Development of an information and analytical system for modeling the demographic situation in the Russian Federation

Denis R. Kalugin[1], Yuriy A. Leonov [1], Rodion A. Filippov [1] and Lyudmila B. Filippova [1]

[1] *Bryansk State Technical University, Bryansk 241035, Russia*

Abstract
This article is devoted to the study of demographic statistics indicators for the subjects of the Russian Federation. The article describes the relevance of this research, as well as the advantages and disadvantages of the similar software solutions: the IAS "Demography", the Rosstat data showcase, the automated IS "Municipal Population Register". The functional scheme of the developed IAS is presented both in the form of a single functional block and in the form of its decomposition into five modules responsible for collecting analytical data, interacting with the database, identifying correlations, predicting future indicators and visualizing data. Based on the obtained data, it is possible to perform extrapolation, interpolation, and regression analysis of the data. The authors introduce a block diagram of the algorithm for predicting demographic statistics. This algorithm is used to calculate the forecast of the demographic situation in the Russian Federation based on the information available in the database, taking into account user settings. The functional capabilities of the developed information system are considered. The authors use the data of the Federal State Statistics Service (Rosstat) as the material base for the study. The methods of DataMining were used as a theoretical basis for the development of the prediction algorithm. The developed information system provides the analyst with tools for analyzing the dynamics, development trends and identifying the correlation of demographic indicators, as well as building a forecast of their development in the coming years.

Keywords
Demography, demographic situation modeling, demographic indicators development dynamics, information and analytical system, method of least squares

1. Introduction

The study of the demographic situation is one of the global problems of modern statistics and requires a deep and comprehensive study. Demographic processes influence the course of all other social processes and at the same time develop under their influence.

To maintain the necessary level of reproduction of the population, a developed social base is necessary, which stimulates the birth rate and affects the increase in average life expectancy. Without taking into account the demographic factors of the development of society, it is impossible to build an optimal social policy.

In turn, thanks to a reliable demographic forecast, the executive authorities have data on the qualitative and quantitative composition of the population of both municipalities and the country as a whole for several years to come. This makes it possible to plan the necessary activities, such as: output of products and their quantity; construction of the necessary number of schools, kindergartens, higher educational institutions, hospitals and shops; provision of the necessary number of jobs, etc.

There has been depopulation of the Russian population in recent years. Experts identify the following reasons for this phenomenon: low birth rate, high mortality, loss of traditional foundations of a strong family, low level of domestic health care, bad habits of the nation (alcoholism, drug addiction), low income, housing problem, infertility of women as a result of abortions, early mortality due to accidents, murders and suicides, overload of negative information which creates an oppressive and disturbing atmosphere in society that in its turn leads to constant stress, departure abroad of citizens of reproductive age [1, 2].

As a result, it becomes obvious that there is a need to create high-quality methods for modeling and forecasting demographic indicators.

1.1.   Analysis of the results of previous works

To determine the functional requirements for our own IAS, the following software solutions were considered: the IAS "Demography", the Rosstat data showcase, the automated information system "Municipal Population Register".

The IAS Demography is a software product that implements a full cycle of work with medical certificates to automate the formation of medical certificates upon the birth/death of citizens and the accumulation of this data for further analysis.

On the official website of Rosstat, demographic indicators can be viewed using special data showcases that provide a wide range of opportunities for analyzing statistical information.

The IS "Municipal Population Register" is intended for the formation, maintenance and use of a single database for accounting and registration of information about citizens of the municipality and their families registered at the place of residence or at the place of stay within the municipality.

When considering the above-described software solutions, it is established that each of these tools is focused on its own subject area. As a result, their functionality varies greatly, depending on the requirements of the end user.

At the same time, it is not possible to find information about any applied software solutions that allow solving the problem of predicting demographic indicators. This may be due to the fact that this issue, as a rule, is not of interest to ordinary citizens, and therefore making such projects publicly available is not a commercially viable solution. As a result, these projects are used and developed directly in research centers to solve specific internal problems.

During the analysis of the existing software solutions, it was found that at the moment there is no single software system that can fully solve the task.

However, a number of common functional features can be distinguished among them: the use of a client-server architecture, the presence of a role model for implementing access, filtering stored information, generating tables and reports based on available data, exporting statistics in generally accepted formats, and presenting data of interest in the form of graphs.

The aim of the work is to develop an information and analytical system for modeling the demographic situation in the Russian Federation. The task of the system is to assist the analyst in conducting demographic research as an element of comprehensive long-term socio-economic planning.

2. Research methods and materials
2.1.   Choosing a data source

When choosing the source, the following resources were examined: the website of the Federal State Statistics Service (Rosstat); the website of the International Labour Organization (ILO); the website of the United Nations Statistics Division (UNdata).

As a result of the analysis of resources, the site of Rosstat was chosen, despite the fact that it has the least convenient system of access to data from all the options considered and there is no stable form of reporting. However, the information provided is the most relevant and fully meets the stated subject area of the developed IAS.

The reason for the rejection of the most convenient sources, in terms of access and a formalized stable data structure, is explained by the following factors.

The statistics on the website of the International Labour Organization are aimed at analyzing the labor activity of the population. Demographic indicators are a way to demonstrate the conditions in the labor market, but do not give a complete picture of the demographic situation in the country. This information does not fully meet the requirements for the developed IAS.

On the website of the UN Statistics Division, the necessary data that meet the requirements for the developed IAS were presented. However, their update for the Russian Federation stopped in 2013, which greatly restricts the potential user when conducting the analysis, and the developer when building a forecasting algorithm. Thus, the data from the site was not used, due to the low relevance.

2.2. Development of a forecasting algorithm

When developing demographic forecasts, the following groups of methods are most often used:
1. extrapolation methods;
2. economic and mathematical methods that make it possible to develop multi-factor dynamic models;
3. age movements – a method for predicting the future composition of the population;
4. cohort analysis that compares the behavior of similar groups of people;
5. methods of expert assessments.

Each of these methods has its own advantages and disadvantages, and they can be used together to increase efficiency [3].

The task can be reduced to a regression problem, which is classically solved using the least squares method (LSM), which is the basis of the developed algorithm [4, 5].

In the process of calculating the forecast based on the LS, each subject of the Russian Federation is considered separately, as an independent set of data, since the values of indicators and their dynamics, for example, for the city of Moscow will be very different from the values of the city of Bryansk.

Within one subject of the Russian Federation, the forecast is calculated using a recursive algorithm (Figure 1) consisting of two methods: the method that calculates the forecast (Predict) and the method that identifies correlating indicators Correlation).

Predict calculates the forecast for the chosen indicator using the LSM, based on a set of "reference" values that are known for the forecast period. On the first recursive call (or in the absence of sufficiently correlated indicators), the set of "reference" values is the time period for which the forecast is calculated. After calculating the set of values for the next indicator, it is removed from the general list of indicators for which the forecast is calculated. Next, Predict calls the Correlation method to choose the next predicted indicator.
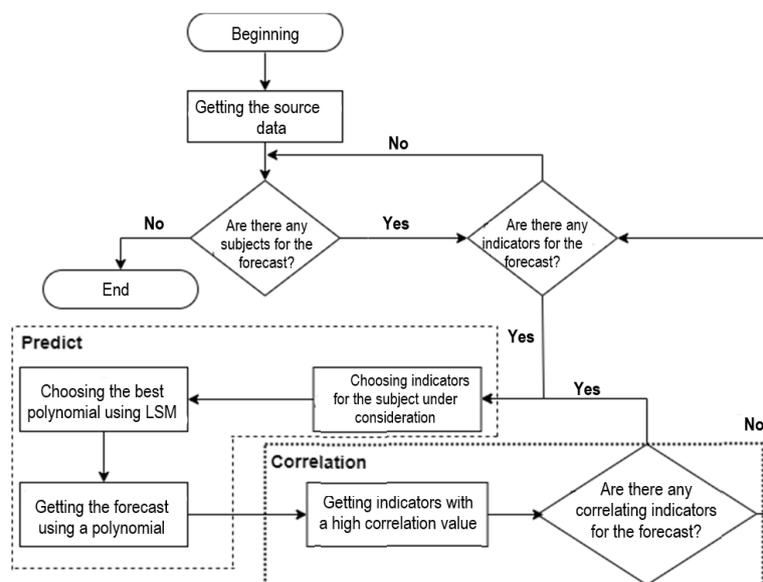

Figure 1. Prediction algorithm

Correlation determines the degree of correlation between the indicator for which the Forecast method calculated the forecast and those for which the forecast is yet unknown. As a result, we get a list of indicators with a high correlation value. If this list is not empty, the Predict method is called for each of these indicators, but the predicted data of the calculated indicator is now in the form of a set of "reference" values.

Thus, these two methods call each other, until the forecast for all indicators is calculated. After that, the operation is repeated for the next subject of the Russian Federation, and the algorithm starts again.

2.3.   Development of the IAS functional scheme

If we consider the system as a single functional unit (Figure 2), then it receives user-defined forecast settings as an input and an Excel file with new data. The system is managed by user instructions. The database receives and records data using the database server. The interaction with the user is carried out via the customer interface. As a result of data processing on the application server, the following can be obtained: socio-economic indicators for entering them into the database, tabular and graphical representations, predicted indicators, data on the correlation of indicators, files to export.
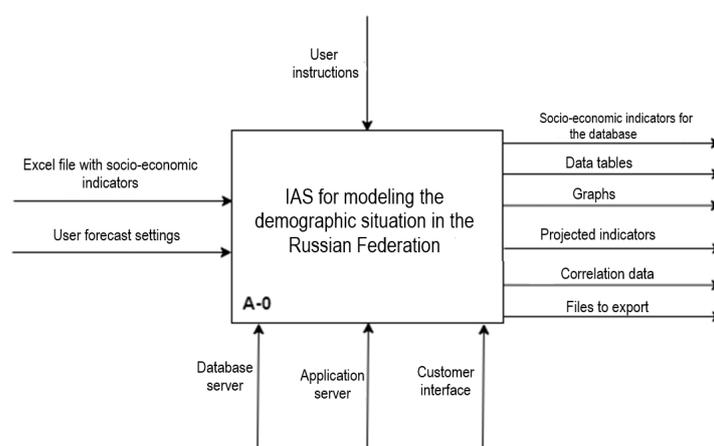

Figure 2. Functional diagram (level of detail A-0)

When detailing the functional scheme of level A-0, it is divided into the following functional blocks (Figure 3): the module for collecting analytical data, the module for working with the database, the module for detecting correlation, the forecasting module and the visualization module.

Each of the modules interacts with the database module to get the stored information or update it. The analytical data collection (A1) and visualization (A5) modules require direct interaction with the user, to load new indicators or to get settings for the graphical representation of information. The database management module (A 2) sends automatically generated SQL queries to the database server and returns the result of processing these queries to the customer. The correlation detection (A3) and prediction (A4) modules run on the application server and are used to implement the prediction algorithm (Figure 1).

2.4.   Mathematical model of the database

The mathematical model of the database is represented by the following entities: measurements of indicators in the subject for a certain year, subjects of the Russian Federation, indicators, a group of indicators, data on correlation, predicted values of indicators, system users.

The most important information for the forecast is the measurement of indicators by subjects and years (Measurement):

$$Measurement = \{SUB, I, Y, V\} \tag{1}$$

where SUB is the subject of the Russian Federation ($SUB \in Subject$), I is the indicator ($I \in Indicator$), Y is the year in which the measurement was made, V is the value of the indicator.

To represent measurement data, you need reference books that contain information about subjects (*Subject*) and indicators (*Indicator*)

$$Indicator = \{T, S, B\}, \tag{2}$$

where T is the name of the section, S is the section to which the indicator ($S \in Section$) belongs, and B is the parent indicator ($B \in Indicator \cup \emptyset$).
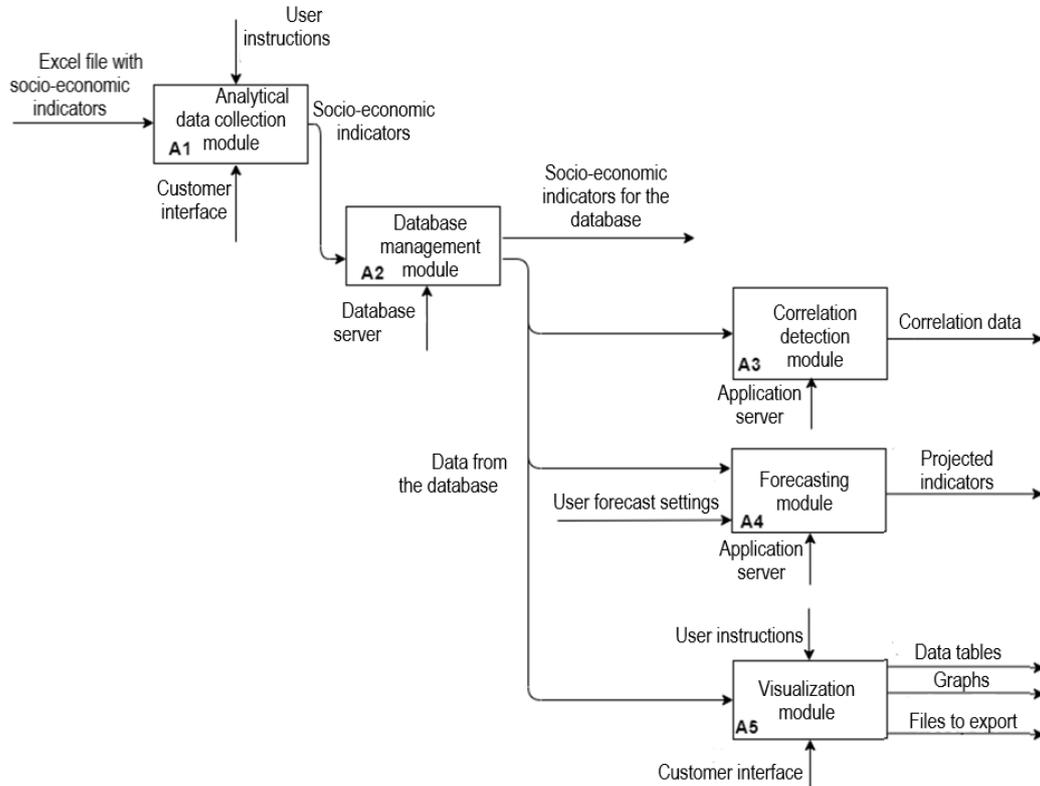

Figure 3. Functional diagram (decomposition of block A-0)

Each of the reference books can have a hierarchical structure (since some subjects can be components of others, and also with indicators).

In turn, the measurements relate to their socio-economic area, which will be stored in a separate reference book (*Section*).

Since the correlation values of indicators are often used in the construction of the forecast, it is advisable to store these values in the database (CorrelationValue), so as not to recalculate them at each step of the recursive algorithm (Figure 1).

Since the access to the data obtained as a result of the forecast will occur much more often than updating this data, it is necessary to save this data in the database (Prediction). The structure of the predicted indicators is similar to the structure of the measurements (Measurement), since it will store a similar set of fields.

Since the system must implement a role-based access model, it is necessary to store information about users (User).

2.5. Development of customer-server architecture

To provide access to the system from various devices, it is necessary to develop it according to the customer-server architecture. In this regard, it is advisable to use the classic three-layer architecture of the customer-server application, which includes: the customer application, the application server and the database server [6].

For the system under development, the designated components will carry the following functions:

- Customer – receiving data from the application server, presenting data to the user, configuring the view, receiving data from the user, sending queries to the server;
- Application server – receiving data and queries from the customer, sending data to the customer, sending SQL queries to the database server and receiving responses from it, business logic of data processing (calculation of the forecast, calculation of correlation, formation of an SQL query based on the data received from the customer);
- Database server – receiving a query from the application server, working with the database, sending a response to the application server.

The system provides two types of customer interfaces: desktop customer and mobile customer.

The server part is the most complex and voluminous part of the system, since all the business logic is concentrated here, except for converting data into a view that is suitable for presentation, which is implemented separately on each of the customer interfaces.

The application server is implemented as a RESTful service.

3. Research results and discussion

The result of the study was an information and analytical system for modeling the demographic situation in the Russian Federation.

The developed IAS allows you to make a forecast of socio-economic indicators for various subjects of the Russian Federation.

For example, in the graph (Figure 4), you can see the population growth, both in working age and younger / older working age. This increase occurs against the background of a sharp decrease in the migration growth of the population, which allows us to conclude that the number of indigenous people is increasing.
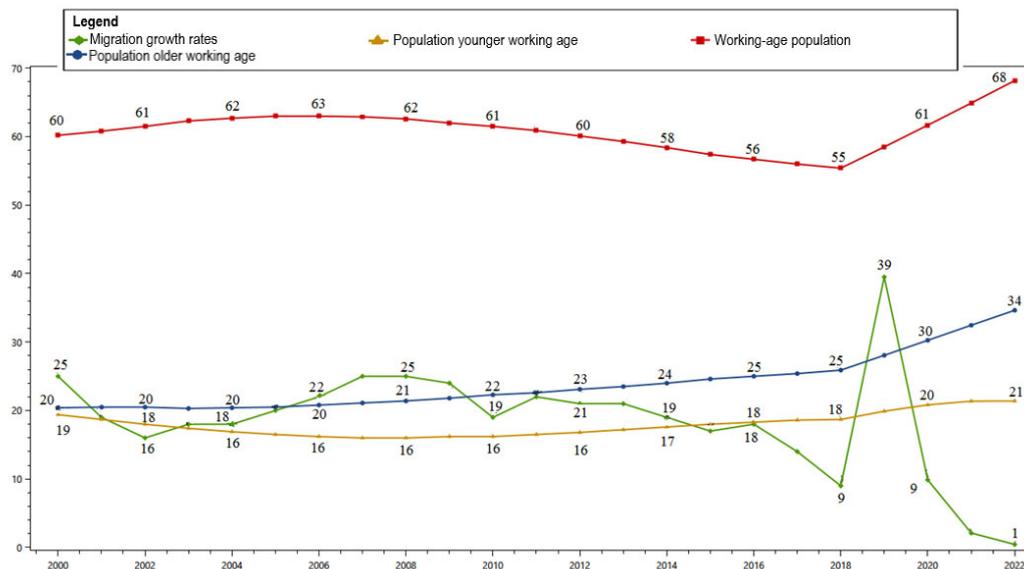

Figure 4. Age composition of the population and migration rate

At the same time, we can see a sharp increase in the proportion of the urban population (Figure 5), which may indicate that more and more people prefer to live in cities.
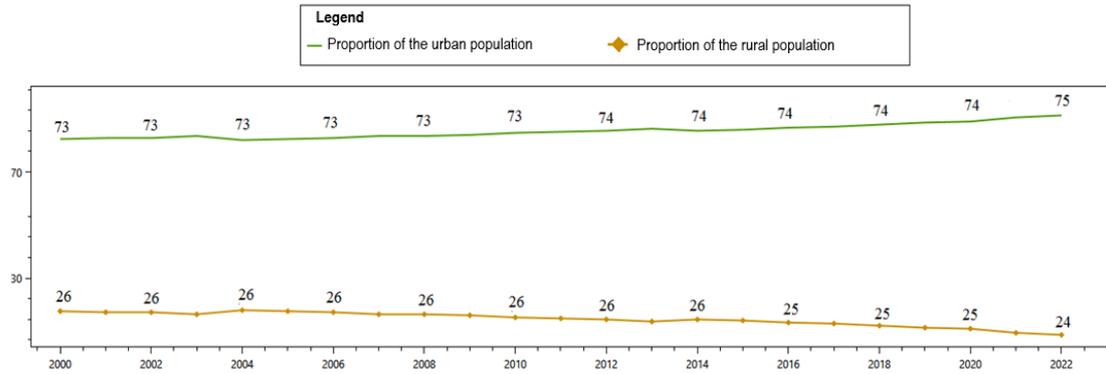
138

Figure 5. Proportions of urban and rural population

The disadvantages of the developed forecasting algorithm include the fact that some of the predicted indicators, such as the total population (Figure 6), show a clearly overestimated value relative to the overall trend.
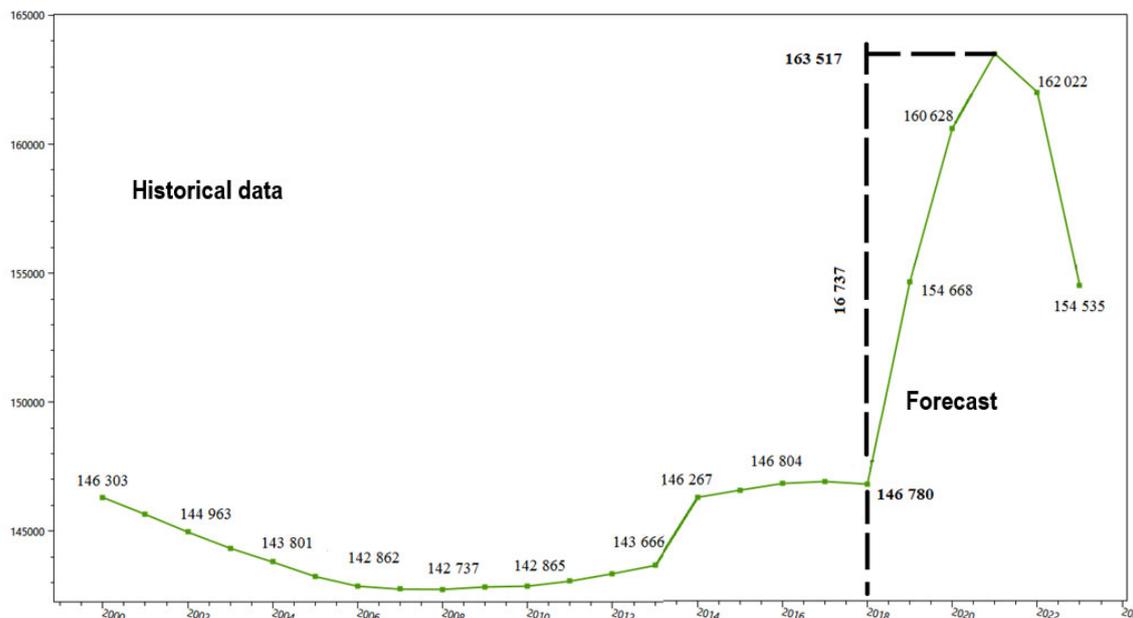

Figure 6. Total population

One of the advantages of the developed algorithm is that it is not tied to specific data, but only to the structure. This means that it is quite easy to enter new socio-economic indicators into the system and calculate the forecast based on a larger set of source data.

4. Conclusion

As a result of the study, the authors conducted a review of programs related to the processing of demographic statistics, described the materials and methods of the study, summarized the results of the study, and developed an algorithm for predicting demographic and socio-economic indicators in the Russian Federation. The prediction algorithm is able to work with any data that is suitable in structure, which allows you to use it to predict other data.

To model the demographic situation in the Russian Federation, an information and analytical system (IAS) was developed. The system has a customer-server architecture, which made it possible to bring information about statistics to the database server, and most of the logic to the application server, to

139

provide easy access to data using various interfaces. Two system access interfaces were also implemented: for PC and mobile devices.

Further development of the system can contribute to the creation of new models that allow us to make more reliable forecasts for various indicators, including those that go beyond demographic statistics.

5. References

[1] Course on demography and population statistic, Novosibirsk: Siberian University Publishing House, Normatika, p. 185, 2016. ISBN 978-5-379-01880-1. Text: electronic. Electronic library system IPR BOOKS: [website]. URL: http://www.iprbookshop.ru/65171.html. (accessed: 31.01.2021). Access mode: for authorized users.

[2] Yu. N. Solovarova, Demography: educational and methodical manual, Kazan: Kazan National Research Technological University, p. 108, 2019. ISBN 978-5-7882-2578-4. Text: electronic. Electronic library system IPR BOOKS: [website]. URL: http://www.iprbookshop.ru/100527.html. (accessed: 31.01.2021). Access mode: for authorized users.

[3] I. A. Chubukova, Data Mining, Electronic resource, Electronic text data, M.: Internet University of Information Technologies (INTUIT), p. 470, 2016. Access mode: http://www.iprbookshop.ru/56315.html. - ELS "IPRbooks".

[4] D. V. Aleksandrov, Modeling and analysis of business processes, Electronic resource: textbook, Electronic text data, Saratov: IPR Media, p. 226, 2017. Access mode: http://www.iprbookshop.ru/61086.html.

[5] V. I. Brezgin, Modeling of business processes with AllFusion Process Modeler 4.1. Part 1 [Electronic resource]: workbook, Electronic text data, Yekaterinburg: Ural Federal University, p. 80, 2015. Access mode: http://www.iprbookshop.ru/66174.html.

[6] T.V. Alekseeva, et al., Information analytical systems, [Electronic resource]: textbook, Electronic text data. Moscow: Moscow Financial and Industrial University "Synergy", p. 384, 2013. Access mode: http://www.iprbookshop.ru/17015.html. - ELS "IPRbooks".

[7] V. S. Belov, Information and analytical systems. Fundamentals of design and application [Electronic resource]: textbook, Electronic text data, Moscow: Eurasian Open Institute, p. 112, 2010. Access mode: http://www.iprbookshop.ru/10678.html. - ELS " IPRbooks»

[8] O. E. Baklanova, Information systems, Electronic resource: textbook, Electronic text data, M.: Eurasian Open Institute, p. 290, 2008. Access mode: http://www.iprbookshop.ru/10682.html. - ELS "IPRbooks".

[9] A. V. Boychenko, et al. Fundamentals of open information system, Electronic text data, M.: Eurasian Open Institute, Moscow State University of Economics, Statistics and Informatics, p. 160, 2004. Access mode: http://www.iprbookshop.ru/ 11043.html. - ELS "IPRbooks".

[10] M. S. Gasparian, et al. Information systems and technologies, Electronic text data, M.: Eurasian Open Institute, p. 370, 2011. Access mode: http://www.iprbookshop.ru/10680.html. - ELS " IPRbooks».

[11] A. A. Kuz'menko, D. Ye. Kondrashin, Metody i podkhody k razrabotke sistemy avtomatizirovannogo analiza dinamiki izmeneniya ploshchadi lesnykh nasazhdeniy na osnove metodov avtomaticheskogo raspoznavaniya obrazov, Bryansk, Ergodizayn 6 (2019) 230-240.

[12] V. M. Kozhukhar, V. I. Averchenkov, A. G. Podvesovskiy, A. S. Sazonova, Monitoring i prognozirovaniye regional'noy potrebnosti v spetsialistakh vysshey nauchnoy kvalifikatsii: monografiya, Bryanskiy gosudarstvennyy tekhnicheskiy universitet, 2010.

[13] A. S. Sazonova, L. Bb. Filippova, R. A. Filippov, Teoriya informatsionnykh protsessov i system, Bryanskiy gosudarstvennyy tekhnicheskiy universitet, 2016.

[14] E. A. Leonov, Yu. A. Leonov, Yu. M. Kazakov, L. B. Filippova, Intellectual subsystems for collecting information from the internet to create knowledge bases for self-learning systems, Proceedings of the Second International Scientific Conference "Intelligent Information Technologies for Industry" (IITI'17). IITI 2017, Advances in Intelligent Systems and Computing, 679 (2017) 95-103, Springer, Cham.