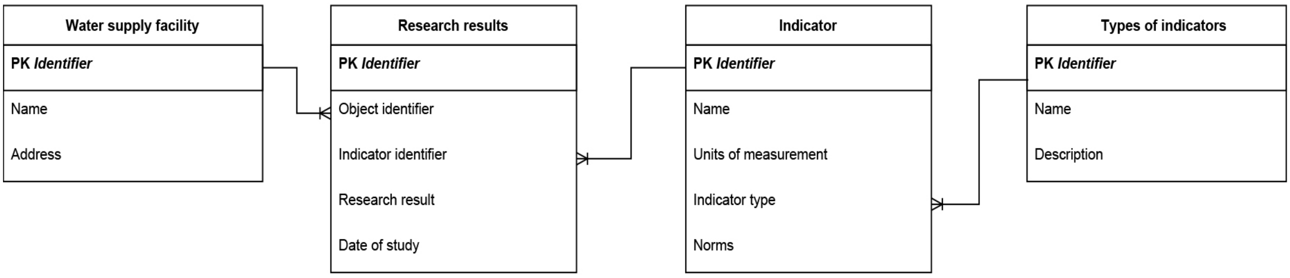

Fig. 2. An example of a universal model for describing water quality at a water supply facility

Table 1

Summary table of comparison results

Criterion name	Traditional approach	Universal approach
Scalability	*It requires significant changes in the structure of the database and the software system that uses it.* To expand the data set, it is necessary to add or delete tables/fields, which requires changes and programs, especially if technologies tied to the DB structure are used	*Does not require changes in the database structure and minimal changes in the software system that uses it.* To expand the data set, it is necessary to simply add rows with new information, without the need to change the database structure and minimal software changes (without the need to change the data model in the program)
Ease of updating data	*Update by one command.* Data about one object is collected in one row of the table, so the update will be carried out within one row	*Update using a set of commands (transaction).* Data about an object can be in multiple rows of a table, so an update may require a set of update requests
Correctness and completeness of information	DB structure created according to all the rules ensures correctness and completeness of information	The database structure created according to all the rules ensures correctness and completeness of information
Absence of empty fields	*There are empty fields in individual records.* Since not all objects have values according to all parameters, some of the cells will be empty, which leads to the presence of redundant data and an increase in the volume of the database	*There are no empty fields.* Only the available values of the indicators are stored in the tables, so there is no redundancy
Number of entities and attributes	*It significantly depends on the detailing of the objects of the subject area.* Corresponds to a set of domain objects, so it changes depending on the level of detail	*Almost does not depend on the detail of the subject area.* Corresponds to a set of top-level abstractions, that is, instead of specific objects, we work with general groups that are immutable

Table 2

The value of statistical indicators for two variants of the DB structure of the water resource monitoring subsystem

Indicator	Traditional approach	Universal approach
Number of entities	9	4
Number of attributes	107	20
The number of objects about which information is stored in the database	20	20
Number of entries	60	195
Filling of tables, %	57	100
Database volume, KiB	144	64
The minimum number of tables participating in the query	3	4
The maximum number of tables participating in the query	9	4
Max execution time for sampling indicators for one object, ms	232	157
Min execution time of sampling indicators for one object, ms	58	15

Note: *Filling of tables* is a coefficient that reflects the percentage of data filling and is calculated by the formula *K*=Number of filled fields/Number of fields*100. *The speed of execution of requests* was determined programmatically based on a series of 10 samples, the minimum and maximum values of the time spent on execution in ms are presented

The index of filling the tables for the traditional design approach can vary depending on the completeness of the study, that is, the more index values, the higher this index. The rate of table filling with the universal approach is always close to 100 %.

In addition, tables with a large number of fields are bulky, which affects the speed of loading and processing data. This is clearly demonstrated by the results of the study, because with the same volume of significant information, the minimum speed of sampling the characteristics of one research object is 3.87 times greater in DB1 than in DB2 (sampling is 3.87 times slower). This difference is clearly presented in the diagram in Fig. 4.

So, the obtained results show that the proposed universal approach to the design of complex DB structures using the example of the CEEEM system made it possible to reduce the DB volume by 2.25 times due to the absence of empty fields. In particular, in the considered example in the work, the database filling factor with the traditional approach is 1.75 times less than with the proposed approach. This difference is clearly presented in the diagram in Fig. 3.

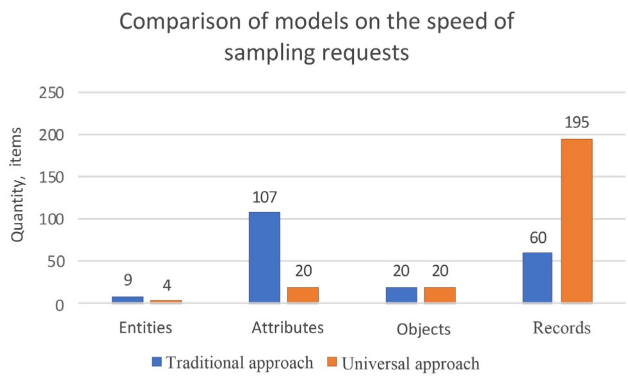

Fig. 3. Comparison of approaches

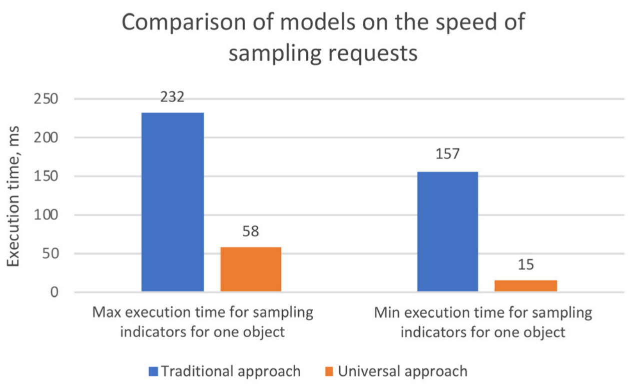

Fig. 4. Comparison of database models by query speed

Also, the proposed approach to DB design made it possible to speed up data loading and processing. For example, for the minimum speed of sampling the characteristics of one research object, it is 3.87 times more in a DB with a universal approach. This difference is clearly presented in the diagram in Fig. 4.

The obtained results make it possible to evaluate different approaches to the design of databases for integrated monitoring systems. Research results can help determine the most suitable structure of databases for the formation of a data bank at the level of the enterprise, region and country. The main limitations of the study are the closure of access to the data of some enterprises and departments due to the introduction of martial law in the country. Given this, research was conducted on environmental monitoring data that are open. Therefore, it is possible to believe that conducting the research in the conditions of war did not affect the obtained results. Further research into database design approaches could focus on speeding up data loading and processing, which is critical when developing web-based monitoring systems.

4. Conclusions

The work analyzed the traditional and unified approaches to the design of the database structure according to the following criteria: scalability, ease of updating data, absence of empty fields, volume of the database, number of tables and fields, ease and speed of execution of requests for sampling a set of indicators of the research object.

Research has shown that the traditional approach produces a database structure that is easy to understand by containing specific attribute names. This database structure is convenient for displaying and updating data, as it is done within one table and does not require complex data preprocessing mechanisms. But at the same time, it is rather cumbersome and slow to work, due to the presence of empty fields and a large number of attributes. Changing the set of entities and attributes requires reorganization of the structure, which in turn leads to changes in the program code. Due to the focus on a specific subject area, the traditional approach is suitable for systems that have a narrow focus.

A universal approach to database structure design generates a fixed set of tables that cover a wide range of tasks within the subject area. In such databases, displaying and updating data requires specialized algorithms. For example, displaying information about one object may require sampling from several tables, and updating may require executing a set of update commands. Such a database contains almost no empty fields and stores only the available information, so it is smaller in volume than a similar database obtained by the traditional approach. This affects the speed of data loading and execution of information sampling requests. Changing the set of entities and attributes usually does not affect the operation of software systems that use a universal structure database. That is why this approach is suitable for complex software complexes that do not have a narrow focus within the subject area.

The described structure of the database is successfully used in the system of complex eco-energy-economic monitoring.

The proposed structure of the database can serve as an effective basis for the formation of an electronic data bank at the level of the enterprise, region and country.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, including financial, personal, authorship, or any other nature that could affect the research and its results presented in this article.

Financing

The study was conducted without financial support.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the presented work.

References

1. Slipchenko, V. H., Poliahushko, L. H., Krush, O. Ye. (2021). The system of complex eco-energy-economic monitoring to optimize management decisions (region, district and city). *Visnik of the Volodymyr Dahl East Ukrainian National University, 4 (268),* 13–20. doi: https://doi.org/10.33216/1998-7927-2021-268-4-13-20
2. Slipchenko, V., Poliahushko, L., Krush, O. (2022). Developing an automated system for collecting various information in the system of complex eco-energy-economic monitoring. *Technology Audit and Production Reserves, 3 (2 (65)),* 11–18. doi: https://doi.org/10.15587/2706-5448.2022.259069
3. Xia, J., Lin, L., Lin, J., Nehal, L. (2014). Development of a GIS-Based Decision Support System for Diagnosis of River System Health and Restoration. *Water, 6 (10),* 3136–3151. doi: https://doi.org/10.3390/w6103136

4. Kuznichenko, S. D., Buchynska, I. V. (2021). *Modeli, metody ta instrumentalni zasoby bahatokryterialnoho analizu rishen v heoinformatsiinykh systemakh.* Zhytomyr: TOV «505», 202.
5. Halchenko, N. P., Kozar, V. I. (2019).The structure of the data bases for the provision of geoinformation monitoring of the nature reserve fund lands. *Ekolohichna bezpeka, 1 (27),* 32–37.
6. Saranenko, I. I. (2018). Stvorennia bazy danykh ekolohichnoho stanu hruntiv u MICROSOFT ACCESS. *Ekolohichni nauky, 2 (21),* 74–80.
7. Sanhinova, O. V., Kraieva, K. O., Andriiuk, V. K., Arkhipova, A. O. (2018). Database of water bodies and services of the water quality monitoring system. *Computer modeling for chemistry, technologies and sustainable development systems.* Kyiv: KPI im. Ihoria Sikorskoho, 214–216.
8. *Ekolohichnyi monitorynh dovkillia.* Ministerstvo zakhystu dovkillia ta pryrodnykh resursiv Ukrainy. Available at: https://mepr.gov.ua/diyalnist/napryamky/ekologichnyj-monitoryng/ekologichnyj-monitoryng-dovkillya/ Last accessed: 15.03.2023
9. Vangu, G. M., Croitoru, A., Mitrache, M., Dima, N. (2023). Design of a GIS Database for Surface Mining. *Journal of Applied Engineering Sciences, 13 (2),* 289–296. doi: https://doi.org/10.2478/jaes-2023-0037
10. *«Vidkryte dovkillia».* Ofitsiinyi portal ministerstva zakhystu dovkillia ta pryrodnykh resursiv. Available at: http://openenvironment.org.ua/index.htm?sl=UA Last accessed: 15.03.2023
11. *What is Risk Based Corrective Action (RBCA)?* Atlantic RBCA. Available at: https://atlanticrbca.com/faq/what-is-risk-based-corrective-action-rbca/ Last accessed: 14.09.2021
12. *Pro zatverdzhennia Derzhavnykh sanitarnykh norm ta pravyl «Hihiienichni vymohy do vody pytnoi, pryznachenoi dlia spozhyvannia liudynoiu» (DSanPiN 2.2.4-171-10)* (2010). Nakaz MOZ Ukrainy No. 400. 12.05.2010. Available at: https://zakon.rada.gov.ua/laws/show/z0452-10#Text Last accessed: 15.03.2023
13. *Osnovy rozrobky baz danykh.* Microsoft. Available at: https://support.microsoft.com/uk-ua/topic/%D0%BE%D1%81%D0%BD%D0%BE%D0%B2%D0%B8-%D1%80%D0%BE%D0%B7%D1%80%D0%BE%D0%B1%D0%BA%D0%B8-%D0%B1%D0%B0%D0%B7-%D0%B4%D0%B0%D0%BD%D0%B8%D1%85-eb2159cf-1e30-401a-8084-bd4f9c9ca1f5#bmterms Last accessed: 15.03.2023
14. *Shchodo yakosti pytnoi vody.* Sait Kyivvodokanal. Available at: https://www.vodokanal.kiev.ua/shhodo-yakosti-pytnoi-vody Last accessed: 16.04.2024

Volodymyr Slipchenko, *Doctor of Technical Sciences, Professor, Department of Digital Technologies in Energy, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine, ORCID: https://orcid.org/0000-0002-3405-0781*

✉***Liubov Poliahushko,*** *PhD, Department of Digital Technologies in Energy, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine, ORCID: https://orcid.org/0000-0003-3287-8523, e-mail: liubovpoliagushko@gmail.com*

Olha Krush, *Department of Digital Technologies in Energy, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine, ORCID: https://orcid.org/0000-0001-8792-2515*

Volodymyr Rudyk, *Department of Digital Technologies in Energy, National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine, ORCID: https://orcid.org/0009-0004-4774-6579*

✉*Corresponding author*