Using Visualization Tools to Search for Contradictions in Planning Results

A.A. Zakharova[1,A], E.V. Vekhter[2,B], A.V. Shklyar[3,B]

[A] Bryansk State Technical University
[B] Tomsk Polytechnic University

[1] ORCID: 0000-0003-4221-7710, zaa@tu-bryansk.ru
[2] ORCID: 0000-0003-0604-0399, vehter@tpu.ru
[3] ORCID: 0000-0003-4442-7420, shklyarav@tpu.ru

Abstract

When designing most types of activities, as well as when needed to audit the results achieved, a problem arises to analyze simultaneously data included in the descriptions of individual elements of this activity. The purpose of such analysis may be to identify contradictions among disparate information descriptions of processes included in the general activity, as well as to search for possible ways of improving the results achieved.

The paper presents the results of a study of visualization potential in the analysis of large volumes of heterogeneous data. The results obtained have been used in the development of visual research tools which help to improve the design of complex processes and to solve the applied problem of planning the educational process.

Keywords: visual analytics, visual model, data analysis, visual interpretation, visual perception.

1. Introduction

During the design of various processes, including those associated with the presence of a large number of internal relationships, the volume of heterogeneous data is formed, i.e. a set of heterogeneous data characterizing the entire process under study and its individual elements (Fig. 1). The interpretation of such data is aimed at identifying implicit patterns and contradictions in heterogeneous descriptions of processes [1].

Fig. 1 Standard representation of educational process parameters

A large amount of input data and a limitation of available resources (computational, time, human) often become problems of practical analysis of heterogeneous data during designing and controlling results of various processes. The search for new analysis tools is conditioned by the need to analyze a large number of relationships among elements of the general process [2]. Traditional means have several disadvantages, which include high qualification requirements to experts involved in the analysis and the need for interdisciplinary interaction among them. Besides, to improve the achieved results it is necessary to eliminate internal contradictions in the studied data, the source of which is errors made at the process planning stage. The subjective nature of such contradictions complicates their detection based on pre-formulated criteria.

2. Interpretation and visualization

The amount of data and a number of additional information influence the data interpretation process. Some information is not directly related to the current problem. At the stage of preliminary analysis of the initial data or during the problem solving process the analyst gets a lot of additional information (intermediate solutions, verification data, classification and clustering results, etc.). In addition, the process of researching and interpreting data which requires significant time resources is associated with the need to study new data – replacing or complementing the old ones [3].

History of data acquisition, methods of their collection, the presence of errors, partial lack of data, etc. can have a significant impact on the research objective achievement. Metadata, i.e. second-order data, from the point of view of interpretation, play the role of an additional source of information for the researcher and are a part of heterogeneous data. Thus, the peculiarity of studying heterogeneous data at the user's disposal when solving a planning problem is the need to compare many heterogeneous descriptions of individual elements of the process, to take into account intermediate versions of solutions and to update interpreted data during analysis. As an advanced way to overcome these difficulties, it is proposed to use data visualization tools capable of organizing an interactive communication between a user and initial data [4].

Interpretation of heterogeneous data is a process of obtaining new information and it is characterized by the amount of resources needed for this. This condition is formed by the requirements for the data interpretation tools. Interactive visualization used to represent and interpret the source data makes it possible to use the potential of visual perception to compensate for these difficulties. An important feature of visual perception is the ability to compare several events (options), as well as the credibility of the received information, which is the basis for making necessary decisions.

Visual image of data can be examined in various ways and, therefore, can give the user answers to various questions. In the simplest case, visualization provides confirmation or refutation of the already formulated answer to the question and does not require additional understanding of the results interpretation. This allows us to highlight the illustrative role of visualization, which has a pronounced practical orientation, for example, demonstration of the results of solving analysis problems obtained by other means. This type of visualization is characterized by the development in the direction of automated model construction, as well as the search for visual representation forms reducing interpretation time.

The paper presents methods for more complex use of visual models based on the direct interaction between the researcher and the visualization tool. The opportunity to change the state of the visual model while it is used corresponds to the construction of a controlled data model and creates a wide range of advantages, which include the targeted use of the model state to monitor the control action results [5].

3. Formal statement of the problem

The analyzed process is an ordered set of elements $O=\{O_1..O_n\}$, where O_i is an elementary process defined by a set of unique properties [6]. Meanwhile, properties of the element O_i are divided into two subsets, based on their functional differences:

$$O_i = \{P_i\} = \{In_x, Out_y\},$$

where the subset $\{In_x\}$ is process requirements, so-called incoming connections of O_i element, and the subset $\{Out_y\}$ is results of the process, so-called outgoing connections of O_i element. Connections of O_i element determined in this way are its main characteristics which ensure interaction with other elements and an achievement of the goal of the main process.

In accordance with these definitions, the goal of the main process can be represented as a composition of outgoing connections, characterized by some parameters, for example, the number of outgoing connections (results) and some weighting factors. Thus, the potency of set $O=\{O_1..O_n\}$ is defined by the predetermined goal $G=\{Out\}$ and the available resources.

The formal goal of solving the planning problem is formulated as following: to find the set O that provides targeted or best results and meets all boundary requirements:

$$O=\{O_n : n = min, G=max\}.$$

In the simplest case, when properties of elements O_i are specified, the goal of the solution is to order the set O which ensures the fulfillment of the boundary requirements. In the opposite case, the solution to the problem is a new set $O^* = \{In^*, Out^*\}$.

User's abilities

Human reasoning abilities are diverse and insufficiently studied [7], so, the following are highlighted for data visualization:

- Adaptability, i.e. readiness to include new information in the system of existing concepts and their categories. A useful consequence of this property is the ability to perceive rapidly changing visual information, including observing scenes that are complex sets of objects and events, as well as identifying interactions, including new ones.
- Attention selectiveness, which excludes details from the perceptual data set that are not in the focus of attention. This optimizes activity in conditions of limited resources (for example, time).
- A two-stage perception mechanism, which, interacting with accumulated experience, creates a flexible process of the operational comprehension of a visual image.

The need to use perceptual abilities, characterized by empirical principles, sets to a developer a task of determination of the set of user characteristics that are purposefully involved in the data visual image interpretation [8]. Selection can be based on a preliminary assessment of available resources, which include:

- Characteristics of a potential user, whose involvement in the process of data interpretation does not create prerequisites for increasing its duration.
- Computing resources, which allow obtaining visual images of data, if they meet the requirement of interactivity with them.
- Time resources that determine speed of construction and interpretation of visual data images.
- Additional requirements arising from the statement of the research task, including: the user's prior knowledge, his qualifications, probable features of perception, etc.

The set of the user's characteristics involved in the cognitive interpretation of the visual image of the studied data can be determined by generalization of existing schemes of the visualization process [9]. Based on the obtained set, three groups of visualization tools are distinguished, which differ in functional purpose and in methods of practical implementation:

- Observation. It is obtaining visual information (perception of colors, space, movements, allocation of groups, recognition of forms, signs).
- Search. Identifying relevant objects and processes (spatial thinking, prior awareness, motivation) in the initial visual information.

- Formulation. Formulating hypothesis of the answer to the research question (experience of using visual analytics, ability to study and apply new language systems), formalization of new information.

Development of computer visualization technologies and their continuous complication create difficulties in interaction between a user and the developed visualization tools [10]. This circumstance becomes critical in a situation where visualization tools provide cooperative participation of a group of researchers or are a way of exchanging information between experts with different levels of training or area of specialization. There is a need to choose between training users to use new visualization tools or involving existing visual communication skills in interpreting data.

4. Visualization tools for the planning problem

Based on the formulated requirements for visualization tools, which are necessary to solve the planning problem, a system for interactive presentation of data included in the description of an arbitrary educational program has been developed. Three-dimensional visual model is proposed as a visualization tool designed to solve the problem. It forms a visual image of information objects included in the source data (Fig. 2). An information object is an element of an educational program (an academic discipline, a course, program section). Each such object is an array of data (name, course, duration, content, incoming requirements, planned results), including variables of different types.

Software has been developed that allows acquiring an interpretable visual image with the use of Autodesk 3ds Max package visualizers. The algorithm for constructing a visual image is implemented using Maxscript language; therefore, it is portable and can be easily adapted to new visualization technical capabilities. Visualization interface only partially uses Autodesk 3ds Max environment and can be adapted to a particular user's needs.
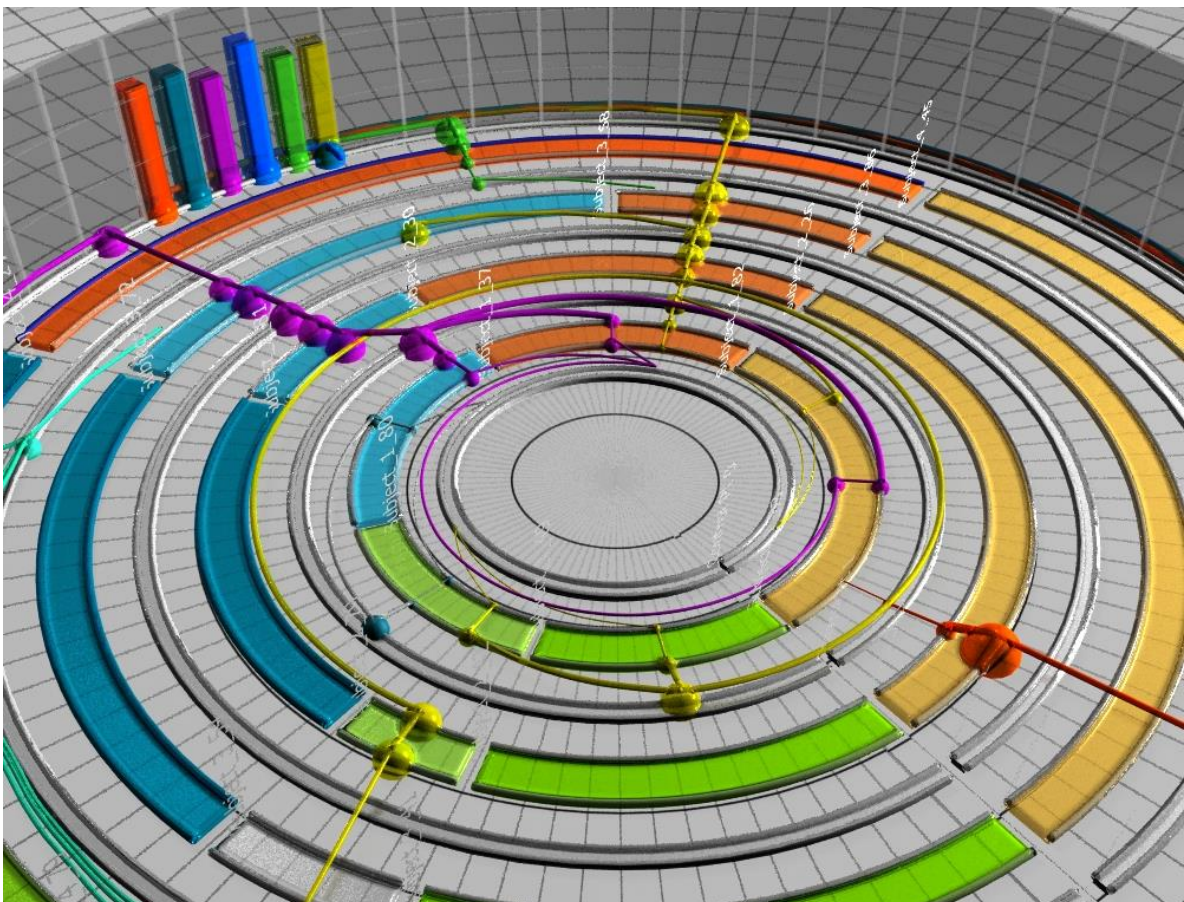

Fig. 2 Tool for interactive representation of data, included in the description of an arbitrary educational program.

Reasonable interactive management system creates conditions for setting new research questions and receiving answers quickly, accelerating achievement of the analysis goal. Consequently, interactive features of the visualization tool determine sequence and logic of a researcher's reasoning.

To reduce the training period required to become familiar with the new interpretation tool, it is proposed to use a representation metaphor based on traditional methods of visualizing tabular data (charts, graphs). A cylindrical coordinate system is defined in 3D space of the visualization tool, that allows each point in space to be matched with three values: training time, load, result. The scale of measurement units along time and result axes can be arbitrary; the load is measured as percent of the maximum possible. The opportunity to compare objects is realized through the use of color coding, which rules can be changed in accordance with perceptional characteristics of a particular user.

The radial direction of the time axis allows visualizing data of educational programs of any duration (bachelor's degree, specialist's program and master's degree). Program steps which correspond to the given time intervals (years, semesters) are divided by concentric circular elements used to represent the accumulated results. Each concentric element is a reference scale (0-100%) with a common starting point. The proposed structure provides presentation of an increasing number of learning results without worsening general perception of data.

The information object provides to an observer an opportunity to interpret visual attributes as values of the corresponding parameters: color is an identification attribute, dimensions are load ones, position corresponds to the training period. Information about incoming requirements and planned learning results is presented as links between information objects. Attributes of such connections are their direction and quantity, which correspond to the source data. In accordance with characteristics of the subject area, incoming (requirements) and outgoing (results) communications have the same type, i.e. they can be interpreted as created or developed competencies.

For visual representation of accumulated learning results, simultaneous use of two expressive means is proposed: a rating scale of effectiveness and visual scaling of links. In the first case, it becomes possible to efficiently use the three-dimensional space of the visual model, in the second one – the assessment of training results in the educational program can occur while interpreting two-dimensional visualization.
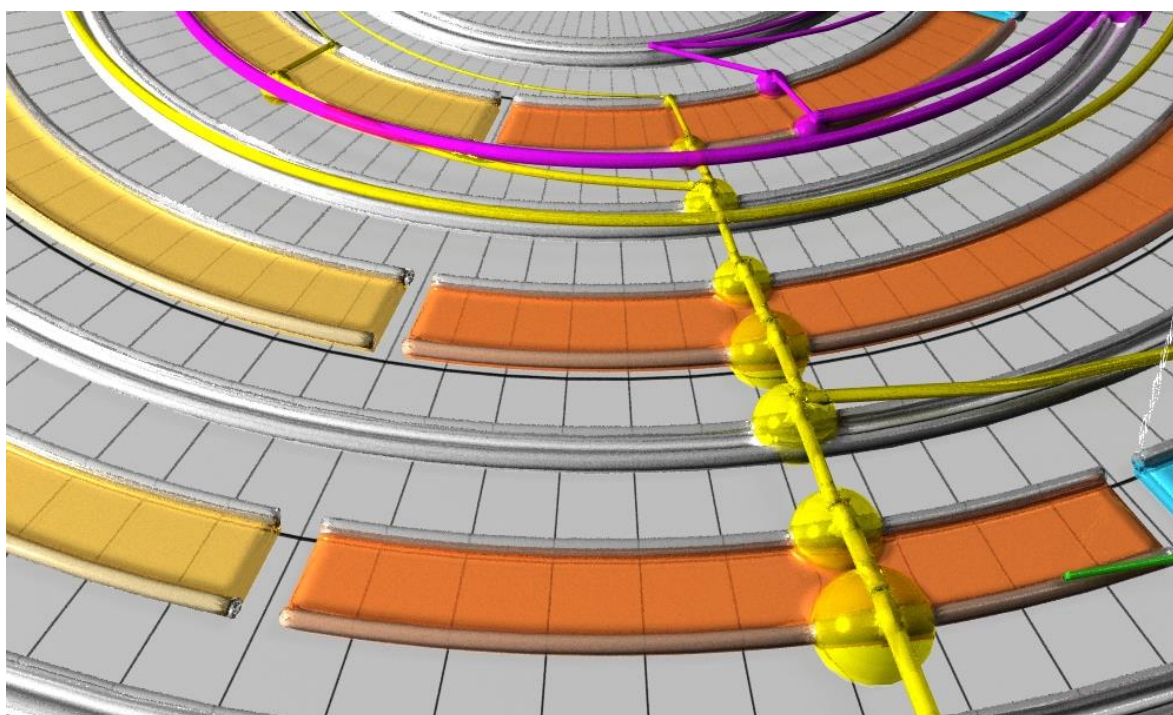

Fig. 3 Visualization of accumulated results.

A method for visualizing intermediate and general results in the form of result summing profiles was developed. A profile has a form of a histogram representing accumulated results as a percentage of planned values. To simplify the image of the studied data, elements of the result presentation can be temporarily excluded from the visual model.

Visual objects, which present to a user data of the achieved learning results, perform one of the main functions, which consists in searching for options of individual element characteristics satisfying the planning goal. Thus, the use of visual representation elements as a system of interactive control of the state of the visualization tool is proposed. The interface of interaction with the visualization tool created in this case becomes the basis for cognitive interpretation of the source data images taking into account individual characteristics of a user's thinking.

5. Contradictions in planning results

Dynamic correction of planned results by detection and elimination of contradictions made at the planning stage can be deemed a promising option for employing user interaction with the developed visualization tool.

The quickest search occurs for contradictions of several types:

- Chronological discrepancy between incoming requirements of the information object and the results achieved (Fig. 4). Contradiction arises if the input of the information object requires data that will be received later. It is visualized as a line of accumulated results going in the opposite direction.
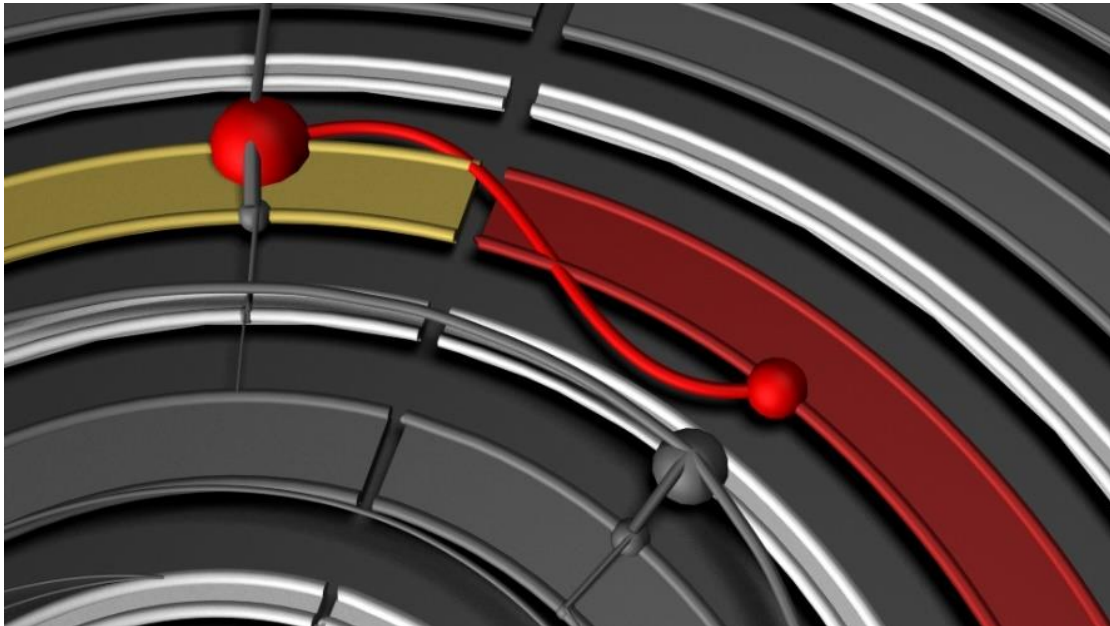

Fig. 4 Chronological discrepancy.

- Lack of input data: incoming requirements of the information object are not provided with results, i.e. the requested data are missing. It is detected as an open input of an information object (Fig. 5).
- Unreasonable use of existing resources. It is a result of interpretation of result lines that are not presented at the output of the overall process.
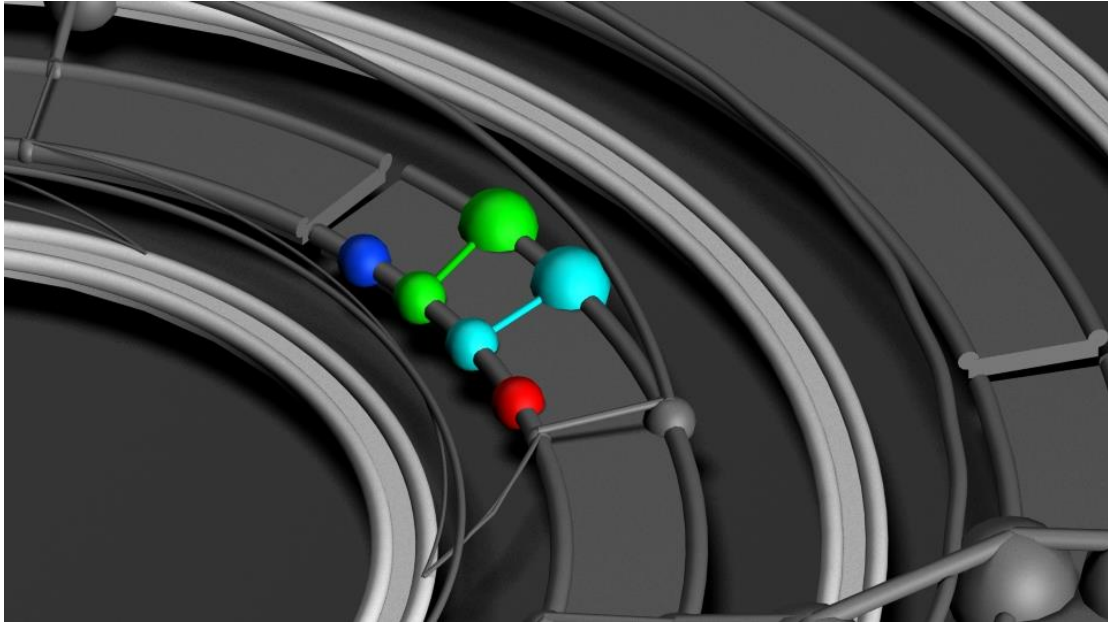
Fig. 5 Lack of input data

- Degradation of results. The proposed visualization tool takes into account possibility of reducing the level of the achieved result over the course of time. The results present in the final profile that do not take into account the level of their degradation are considered a contradiction.
- Duplication of results. Resources are used for repeatedly obtaining similar results. The visual image of such a contradiction is diverging lines of cumulative results.

Performance assessment

Application of the developed visualization tool has made it possible to obtain the assessment of advantages of the proposed approach in comparison with traditional methods of interpretation and verification of heterogeneous data contained in documents regulating educational programs.

Table 1. Comparison of duration of data research stages.

Comparison parameter	Traditional approach	Visual analysis
Training	Previous experience	Less than 5 min
Forming general idea	15-20 min	Less than 2 min
Searching contradictions	up to 30 min	1-5 min
Changing goal	Impossible	Interactive
Managing data	Impossible	10-20 min

The proposed tool for visual analytics of educational environment can be supplemented with experience maintainability when already formed versions of educational trajectories and corresponding real results are saved. Opportunities of advanced planning for periods of varying duration are expanded by analyzing factors that influence deviation of real values of achieving professional competencies from those planned in educational program design.

6. Conclusion

The use of visualization tools as means of operational data research in planning tasks is an example of visual analysis of heterogeneous data. The obtained advantages include high speed of visual perception, which is necessary for simultaneous comparison of large amount of disparate facts, as well as the opportunity to interpret data not only in numerical form.

The technique of using visualization tools allows achieving the research goal without interpreting numerical values. Thus, an approach to solving problems of this type solely based on

visualization tools is proposed. The visualization tool has a low dependence on specifics of the problem subject area. Thus, the experience of using the proposed means of interpretation can be involved in solving other problems. Low requirements for users' preliminary preparedness make it possible to use visualization tools in absence of special training for solving data interpretation problems which have an interdisciplinary nature.

References

1. Guo H. [et al.]. A Case Study Using Visualization Interaction Logs and Insight Metrics to Understand How Analysts Arrive at Insights // IEEE Transactions on Visualization and Computer Graphics, 2016, Vol. 22, Issue 1, pp. 51–60. doi: 10.1109/TVCG.2015.2467613
2. Zakharova A.A., Vekhter E.V., Shklyar A.V., Zavyalov D.A. Visual Detection of Internal Patterns in the Empirical Data // // A. Kravets et al. (Eds.): CIT&DS 2017, Communications in Computer and Information Science, Vol. 754, pp. 215-230. doi: 10.1007/978-3-319-65551-2_16
3. Manakov D., Averbukh V. Verification of Visualization // Scientific Visualization, 2016, Vol. 8, No. 1, pp. 58–94.
4. Blascheck T. [et al.]. VA2: A Visual Analytics Approach for Evaluating Visual Analytics Applications // IEEE Transactions on Visualization and Computer Graphics, 2016, Vol. 22, Issue 1, pp. 61–70. doi: 10.1109/TVCG.2015.2467871
5. Zakharova A.A., Vekhter E.V., Shklyar A.V. Methods of Solving Problems of Data Analysis Using Analytical Visual Models // Scientific Visualization, 2017, Vol. 9, No. 4, pp. 78–88. doi: 10.26583/sv.9.4.08
6. Zakharova A.A., Shklyar A.V., Rizen Y.S. Measurable Features of Visualization Tasks // Scientific Visualization, 2016, Vol. 8, No. 1, pp. 95–107.
7. Biederman I. Recognition-by-Components: A Theory of Human Image Understanding // Psychological Review, 1987, Vol. 94, No. 2, pp. 115–147.
8. Chen H. [et al.]. Uncertainty-Aware Multidimensional Ensemble Data Visualization and Exploration // IEEE Transactions on Visualization and Computer Graphics, 2015, Vol. 21, Issue 9, pp. 1072–1086. doi: 10.1109/TVCG.2015.2410278
9. Van Wijk J.J. The Value of Visualization // Proceedings of VIS 05. IEEE Visualization, 2005, doi: 10.1109/VISUAL.2005.1532781
10. Chen C. Top 10 Unsolved Information Visualization Problems // IEEE Computer Graphics and Applications, 2005, Vol. 25 Issue 4, pp. 12–16. doi: 10.1109/MCG.2005.91