The Conception of an Intelligent System Formation with the Choice of Methods and Tools for Data Analysis for Information Processing

A.V. Koskin, A.A. Mitin, A.V. Artemov

Abstract— The article describes the conception of an intellectual information and analytical system for supporting decision-making. Here is presented the structure of such system, its constituent components and their functional capabilities. In this work we have also formulated the requirements for intelligent information and analytical system.

Keywords: information-analytical system, intellectual system, data analysis, data processing.

I. INTRODUCTION

Nowadays information flows which need to be efficiently processed, analyzed and interpreted in order to obtain high-quality and reliable information have dramatically increased in connection with the increase in the volume of data, the development and improvement of technologies for their storage and processing and the development of the "big data" technology. For these reasons the issues of building information and analytical systems are of particular relevance in the field of information technology. Information-analytical systems (IAS) as a class of information systems are focused on solving decision-making problems on the basis of heterogeneous data stored in various sources. IAS is a modern high-performance tool for making tactical, strategic and operational decisions based on providing all necessary data set to users responsible for making decisions.

The analysis requires access to a large amount of data and the purpose of the analysis is usually to identify certain trends: first, historical data is analyzed and then the identified trend is extrapolated for a future period of time.

The main requirements for IAS are:

1. Access to large amounts of data stored in various sources.
2. Interaction with external data processing systems and data sources.
3. Availability of means for checking consistency and completeness of data.
4. Availability of viewers, data interpretation and render features so as report generation ones.

One of the strategic directions for the development of information and analytical systems is their intellectualization which consists in building formalized data processing procedures and choosing a method for their analyzing, interpreting and presenting information in the form of some indicators and their further use directly to make decisions in the target area, i.e. selecting one of the alternative user options.

Thus, an intelligent information-analytical system (IIAS) is an information-analytical system that is based on the concept of choosing data processing procedures and methods, tools and technologies for analyzing data depending on the qualitative content and structured data in the decision-making process.

Intelligent data analysis is more often implemented by autonomous software systems due to the complexity of the tasks that are to solve. At the same time, OLAP-systems partially perform the most common and easily implemented data mining functions. Intelligent information-analytical systems are a kind of add-on for traditional data processing systems (DPS). They use the data provided by DPS as well as additional information and data for analysis which is conducted with the use of own applications and the results of which may actually be knowledge, i.e. structured information containing estimates of the relationship between the parameters of the control object description.

In a generalized form the structure of the intellectual information-analytical system is presented in Figure 1.

Revised Manuscript Received on April 12, 2019.

A.V. KOSKIN, Federal State Budgetary Educational Institution of Higher Education «Orel State University named after I.S. Turgenev», Doctor of Technical Sciences, Professor at the Department of the Information Systems. tel.: +79065692020 (E-mail: koskin@ostu.ru)

A.A. MITIN, A.V. Federal State Budgetary Educational Institution of Higher Education «Orel State University named after I.S. Turgenev», Candidate of Technical Sciences, Associate Professor at the Department of Information Systems. tel.: +79065692020 (E-mail: mcc77@yandex.ru)

ARTEMOV, Federal State Budgetary Educational Institution of Higher Education «Orel State University named after I.S. Turgenev», Candidate of Technical Sciences, Associate Professor at the Department of Software Engineering. tel.: +79065692020 (E-mail: avladar1984@yandex.ru)

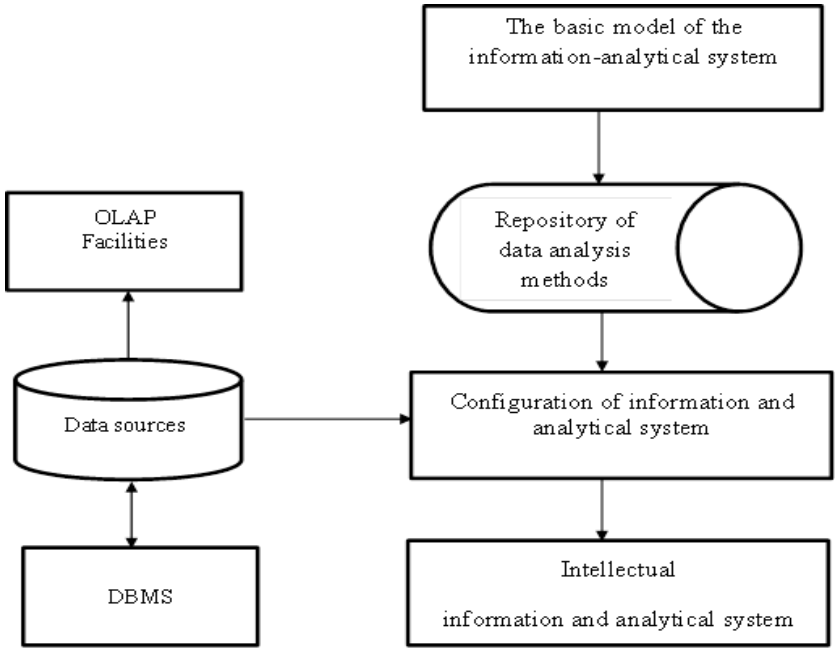

Figure 1 – The structure of the intellectual information-analytical system

The basic model of the information-analytical system includes:

1. Means of importing and downloading data from operating databases and other information sources that interact with various operating systems and DBMS.
2. Data conversion tools that implement validation check, transformation of structures and aggregation.
3. A set of software tools that perform operational functions of operational analysis.
4. Means of graphical and visual reporting intended for an end user.
5. Means of remote access, multiple access, ensuring work in distributed mode.
6. Object and process modeling tools.
7. Administration tools.

ANAYALATICAL RESULTS & DISCUSSIONS

The repository of data analysis methods should include a set of models, methods and technologies that can be used to obtain some of the patterns for data stored in data sources. At the same time, the models can be in the form of mathematical models or can be expert systems based on a logical model of the subject domain implemented in the form of a knowledge base and inference mechanism. The following data analysis methods can be represented in the repository: association, classification, clustering, forecasting, the method of successive models, neural network methods and technologies, decision tree methods, case-based methods and a number of other statistical methods [1].

Data sources are structured and unstructured data sets that include descriptions or characteristics of some objects or processes. In general, the data needed to support decision-making can be divided into the following types:

1) primary data that characterize quantitatively the state of the control object (process);
2) the results of the primary data processing which is performed according to the algorithms adopted in a particular system on the basis of its functional purpose representing aggregated (averaged) data;
3) generalized indicators of the object functioning for a certain period of time characterizing the effectiveness of its target functioning [2].

In general, the task that needs to be solved in an intelligent information-analytical system can be formulated as the data conversion like in 1), 2) and 3) into aggregate indicators of the control object properties which cannot be directly measured. Then these indicators can be interpreted as some quality of the control object as a whole or its certain properties. [2].

Typical tasks of OLAP tools are:

1. Providing the completeness and validity of stored data - “data cleansing” including consistency check.
2. Providing access to complex multidimensional data in any given section (the formation of complex queries).
3. Providing the display of complex multidimensional data in a readable form [3].

The configuration of the information-analytical system is formed by selecting methods and analysis technologies from the repository. Then the data are processed and analyzed and the results are interpreted in accordance with the procedures of the selected method. Next the results are visualized and presented to the user.

Primary data mining tasks include:

- classification (recognition) of data and situations;
- data clustering, i.e. splitting data arrays into groups (clusters) according to the proximity of feature value structures;
- construction of generalized indicators of efficiency and quality;
- forecasting [3].

The use of existing methods for processing multidimensional data revealed a number of shortcomings in their practical use. One of the most important among them is the practical impossibility of

taking into account the uncertainties in the source data in processing them, assessing the effectiveness of the system and predicting the development of objects without making assumptions which verification is simply impossible or expensive.

In the framework of existing approaches the removal of uncertainty is achieved by extending the axiomatics of probability theory to the formalization of the life processes of objects. Subjectivity in the use of a probabilistic measure leads to discrepancies obtained in the models of results and experimental data. The analysis of methods for accounting the uncertainty in the construction of generalized characteristics of control objects has shown that the use of the fundamental principle of entropy maximum allows to get a number of important advantages. At the same time there are no formal and computational difficulties in solving problems of identifying the states of objects described by multidimensional data for two reasons:

1. entropy being a measure of uncertainty has such property that at the same time its maximum value is a measure of proximity having the basic properties of a metric.
2. the obtained formal representations of entropy allow us to develop simple and effective computational algorithms for estimating generalized characteristics of objects. [2].

Thus in order to perform decision-making support functions the IIAS should provide:

- preliminary analytical processing aimed at obtaining aggregates of the process under consideration;
- intellectual analytical processing which consists in the choice of a method and technology of analysis for data depending on their quality content and features and the construction of integral indicators of the process under consideration for any time period;
- presentation of data and results in any given section in graphical form.

REFERENCES

1. Mitin A.A. Methods and tools for data mining. - Information systems and technologies № 1 (105) 2018. - Orel: Orel State University named after I.S. Turgenev, 2018. - pp. 34-38.
2. Ananyev N.S. Methods and tools for data analysis in the systems of decision-making support : Thesis for the degree of candidate of technical sciences: 05.25.05: Moscow, 2005. - 157 pp. Russian State Library, Thesis Department, 61: 05-5 / 3369.
3. Belov V.S. Information and analytical systems. Basics of design and application. Tutorial. - M: Moscow State University of Economics, Statistics and Informatics, 2004.
4. Filippov V.A. Data Mining: Methods and Tools [Text] / M .: Editorial URSS. - 2001. - 52 p.
5. Osipov G.S. Methods of artificial intelligence [Text] / Fizmatlit, 2011. - 296 p.
6. Volkov V.N., Stychuk A.A., Mitin A.A. The development of the information model for the presentation of data on e-service regulations. - Information systems and technologies № 4 (84) 2014. - Orel: Orel State University named after I.S. Turgenev, 2014. - pp. 21-30.
7. Eltsova N.S., Eremenko V.T., Kos'kin A.V., Mishin D.S., Mishin Ya.D. Interpretation of models of processes for obtaining and processing information in the portals of the executive authorities on the basis of operator schemes. - Information systems and technologies No. 4 (96) 2016. - Orel: Orel State University named after I.S. Turgenev, 2016 - pp. 42-50.
8. Koskin A.V., Uzharinsky A.Yu. Data access mechanisms based on a single integrating data scheme. - Information systems and technologies № 1 (87) 2015. - Orel: Orel State University named after I.S. Turgenev, 2015. - pp. 38-48.
9. Koskin A.V., Uzharinsky A.Yu. Methods of forming an integrating data model based on the available heterogeneous data sources. - Information systems and technologies № 2 (82) 2014. - Orel: Orel State University named after I.S. Turgenev, 2014 - pp. 19-27.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication