

Intellectual Analysis of Spatial Data for Information Support in Management of Technical Component of Complex Distributed Information Systems

Elena Brekotkina¹, Ramil Gilyazov¹, Sergey Pavlov¹, Vladislav Trubin¹
and Olga Khristodulo¹

¹Ufa State Aviation Technical University, K. Marx Street, 12, Ufa, 450008, Russia

Abstract

Management of complex distributed systems, at present, is carried out using information systems designed to collect and manage a large amount of heterogeneous information of an organization or institution. The technical component of such information systems is itself a complex distributed system, and information about the location and mutual arrangement of its individual components, as well as the location of the associated infrastructure and main users, is essential for its high-quality functioning. A method of spatial data mining that suggested in this paper, proposed for formalizing the process of creating specialized databases for the implementation (creation) of individual subsystems as part of complex distributed information systems. The results of the application of this method for the construction of one of the subsystems, namely, the subsystem of information support for the management of the technical component of the information system, as part of a complex distributed information system, are presented on the example of the geoinformation system of the Ufa State Aviation Technical University. The proposed approach to intelligent analysis and representation of spatial data, consisting in a formalized description of various thematic data applied to solving specific problems of information management support in complex distributed systems, can be applied to the construction of complex information systems in significantly distributed organizational and infrastructure.

Keywords

Spatial data, complex distributed systems, data mining, information management support, technical component of information systems

1. Introduction

Modern approaches to organizing the management of complex distributed systems (CDS), which are characterized by a significant removal of their subsystems and objects, are based on the use of information systems (IS) designed for the collection, management and intellectual analysis of large amounts of heterogeneous information about the state and functioning of the CDS and its environment. On the one hand, the IS is an integral component (subsystem) of CDS, on the other hand, it represents CPC itself, as its main components are distributed in separate subsystems of CDS, and needs to analyze the state, support the performance and management. One of the most important subsystems of such IS is its technical component, which includes a variety of computing resources (servers, stationary computers, mobile devices), input-output devices (printers, monitors, scanners, multifunction devices), united in a computer network by special means (communication lines, routers, switches, etc.).

Obviously, the technical component of the IS is also the CDS and to ensure its quality creation and operation, the information about the location and mutual location of its individual components, as well

GraphiCon 2021: 31st International Conference on Computer Graphics and Vision, September 27-30, 2021, Nizhny Novgorod, Russia

EMAIL: brekotkina@mail.ru (E. Brekotkina); verols335@gmail.com (R. Gilyazov); psvgis@mail.ru (S. Pavlov); refracto@yandex.ru (V. Trubin); o-hristodulo@mail.ru (O. Khristodulo)

ORCID: 0000-0002-2281-8351 (E. Brekotkina); 0000-0003-4918-013X (R. Gilyazov); 0000-0001-9672-7623 (S. Pavlov); 0000-0002-8787-9833 (V. Trubin); 0000-0002-3987-6582 (O. Khristodulo)



© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

as the location of the associated infrastructure (floor plans, power supply network, etc.) and the main users (consumers) of its functioning is essential, which leads to the need for automated processing of these types of information, which is called spatial, using geoinformation systems.

This paper considers some approaches to the analysis, structuring and organization of heterogeneous (spatial and attributive) data for information management support of the technical component of complex distributed information systems on the example of the Ufa State Aviation Technical University (USATU), which includes many remote from each other (up to 300 km) departments, each of which operates the appropriate IS components, including their technical component.

2. Intellectual analysis of spatial data for information support

For effective use of spatial information in the management of the university, GIS USATU is created [1], whose structure is shown in Fig. 1, where as one of the subsystems, the subsystem of information management support of the technical component (IMSTC) is selected. The central (most important) part of the system is the database, which includes the main types of planning and cartographic information: maps of territories, floor plans of all buildings, plans of individual rooms and structures, which include the main information layers of engineering communications, as well as their attributive data.

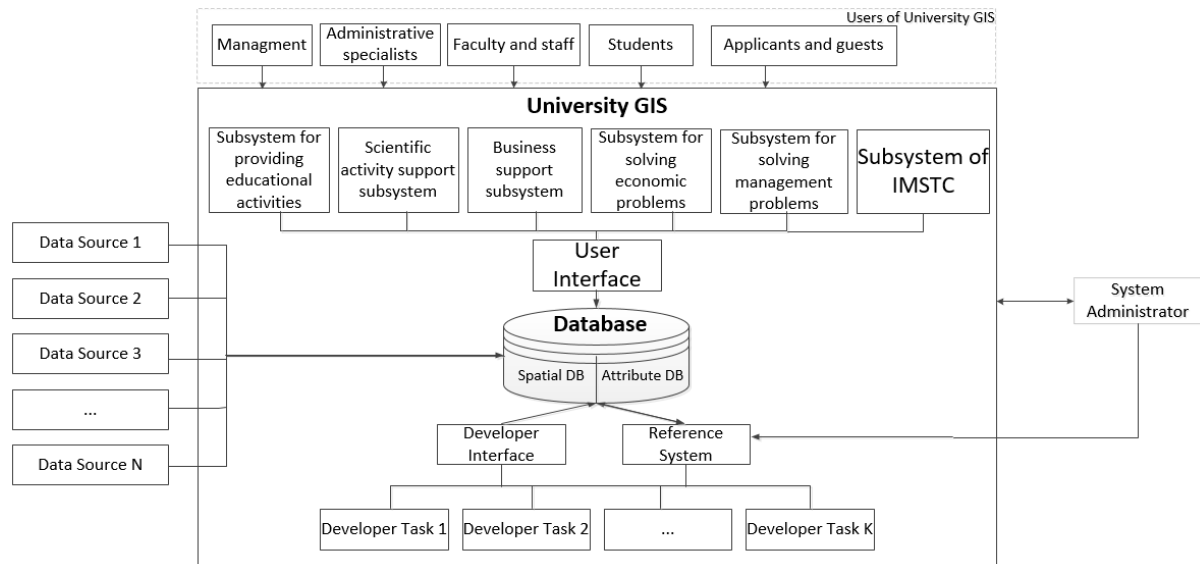


Figure 1: Logical structure of GIS USATU

Let's denote the entire USATU GIS database

$$DB = \{DBPC, ADB\}, \quad (1)$$

where DBPC is a database (or set) of planning and cartographic data, which includes geographical maps of the territories where the facilities and units of USATU are located, as well as schemes of individual buildings and structures [3, 4]. Hereinafter, for the sake of brevity, maps and plans will be referred to the same way, i.e.

$$DBPC = \{K^i\}, i = \overline{1, n}, \quad (2)$$

where n is the number of maps and plans needed to manage the university. The order in which the cartographic data are arranged in the DBPC set is not essential, so for certainty let's arrange them in order of increasing degree of detail of spatial objects:

- K^1 – map of the Republic of Bashkortostan with USATU objects on it;
- K^2 – map of the city of Ufa;
- K^3 – map of Kumertau;
- K^4 – map of Ishimbay;
- Maps K^2, K^3, K^4 show the location of USATU facilities;

- K^5 – master plan of the university buildings (complex placement of educational buildings of the university);
- K^6 – master plan of the campus (complex accommodation of university dormitories).

Under some numbers in this set are floor plans of all buildings and plans of individual rooms:

- K^{11} – plan of the first floor of the first campus building;
- K^{12} – plan of the second floor of the first campus building;
-
- K^{61} – plan of the first floor of the sixth campus building;
- K^{62} – plan of the second floor of the sixth campus building;
-
- K^{6101} – plan of classroom 101 of the sixth campus building;
- K^{6102} – plan of classroom 102 of the sixth campus building, and so on.

All in all, there are more than two hundred elements, i.e., digitized maps and plans of USATU facilities, in the DBPC today.

Each map from (2) is a set of thematic layers containing a set of spatial objects of the same type (in their physical meaning and in the way they are represented).

$$K^i = \{S_j^i\}, \quad j = \overline{1, m_i}, \quad (3)$$

for all $i = \overline{1, n}$, here m_i – the number of thematic layers of the i-th map.

To solve each specific task of university management, a certain set of maps and their layers is used, and part of the maps and layers is used to solve several (more than one, but not necessarily all) tasks, and the other part is used to solve only one task [2]. In works [1, 2] authors of this article considered in detail the problem of organization of storage of such data on the basis of formal analysis of their belonging to different objects and subsystems of CDS. In these studies, the main emphasis is placed on ensuring the consistency of all data stored in a distributed database, necessary for the management of specific CDS, by storing only one instance of data with preservation of all semantic relations. In this paper we consider the problem of maintaining the integrity and consistency of these data when they are used repeatedly by a large number of users (people or tasks) independent of each other.

For the first time, the proposed approach was used by the authors to solve the problem of information security management (ISM) of CDS [2]. As follows from (3) by the time of creating a new subsystem in the general management system of some CDS (in this article as an example, but without violating the generality of reasoning, it is USATU), each such map of (2) consists of m_i thematic layers needed to solve current management problems[5]. For each specific task is used, as a rule, not all, but only part of the available layers included in the set (3) under specific numbers, for example, to solve the problem of public safety management on the territory of the university (its main part in the city of Ufa) the following layers are used:

$$\begin{aligned} S_{j_1}^5 & - \text{the layout of buildings and structures of the campus;} \\ S_{j_2}^5 & - \text{scheme of electrical communications;} \\ S_{j_3}^5 & - \text{layout of video cameras.} \end{aligned} \quad (4)$$

Let's introduce the set:

$$G_p^i = \{j_1, j_2, \dots, j_{l_p^i}\}, \quad (5)$$

where l_p^i – the number of layers of the i-th map required to solve the p-th management task, $p = \overline{1, Q}$, Q – the total number of tasks (subsystems) included in the system (in our case – GIS USATU). Note that

$$l_p^i \leq m_i, \quad (6)$$

that is, the number of layers of each map used to solve the p-th problem is less than the number of all map layers. For the layers used in (4) the set G_p^i will have the form:

$$G_3^5 = \{j_1, j_2, j_3\}, \quad (7)$$

indexes j_1, j_2 and j_3 have certain numerical values in each case, but this is not essential for formalizing the process of data analysis.

To solve a newly arisen (or newly set and ready for implementation) management problem (e.g., IMSTC) all necessary maps are supplemented with new thematic layers [6] in the number m_i^1 , for each $i = \overline{1, n}$. It should be noted that for each new task it is possible that

$$m_i^1 = 0, \quad (8)$$

for some values of i , which means that there are no additional thematic layers for this problem in this map.

In order to develop and implement the IMSTC subsystem, it is necessary to add thematic layers containing spatial information about the university computing infrastructure objects to the base of plan-map data. These additions refer to those maps from (2), which can display the corresponding spatial objects, so in Fig. 2 is a plan of equipment placement on one of the floors of the campus building, and in Fig. 3 is a layer-by-layer image of these objects in one of the rooms.

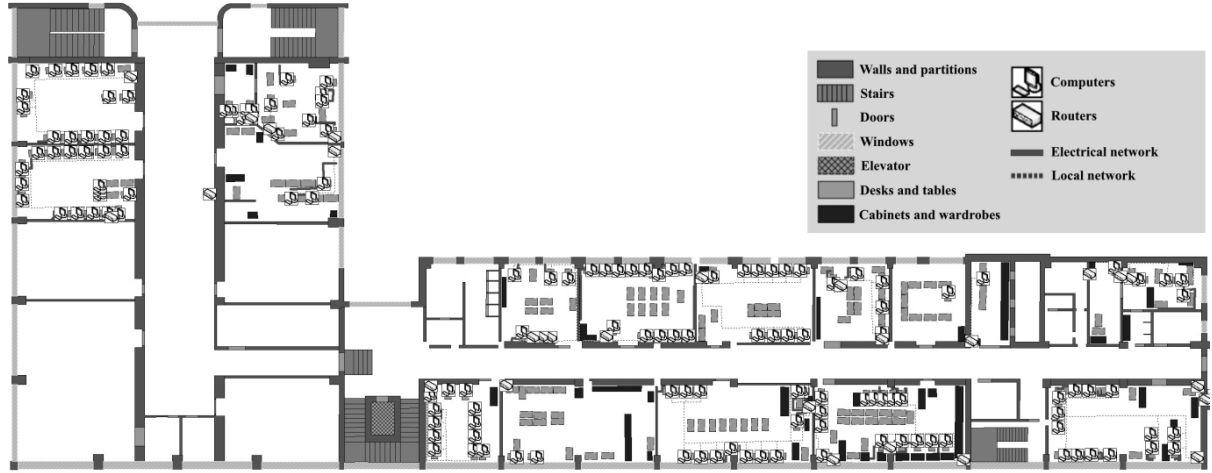


Figure 2: An example of a digital plan of the 3rd floor of sixth campus building with the placement of computing infrastructure

The information shown in Fig. 3 corresponds to the addition of the map $K^{6,315}$ new layers in the number $m^0 = m_{6,315}^1 = 5$ and using one existing layer, which is included in the set of all layers of this map $m = m_{6,315}$ under the number j_1 [7]:

- $S_{j_1}^{6,315}$ – room boundaries;
- $S_{m+1}^{6,315}$ – layout plan for point objects of a computer network (routers, switches, computers);
- $S_{m+2}^{6,315}$ – layout plan for linear objects of a computer network (communication lines);
- $S_{m+3}^{6,315}$ – furniture layout plan;
- $S_{m+4}^{6,315}$ – layout plan for point objects of the power supply network;
- $S_{m+5}^{6,315}$ – layout plan for linear objects of the power supply network.

And the set G_p^i , fixing the layers of this map to create a subsystem of IMSTC, which has some number $p = p_0$ in the general list of GIS USATU tasks, will have the form

$$G_{p_0}^{6,315} = \{j_1, m+1, m+2, \dots, m+m_0\}. \quad (9)$$

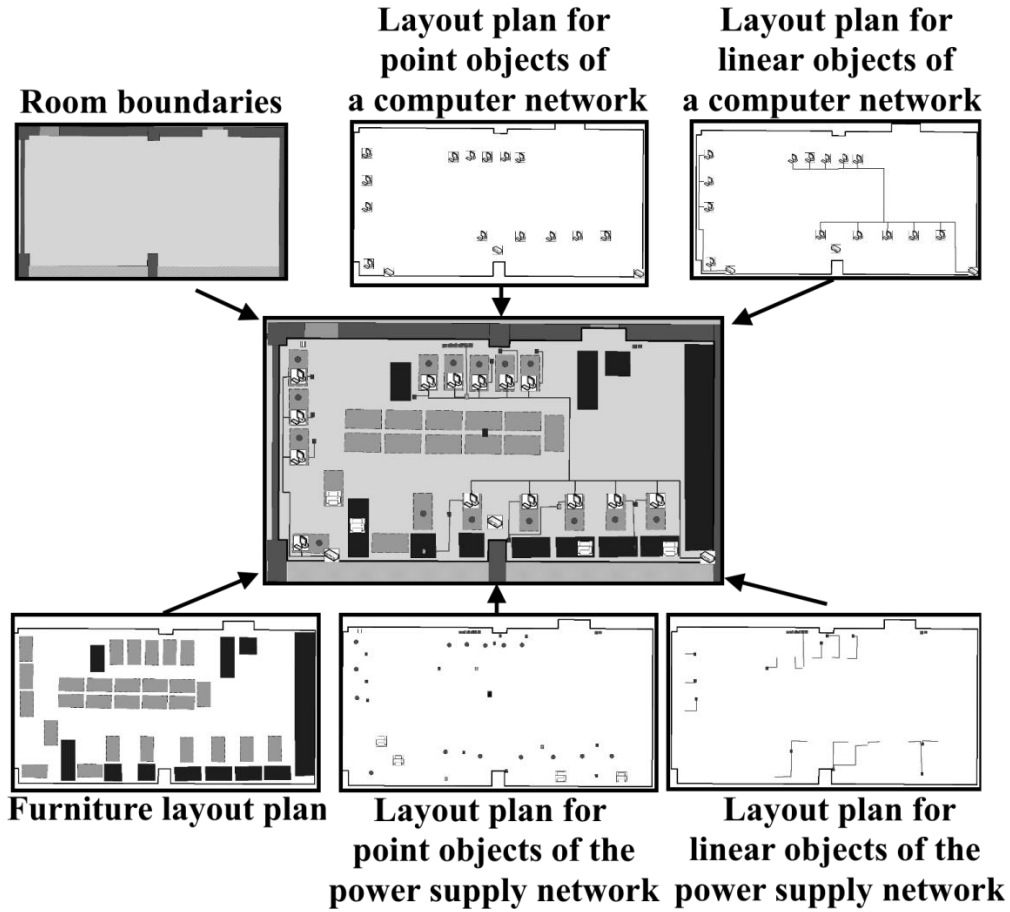


Figure 3: An example of a digital layout of classroom 6-315 with placement of computing infrastructure

Similarly, in all maps of floors and rooms is the addition of thematic layers with objects of computing infrastructure, which were created on the basis of layouts of equipment and passports of educational laboratories, supplemented by the results of the inspection of premises, and for each of these maps is formed set of used map layers, similar to (9), and the entire set of numbers (list) layers of all maps from (2) for this subsystem will be determined by the ratio

$$G_{p_0} = \bigcup_{i=1}^n G_{p_0}^i. \quad (10)$$

All spatial objects are described both by their location reflected in the corresponding maps from (2) and by some attributive characteristics stored in the ADB from (1). At the same time, attributive information about objects of each thematic layer of each map is stored in one (without taking into account normalization) relation [8, 9], so that each layer of each map corresponds to exactly one relation. By analogy with the notation of layers, let us denote each relation by R_j^i , then the whole attributive database is a set of such relations:

$$ADB = \{R_j^i\}, \quad (11)$$

where still $i = \overline{1, n}$, and $j = \overline{1, m_i}$.

Further, by analogy with the formation and use of the used cartographic information on the basis of the already constructed sets G_p^i , a set of used or newly added relations is formed:

- $R_{j_1}^{6\ 315}$ – attributive information about the room;
- $R_{m+1}^{6\ 315}$ – attributive information for point objects of a computer network (routers, switches, computers);

- $R_{m+2}^{6\ 315}$ – attributive information for linear objects of a computer network (communication lines);
- $R_{m+3}^{6\ 315}$ – attributive information about a furniture;
- $R_{m+4}^{6\ 315}$ – attributive information for point objects of the power supply network;
- $R_{m+5}^{6\ 315}$ – attributive information for linear objects of the power supply network.

The relationship of relations with the attributive information about the objects of computing infrastructure is presented in Fig. 4 in the form of an information model, and Fig. 5 shows the data of one of the elements of this model. Thus, to solve the problem of intelligent analysis of spatial information and the formation of specialized databases (SDB) based on the results of this analysis to solve individual management tasks (subsystems) of CDS, for example, for the task of information support for the management of technical component of such systems, an algorithm for the formation of each of these tasks in set G_p , which can serve as a mapping function of the entire database system in this SDB is proposed.

$$DB \xrightarrow{G_p} SDB_p \quad (12)$$

for each $p = \overline{1, Q}$.

This approach is used in the design, creation and use of the spatial database in the GIS USATU, registered in the Russian Fund of programs and databases [10], including a common database for the entire system and specialized databases for some tasks (subsystems).

The use of spatial, including - attributive data on the objects of computing infrastructure to solve some problems of university management (initially in information and reference mode) is based on a specialized interface developed in the GIS software environment "InGeo". [11].

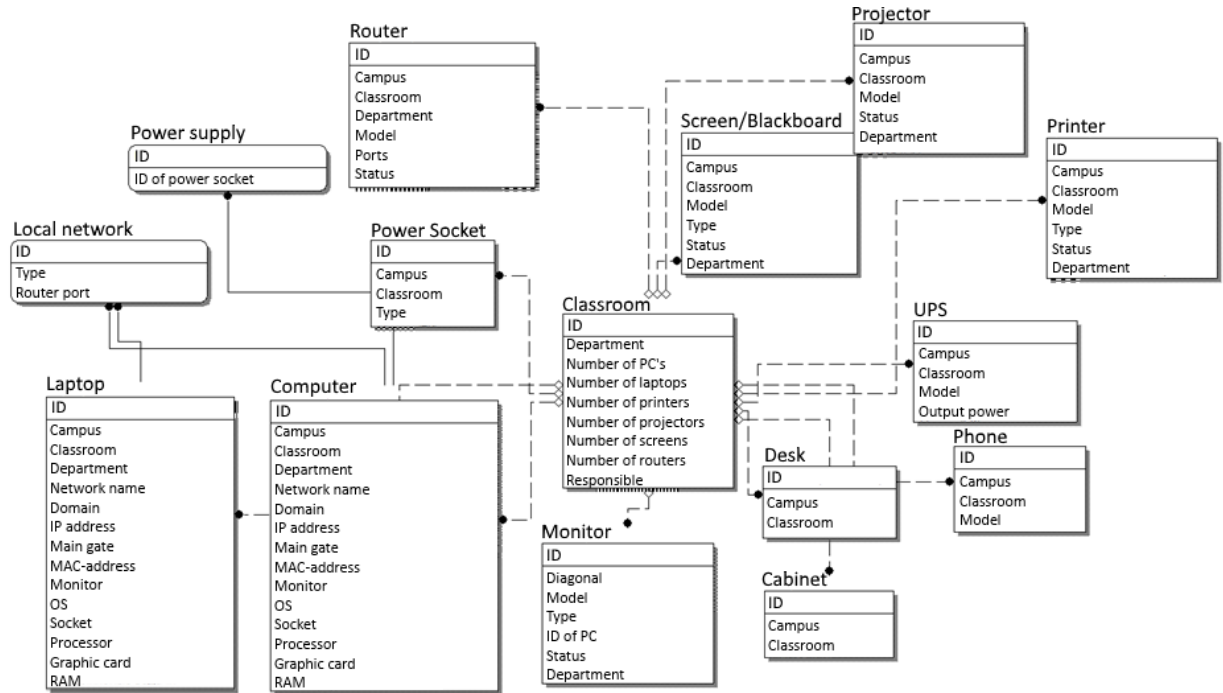


Figure 4: Information model of computing infrastructure

Object Info	
Inventory number	1101341103
Campus	6
Room	309
Department	GIS
Network name	309-10
Domain	gis.ugatu.ac.ru
IP address	172.20.39.10
Main gate	172.20.1.1
MAC-address	40-B0-76-9F-5D-F4
Monitor	19"
OS	Windows 10
Socket	1151
Processor	Pentium-G5400
Graphics card	Intel HD Graphics 610
RAM	4 Gb
Responsible	Zolotareva S.
Task	Laboratory works

Figure 5: Example of filling in the information about the "Computer" object

3. Conclusion

The proposed approach to intelligent analysis and representation of spatial data, consisting in a formalized description of various thematic data applied to solving specific problems of information management support in complex distributed systems, can be applied to the construction of complex information (especially – geoinformation) systems in significantly distributed organizational (institutions, enterprises, authorities, etc.) and infrastructure (pipelines, electricity, heat and gas supply networks, transport, etc.). Application of this approach to the construction of a distributed subsystem (information support for technical component management) of a complex distributed information system GIS USATU, which significantly reduced the time of creation and the number of errors in the design of the entire presented system.

4. Acknowledgements

The reported study was funded by RFBR, project number 20-08-00301.

5. References

- [1] E.S. Brekotkina, A.S. Pavlov, S.V. Pavlov, O.I. Khristodulo, V.D. Trubin, Analysis and synthesis of spatial data structures for managing complex distributed systems, in: Proceedings of the 8th All-Russian Scientific Conference with International Participation "Information Technologies and Systems", Khanty-Mansiysk, Russia, 2020, pp. 120-124.
- [2] E.S. Brekotkina, M.B. Guzairov, A.S. Pavlov, S.V. Pavlov, O.I. Khristodulo, Information support for managing the vulnerability of complex distributed systems based on the processing of spatial information, Modeling, optimization and information technology 8 (2020). DOI: 10.26102/2310-6018/2020.29.2.023
- [3] R.A. Bagutdinov, M.F. Stepanov, Methods of integration, reduction of sizes and normalization of processing of heterogeneous and multi-scale data, International Journal of Open Information Technologies 9 (2021) 39-44.

- [4] K.N. Markov, Structures, functionality, and characteristics of implementation of distributed geo-information processing in the MGS-framework development environment, *Geoinformatics* 1 (2010) 22-29.
- [5] I.S. Rizaev, *Geoinformation systems*, Kazan National Research Technical University, Kazan, Russia, 2013.
- [6] S.L. Beliacov, M.L. Beliacova, I.N. Rozenberg, Integrity constraint in visualizing spatial database, *IzvestiyaSFedU. Engineering Sciences* 5 (2013) 138-143.
- [7] E.S. Antonov, D.V. Lisitsky, S.S. Yankelevich. Theoretical and methodological representation of the direct transition from geoinformation to geoscience, *Bulletin of SGUGIT* 26 (2021) 82–90. DOI: 10.33764/2411-1759-2021-26-2-82-90
- [8] M.A. Ahmad, I. Tvoroshenko, J.H. Baker, L. Kochura, V. Lyashenko, Interactive Geoinformation Three-Dimensional Model of a Landscape Park Using Geoinformatics Tools, *International Journal on Advanced Science Engineering and Information Technology* 10 (2020). DOI: 10.18517/ijaseit.10.5.11298
- [9] Yu.B. Gritsenko, O.I. Zhukovskiy, Typical architectural decisions of information-analytical systems of support of the state and municipal management, *Quality. Innovation. Education* 1 (2012) 73–82.
- [10] Certificate of state registration of the database No. 2020621703. Database of the geoinformation system of the Ufa State Aviation Technical University (Ingeo). Application No. 2020621557: receipt. 09/07/2020: registration. 09/18/2020 / L.V. Akhmadieva, R.A. Gilyazov, V.D. Trubin, K.D. Shinova, 2020.
- [11] GIS "Ingeo". Center for System Research "Integro", 2021. URL: http://www.integro.ru/projects/gis/main_gis.htm