

The complex methodology for processing heterogeneous data in intelligent decision support systems is developed. This method is made to increase the efficiency of processing heterogeneous data in intelligent decision support systems. The complex methodology consists of the following interrelated procedures: heterogeneous data storing model; heterogeneous data synchronization algorithm; heterogeneous data separation algorithm; heterogeneous data indexing algorithm. The model of storing heterogeneous intelligence data, which is the basis of the methodology, differs in the presence of templates of intelligence objects and parameter templates of intelligence objects. Templates allow storing both unstructured heterogeneous intelligence data and structured intelligence data according to a defined pattern, which reduces the time to access the data. In the heterogeneous intelligence data storage model, a heterogeneous intelligence data synchronization algorithm, heterogeneous intelligence data separation algorithm and heterogeneous intelligence data indexing algorithm are developed. The development of the proposed technique is due to the need to increase the efficiency of processing various information types in intelligent decision support systems with acceptable computational complexity. The proposed method allows increasing the efficiency of intelligent decision support systems through integrated processing of data circulating in them. The proposed method allows increasing the efficiency of information processing in decision support systems from 16 to 20 % depending on the amount of information about the monitoring object

Keywords: *decision support system, monitoring object, different types of data, computational complexity, information processing, type of information*

Received date 15.06.2020
Accepted date 17.07.2020
Published date 31.08.2020

1. Introduction

The rapid development of the number of information sources in the world significantly complicates the process of data extraction and necessitates their integration and

processing of data coming from various technical devices of information extraction [1, 2].

The main advantage of information retrieval technologies over other information technologies is the ability to combine different types of information from different da-

UDC 004.89

DOI: 10.15587/1729-4061.2020.208554

DEVELOPMENT OF COMPLEX METHODOLOGY OF PROCESSING HETEROGENEOUS DATA IN INTELLIGENT DECISION SUPPORT SYSTEMS

P. Zuiev

PhD, Head Deputy
General Staff of the Armed Forces of Ukraine
Povitroflotsky ave., 6, Kyiv, Ukraine, 03168

R. Zhyvotovskiy

PhD, Senior Researcher, Head of Research Department*

O. Zvieriev

PhD, Associate Professor, Researcher**

S. Hatsenko

PhD, Deputy Head of Department**

E-mail: gans1501@ukr.net

V. Kuprii

PhD, Associate Professor, Head Specialist of

Department Military Unit A 0135

Povitroflotsky ave., 6, Kyiv, Ukraine, 03168

O. Nakonechnyi

Adjunct

Institute for Support of Troops (Forces) and Information Technologies

Ivan Chernyakhovsky National Defense University of Ukraine

Povitroflotskiy ave., 28, Kyiv, Ukraine, 03049

M. Adamenko

PhD, Head of Laboratory

Research Laboratory**

A. Shyshatskiy

PhD, Senior Researcher

Research Department of Electronic Warfare Development

Central Scientifically-Research Institute of Arming and Military Equipment

of the Armed Forces of Ukraine

Povitroflotskiy ave., 28, Kyiv, Ukraine, 03168

E-mail: ierikon13@gmail.com

Ye. Neroznak

Adjunct***

V. Velychko

Lecturer***

*Research Department of the Development

of Anti-Aircraft Missile Systems and Complexes

Central Scientifically-Research Institute of Arming and

Military Equipment of the Armed Forces of Ukraine

Povitroflotskiy ave., 28, Kyiv, Ukraine, 03168

**Department of Intelligence

Ivan Chernyakhovsky National Defense University of Ukraine

Povitroflotskiy ave., 28, Kyiv, Ukraine, 03049

***Department of Automated Control Systems

Military institute of telecommunications and informatization named after Heroes of Kruty

Moskovsky str., 45/1, Kyiv, Ukraine, 01011

Copyright © 2020, P. Zuiev, R. Zhyvotovskiy, O. Zvieriev, S. Hatsenko, V. Kuprii,

O. Nakonechnyi, M. Adamenko, A. Shyshatskiy, Ye. Neroznak, V. Velychko

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0>)

tabases with the possibility of their graphical analysis and visualization [1, 2].

Processing of different types of data from various sources of information extraction requires significant computational operations with strict restrictions on the time of calculations.

This leads to the search for new scientific approaches to processing different information types to increase the efficiency of information systems for information retrieval.

One of the ways to increase the efficiency of information systems for information retrieval is to increase the efficiency of decision support systems (DSS). DSS were especially widespread in the processing of large data sets, providing information support to the decision-making process of decision-makers.

The creation of intelligent DSS has become a natural continuation of the widespread use of classical DSS. Intelligent DSS provide information support for all production processes and services of enterprises (organizations, institutions).

Also, these intelligent DSS have been widely used to solve specific tasks of military purpose, for example [1, 2]:

- planning the deployment, operation of communication systems and data transmission;
- automation of troops and weapons control;
- collection, processing and generalization of intelligence information on the state of intelligence objects, etc.

When developing and making decisions in automated control systems (ACS), corresponding to the operational environment, the main role is played by their justification for the decision-maker (DM) information and knowledge. Information and knowledge must meet the requirements of completeness, reliability, adequacy and consistency.

These features force the use of currently very developed software and hardware and new information technologies – DSS. Extensive and effective use of these tools at control points at different levels is becoming a vital component for making informed decisions in the ACS.

To sum up, it is necessary to solve an urgent scientific problem, which consists in the development of a comprehensive methodology for processing various types of data in intelligent decision support systems.

2. Literature review and problem statement

In [3], the analysis of known methods of processing various types of information is carried out. Also, it was found that the researchers identify a number of problems associated with the process of data extraction for further analytical processing:

- data in sources, as a rule, are presented in various formats, coding and forms, thus the of analytical problems involves the use of a uniform, universal format that will be supported by a data warehouse and analytical applications;
- for excessively detailed data contained in the sources, it is necessary to carry out cleaning and generalization. In this case, the methods and algorithms that are designed for this purpose are often more complex than the analysis algorithms themselves;
- lack of integrated use of information processing and distribution methods.

In [4], the development of a generalized metric in the problem of analyzing multidimensional data with different

types of features is carried out. The essence of the proposed metric is that it allows you to build clustering, classification and association algorithms based on it, using classical processing methods.

The proposed metric is designed to assess the proximity of objects with given features, which allows reducing it to scalar numerical values. This makes it possible to reduce the problem to the classical numerical form and provides a fundamental opportunity to apply known methods and algorithms. However, this metric does not allow effective functioning in conditions of insufficient computing resources.

In [5], an approach to in-depth analysis of various types of data affecting the energy efficiency of buildings based on the representation of the hierarchy of factors in the form of a multidimensional cube with different levels of abstraction is proposed. This approach allows building a multi-level description of the object, but does not take into account the uncertainty about the state of the monitoring object, which does not allow a full assessment of its condition. This approach is focused on the use of sufficient computing resources.

In [6], an approach to processing various types of data obtained from an unmanned aerial vehicle, implemented in the GRASS GIS software environment, is presented. This approach is based on three-dimensional raster methods of image processing with a subsequent reduction of their redundancy. However, this approach is intended only for processing graphical information and does not take into account the type of uncertainty about the state of the monitoring object.

In [7], the description of the work of spatial and temporal degradation of soil erosion is carried out. The description of the soil erosion process is described using the method of erosion potential. This method is based on the analytical processing of various types of data on the factors influencing the erosion process. The method is characterized by a high degree of reliability, less computational complexity, simplicity and is suitable for use in GIS. However, this approach is intended only for processing various types of cartographic information with a sufficient number of available computing resources. This feature limits its scope.

In [8], the method of binary classification of social network users based on the method of logical regression is given. This method allows processing a variety of information about users of social networks. However, this method requires significant computing resources and does not take into account the uncertainty about the state of the monitoring object (in this case, social network users).

In [9], the problem of processing information from heterogeneous technical monitoring devices is considered. As a possible solution to the problem, the application of a generalized method of information processing based on the method of clustering of territorially combined monitoring information sources and the use of a frame model of the knowledge base for identifying monitoring objects is proposed. The clustering technique is formed on the basis of the Lance-Williams hierarchical agglomerative procedure using the Ward metric. The frame model of the knowledge base is built using object-oriented modeling tools. The disadvantages of the proposed generalized methodology include not taking into account the relative importance of events and the inability to work in a shortage of computing resources. Also, the disadvantages of this technique include the inability to redistribute computing resources between elements to increase the efficiency of information processing.

In [10], a method for processing different types of acoustic information from different sources of origin to identify the level of information security of unmanned autonomous objects is developed. This method is based on the use of a two-layer neural network with sigmoid hidden neurons. However, this method is intended only for working with acoustic information, requires significant computing resources and does not take into account the degree of uncertainty about the state of the monitoring object.

In [11], a method for determining the type of signal modulation based on a convolutional neural network by analyzing various signal parameters is proposed. This method is highly effective, but can only be used to solve radio monitoring problems, requires significant computing resources and does not take into account the degree of uncertainty about the state of the monitoring object.

In [12], a method of signal identification for unmanned aerial vehicles is proposed. The method is based on an artificial neural network and uses a knowledge base of radio wave propagation taking into account geographical coordinates. The disadvantages of this method are that the method is limited to solving radio monitoring problems, requires significant computing resources and does not take into account the degree of uncertainty about the state of the monitoring object.

In [13], the development of a methodology for correcting general geometric and topological errors in geographic information systems is carried out. This methodology is made to correct errors that occur while converting different types of data in geographic information systems from analog to digital. However, this methodology is not intended to work with other types of information than geometric, requires significant computing resources and does not take into account the degree of uncertainty about the state of the monitoring object.

In [14], an approach to the conversion of aerial photographs and satellite images is proposed, which is based on the results of their geomorphological, geobotanical, reclamation, erosion and other surveys in geographic information systems into a digital landscape model. However, this approach does not take into account other types of information circulating in geographic information systems and does not take into account uncertainty about the state of the monitoring object.

In [15], an intelligent system for processing various types of data circulating in geographic information systems is proposed. This intelligent system is designed to solve problems of geological exploration in geographic information systems. The essence of this approach is that the components of the target mineral system, which are compared, are converted into a set of maps, which leads to automatic updating of the map. However, this intelligent system is designed only to solve geological exploration problems and requires significant computing resources.

In [16], an approach to monitoring the state of the power grid using a geographic information system is proposed. The essence of this approach is that on the basis of comprehensive processing of information about the state of the grid, its integration occurs. However, this intelligent system is designed solely to monitor the state of the grid, requires significant computing resources and does not take into account the uncertainty about the state of the monitoring object.

In [17], the architecture of the integrated GIS platform architecture is proposed, which is designed to meet the requirements of processing and analysis of real-time space-time data using cloud computing. This platform does not take into account the relative importance of events that have

to be analyzed and further processed and requires significant computing resources.

In [18], an approach to using GIS to process and present geotechnical data in formats that are useful to engineers, planners and land management professionals is described. This approach significantly reduces the time for processing data circulating in the GIS. However, this approach is intended exclusively for use in land management without taking into account the importance of information circulating in the system. This approach requires significant computing resources, is unable to distribute information to improve the efficiency of its processing.

The work [19] considers the use of GIS for processing various types of spatial data to optimize urban planning. This approach allows increasing the energy efficiency of urban planning, optimizing transport links and proposing strategies for city development. The proposed approach requires significant computing resources and does not take into account the degree of uncertainty of the data about the monitoring object.

In the work [20], the approach to processing various data types on the basis of GIS is presented. The essence of the proposed approach is to analyze the energy efficiency of buildings by technical, economic, environmental and social criteria. Energy balance and spatial analysis were performed on the basis of a geographic information system. This approach requires significant computing resources and complete information about the status of the monitoring object.

The work [21] presents an approach to processing various data on the state of the communication channel. The essence of the proposed approach consists in assessing the state of the communication channel and three indicators with different dimensions. This approach allows an integrated assessment of the state of the communication channel using fuzzy logic. This approach requires complete information about the state of the channel and accumulates an evaluation error during operation.

The analysis showed that the known methods (techniques):

- have great computational complexity;
- do not take into account the degree of awareness of the state of the monitoring object;
- do not index heterogeneous data;
- do not make synchronization of heterogeneous data;
- do not allow complex processing and distribution of information about the state of the monitoring object.

Therefore, it is necessary to develop a comprehensive methodology for processing various types of data in intelligent decision support systems, which can effectively process and distribute large data sets in conditions of uncertainty, as well as shortage of computing resources.

3. The aim and objectives of the study

The aim of the study is to develop a comprehensive methodology for processing different types of data in intelligent decision support systems in the conditions of different types of data and uncertainty about the state of the monitoring object.

To achieve the aim, the following objectives were set:

- to develop a model for storing heterogeneous data in accordance with the concept of data lake;
- to develop an algorithm for synchronizing heterogeneous data;

- to develop an algorithm for separating heterogeneous data;
- to develop an algorithm for indexing heterogeneous data.

4. Development of a model for storing heterogeneous data

A comprehensive methodology for processing different types of data in intelligent decision support systems in the conditions of different types of data and uncertainty about the state of the monitoring object consists of the following interrelated procedures:

- a model for storing heterogeneous data according to the concept of data lake;
- an algorithm for synchronizing heterogeneous data;
- an algorithm for separating heterogeneous data;
- an algorithm for indexing heterogeneous data.

These types of intelligence information are combined into the concept of data, due to the diversity of the processed information (geographical, textual, video, photo) as a decision support system.

In computer science and information technology, data are defined as:

- subject to multiple interpretation of information presentation in a formalized form suitable for transmission or processing [23–25];
- forms of information presentation that information systems and their users deal with.

Thus, in decision support systems, the problem is complicated by the heterogeneity of data. Based on the analysis of the processes that occur in data processing, the following classification of data types used in decision support systems is identified. Table 1 shows the relationship of types of intelligence information to the corresponding classification of data types used in decision support systems.

According to the location of data sources, we should talk about concentrated data (source of information) and geographically distributed sources. The area involved in data collection and processing is called data fusion.

Accordingly, the data were classified by various criteria used in processing in decision support systems. The developed classification is presented in Fig. 1.

Depending on the data structure, data can be divided into 3 main classes: structured, unstructured and semi-structured.

Structured data are often stored in databases, where they are already available and processed in a fixed format. Otherwise, unstructured data are data of unknown structure. This form of data is characterized by a number of difficulties in processing and retrieving useful information. Typical examples of unstructured data are those that contain a combination of plain text files, images, and video files.

Table 1

Ratio of information types to the corresponding classification of data types used in DSS

Types of intelligence information	Types of DSS data
Sensor-type machine data (time parameters of ground, air, and sea-based radar stations) were obtained by means of radioelectronic intelligence.	sensors data
Data that capture the action, event, or state of an object at a particular point in time (semantic data about changes in the state of objects).	log data
Visual data – images or videos that record the behavior of the target system or systems in the operating environment. Visual data directly or indirectly characterize the state of the system (video surveillance data, photographs from unmanned aerial vehicles, photographs obtained from space exploration. Geospatial data, event data, which have two geographical coordinates (latitude and longitude) as the main attributes and timestamp attributes, maps.	geospatial data
Textual data are unstructured data that are messages in natural language and usually a reflection of a particular situation.	textual data
Data from Internet sources – open source unstructured data, representing a set of textual data, expressions (news, messages), describing, characterizing the behavior of individuals, groups, states, countries, coalitions.	open source data

Semi-structured data contain the described properties of structured and unstructured data. Increasing the number of data sources inevitably leads to an increase in data volume. As a result, the concept of “big data” has emerged. Big data are characterized by volume, speed of receipt, variety and variability.

To solve this problem, the concept of data storage by the principle of “data lake” is proposed. The idea of “data lake” is to store raw, unstructured, or semi-structured data in original format until needed. This will allow users to request smaller, more up-to-date and more flexible data sets. As a result, query time can be reduced to work in a data store, data warehouse or relational database.

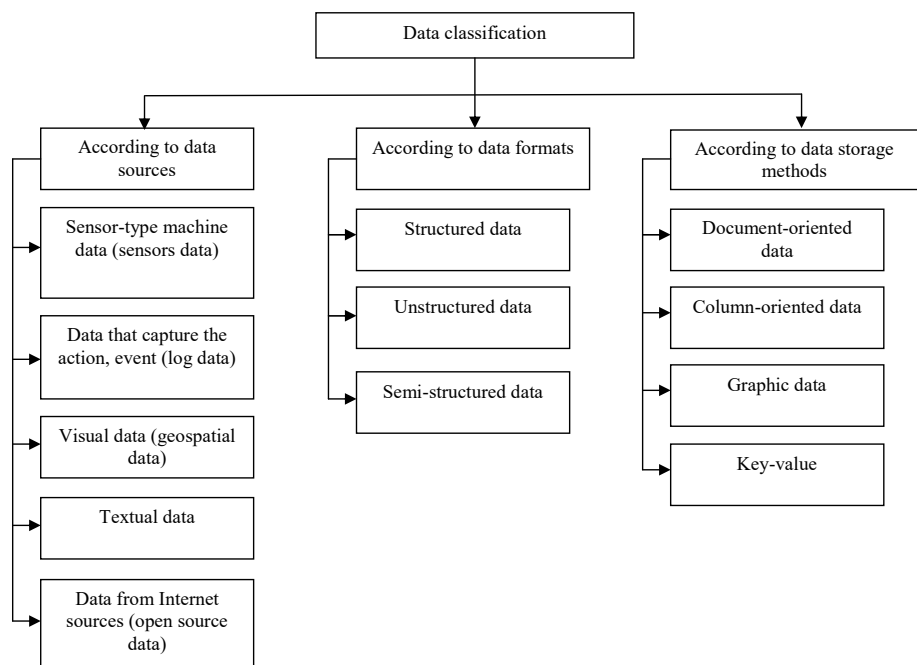


Fig. 1. Classification of data by the criteria used for DSS processing

As a result of the system analysis of the data collection process and structure, a model for storing heterogeneous intelligence data in accordance with the data lake concept is proposed.

$$S = \left\langle \{DT\}_{i=1}^{n_{ST}}, \{SS\}_{j=1}^{m_{ST}}, \{E\}_{k=1}^{p_E}, IS, DS \right\rangle, \tag{1}$$

where $\{DT\}_{i=1}^{n_{ST}}$ is the set of intelligence data templates; n_{ST} is the number of intelligence data templates, $\{SS\}_{j=1}^{m_{ST}}$ is the methods of breaking up heterogeneous intelligence data, m_{ST} is the number of intelligence data types; $\{E\}_{k=1}^{p_E}$ is the set of performers of intelligence gathering tasks; IS is the method of data indexing in the data lake storage; DS is the lake structure of heterogeneous intelligence data.

Fig. 2 shows a model for storing heterogeneous data.

Let's consider the components of the model. The Object Template component is developed to template an intelligence object O_i , where $i=1, \dots, n$. Each intelligence object can have multiple sources of intelligence data sources M_{ds} , which are created according to the Data Source Template. We suppose that there is some intelligence object, the structure of which can be presented as follows:

$$O = \langle ds_1, ds_2, \dots, ds_m \rangle. \tag{2}$$

Each data source can have many parameters with different types of intelligence data M_i according to the Parameter Template templates. The structure of such a source of intelligence data ds is presented as follows:

$$ds = \langle p_1, p_2, \dots, p_k \rangle. \tag{3}$$

The model of storing heterogeneous intelligence data, which is the basis of the methodology, differs from the existing ones by the presence of templates of intelligence objects and templates of intelligence object parameters. Templates allow distributed storage of both unstructured heterogeneous intelligence data and structured intelligence data according to a specific pattern, which reduces the time to access the data.

In order to increase the efficiency of synchronization of heterogeneous data, an algorithm for synchronization of heterogeneous data was developed.

5. Development of an algorithm for synchronizing heterogeneous data

The algorithm for synchronizing heterogeneous data is shown in Fig. 3.

1. A set of intelligence data P is obtained. At this stage, the value of uncertainty about information sources is taken into account, according to expressions (10)–(12) [26].

2. Create empty $L_i \langle t, d \rangle$ lists to store $L_i \langle t, d \rangle = f$, $i=1, n$ objects, where n is the number of intelligence object templates, t is the data d generation time (d in the form of *JSONObject* (JavaScript Object Notation)).

3. For each element of the obtained data package, make the following steps:

3. 1. Identify the object by the key parameters included in the obtained data.

3. 2. If the object is not found in the L_k list that corresponds to this object according to the template, $k=1, n$ then,

3. 2. 1. Create a new *JSONObject*.

3. 2. 2. Save the obtained data to the created *JSONObject* according to the object template.

3. 2. 3. Add the created object to the L_k list.

Otherwise:

3. 2. 4. Sort all data about the found object by the *timestamp* time T of the object data block generation.

3. 2. 5. Save the obtained data to the created *JSONObject* according to the template of the intelligence object.

3. 2. 6. Output the L list.

Each item in the resulting L list is a block of intelligence data in *JSONObject* format. Data on the intelligence object can be heterogeneous, which in some situations leads to significant difficulties in further analysis. Therefore, for ease of processing heterogeneous intelligence data, this paper proposes an algorithm for separating heterogeneous intelligence data, which is based on a parameter template.

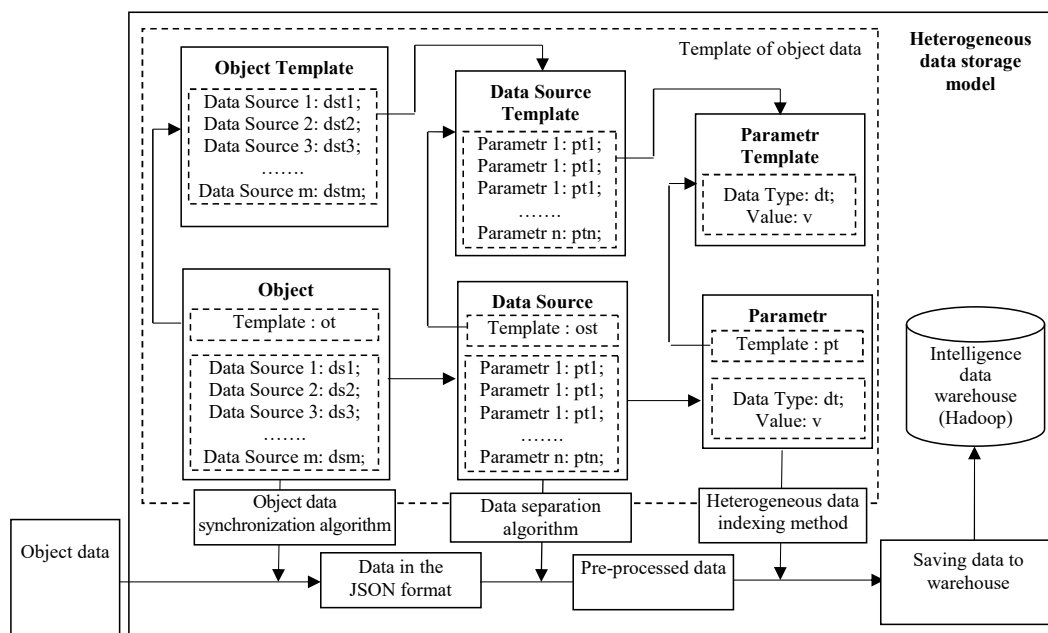


Fig. 2. Heterogeneous data storage model

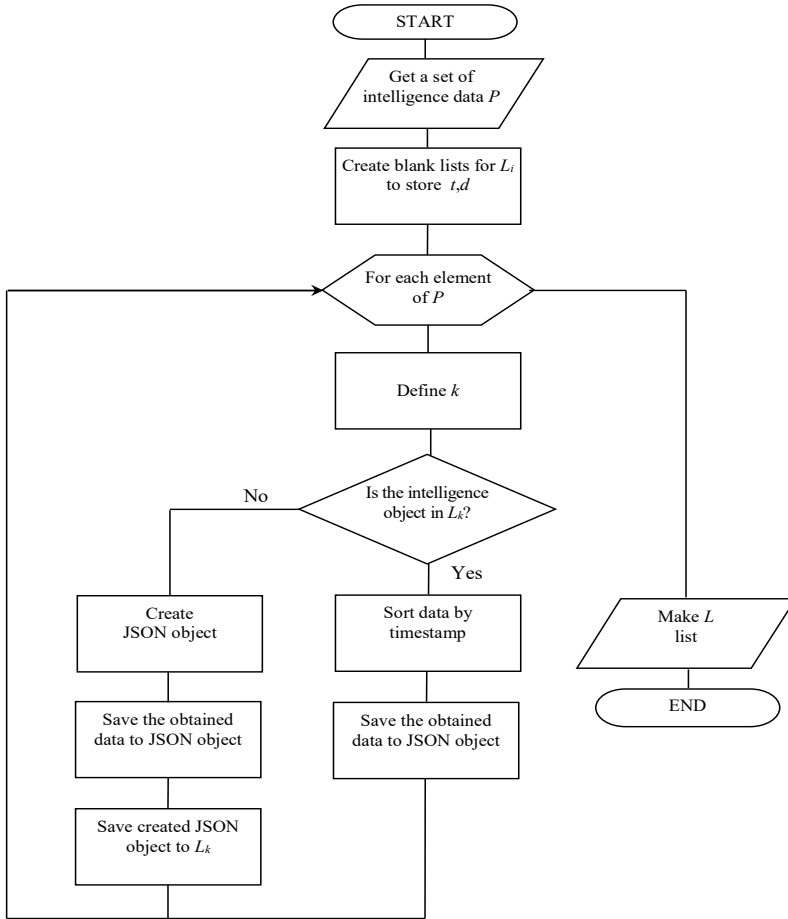


Fig. 3. Algorithm for synchronizing heterogeneous data

6. Development of an algorithm for separating heterogeneous data

The algorithm for separating heterogeneous data is shown in Fig. 4.

1. The L list (result of executing the algorithm of heterogeneous intelligence data synchronization) is obtained.

2. Create an empty $R\langle JSONObject \rangle$ list, which contains data about the intelligence object according to the "Parameter Template".

3. For each element of the L list must be made:

3. 1. Divide the items according to the "Data Source Template" and get an L_s list.

3. 2. For each item of L_s list must be made:

3. 2. 1. Divide the items according to the "Parameter Template" and add it to the L_p list.

3. 2. 2. For each item of the L_p list:

3. 2. 2. 1. Create a new $JSONObject$.

3. 2. 2. 2. Save the obtained data to the created $JSONObject$ according to the parameter template.

3. 2. 2. 3. Add the object to $R\langle JSONObject \rangle$ list.

4. Return $R\langle JSONObject \rangle$ list.

7. Development of an algorithm for indexing heterogeneous data

To increase the efficiency of collection and storage of heterogeneous data, in the heterogeneous data

storage model, an algorithm for indexing heterogeneous data is proposed. The efficient storage structure of heterogeneous data of data lakes allows convenient and fast access to storage for batch data processing.

Each file in the storage of heterogeneous intelligence data is a pair $\langle t,v \rangle$, where t is the intelligence data generation time, v is the value of intelligence data.

Fig. 5 shows the created file that is indexed in the storage in accordance with the principle of indexing heterogeneous intelligence data in the data warehouse.

The explanation of the application of the proposed method of indexing heterogeneous intelligence data.

1. $The_root_to_store_heterogeneous_data$ is the root directory for storing heterogeneous data.

2. ot_IDn is the identifier of the n object template.

3. od_IDm is the identifier of the unique m object.

4. ds_IDk is the identifier of the unique source of this object to k .

5. p_IDl is the identifier of the unique parameter of l source.

6. f_IDp is the identifier of the unique file of p parameter.

For ease of reducing the data processing time, the file is given a name that contains the data block generation time t .

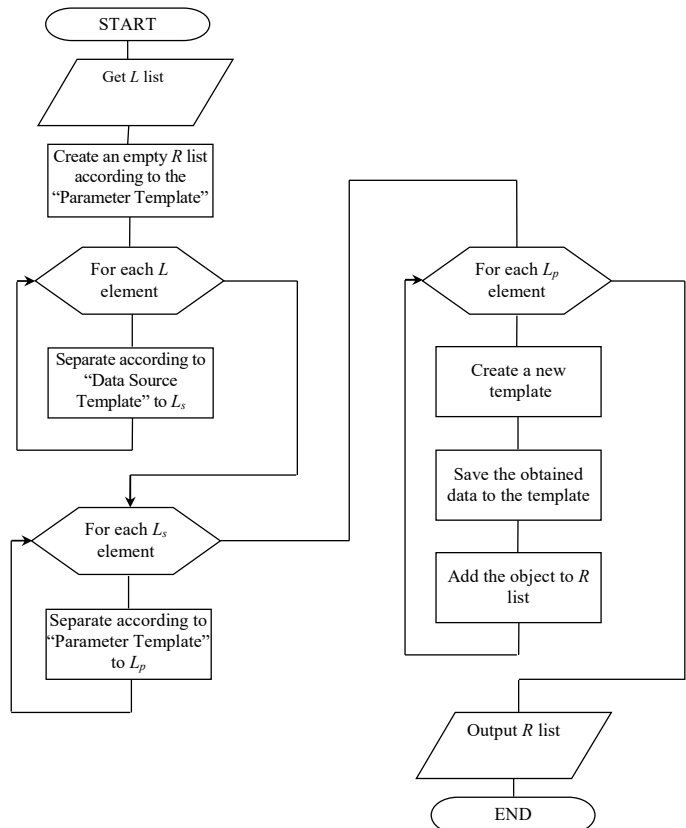


Fig. 4. Algorithm for heterogeneous intelligence data separation

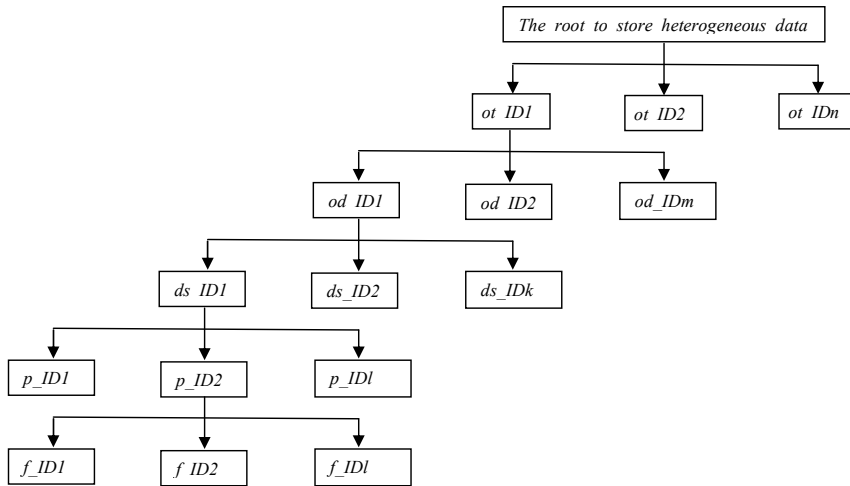


Fig. 5. Principle of indexing heterogeneous intelligence data in the data warehouse

8. Discussion of the results of the development of a comprehensive methodology for processing different data types

Let's make the modeling of the complex methodology of processing data of different types in accordance with the algorithms of procedures in Fig. 2–5 and expressions (1)–(3). The technique of processing data of various types in intelligent decision support systems is proposed. The proposed technique is modeled in the Umbrello UML Modeller software environment (in terms of processing different data types) and MathCad 14 (in terms of computational complexity estimating).

The standard notation for modeling large information systems based on object-oriented methodology is the Unified Modeling Language (UML). One of the available tools is the Umbrello UML Modeller environment, which meets two key requirements: free and cross-platform. This application is free software made to build UML diagrams and supports all their standard types. Using the Umbrello UML Modeller environment, a frame knowledge base was built to identify the monitoring objects. The frame knowledge base was built as a UML class diagram, where each frame is represented as a class with its own attributes and procedures.

For this purpose, the method was modeled to determine the number of computing operations required for this method.

Initial data for assessing the condition of the monitoring object using the proposed method:

- the number of sources of information about the condition of the monitoring object is 3 (radio monitoring, remote sensing devices and unmanned aerial vehicles);
- the number of information signs, which determine the state of the monitoring monitoring is 13 (affiliation, type of organizational formation, priority, minimum width on the front, maximum width on the front, number of personnel, minimum depth on the flank, maximum depth on the flank, total number of personnel, number of samples of armaments and military equipment, the number of types of pieces of armaments and military equipment, number of communication devices, type of communication devices);
- the options for organizational and staff formations are company, battalion and brigade.

To make the experiment, 8 sample frames were used in this method: “Monitoring object”, “Unit”, “Operational and

tactical grouping of troops”, “Brigade”, “Battalion”, “Company”, “Detected object”, “Selected cluster”. Each of the sample frames (except for “Monitoring object” and “Unit”) corresponds to instance frames in the form of information about specific divisions. The constructed dependency has a clear hierarchical structure and 3 types of relations between frames: generalization, association and dependence.

The “Subdivision” frame is related to the generalization relationship with the “Monitoring object” frame and is its descendant. The “Operational and tactical grouping of troops”, “Brigade”, “Battalion”, “Company” frames are connected with the “Unit” frame and are its descendants. While filling the

knowledge base, instance frames are created with a structure similar to the mentioned frames, but the slots of which are filled according to the information about specific divisions.

The “Detected object” frame is a generalized form that is filled at monitoring points (posts) and contains information about the class, type, coordinates and number (in the case of a group object). The specified frame is associated with a relationship to the “Selected cluster” frame. The slots of the “Selected cluster” frame are filled as a result of the procedure of clustering territorially combined monitoring information sources, the implementation method of which is given above. To determine the values of the slots, the attached procedures centerKoordFinding (finding the coordinates of the center of the selected cluster) and clusterSizeFinding (finding the size of the selected cluster) are used. The resulting set of selected clusters, the information of which is represented by a set of instance frames of the type “Selected cluster”, is subjected to the procedure of identification of monitoring objects. During the identification, the frames corresponding to the selected clusters are compared with the reference frames contained in the knowledge base.

The result is a conclusion about the cluster correspondence to a specific unit (or identification of a new monitoring object) and its current state.

Primary processing of information from information monitoring sources, filling of frames of the type “Detected object”.

Clustering of identified monitoring information sources in accordance with the filling of frames of the “Selected cluster” type. Identification of selected clusters and formation of conclusions.

Based on these initial data, we obtain the distances between the points given in Table 1.

Fig. 6 shows graphical estimates of the efficiency of data processing by the criterion of the number of computing operations of the method proposed in the paper. The comparison of the developed complex technique was carried out by comparing its efficiency with the known ones presented in [9, 10, 12]. Denote these works 1, 2 and 3, respectively.

From the given graphic dependencies, it is seen that the proposed technique allows increasing the efficiency of information processing (reducing the number of computing operations) from 16 to 20 % depending on the amount of information about the monitoring object.

Table 1

Distances between the points

Fragment of the matrix of distances between points	1	2	3	4	5	6	7	8	9	10
1	0	6.18E+08	2.58E+08	2.08E+08	5.18E+08	1.35E+09	2.29E+08	2.85E+06	1.21E+09	2.01E+08
2		0	5.50E+08	1.10E+08	4.16E+08	7.28E+08	1.04E+09	5.61E+08	3.63E+08	1.20E+08
3			0	2.77E+08	1.10E+09	2.06E+09	9.70E+08	2.88E+08	1.63E+09	2.28E+08
4				0	3.08E+08	8.33E+08	5.45E+08	1.75E+08	5.69E+08	3.17E+06
5					0	1.98E+08	3.43E+08	4.46E+08	2.25E+08	3.72E+08
6						0	9.79E+08	1.23E+09	8.98E+07	9.38E+08
7							0	2.07E+08	1.12E+09	5.88E+08
8								0	1.10E+09	1.73E+08
9									0	6.47E+08
10										0

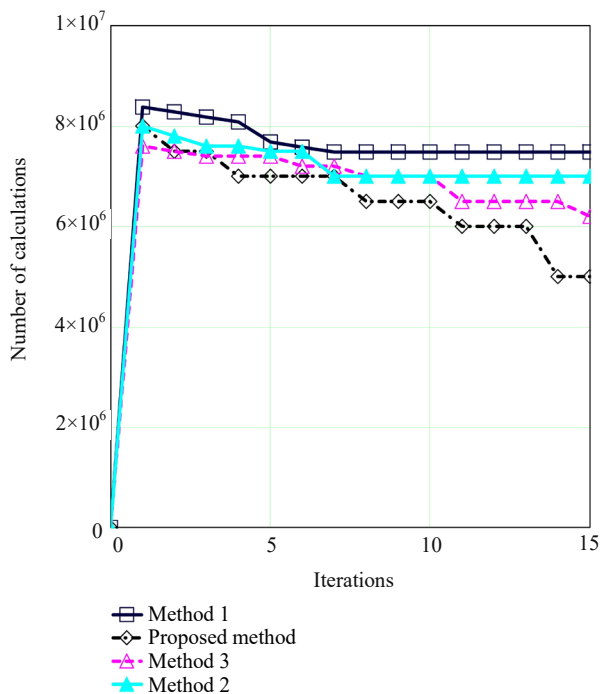


Fig. 6. Results of evaluating the effectiveness of the proposed algorithm

This gain is explained by a complex combination of not only taking into account different types of data on the state of the monitoring object, but also the procedures of synchronizing heterogeneous data, separating heterogeneous data and indexing heterogeneous data.

The difference of the proposed technique:

- allows high-quality processing of large arrays of different types of data that have a numerical and quantitative nature;
- has less computing complexity due to the redistribution of computing operations;
- takes into account the degree of awareness about the condition of the monitoring object;
- allows complex processing and distribution of information about the state of the monitoring object.

This method allows increasing the efficiency of processing different types of data by 16–20%. The advantages of this technique are:

- the possibility of rational allocation of resources of information systems to solve problems of responding to different events, taking into account their relative importance;

- minimization of the total time required to perform the task of responding to another circumstance;

- limiting the degree of human participation in the management cycle of complex processing resources, as well as automatic determination of the option of forming scenarios for solving monitoring tasks.

The limitations of this method include the need for communication channels with high reliability of information transmission and minimal delay. This is due to the need to process information in close to real-time mode and high requirements for the reliability of information circulating in special-purpose decision support systems.

The disadvantages of this method include the need to process large data sets to determine the state of the monitoring object.

The practical significance of the developed technique is the significant increase of the efficiency of integrated data processing in automated control systems.

It is proposed to use the proposed methodology in ACS DSS of artillery units, special-purpose geographic information systems, ACS DSS of aviation and air defense, as well as ACS DSS of logistics of the Armed Forces of Ukraine in the development of software for processing heterogeneous data and decision-making.

This research is a further development of research conducted by the authors, aimed at developing methodological principles for improving the efficiency of data processing in special-purpose geographic information systems, published earlier [26, 27]. Areas of further research should be aimed at reducing computing costs when processing various types of data in special-purpose information systems.

9. Conclusions

1. The model for storing heterogeneous data in accordance with the concept of data lake is developed in the paper. The model of storing heterogeneous intelligence data, which is the basis of the methodology, differs from the existing ones in templates of intelligence objects and parameter templates of intelligence objects, which allow distributed storage of both unstructured heterogeneous intelligence data and structured intelligence data according to a certain scheme to reduce time spent on data access.

2. The algorithm for synchronizing heterogeneous data is developed and it takes into account the type of uncertainty about information sources and does not require editing da-

tabase volumes and does not require accounting for changes in objects in a separate storage.

3. The algorithm for separating heterogeneous data, which allows the separation of both processed and unprocessed data with different units of measurement and origin is developed. It means that the separation occurs on a set of selection parameters that characterize the intelligence object.

4. The algorithm of indexing heterogeneous data, which allows indexing heterogeneous data (processed and unprocessed) for their batch processing is developed. Taken together, these scientific results are a comprehensive method of processing different data in intelligent decision support systems, which increases the efficiency of processing different types of information (reduces computational complexity) taking into account uncertainty about the state of the monitoring object, and allow complex processing and distribution of information on the condition of the monitoring object.

Acknowledgments

The author's team is grateful for the help to:

Honored worker of science and technology of Ukraine, Doctor of Technical Sciences, Professor Slyusar Vadym – chief researcher of the Central Research Institute of Armaments and Military Equipment of the Armed Forces of Ukraine.

Doctor of Technical Sciences, senior researcher Sova Oleg – head of the “Automated control systems” Department of the Military Institute of Telecommunications and Informatization named after Heroes of Kruty.

Doctor of Technical Sciences, Senior Researcher Zhuravsky Yuri – leading researcher at the Research Center of the Zhytomyr Military Institute named after S. P. Korolev.

PhD, Associate Professor Bashkirov Oleksandr – leading researcher of the Central Research Institute of Armaments and Military Equipment of the Armed Forces of Ukraine.

References

- Makridenko, L. A., Volkov, S. N., Hodnenko, V. P. (2010). Kontseptual'nye voprosy sozdaniya i primeneniya malyh kosmicheskikh apparatov. *Voprosy elektromehani, 114*, 15–26.
- Bashkirov, O. M., Kostina, O. M., Shishats'kiy, A. V. (2015). Development of integrated communication systems and data transfer for the needs of the Armed Forces. *Weapons and military equipment, 5* (1), 35–39.
- Trotsenko, R. V., Bolotov, M. V. (2014). Data extraction process for heterogeneous sources. *Privolzhskiy nauchnyy vestnik, 12-1* (40), 52–54.
- Bodyanskiy, E., Strukov, V., Uzlov, D. (2017). Generalized metrics in the problem of analysis of multidimensional data with different scales. *Zbirnyk naukovykh prats Kharkivskoho universytetu Povitrianykh Syl, 3*, 98–101.
- Noh, B., Son, J., Park, H., Chang, S. (2017). In-Depth Analysis of Energy Efficiency Related Factors in Commercial Buildings Using Data Cube and Association Rule Mining. *Sustainability, 9* (11), 2119. doi: <https://doi.org/10.3390/su9112119>
- Petras, V., Petrasova, A., Jeziorska, J., Mitasova, H. (2016). Processing UAV and lidar point clouds in Grass GIS. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B7*, 945–952. doi: <https://doi.org/10.5194/isprs-archives-xli-b7-945-2016>
- S Polovina, S., Radic, B., Ristic, R., Milcanovic, V. (2016). Spatial and temporal analysis of natural resources degradation in the Likodra River watershed. *Glasnik Sumarskog Fakulteta, 114*, 169–188. doi: <https://doi.org/10.2298/gsf1614169p>
- Poryadin, I., Smirnova, E. (2017). Binary Classification Method of Social Network Users. *Science and Education of the Bauman MSTU, 17* (02), 121–137. doi: <https://doi.org/10.7463/0217.0000915>
- Tymchuk, S. (2017). Methods of Complex Data Processing from Technical Means of Monitoring. *Path of Science, 3* (3), 4.1–4.9. doi: <https://doi.org/10.22178/pos.20-4>
- Semenov, V. V., Lebedev, I. S. (2019). Processing of signal information in problems of monitoring information security of unmanned autonomous objects. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics, 19* (3), 492–498. doi: <https://doi.org/10.17586/2226-1494-2019-19-3-492-498>
- Zhou, S., Yin, Z., Wu, Z., Chen, Y., Zhao, N., Yang, Z. (2019). A robust modulation classification method using convolutional neural networks. *EURASIP Journal on Advances in Signal Processing, 2019* (1). doi: <https://doi.org/10.1186/s13634-019-0616-6>
- Zhang, D., Ding, W., Zhang, B., Xie, C., Li, H., Liu, C., Han, J. (2018). Automatic Modulation Classification Based on Deep Learning for Unmanned Aerial Vehicles. *Sensors, 18* (3), 924. doi: <https://doi.org/10.3390/s18030924>
- Kukulka, A., Salata, T., Cegielska, K., Szylar, M. (2018). Methodology of evaluation and correction of geometric data topology in QGIS software. *Acta Scientiarum Polonorum Formatio Circumietus, 17* (1), 125–138. doi: <https://doi.org/10.15576/asp.fc/2018.17.1.125>
- Rulev, A., Yuferev, V. (2015). Theory of geoinformatic mapping of erosion geomorphological systems. *Vestnik Volgogradskogo Gosudarstvennogo Universiteta. Seriya 11. Estestvennye Nauki, 4*, 62–67. doi: <https://doi.org/10.15688/jvolsu11.2015.4.7>
- Yousefi, M., Kreuzer, O. P., Nykänen, V., Hronsky, J. M. A. (2019). Exploration information systems – A proposal for the future use of GIS in mineral exploration targeting. *Ore Geology Reviews, 111*, 103005. doi: <https://doi.org/10.1016/j.oregeor-ev.2019.103005>
- Ashkezari, A. D., Hosseinzadeh, N., Chebli, A., Albadi, M. (2018). Development of an enterprise Geographic Information System (GIS) integrated with smart grid. *Sustainable Energy, Grids and Networks, 14*, 25–34. doi: <https://doi.org/10.1016/j.segan.2018.02.001>

17. Wang, S., Zhong, Y., Wang, E. (2019). An integrated GIS platform architecture for spatiotemporal big data. *Future Generation Computer Systems*, 94, 160–172. doi: <https://doi.org/10.1016/j.future.2018.10.034>
18. Wan-Mohamad, W. N. S., Abdul-Ghani, A. N. (2011). The Use of Geographic Information System (GIS) for Geotechnical Data Processing and Presentation. *Procedia Engineering*, 20, 397–406. doi: <https://doi.org/10.1016/j.proeng.2011.11.182>
19. Pedro, J., Silva, C., Pinheiro, M. D. (2019). Integrating GIS spatial dimension into BREEAM communities sustainability assessment to support urban planning policies, Lisbon case study. *Land Use Policy*, 83, 424–434. doi: <https://doi.org/10.1016/j.landusepol.2019.02.003>
20. Mokhtara, C., Negrou, B., Settou, N., Gouareh, A., Settou, B. (2019). Pathways to plus-energy buildings in Algeria: design optimization method based on GIS and multi-criteria decision-making. *Energy Procedia*, 162, 171–180. doi: <https://doi.org/10.1016/j.egypro.2019.04.019>
21. Kalantaievska, S., Pievtsov, H., Kuvshynov, O., Shyshatskyi, A., Yarosh, S., Gatsenko, S. et. al. (2018). Method of integral estimation of channel state in the multiantenna radio communication systems. *Eastern-European Journal of Enterprise Technologies*, 5 (9 (95)), 60–76. doi: <https://doi.org/10.15587/1729-4061.2018.144085>
22. Karin, S. A. (2012). Integration in the single information space of heterogeneous geospatial data. *Informatsionno-upravlyayushchie sistemy*, 2, 89–94.
23. Karin, S. A. (2014). Developing a domain-specific ontology in spatial data processing systems. *Informatsionno-upravlyayushchie sistemy*, 4, 78–84.
24. Belousov, S. M. (2006). Matematicheskaya model' mnogopotchnoy sistemy massovogo obsluzhivaniya, upravlyaemoy planirovshchikom resursov. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Ser.: Informatsionnye tehnologii*, 4 (1), 14–26.
25. Karin, C. A., Dudin, E. A. (2014). Podhody k sozdaniyu raspredelennoy sistemy sbora, hraneniya i obrabotki geoprostranstvennyh dannyh. *Informatsiya i kosmos*, 3, 46–51.
26. Koshlan, A., Salnikova, O., Chekhovska, M., Zhyvotovskiy, R., Prokopenko, Y., Hurskiy, T. et. al. (2019). Development of an algorithm for complex processing of geospatial data in the special-purpose geoinformation system in conditions of diversity and uncertainty of data. *Eastern-European Journal of Enterprise Technologies*, 5 (9 (101)), 35–45. doi: <https://doi.org/10.15587/1729-4061.2019.180197>
27. Kuchuk, N., Mohammed, A. S., Shyshatskyi, A., Nalapko, O. (2019). The method of improving the efficiency of routes selection in networks of connection with the possibility of self-organization. *International Journal of Advanced Trends in Computer Science and Engineering*, 8 (1.2), 1–6. Available at: <http://www.warse.org/IJATCSE/static/pdf/file/ijatcse01812sl2019.pdf>