Indira Uvalieva[1]**, Saule Smailova**[2]**, Fatma Tarifa**[3]

DEVELOPMENT OF INFORMATION ANALYSIS SOFTWARE FOR THE MONITORING OF DISTRIBUTED OBJECTS WITHIN SOCIOECONOMIC SYSTEMS

This paper describes software for monitoring of distributed objects based on a multi-level and multi-dimensional model. The developed monitoring system allows for different points of view, criteria and management objectives which will improve the efficiency of distributed objects.

Keywords: *process monitoring; rapid data analysis; monitoring software; data mining; distributed objects.*

Індіра Увалієва, Саулє Смаілова, Фатьма Таріфа

РОЗВИТОК ІНФОРМАЦІЙНО-АНАЛІТИЧНОГО ІНСТРУМЕНТАРІЮ МОНІТОРИНГУ РОЗПОДІЛЕНИХ ОБ’ЄКТІВ У МЕЖАХ СОЦІАЛЬНО-ЕКОНОМІЧНИХ СИСТЕМ

У статті описано інструментарій моніторингу розподілених об’єктів на основі багаторівневої та багатоаспектної моделі. Розроблена система моніторингу враховує різні точки зору, критерії, цілі управління, що дозволить підвищити ефективність діяльності розподілених об’єктів.

Ключові слова: *моніторинг процесів; оперативний аналіз даних; програмне забезпечення моніторингу; інтелектуальний аналіз даних; розподілені об’єкти.*

Форм. *7.* ***Рис.*** *6.* ***Літ.*** *13.*

Индира Увалиева, Сауле Смаилова, Фатьма Тарифа

РАЗВИТИЕ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОГО ИНСТРУМЕНТАРИЯ ДЛЯ МОНИТОРИНГА РАСПРЕДЕЛЕННЫХ ОБЪЕКТОВ В РАМКАХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ СИСТЕМ

В статье описан инструментаций мониторинга распределенных объектов на основе многоуровневой и многоаспектной модели. Разработанная система мониторинга учитывает различные точки зрения, критерии, цели управления, что позволит повысить эффективность деятельности распределенных объектов.

Ключевые слова: *мониторинг процессов; оперативный анализ данных; программное обеспечение мониторинга; интеллектуальный анализ данных; распределенные объекты.*

Introduction. Decision-making management in various areas of human activity is constantly reaching new levels of complexity. This is due, primarily, to continuous growth of volume and dynamics of information flows that require careful analysis on the part of decision makers (DM). With rapid growth of information dynamics and structural complexity of contemporary society, the development of common principles, models and methods for monitoring of distributed objects basing on geographical location factors is an extremely urgent task.

Effectiveness of solutions directly depends on the level and quality of automation of key processes associated with observation, assessment, monitoring and analysis within the boundaries of subject area. In order to achieve this, 3 key issues must be addressed:

[1] Serikbaev East Kazakhstan State Technical University, Kazakhstan.
[2] Serikbaev East Kazakhstan State Technical University, Kazakhstan.
[3] Universite Libre de Tunis, Tunisia.

- the periodic problem of gathering raw statistical data describing the subject area, both in terms of time and in different structural aspects;
- the problem of developing logical and physical structures for computer representation of statistical information;
- the problem of developing software and hardware for processing statistical information and implementation of various types of control algorithms, simulation, optimization etc.

The object of this research is heterogeneous branching statistics of distributed objects. The subject of our research relates to the models and methods for monitoring processes of territorially distributed objects using multivariate databases of statistical data and intelligent technologies.

The aim of this study is to improve the effectiveness of territorially distributed objects by creating data analytical monitoring systems.

The following tasks have been set in order to achieve this goal:
- to develop a concept for monitoring of territorially distributed objects;
- to develop a multi-layered objects monitoring model;
- to develop a method for automated rapid collection and presentation of multidimensional data for monitoring objects;
- to develop a monitoring methodology based on the integration of intelligent technologies;
- to outline the functional support and structure of information system that monitors distributed objects.

In order to conduct this study the following methods have been used: the control theory (Novikov, 2007), the graph theory (Diestel, 2005), geographic information technology (Bugayevskiy, 2000), online analytical processing (OLAP) (Barsegyan, 2007), data mining and object-oriented design (Han, 2012).

Description of the existing monitoring approaches. The main functional objective of the majority of known monitoring systems – integrated support for regulatory bodies with complete, relevant and reliable information on the processes occurring in the subject area.

As the result of the general survey, the following common problems and shortcomings of the existing systems can be highlighted (Bozhday, 2009):
- low level of system integration during the analysis of processes of various nature;
- the lack of integrated monitoring models that take into account inter-branch collaboration;
- the lack of integration technology of various information-based approaches to monitoring and spatial analysis within a single system.

The concept and model for monitoring of territorially distributed objects. Territorially distributed objects (TDO) are to be understood as an array of elements of the socioeconomic system; system integrity within the framework of spatial and temporal scales. Since any socioeconomic system of regional (republican) scale is open and includes objects of different spatial and thematic, the following components of monitoring process are highlighted: level; subjects; informational space; territory (region).

To describe the forms of functions of territorially distributed objects monitoring different levels of monitoring are identified. Aspects of monitoring include: territorial, informational and organizational.

The theme of monitoring refers to different public and productive socioeconomic sectors. For example, public health, economics, demography, education etc.

Information space of a monitoring system consists of the flows of information on various subjects and aspects. 3 groups of information flows were created in order to implement the monitoring system: internal, external and control. These 3 groups form the information space of the monitoring system within the confines of this project.

Since monitoring is interdisciplinary and multifaceted, it was decided to consider geographical factors of monitored objects. Geographical location determines the structural boundaries of monitored objects. Thus, monitoring is implemented for territorially distributed objects.

The following principles of territorially distributed objects monitoring are identified (Bozhday, 2009):

- topical invariance in the field of monitoring;
- openness and interdependence of monitoring aspects;
- continuity between monitoring and geographic location of objects.

In order to create a generic monitoring model 4 areas of monitoring are identified: system-wide; logical; spatial and physical. Let us describe these aspects briefly:

- the system-wide aspect allows consider the monitoring process from the perspective of informational and organizational relationships between objects;
- the logical aspect allows consider the process of monitoring in terms of data sharing;
- the spatial aspect allows considering the monitoring process taking into account spatial and geographic location;
- the physical aspect allows considering the monitoring process taking into account the telecommunication environment for data exchange during monitoring.

The 4 aspects of monitoring mentioned above allow the formation of a generalized model with 4 distinct levels: model on the abstract-mathematical level, model on the informational logic level, model on the informational space level, model of the physical layer. A diagram of a multilevel model is shown in Figure 1.

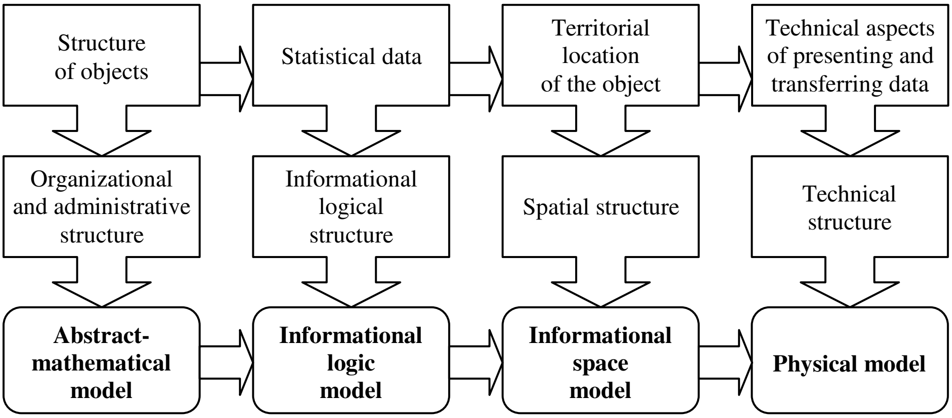

Figure 1. **Diagram of a multilevel model,** *authors' development*

The mathematical toolkit of the theory of graphs was used to construct a generalized multilevel monitoring model. The model is represented as a hypergraph structure the hyperedges of which are the result of the classification of monitoring objects according to specific tasks. A generalized multilevel monitoring model takes the following the form:

$$M = MM + ML + MI + MP, \quad (1)$$

where MM – the model on the abstract-mathematical level; ML – the model on the informational logic level; MI – the model on the informational space level; MP – the model of the physical layer.

The model on the abstract-mathematical level is represented as an MM hypergraph, consisting of two sets and a predicate:

$$MM = (V, U, P). \quad (2)$$

The set finishes the hypergraph structure on the vertex level:

$$V = \{v_i\}, i = 1, 2, .., N, \quad (3)$$

where N is the total number of peaks corresponding to the number of elements in monitoring.

Set U has variable power and describes a multilayer structure of a hypergraph at the level of hyperedges:

$$U = \{u_j\}_f, j = 1, 2, .., KS_f, f = 1, 2, .., F; \quad (4)$$

$$KS_f \le |u_{jf}|; \quad (5)$$

$$\mu_f = \frac{Z_{\max}(k_f) - Z_{\min}(k_f)}{KS_f}; \quad (6)$$

$$D_j = \left[\frac{Z_{\max}(k_f) + \mu_f \times (j-1)}{Z_{\max}(k_f) + \mu_f \times j} \right], \quad (7)$$

where V is the set of vertices of the hypergraph; N is the total number of vertices; U is the set of hyperedges; P is a binary predicate that determines the incidence of vertices and hyperedges of each layer; $|u_{jf}|$ – cardinality of the set of nodes incident hyperedge u_{jf}; KS_f – the coefficient of classification problems for each fth criterion of the selected; $Z_{\max}(k_f)$ and $Z_{\min}(k_f)$ – respectively, the maximum and the minimum of a set of values corresponding to the criterion k_f; μ_f – a step that determines the range of objects in the criteria set $Z(k_f)$, entering the hyperedges fth layer; D_j – the range of objects in the criteria set $Z(k_f)$, falling within the hyperedge u_{jf}.

The applications of hypergraph models for the creation of monitoring systems are described in (Avdeyeva, 2013). However, these models did not take into account the simultaneous multiplicity of possible perspectives on the classification of the relationship between participants and monitored objects (only one dynamically changing hyperedge layer was used, i.e. one perspective at a time). The proposed model has 3 major advantages:

- allows simultaneously analyse the relationship of the same structural elements in various dimensions and from different perspectives;
- retains the systemic unity of monitoring real objects;

- takes into account temporal dimension.

Application of fuzzy cognitive maps to monitoring of distributed objects. A cognitive map was constructed in order to formalize the monitoring system of distributed object as a set of concepts, display its systemic factors and identify cause-effect relationship between them. The effectiveness of using the existing methods of construction and application of cognitive maps as models for monitoring TDO is determined by the following factors (Bozhday, 2009):

- the ability to visualize the monitoring process;
- the lack of the necessity for concepts prespecification;
- constructability, visibility and relative ease of interpretation of causal relationships between the concepts;
- the ability to integrate the results of the analysis with appraisement methods.

An algorithm for constructing a cognitive map for the task of monitoring TDO is shown in Figure 2.

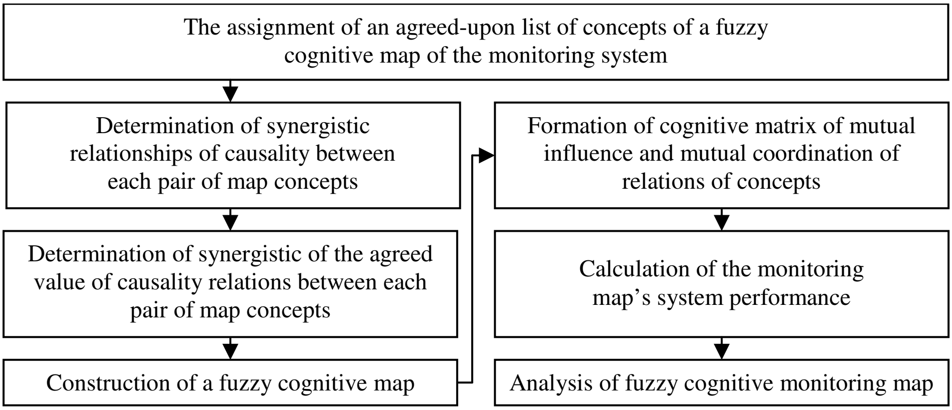

Figure 2. **An algorithm for constructing a cognitive map,** *authors' development*

The method of cataloguing and verification of data used for monitoring. The key issue in the functioning of any monitoring or information analytical system is the collection of relevant and reliable raw data. The practical value of this system will be primarily evaluated basing on these factors. The 5 repetitive key steps to automating the process of gathering raw statistical data is show in Figure 3.

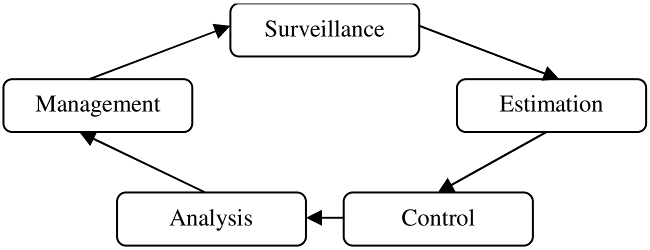

Figure 3. **Stages in data collection for monitoring,** *authors' development*

Automation of the cataloguing stage is carried out through the creation and management of a cataloguing meta-database (CMDB). The search automation of data sources are realised through the developed system of CMDB's request handling. The architectural concept of CMDB is based on the clearance of 3 key stages (Bozhday, 2009):

- identification of data sources, search and evaluation;
- formalized description and storage of all material processed at the previous stage;
- direct creation of CMDB and the development of a mechanism for addressing queries.

The stage of formalizing the requirements, the informational support of monitoring software must adhere to, involves a detailed analysis of goals and objectives of monitoring as well as the requirements for its thematic, technical, organizational, and economic components.

For practical implementation of the principles of integrated automation of stages of raw data collection for monitoring, the use of the methods of remote and online collection is encouraged. A diagram of organizational and technical cooperation for raw data collection is shown in Figure 4.

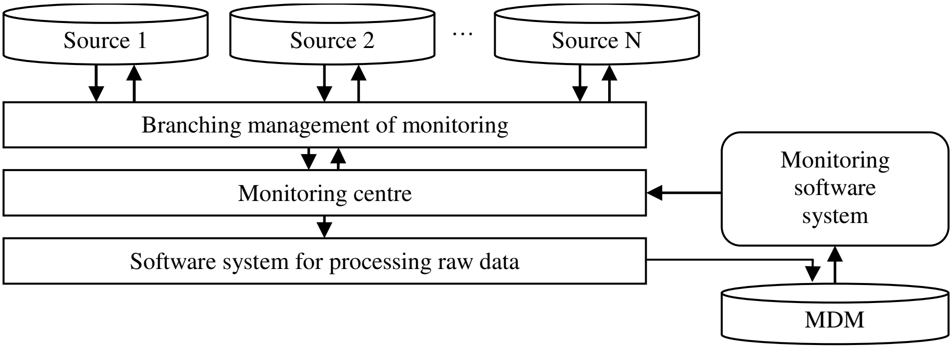

Figure 4. **The diagram of organizational and technical cooperation for raw data collection,** *authors' development*

The monitoring centre redirects to relevant management branch of data acquisition (or the web-resource address) which, in turn, distributes them to subordinate data sources. Streams of raw data are automatically generated basing on the results of the data collection module. This data is accumulated in the management branch in the form of a single archive and are redirected to the monitoring centre. The software system for processing raw data processes the received archives and integrates them into the unified structure of the multidimensional database (MDB). All further thematic sampling for informational support of the monitoring software system is made from the MDB.

Verification procedures must be implemented in order to verify the semantic and logical accuracy of data. Verification procedures are also required for automated and manual removal of inconsistencies, redundancies and incompleteness.

This study has identified the following 5 classes of invalid and/or erroneous data (Paklin, 2013):

- induced incorrect data – data errors inherited through the acquisition of third-party database (DB);

- logical factor — errors in data due to incompatibilities or improper integration of different logical database models that serve as sources of data for MDB;
- human factor — errors in data that occur due to errors of those responsible for providing raw information;
- technical factor — errors in data due to technical failures in storage, transmission, processing and data conversion;
- organizational factor — errors in data from the violation of organizational rules of raw data collection.

Technology verification and data cleansing involves several steps:

- data analysis — identifying types of errors and inconsistencies to be removed;
- specification of order and rules of data conversion;
- affirmation of conversion rules on the example of copies;
- data conversion, counter flow of cleaned data.

Using the discussed verification technology will optimize the raw data sets both in terms of volume and in terms of contents, a prerequisite for their inclusion into the concept of monitoring methods of rapid data analysis (OLAP) and data mining.

Method of monitoring territorially distributed objects on the basis of intelligent technologies. The following functions were included as the key ones in the TDO monitoring system:

- surveillance — a focused and regularly repeated process of collecting raw data as well as their presentation and storage within the selected logical model and a specially designed physical storage;
- estimation — ranking and classification of raw information using subjective criteria related to the subject area of monitoring and its objectives;
- control — the process of periodic inspection of system parameters, ensures timely achievement of the goals set;
- analysis — a wide range of scientific methods to extract information on the structure and behaviour of the subjects of monitoring on the basis of raw data;
- management — a set of functions designed to support decision-making and help decision-makers in complex multifactorial conditions for full and objective analysis of the subject area.

The main components of the analysis function within the context of a monitoring system are mathematical modelling (based on a selected mathematical model of object monitoring) (Orlovskiy, 1981), statistical analysis (based on the methods of mathematical statistics), rapid analysis based on OLAP technology (implementation of queries addressed to a multidimensional database, the construction of multicriteria slices of data and information samples, abstraction) (Barsegyan, 2007), data mining (the process of discovering in raw data sets previously unknown knowledge which is non-trivial, practically useful, open for interpretation and, most importantly, required for decision-making in various spheres of human activity) (Pyatetskiy-Shapiro, 2009), GIS spatial analysis.

An identified set of functions that describe the iterative steps of TDO monitoring techniques IA is shown in Figure 5.

This is what the arrangement of interactions within the framework of the monitoring method based on the integration of OLAP technology (Makarychev, 2010), data mining (Witten et al., 2011) and geographic information systems (GIS) appear as: the construction of interdisciplinary MDB subsystem is implemented by means of statistical and OLAP analysis. Once done, the mathematical modelling subsystem steps in and constructs a hypergraph model of a discrete mathematical level (based on the criteria defined by a decision-maker).

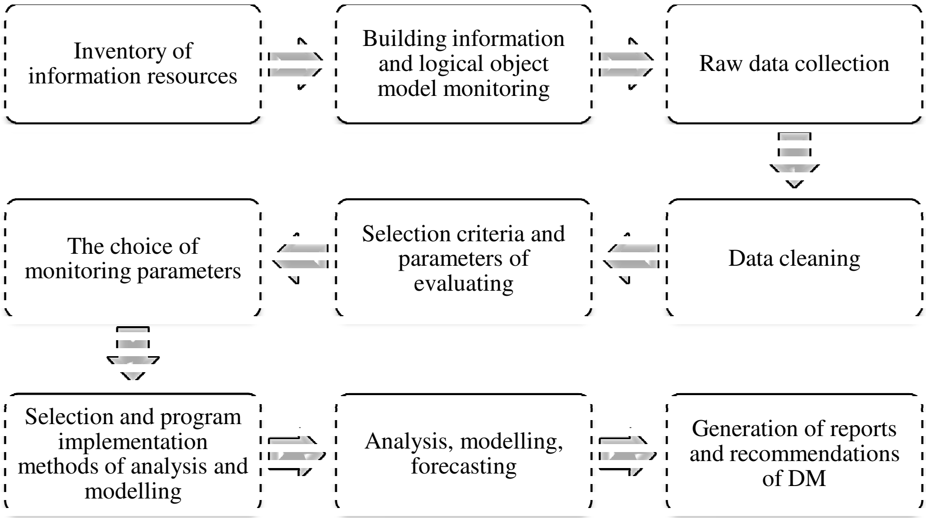

Figure 5. **Steps ib TDO monitoring techniques IA,** *authors' development*

The proposed scheme of interaction of the TDO monitoring method based on the integration of OLAP technology, data mining and GIS determines its structural implementation, features and scientific innovation. They are the invariance of the structure to the specifics of different sectors of monitoring or socioeconomic problems of management, they allow for the integration of advanced technologies for monitoring in a single system and they create the convenience of a territorial scaling of the system. As part of the structural implementation of this method structural schemes and algorithms of subsystems for raw data collection, mathematical modelling, operational, statistical and predictive analyses, spatial analysis have been developed.

The structure of the information system for monitoring of distributed objects. The proposed mathematical model of multilevel monitoring is implemented in the form of an information system which includes the following organizational and informational software subsystems:

- the inventory subsystem and the raw data collection subsystem implement the functions of observation;
- the statistical and OLAP-analysis subsystem is responsible for providing the functions of assessment, monitoring, statistical and operational analysis;
- the mathematical modelling subsystem, the basis of which is the abstract mathematical level of the mathematical model for monitoring. This subsystem's function is the analysis and modelling on the basis of the graph theory algorithms;
- the data mining subsystem provides the analysis of data in order to detect hidden patterns and correlations;

- the spatial analysis subsystem consists of digital cartographic area foundation and provides spatial analysis functions;
- the report generating subsystem implements the management function by providing decision-makers with monitoring reports of 3 types: statistical reports, text reports and cartographic reports;
- the multidimensional database (MDB) includes consolidated volume of statistics of all thematic layers of monitoring;
- the core provides the system integration of all these abovementioned subsystems.

The diagram of the structural implementation of a monitoring method based on the integration of OLAP technology, data mining and GIS is shown in Figure 6.

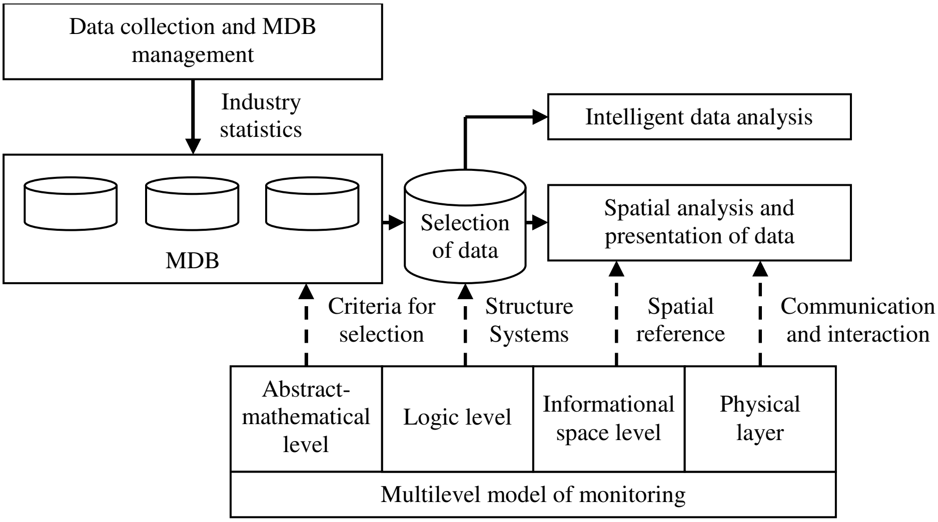

Figure 6. **Structural implementation of the monitoring method based on intelligent technologies,** *authors' development*

Functional System of Software for monitoring of distributed objects. The TBO monitoring information system (IS) is intended for observation, evaluation, monitoring, analysis and management of territorially distributed objects in order to improve the effectiveness of their functions.

The subsystem for raw data collection performs 4 basic tasks:
- generation of data collection forms;
- accumulation of raw data;
- cleansing and verification of data;
- data integration.

The functioning of the subsystem for mathematical modelling relies on the accomplishment of 3 key objectives:
- determine the quantity and composition of the selection criteria of multidimensional database;
- generate a hypergraph model using the selection criteria;
- perform the modelling procedure using the obtained hypergraph and a specific set of algorithms.

The main stages in the functioning of rapid statistical and data mining subsystems are:

- specification of selection criteria from the multidimensional database and generation of a hypergraph model of a discrete mathematical level;
- transition to an informational logical level of the model;
- formation of a multidimensional sub-cube of selection that includes data corresponding to the selected set of criteria;
- selection of methods for the analysis and development of appropriate procedures included in the functional arsenal of the subsystems.

Conclusion. As a result of theoretical and practical research, a multi-level, multifaceted model of monitoring has been proposed. This model allows for formalizing and taking into account real-time monitoring of the specifics of different perspectives, criteria and management objectives. This model has been implemented as a software solution — an information system that provides observation, evaluation, monitoring, analysis and management functions. A distinctive feature of the proposed approach is a clear algorithmic formalization of basic monitoring functions.

References:

Авдеева З.К., Коврига С.В. Диагностирование проблемных ситуаций в развитии сложных систем на основе когнитивных карт // Управление большими системами.– 2013.– Вып. 42. – С. 5–28.

Барсегян А.А., Куприянов М.С., Степаненко В.В. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. – 2-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2007. – 384 с.

Бождай А.С. Комплексная инфраструктура территории: методы и модели информационного мониторинга // Информационные технологии.– 2009.– №9. – С. 57–63.

Бождай А.С. Концепция мониторинга комплексной инфраструктуры территории на примере системы послевузовского образования // Известия Волгоградского государственного технического университета: Межвуз. сб. науч. ст.– Серия: Актуальные проблемы управления, вычислительной техники и информатики в технических системах.– 2009.– Вып. 7, Т. 12. – С. 96–98.

Бугаевский Л.М., Цветков В.Я. Геоинформационные системы. – М.: Златоуст, 2000. – 222 с.

Макарычев П.П., Афонин А.Ю. Оперативный и интеллектуальный анализ данных: Учеб. пособие. – Пенза: ПГУ, 2010. – 156 с.

Новиков Д.А. Теория управления организационными системами. – 2-е изд. – М.: Физматлит, 2007. – 584 с.

Орловский С.А. Принятие решений при нечеткой исходной информации. – М.: Наука, 1981. – 206 с.

Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных к знаниям. – 2-е изд. – СПб.: Питер, 2013. – 704 с.

Пятецкий-Шапиро Г. Data Mining и перегрузка информацией // Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод и др. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – С. 13–14.

Diestel, R. (2005). Graph Theory, Electronic Edition. NY: Springer-Verlag. 422 p.

Han, J., Kamber, M., Pei, J. (2012). Data mining: concepts and techniques. 3rd ed. Morgan Kaufmann/Elsevier. 744 p.

Witten, I.H., Frank, E., Hall, M.A. (2011). Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edition. Morgan Kaufmann. 664 p.

Стаття надійшла до редакції 17.12.2014.