УДК 378 DOI 10.29141/2949-477X-2023-2-3-4 EDN DPVJWP

O. L. Kopnova[1], A. M. Aitymova[1], G. M. Abildinova[2]

[1]Non-profit limited company "Manash Kozybayev North Kazakhstan university", Petropavlovsk, Republic of Kazakhstan
[2]L. N. Gumilyov Eurasian National University, Astana, Republic of Kazakhstan

System analytics of data redundancy of corporate information systems using the theory of symmetry

Abstract. In this article, we will consider an example of the analysis of a corporate information system in order to eliminate the data redundancy of a corporate information system using the symmetry method. The proposed analysis may be of interest to both analysts of universities and large enterprises whose corporate information system consists of several integrated systems. The main types of information systems integration are described. The methodology presents an algorithm for analyzing the redundancy of information systems data in the outline of a corporate information system. The given example describes what problems there may be when merging information systems. The presented data analysis algorithm can be scaled to the corporate information system of any enterprise or organization.

Key words: symmetry theory; data analysis; corporate information system.

Paper submitted: September 11, 2023.

For citation: Kopnova O. L., Aitymova A. M., Abildinova G. M. System analytics of data redundancy of corporate information systems using the theory of symmetry. Digital models and solutions. 2023. Vol. 2, no. 3. Pp. 51–60. DOI: 10.29141/2949-477X-2023-2-3-4. EDN: DPVJWP.

О. Л. Копнова[1], А. М. Айтымова[1], Г. М. Абильдинова[2]

[1]Северо-Казахстанский университет им. М. Козыбаева, г. Петропавловск, Республика Казахстан
[2]Евразийский национальный университет им. Л.Н. Гумилева, г. Астана, Республика Казахстан

Системная аналитика избыточности данных корпоративных информационных систем с использованием теории симметрии

Аннотация. В статье представлен анализ корпоративной информационной системы, целью которой является устранение избыточности данных методом симметрии. Предлагаемый анализ может быть интересен как аналитикам вузов, так и крупным

предприятиям, корпоративная информационная система которых состоит из нескольких интегрированных систем. Описаны основные виды и методика интеграции информационных систем. Методика представляет собой алгоритм анализа избыточности данных информационных систем в структуре корпоративной информационной системы. Пример показывает, какие проблемы могут возникнуть при объединении информационных систем. Представленный алгоритм анализа данных может быть масштабирован на корпоративную информационную систему любого предприятия или организации.

Ключевые слова: теория симметрии; анализ данных; корпоративная информационная система.

Дата поступления статьи: 11 сентября 2023 г.

Для цитирования: Kopnova O. L., Aitymova A. M., Abildinova G. M. System analytics of data redundancy of corporate information systems using the theory of symmetry // Цифровые модели и решения. 2023. Т. 2, № 3. С. 51–60. DOI: 10.29141/2949-477X-2023-2-3-4. EDN: DPVJWP.

Introduction

In his article M. F. Karavai examines the application of the theory of symmetry to the analysis and synthesis of fault-tolerant systems [1; 2]. Despite the fact that more than forty methods described by him have passed since the moment of research, they claim to be effective. This can be judged on the basis of a large number of articles, the authors of which apply the above-mentioned methods to the analysis of various production problems [3]. For example, H. Kopetz [4] investigates mechanisms for localizing failures and detecting errors in distributed safety-critical systems with time synchronization. G. Bauer also applies the theory of symmetry for software-hardware synchronization of the application system and data backup [5].

Recent research by colleagues [6–9] shows that applying the theory of symmetry can successfully manage the distribution of parking spaces, which is especially important in the context of globalization and mass tourism.

This article proposes the author's method for analysing data redundancy in the outline of a corporate information system using the theory of symmetry. The analysis is done based on the graph of the corporate information system. The vertices of the graph are direct information systems integrated into the outline of the corporate information system. The edges of the graph are the data that supports the integration of information systems.

Some organizations, in addition to their own information systems, have a number of third-party ones. To integrate data between such information systems, there are several methods that programmers use and which are described in the article. if an organization has a complex information system consisting of modules (information systems) and its functionality in the context of the entire corporate information system can be considered as one information system, we will consider it as one vertex of the graph.

However, it is possible to scale the reasoning given in the article, as well for the analysis of data redundancy within the framework of such a “large” information system. In this knocking, the vertices of the graph will be the modules of the analysed information system, and the edges will be the key fields along which the data links occur in the tables of the corresponding modules. As practice has shown, such analysis is advisable if the information system is the author's development of the organization, the modules were added in stages depending on the new requests of the organization's external environment and different teams of programmers. In such cases, the possibility of data redundancy within the information system is not excluded.

Materials and methods

Very often corporate information systems consist of information systems that are different in their structure and content, each of which performs its own functions. To combine this information into a single system and build an analytical system, data integration between systems is often used. When integrating the system with each other, programmers use the method of connecting databases through common data tables. Thus, each system must have a key field, for example, an individual user number (id_employee), or fields suitable for data integration.

Information systems are added to the outline of the corporate information system, since the requirements for software often change, and market and environment offer new technologies. Some systems stop working in the loop of the corporate information system due to their obsolescence, others continue to function for reasons of convenience. So, after some time, every corporate information system becomes redundant, information within the information systems in its contour begins to duplicate. This often leads to data analysis problems, the complexity of reporting and the collapse of the corporate information system. Leaders of the organization begin to look for opportunities to move to other, more popular systems. However, this often only exacerbates the situation, since that unique data about the organization's work can be lost. In this case, the head of the organization is faced with a choice: to continue working within the framework of data redundancy, or to get rid of information systems that have become obsolete, but lose some of the data.

This article provides an example of the analysis of the corporate information system of Kozybayev University for the purpose of eliminating data redundancy using the theory of symmetry and analysis of the system's fault tolerance. A model for constructing an analytical system for more efficient work of the university is also proposed.

In theory, it is usually accepted to adhere to the rule of designing fault-tolerant systems by pre-selecting the architecture of the target system, taking into account the appropriate element base and the required performance. Then the real-time requirements are evaluated and the question of choosing the architecture of the redundant system is decided, taking into account the need to detect and search for failures and restore the system. However, as practice has shown, when building an analytical system based on a corporate information system of an enterprise, it is not always, or rather never possible, to choose an already ideal architecture of the target system. Accordingly, the construction of the analytical system is performed on the basis of the existing architecture, since the corporate information system of the enterprise at the time of analysis has already been formed.

Let us define a corporate information system as a set of integrated information systems that appear as certain production issues in an organization are resolved.

In many organizations, within the framework of the corporate information system, one can count from five or more information systems. These include systems for recording working hours, personnel records and accounting. Such systems can be linked into a single corporate information system by the key field employee (id_employee), which in all these information systems must uniquely identify a person. Other systems that store data about an organization's manufacturing processes can be linked by key process fields stored in a database. But then any corporate information system can be represented in the form of a graph, the vertices of which are information systems, and the edges are information communication channels, for example, id_employee.

Results

Let's build a graph of a corporate information system by analysing the functionality of the information systems integrated into it.

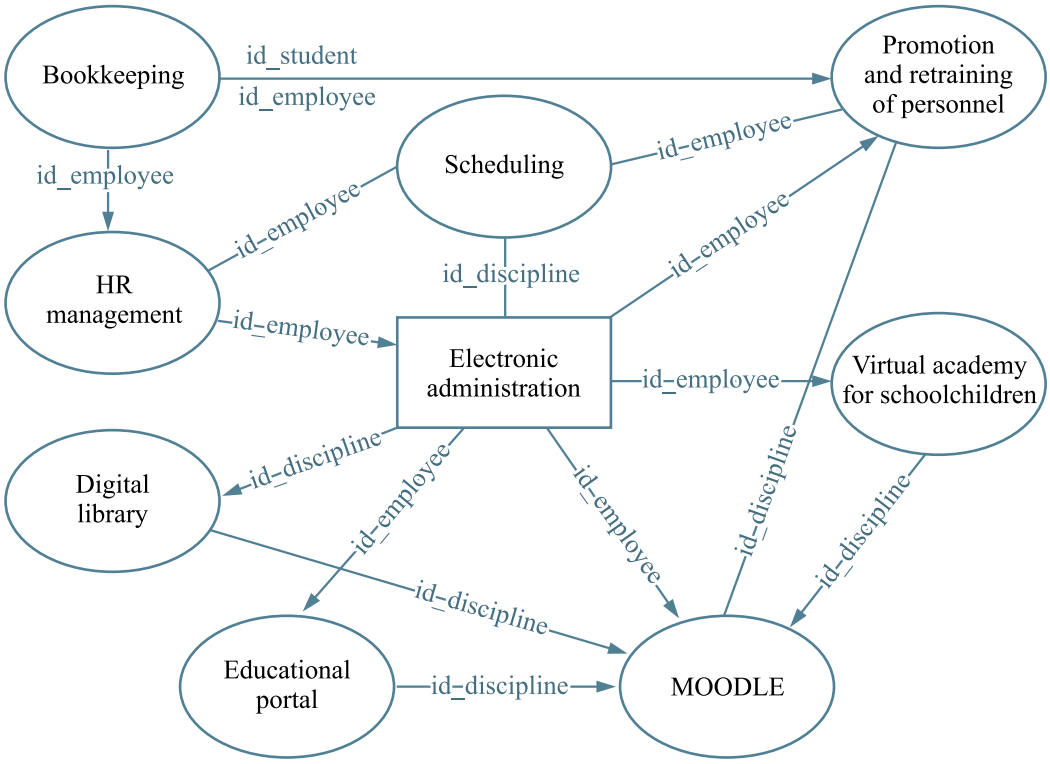

Fig. 1. Representation of the corporate information system of a university in the form of a graph

Figure 1 shows the graph of a university corporate information system. The ovals show simple information systems, each of which was purchased, installed and configured, or written by the programmers of the organization to solve the educational problems of the university.

The electronic administration is a modular information system consisting, in fact, of several information systems with a single database. Modules of the electronic administration: curricula, educational programs, academic calendar, registrar's office, classroom fund, arrears for services rendered, book supply, electronic document management, admission commission, student

contingent, student certification, and so on. That is why in the figure we have designated this vertex of the graph differently. However, from the point of view of the graph, this is the same vertex as any other.

Moodle is a system for organizing distance interaction with students. The latest in the corporate information system of the university appeared in connection with the introduction of COVID19 quarantine measures. It is very popular due to open source, scalability and constantly updated software versions. For greater convenience and interaction with students, the BigBlueButton module has been added, which allows you to conduct online classes without time limits.

Electronic library is an information system that stores electronic materials for students in the context of their specialty.

The educational portal is one of the first information systems that was created for students' navigation and stores links to most of the information materials, such as an electronic library, curricula, academic calendar, etc.

The virtual academy for schoolchildren was originally created as an information system of teaching materials for students wishing to gain additional knowledge in basic subjects. For some time, the information system was out of date and dropped out of the corporate information system, but revived after the introduction of COVID-19 quarantine measures.

Promotion and retraining of personnel is an information system that stores data about course participants, certificates and payment data.

The edges of the graph are links between information systems. Figure 1 shows two types of links between information systems-directed (shown by arrows) and non-directed (shown as a straight line). The direction determines the order of borrowing data in information systems. For example, to update data in the Moodle information system once a year, before the start of training sessions, programmers import data from the electronic administration (groups, students, disciplines) from the electronic administration to the distance learning system. Further changes in the contingent, the unification of disciplines, the addition of new courses and other changes related to the organization of the educational process are carried out by the employees of the corresponding department. Therefore, on the graph, this relationship is indicated in the form of an arrow. Of course, given the fact that Moodle is an open source system, and the electronic administration is the university's own development, full integration between databases is possible. However, in conditions of data redundancy, we get the problem of describing the integration procedure, which is reduced to solving a number of organizational issues.

In the works of M. F. Karavai showed that the fault tolerance of a system is closely related to the possibilities of representing systems in the form of a graph or spatial figures with a certain mathematical symmetry [1; 2]. The analysis of this symmetry allows not only to give a complete description of the fault tolerance of an arbitrary architecture system, but also to synthesize a fault tolerant system with a minimum redundancy [10–13].

Minimizing redundancy is one of the main design goals for resilient systems. Fault tolerance is usually considered as the properties of a system to maintain its logical structure despite failures of individual components of the system, which is achieved through the use of redundant resources of software hardware and time.

So, as a result of a rather superficial analysis, one can notice the data redundancy of the corporate information system can be eliminated by combining a number of information systems.

However, the question of data loss remains since it is impractical to transfer obsolete data to a new system.

After analysing Figure 1, we can conclude that the virtual academy of schoolchildren, the educational portal, the electronic library, the promotion and retraining of personnel have identical weights of the edges of the graph. Also, part of the functionality of information systems can be transferred to the Moodle information system. By and large, the educational portal lost its functionality after being included in the corporate outline of the Moodle information system. However, not always one information system can ergonomically replace another; often, certain efforts are required on the part of programmers of both the organization and software developers [14–15].

The virtual academy for schoolchildren is also easily integrated into this information system, since they have common functionality. Differences in work only in the age of students and the method of presentation of materials. The information system "Upgrading and retraining of personnel" has a wider amount of data. In addition to teaching materials, it stores data on the previous education of students, data on payment for the course and the transfer of disciplines of the courses taken within the framework of other courses. Modernization of the Moodle remote interaction system for integration with the "Staff Development and Retraining" system is reduced to the installation and configuration of additional plugins responsible for payment control and further communication with the accounting information system.

As for the electronic library, part of its functionality can be transferred to the Moodle system, however, difficulties arise with the amount of disk space on the server. Therefore, the analytical group made a recommendation to temporarily leave this information system within the corporate information system on those servers that were used for this before solving the problem of integrating the disk capacities of other servers into the Moodle system.

Conclusions

This article provides an example of analysing a corporate information system to eliminate data redundancy using the graph symmetry method.

The authors propose the following algorithm for analysing corporate information systems on heir redundancy.

The first step is to highlight and describe the functionality of each system that is part of the corporate information system.

Next step is describing the principle of interaction between information systems. The following types of integration are possible:

1. Direct database integration of information systems is a situation in which information systems have a common database server and each of them refers to the desired table, and a number of tables can be common;

2. Integration by means of duplicating a table or field – it is used most often when information systems have different types of operating systems serving servers, for example Windows and Linux. Direct access to data is usually hampered by system policy, so programmers usually write an additional module that allows at a certain time to unload data from one system to another during the least activity of users in the system. Or, exceptions are assigned to the

system security policy on both servers. This method may have a larger number of errors, but it solves the problem of interaction between systems.

3. Integration by means of copying data from one system to another is the most ineffective way of including an information system in the outline of a corporate information system, since it leads to data redundancy. However, it is resorted to a necessary event for performing data analysis or in a case where data from a closed information system can be useful for the operation of another information system. Data copying often occurs from closed commercial systems supplied with specialized equipment. An example of such systems is the service systems of turnstile systems, machine tools with programmable chips or electronic measuring systems.

At the third stage, a graph is built, the vertices of which are the information systems of the corporate information system, and the edges are the interactions between the corresponding information systems.

The next step is to analyse the symmetry of information systems with respect to each of the graph vertices, taking into account the weight of the graph edges according to the lemma: "In a 1-fault-tolerant structure, the number of redundant vertices cannot be less than the number of different orbits on all symmetry transformations of this structure".

In conclusion, we would like to note that the reflections given by the authors are a generalization of the experience of work in various organizations, the scope of which is both the provision of educational services, production of technical products and e-commerce. The article provides an example of analysing a corporate information system using the example of Kozybayev University, however, all reasoning is easily scalable to industrial enterprises. So the analysis algorithm can be useful for developers of complex modular information systems, especially if there is no programmers team working on the project.

References

1. Karavay M. F. Theory of symmetry and fault-tolerance. Distributed intelligence systems: selected papers from the IFAC Symposium. Arlington: IEEE, 1991. Pp. 42–54.
2. Karavay M. F. Primenenie teorii simmetrii k analizu i sintezu otkazoustoychivykh sistem [Application of symmetry theory to the analysis and synthesis of fault-tolerant systems]. Avtomatika i telemekhanika [Automation and Telemechanics]. 1996. Iss. 6. Pp. 159–173. URL: https://www.mathnet.ru/rus/at3236. (In Russ.)
3. Kolyeva, N., Kopnova, O., Shaporeva, A. Adaptation Information and analytical system adaptation in the contour of the corporate system of the university. E3S Web of Conferences. 2021. Vol. 270. Art. no. 01037. DOI: https://doi.org/10.1051/e3sconf/202127001037.
4. Kopetz H. Fault containment and error detection in the time-triggered architecture. The Sixth International Symposium on Autonomous Decentralized Systems, 2003 (ISADS 2003). Pisa: IEEE, 2003. Pp. 139–146. DOI: https://doi.org/10.1109/ISADS.2003.1193942.
5. Bauer G., Kopetz H. Transparent redundancy in the time-triggered architecture. Proceeding International Conference on Dependable Systems and Networks (DSN 2000). New York: IEEE, 2000. Pp. 5–13. DOI: https://doi.org/10.1109/ICDSN.2000.857508.
6. Kolyeva N. Organization of multi-access in databases. E3S Web of Conferences. 2021. Vol. 270. Art. no. 01006. DOI: https://doi.org/10.1051/e3sconf/202127001006.

7. Povyshev A. A., Sokolov A. N. Algoritm izbytochnosti dannykh detsentralizovannoy oblachnoy sistemy khraneniya [Data redundancy algorithm for a decentralized cloud storage system]. In: Nauchnyy poisk: materialy 14-y nauch. konf. aspirantov i doktorantov (Chelyabinsk, 19–20 aprelya 2022 g.) [Scientific Search: conference proceedings (Chelyabinsk, April 19–20, 2022)]. Chelyabinsk: SUSU, 2022. Pp. 171–179. EDN: SWPUKK. (In Russ.)
8. Yalçın A. S., Kılıç H. S. ve Delen D. The use of multi-criteria decision-making methods in business analytics: a comprehensive literature review. Technological Forecasting and Social Change. 2022. Vol. 174. Art. no. 121193. DOI: https://doi.org/10.1016/j.techfore.2021.121193.
9. Tonoyan S. A., Vysochanskiy V. A. Podkhody proektirovaniya korporativnogo khranilishcha dannykh na baze platformy SAP Netweaver Business Warehouse [Approaches to designing a corporate data warehouse based on the SAP Netweaver Business Warehouse platform]. Alleya nauki [Science Alley]. 2017. Vol. 3, iss. 15. Pp. 646–652. EDN: YMDSRW. (In Russ.)
10. Reinsel D., Gantz J., Rydning J. Data Age 2025: The Evolution of Data to Life – Critical. Don't Focus on Big Data; Focus on the Data That's Big: an IDC White Paper / IDC; Seagate. 2017. URL: https https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/workforce/ey-seagate-wp-data-age-2025-march-2017.pdf.
11. Kolyeva N. S., Golikov S. N., Panova M. V. Obzor rynka CRM dlya torgovoy kompanii [Overview of the CRM market for a trading company]. Vector Economy. 2023. No. 4(82). URL: http://www.vectoreconomy.ru/images/publications/2023/4/innovationmanagement/Koleva_Golikov_Panova.pdf. EDN: ZXYVJV. (In Russ.)
12. Lovett A., Forbus K. Modeling visual problem solving as analogical reasoning. Psychological Review. Vol. 124, iss. 1. Pp. 60–90. DOI: https://doi.org/10.1037/rev0000039.
13. Dichenko S. A. Kontrol i obespechenie tselostnosti informatsii v sistemakh khraneniya dannykh [Control and security of information integrity in data storage systems]. H&ES Research. 2019. Vol. 11, iss. 1. Pp. 49–57. DOI: https://doi.org/10.24411/2409-5419-2018-10224. EDN: ZAJITB. (In Russ.)
14. Ortega J., Tóth J., Moslem S., Péter T., Duleba S. An integrated approach of analytic hierarchy process and triangular fuzzy sets for analyzing the park-and-ride facility location problem. Symmetry. 2020. Vol. 12. Art. no. 1225. DOI: https://doi.org/10.3390/sym12081225.
15. Starikov T. V., Sopin K. Iu., Dichenko S. A., Samoilenko D. V. Model kontrolya tselostnosti dannykh na osnove pravil postroyeniya koda Rida – Solomona [Data integrity control model based on Reed–Solomon code construction rules]. Avtomatizatsiya protsessov upravleniya [Automation of Control Processes]. 2022. No. 1(67). Pp. 98–105. DOI: https://doi.org/10.35752/1991-2927-2022-1-67-98-105. EDN: TWDPTJ. (In Russ.)

Источники

1. Karavay M. F. Theory of symmetry and fault-tolerance. Distributed intelligence systems: selected papers from the IFAC Symposium. Arlington: IEEE, 1991. P. 42–54.
2. Каравай М. Ф. Применение теории симметрии к анализу и синтезу отказоустойчивых систем // Автоматика и телемеханика. 1996. № 6. С. 159–173. URL: https://www.mathnet.ru/rus/at3236.

3. Kolyeva, N., Kopnova, O., Shaporeva, A. Adaptation Information and analytical system adaptation in the contour of the corporate system of the university // E3S Web of Conferences. 2021. Vol. 270. Art. no. 01037. DOI: https://doi.org/10.1051/e3sconf/202127001037.
4. Kopetz H. Fault containment and error detection in the time-triggered architecture // The Sixth International Symposium on Autonomous Decentralized Systems, 2003 (ISADS 2003). Pisa: IEEE, 2003. P. 139–146. DOI: https://doi.org/10.1109/ISADS.2003.1193942.
5. Bauer G., Kopetz H. Transparent redundancy in the time-triggered architecture. Proceeding International Conference on Dependable Systems and Networks (DSN 2000). New York: IEEE, 2000. P. 5–13. DOI: https://doi.org/10.1109/ICDSN.2000.857508.
6. Kolyeva N. Organization of multi-access in databases // E3S Web of Conferences. 2021. Vol. 270. Art. no. 01006. DOI: https://doi.org/10.1051/e3sconf/202127001006.
7. Повышев А. А., Соколов А. Н. Алгоритм избыточности данных децентрализованной облачной системы хранения // Научный поиск: материалы 14-й науч. конф. аспирантов и докторантов (Челябинск, 19–20 апреля 2022 г.). Челябинск: ЮУрГУ, 2022. С. 171–179. EDN: SWPUKK.
8. Yalçın A. S., Kılıç H. S. ve Delen D. The use of multi-criteria decision-making methods in business analytics: a comprehensive literature review // Technological Forecasting and Social Change. 2022. Vol. 174. Art. no. 121193. DOI: https://doi.org/10.1016/j.techfore.2021.121193.
9. Тоноян С. А., Высочанский В. А. Подходы проектирования корпоративного хранилища данных на базе платформы SAP Netweaver Business Warehouse // Аллея науки. 2017. Т. 3, № 15. С. 646–652. EDN: YMDSRW.
10. Reinsel D., Gantz J., Rydning J. Data Age 2025: The Evolution of Data to Life – Critical. Don't Focus on Big Data; Focus on the Data That's Big: an IDC White Paper / IDC; Seagate. 2017. URL: https https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/workforce/ey-seagate-wp-data-age-2025-march-2017.pdf.
11. Кольева Н. С., Голиков С. Н., Панова М. В. Обзор рынка CRM для торговой компании // Вектор экономики. 2023. № 4(82). URL: http://www.vectoreconomy.ru/images/publications/2023/4/innovationmanagement/Koleva_Golikov_Panova.pdf. EDN: ZXYVJV.
12. Lovett A., Forbus K. Modeling visual problem solving as analogical reasoning. Psychological Review. Vol. 124, iss. 1. P. 60–90. DOI: https://doi.org/10.1037/rev0000039.
13. Диченко С. А. Контроль и обеспечение целостности информации в системах хранения данных // Наукоемкие технологии в космических исследованиях Земли. 2019. Т. 11, № 1. С. 49–57. DOI: https://doi.org/10.24411/2409-5419-2018-10224. EDN: ZAJITB.
14. Ortega J., Tóth J., Moslem S., Péter T., Duleba S. An integrated approach of analytic hierarchy process and triangular fuzzy sets for analyzing the park-and-ride facility location problem || Symmetry. 2020. Vol. 12. Art. no. 1225. DOI: https://doi.org/10.3390/sym12081225.
15. Стариков Т. В., Сопин К. Ю., Диченко С. А., Самойленко Д. В. Модель контроля целостности данных на основе правил построения кода Рида – Соломона // Автоматизация процессов управления. 2022. № 1(67). С. 98–105. DOI: https://doi.org/10.35752/1991-2927-2022-1-67-98-105. EDN: TWDPTJ.

Information about the authors

Oxana L. Kopnova, Master of Engineering and Technology, Senior Lecturer of the Department of Mathematics and Informatics, North Kazakhstan University named after M. Kozybaev, 15000, Republic of Kazakhstan, Petropavlovsk, Pushkina St., 86. E-mail: ok_10_ok@mail.ru.

Aliya M. Aitymova, Master of Pedagogy, Senior Lecturer of the Department of "Theory and Methods of Primary and Preschool Education". North Kazakhstan University named after M. Kozybaev, 15000, Republic of Kazakhstan, Petropavlovsk, Pushkina St., 86. E-mail: aitimova_a_1985@mail.ru.

Gulmira M. Abildinova, Candidate of Pedagogical Sciences, Professor. Eurasian National University named after L.N. Gumilyov. 010000, Republic of Kazakhstan, Astana, Satbaeva St., 2. E-mail: aitimova_a_1985@mail.ru.

Информация об авторах

Копнова Оксана Леонидовна, магистр техники и технологий, старший преподаватель кафедры «Математика и информатика». Северо-Казахстанский университет им. М. Козыбаева, 15000, Республика Казахстан, г. Петропавловск, ул. Пушкина, 86. E-mail:ok_10_ok@mail.ru.

Айтымова Алия Маратовна, магистр педагогики, старший преподаватель кафедры «Теория и методика начального и дошкольного образования». Северо-Казахстанский университет им. М. Козыбаева, 15000, Республика Казахстан, г. Петропавловск, ул. Пушкина, 86. E-mail: aitimova_a_1985@mail.ru.

Абильдинова Гульмира Маратовна, кандидат педагогических наук, профессор Евразийский национальный университет имени Л.Н. Гумилева, 010000, Республика Казахстан, г. Астана, ул. Сатбаева, 2. E-mail: aitimova_a_1985@mail.ru.