SemaVis - A new Approach for Visualizing Semantic Information

K. Nazemi, M. Breyer, D. Burkhardt, C. Stab, J. Kohlhammer

Abstract Information is an indispensable resource today. The access and the interaction with information play more and more a key-role, whereas the amount of accessible information increases. Semantic technologies provide new solutions to structure this important property. One promising way to access the complex semantic structures and the huge amount of data, are visualizations. Today's Semantic Visualization systems offer primarily proprietary solutions for predefined and known users and usage scenarios. The adaptation to other scenarios and users is often cost- and time-consuming. This chapter presents a novel model for a full adaptable and adaptive Semantics Visualization framework. Starting with the introduction of a new visualization model, the implementation of this model will be described. The chapter concludes with selected advantages of the described visualization technology.

1 Introduction

Information constitutes a crucial resource of enterprises. An efficient way for accessing the "enterprise information", e.g. personnel expertise, contact persons or project and resources is important to reduce the cost and time effort. Beside the important role of information in several industrial and knowledge working areas, the everyday knowledge adoption process is mainly based on information acquisition from heterogeneous computer-based multimedia systems.

Semantic technologies provide different methods for annotating the underlying information with an appropriate "meaning". With the characteristics of building sentences in natural languages by subject, predicate and object, information entities are set in relation to other information, attributes or resources. Further the information entities are abstracted in appropriate concepts and categories respectively. The se-

K. Nazemi, M. Breyer, D. Burkhardt, C. Stab, J. Kohlhammer
Fraunhofer Institute for Computer Graphics Research (IGD), Fraunhoferstr. 5, 60489 Darmstadt, Germany, e-mail: kawa.nazemi@igd.fraunhofer.de

mantic structure allows searching, finding and gathering the requested information in an efficient way, whereas the result processing can be enhanced with rule-based and machine-learning methods to provide more insights into the given domain of information.

The existing semantic technologies are able to provide the answer of known questions in an efficient way, provided the questions and the goals of the given search are known. But there are several usage scenarios and cases, where searching has more an exploratory character, and the formulation of the question or the search term is not known explicitly. Further semantic technologies provide with their underlying structure to capture a whole information or knowledge domain and gather information about the context of the information.

The described cases are only examples for the exigency of investigating a complementary discipline for acquiring knowledge by considering and exploiting the semantic structure of the given knowledge domain, namely *Information Visualization*. Different research groups and institutions have worked on accomplishing visualization with the semantic structures. Herewith different terms have arisen for the merged technologies, e.g. Ontology Visualization or Semantic Visualization. Thus these technologies are mostly focusing on visualizing semantic structures, the term Semantics Visualization is appropriate in our opinion.

The *THESEUS Core-Technology-Cluster Innovative User Interfaces and Visualizations (CTC-WP5)* have investigated techniques and concepts to visualize semantic data for heterogeneous users, usage scenarios and semantic characteristics. The challenge of the research work was the conceptualization and development of a technology that is user-oriented on the hand and features the characteristics of a core-technology on the other hand. Therefore the adaptable and adaptive SemaVis Technology was developed, which covers both, the features of a user, use case and context adaptable core-technology and a user-centered and adaptive Semantics Visualization.

In this chapter the *SemaVis* technology will be described, staring with the abstract model of visualization, which subdivides the visualization of semantics data in three abstracted layer. Following the implementation of the described model as a framework will be introduced. The chapter concludes with some advances of the *SemaVis* technology and an application example.

2 The Model of Semantics Visualization

Semantic technologies process data to enrich them with meaning and structure and provide information that can be further processed by machines for a better understanding of the underlying domain. With this enrichment not only the search process is more efficient, the acquisition of implicit information is provided in a simple way. Semantic transitions allow by using various techniques to get information about data relationships that are not explicitly modeled. The transitivity of the relations assumes knowledge about a domain and has to be formalized in a machine readable way. The efficiency of the semantic search and information acquisition is constraint to both, the knowledge of the user (that he knows what he searches for) and the formalization of the semantics. But not in every case, the user really knows what the main target of his search is and which question has to be answered. *Semantics Visualizations* provide the ability to explore in a semantic space and acquire information in answers and questions, because ”There is nothing better than a picture for making you think of questions you had forgotten to ask (even mentally)” [14].

Exploring information and providing the ability for verbalizing questions was one of the main tasks given by the heterogeneous use cases. The challenge was to develop user-centered visualizations as core-technologies to support the heterogeneous users, usage scenarios and especially the Use Cases in the THESEUS research program. To achieve this goal a full adaptable visualization technology was necessary to meet the requirements of a core-technology. Therefore the whole process of the information visualization was investigated and the visualizations and the main user interface were subdivided into different layers of abstraction. Visualizations can be described according to the following characteristics: *what is displayed*, *wherewith is it displayed*, and *how is it displayed*?

To ensure a fine-granular adaptation of the whole visualization process, the model for *Semantics Visualization* has to adapt each of the given visual information separately. Therefore we separated the visual information pipeline into three different layers of abstraction. [6] The basic concept of our model was the reference model for information visualization [2], that processes the data in three sequential but iterative steps of *Data Transformation*, *Visual Mappings* and *View Transformations* from raw data to visualization. Accordingly to this model, we defined the visualization model as transformation pipeline for semantic data with *Semantics*, *Layout* and *Presentation*. [6]

The main goal of this model is to provide in each level of abstraction the ability to adapt the visual appearance and provide a user, use case and task-centered visualization model implementable as core-technology. The main difference of this model to existing models is the fact that human perception and the way how human process visual information played a key-role for the design of the model. We separated the visual information based on the findings in the area of human visual processing. [12, 13, 15] The *Semantics* layer in our model considers the amount and complexity of data to be visualized and transforms the required data to geometric structures. These geometries are analyzed in the *Layout* layer. The visual placement (on the plane) [1] and the closeness to other visual entities are defined in this layer. For each of the defined visual placement or layout algorithms a visual appearance is defined in the *Presentation* layer, the highest level of the model. Further each of these layers can be adapted according to the requirements of the given scenario or users. Figure 1 displays the abstracted model of *SemaVis* according to [6].

The generation of visualization is processed as follows: First the *Semantics* layer analyzes the semantic structure, properties and data. Thereafter appropriate layout algorithms are chosen for the data or for certain parts of the data. Finally the chosen layouts are visualized by the *Presentation* layer. Each of these layers can be manipulated by a simple visualization modeling language and can be adapted to

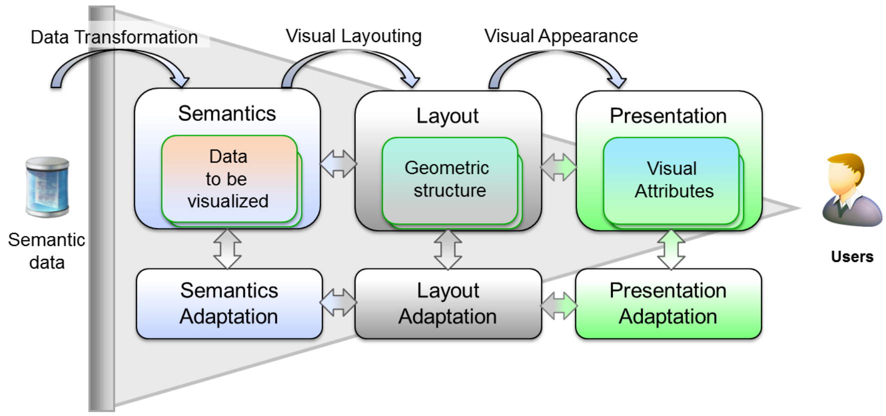

Fig. 1 The abstracted model of *SemaVis*: *Semantics Visualization*

several users, usage scenarios or corporate designs. The main user interface, that may include more than one visualization type, is generated using the same model.

2.1 Semantics - Modeling Data for Visualization

Semantics defines which data is visualized. It contains information about the data (what is the data about), its structure (e.g. hierarchy, incoming and outgoing relations) and implicit and explicit properties. The main task of this layer is to parameterize the data and provide the opportunity to adapt visualization in the lowest level. To visualize semantics data the formalized network-structure of semantic entities, attributes and information have to be transformed in a visualizable model. Additionally, the terminological schema-level has to be considered in the visualization, if semantics formalisms are given. The semantics formalisms intersect the data artifacts to be modeled for visualization. Therefore *SemaVis* applied them as abstracted conceptualizations in *Semantics* layer, covering semantic networks, framebased logics, and description logics [3, 5]. The semantics data model in *SemaVis* is compliant to *lightweight semantics*, consisting of concepts, concept taxonomies, relationships or roles between concepts, and properties describing concepts [4]. The *Semantics* layer models formal axioms, functions, rules, procedures and constraints to model heavyweight formal semantics formalisms [4] for visualizations. This enables to visualize the full scope of formal semantics with SemaVis.

2.2 Layout - Structuring Visual Information

The *Layout* layer is responsible for structuring and transforming the semantic information to produce a geometric structure of the data for visualization. The semantic entities identified and extracted in the *Semantics* layer are systematically analyzed and extended with layout-specific (e.g. spatial) information. *Layout* defines wherewith the semantics will be visualized. Several graphical metaphors, e.g. graph-layout algorithms or space-filling visualizations are chosen, based on the data characteristics derived in the previous step. The analyzed data structure are weighted and classified for visualization. Further adequate layout and placement algorithms (e.g. hierarchical, matrix, sequential of organic layout structure) represent the semantic structure in this layer. Additionally, further impact factors like users' preferences and current tasks are considered in the layout process step for deriving an appropriate geometric representation of the semantic information. It should be noted that only the geometrical layout is defined on this layer, but not yet the visual appearance.

2.3 Presentation - Representing Semantics with Visual Attributes

Presentation is the highest level of the *SemaVis* model and defines the visual appearance of the geometric data model. Based on the outcomes of the layout layer, where the positioning of the visualization entities are computed, this layer adopts the most important visual attributes to describe the entities visually. The visual attributes used in this layer are based on the pioneering work of BERTÍN [1] enhanced by findings of human visual processing abilities [12, 15]. The visual attributes or features adopted in *Presentation* are based on the preattentive information processing. Studies on human perception could show that several visual features drive our attention within a very short time (below 250 ms) to a target (target detection). [12, 13, 15] The *SemaVis Presentation* layer implements the most valuable visual features (e.g. color, size, texture, shape, orientation, position, transparency) to visualize the geometric model and support users in their heterogeneous tasks. Same as the other layer of the model, the *Presentation* layer is adaptable to different scenarios and supports the idea of a core-technology for visualizations.

3 Architecting the Model of Semantics Visualization as Framework

The above described model for *Semantics Visualization* is a theoretic concept of the *SemaVis* framework. The following section gives an overview about the system architecture and the modular implementation of the model.

3.1 System-Architecture

The foundation of the *SemaVis* framework is a modular and expandable system architecture that includes three different layers, (1) data layer, (2) logic layer and (3) model layer (see figure 2). The data layer contains on the one hand the semantics information that is loaded and generalized into an internal data representation for visualization and on the other hand script based files that allow the configuration of the framework for certain application scenarios. In the logic layer the data and the configuration files based on the *Semantics Visualization Markup Language (SVML)* are processed and interpreted for further utilization in the *Semantics Visualization* model. The model layer includes the data model that stores semantic data and offers certain data access methods that provide a uniform access to semantics information. Additionally the model layer contains the configuration model holding configuration data for the certain scenarios and the *Presentation* layer as a model.

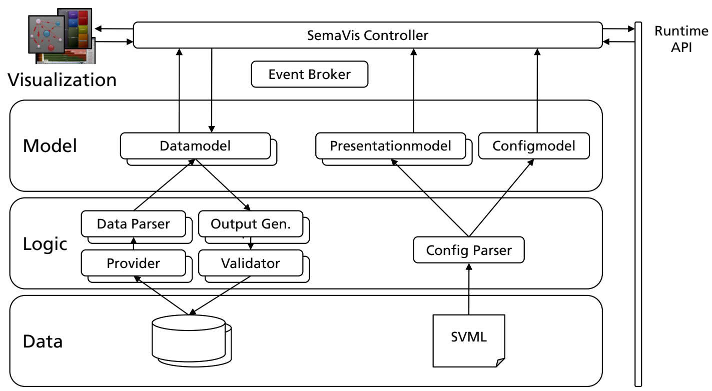

Fig. 2 System architecture of the *SemaVis* framework with arrows representing the information flow between modules.

The data model is responsible for visualization specific data management of semantic knowledge. It provides a generalization for different semantic data types and thus allows a uniform access for the semantics visualizations to the semantic structures. The data provider is responsible for gathering data from a knowledge repository and transfers the semantics information to the data parser that converts the data into the internal representation in the data model. If data entities are changed or edited in the visualization the output generator and the validator transfer the changes to the underlying knowledge repository. The *Presentation* model provides a unification of visual properties (e.g. color, size etc.) that is initially loaded from the configuration file. The event broker constitutes the central hub for each event that controls the state of the visualization and observes interaction events. All described modules are initialized and managed by the SemaVis-Controller that provides all needed models to the visualization components and manages the central processes in the framework and the in- and outgoing commands from the runtime application programming interface (API).

3.2 Modular Integration of Visualization Components

The depicted processing steps from the system-architecture are representing logical modules with a specific task. These modules are containing the logic to process and transform the input data to an output. To control the processing within the modules, the system can be configured statically by a configuration file and dynamically by using the system implemented API. Moreover we define *modules* and *sub-modules*. Modules are representing the essential processing steps from data to the final graphical user interface (GUI) Each module is acting as a standalone application with just the defined input parameters and outputs as result. Furthermore every processed result of a module will be transmitted through each following module. This allows for instance, the consideration of extracted data characteristics from the *Semantics* layer to the *Presentation* layer to set the final visual features. This modular architecture allows the flexible replacement and extension of modules with additional functionality without affecting the general processing pipeline. Due to the standalone conceptualization, it is not required to perform always the whole visualization pipeline. A change needs just the replacement or reconfiguration of the affected module. Among others, a very dynamic influence factor during the visualization process is the interaction of users, which can require a couple of analysis to ensure an efficient user support e.g. to extract the semantics and intension of the interaction. For these steps we defined additionally *sub-modules*, which are dealing with just a single task and are strongly related to just one main module. In contrast to modules they are not formal described, thus the heterogeneous usage has a supporting character.

3.3 Data Integration

SemaVis is intended and designed as a core-technology, thus *SemaVis* is neither domain dependent nor developed for a specific user group. Particularly important is the loosely coupled software design paradigm pursued for the data integration layer, which is responsible for the communication to external data and ontology management systems. For integrating *SemaVis* into various use case scenarios this layer can be adapted, extended, and configured by framework extensions or by settings in the SVML-Markup configuration.

SemaVis support the integration of different data sources offering different data formats simultaneously. Further loading-on-demand, editing, and annotation functionalities are supported. Even gathering data from a specific data source in a given data format and storing changes or annotations to another data management system in a different data format are supported due to the internal representation of the data in the *Semantics* layer. To provide these scopes and functionalities the data integration concept differentiates five responsibilities: *data provider*, *data parser*, *data model*, *output generator* and *validator*.

The data provider implements the connection between *SemaVis* and the data or ontology management system. The most prominent connections like *file-based*, *SparQL* or *Web Services* offered by most of the existing management systems are implemented. Even proprietary connection can be easily implemented with modular characteristics of *SemaVis*.

The data parser converts and transforms the provided data and inserts the information into the internal data model. Due to the announcement of the semantic web community and standardizations formalisms like the *Resource Description Framework (RDF)*, the according schema *RDFS* and *Web Ontology Language (OWL)* are predominant and thus supported. Additionally *SemaVis* supports graph languages like GraphML or the technically lower syntactic levels (e.g. *XML* or *JSON*).

Gathered data are passed to the internal data model which allocates and manages it. Other modules can query specific data artifacts to the model. Further semantic analysis algorithms of the *Semantics* layer identify semantics which may be relevant for the user or the current task. This requires methods for retrieving special artifacts of the formal semantics, which are managed and provided within the internal data model.

For transmitting edited and annotated information the output generator preprocesses this data to the sinks formats. Preconditioned the data or ontology management system supports data evolution, the changes can be inserted into the original data base. Otherwise a different management system can be triggered to store it.

To guarantee only feasible changes, the validator verifies the syntactic correctness. In most application scenarios the validator is closely coupled to the verification mechanisms of the backend system. This computation distribution allows better performance results, especially if large amount of semantic data has to be investigated in the verification process. Due to the fact user demand for rapid feedback a part of the verification process is performed on the client side. So the user can continue his workflow of editing or annotating the semantics data while the system checks iteratively the integrity of the changes.

4 Advances in Semantics Visualization

4.1 Reducing Complexity with Complementary Visualizations

Semantically annotated data contains heterogeneous and complementary attributes for describing the knowledge domain. Existing visualizations cover either one of these attributes or visualize the whole spectrum. Both variations have advantages and disadvantages. While visualizing the whole semantic spectrum is often complex to understand and not always necessary, the reduction to certain semantic characteristics might reduce the information content. Therefore it is necessary to provide a solution where the different semantic characteristic can be visualized in an easy and understandable way.

SemaVis implements a multiple-visualization user interface as a visualization cockpit. [7] The visualization cockpit separates the information units from each other and visualizes this information in separate visualization units. The advantage of the separation of complex information units is obvious, the user of a cockpit is able to perceive the required information very fast and react to the perceived information very fast.*SemaVis* has transformed this idea for visualizing semantics. A set of different visualization techniques was developed; each visualization focuses on certain characteristic of semantic data, e.g. a spatial space-filling visualization was developed to visualize the concept hierarchies [8], graph visualization methods are used for visualizing relation and timeline visualizations for explicit properties [11]. The visualization cockpit metaphor is more than the combination of multiple visualizations. It provides the separation of the semantic information in several view-modes: 1. Aspect-oriented mode: visualizes the same data with different visualizations. The user is able to combine visualization for viewing the same information from different perspectives. The visualizations are linked with each other. Every interaction with one of the visualizations changes the view in the combined one. 2. Visual Comparative mode: In this mode the visualization are not necessarily linked with each other. The user is able to compare certain parts of the same semantics data by using different visualization techniques. 3. Semantics Comparative mode: In this mode the same visualization combinations are used for two different semantics data sets. The user is able to compare two different semantics data bases with the same visualizations. 4. Level-of-Detail mode: In this mode the same visualization technique is used twice on the screen, whereas the presentation layer differs. The user is able to zoom in (semantically or visually) one of the visualizations and keep the overview in the other one.

4.2 Adaptive Semantics Visualization

The layer-based adaptability of *SemaVis* provides an adequate framework for generating heterogeneous user interfaces by manipulating the entire spectrum of the visualization characteristics. This ability is used to further provide an automatic adaptation to several impact factors, e.g. user interaction or data structure. The goal of the automatic adaptation of the visualization is to support the user in her navigation process, during the exploration of information.

Therefore a new approach was conceptualized, where the impact factors are captured and enriched with semantic with a direct link to the semantics of the data. A user interaction for instance, is determined as a three-dimensional conceptualization of the interaction taxonomy. Thereby the semantic taxonomy of the data is used and enriched with a hierarchy of the interaction device and a hierarchical representation of the visualization. These three dimensions allow creating unique interaction semantics, providing a meaning of the users' natural consequence of the visualization operation. [6] Further the characteristics of the data, especially the semantic structure and the explicit and implicit properties, are considered in a similar way. Therefore a data-semantics is generated that provides information about the current visible data. [9] The implicit gathered information is used to personalize both the different layers of the visualizations and the main user interface. While the different layers of the visualization aims to support the navigation and exploration process of the user, the main user interfaces recommends several visualizations for a more efficient way of acquiring information. The transformation of the determined "context semantics" is transformed by an probabilistic algorithm [10] to both, the different graphical layers for adapting the visualization characteristics and to the main user interface for recommending a set of adequate visualizations. Therefore both aspects are considered, the current context (given data and recognized activity) of usage and the preference of the user. We try to comprehend the exploratory adaptation with the following example: A user wants to know, which persons were influenced by the work of the computer scientist *Turing*. He searches in the *SemaVis* search interface for the term "Turing" and enters the exploration space. Because the term *Turing* appears in different knowledge categories, the main user interface recommends a spatial visualization for the concept hierarchy, hereby the concepts uses the presentation layer for highlighting concepts of interest. With the cockpit metaphor, a composition of different visualizations is automatically created, which shows the relevant information for the user separated in several views. A further visualization pictures the relations of the search terms to other data entities, whereas the relation of interest is highlighted and the other relations are put in the background using the transparency feature of the presentation layer. Thus the work of Turing influenced persons even after his lifetime, a further visualization is chosen, which visualizes explicitly the time and semantic relationships. The main display further recommends several visualization with probabilistic quantifiers for the given exploratory search issue. The user is able to choose other visualizations, recommended by the systems or reduce the information space by removing visualizations from the main user interface. Figure 3 displays the described adaptation process for the term *Turing*.

5 Conclusion

Semantic technologies provide a promising way to access information and complex data structures. Today's technologies and research focus on the "machine readability", whereas the human factor is often neglected. *Semantics Visualizations* are promising approaches for providing human-centered solutions for interacting with semantic data. This chapter presented the *SemaVis* technology based on a novel model for a full adaptable and adaptive Semantics Visualization. The model subdivides the visualization pipeline in three visual characteristics and provides consequently a fine-granular adaptation to heterogeneous users and usage scenarios. Further the implementation of the described model as a modular framework tech-

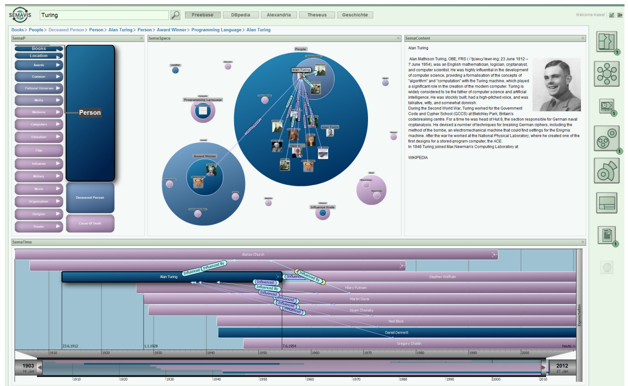

Fig. 3 User-centered adaptation of *SemaVis* for the search term *Turing*.

nology was described. The modularity allows integrating not only visualizations but also recommendation and analysis algorithms into the framework. The advantages of the framework were described with an example of exploratory search.

References

1. Bertin J.: Semiology of graphics. University of Wisconsin Press (1983).
2. Card, S.K., Mackinlay J.D. and Shneiderman, B.: Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann (1999).
3. Fensel, D., Hendler, J., Lieberman, H., Wahlster, W. (eds.): Spinning the Semantic Web. Bringing the World Wide Web to Its Full Potential. MIT Press (2003).
4. Gómez-Pérez, A., Fernández-López, M., Corcho, O.: Ontological Engineering. Springer (2010).
5. Hitzler, P., Krtzsch, M., Rudolph, S., Sure, Y.: Semantic Web: Grundlagen. (Semantic Web: Foundations) Springer (2008).
6. Nazemi, Kawa; Stab, Christian; Kuijper, Arjan: A Reference Model for Adaptive Visualization Systems. In: Jacko, Julie A. (Ed.): Human-Computer Interaction: Part I: Design and Development Approaches. Berlin, Heidelberg, New York: Springer, 2011, LNCS 6761, pp. 480-489.
7. Nazemi, Kawa ; Breyer, Matthias ; Burkhardt, Dirk ; Fellner, Dieter W.: Visualization Cockpit: Orchestration of Multiple Visualizations for Knowledge-Exploration. In: International Journal of Advanced Corporate Learning. (2010), 4, pp. 26-34.
8. Nazemi, Kawa ; Breyer, Matthias ; Hornung, Christoph: SeMap: A Concept for the Visualization of Semantics as Maps. In: HCI International 2009. Proceedings and Posters: With 10 further Associated Conferences. Berlin, Heidelberg, New York: Springer, 2009, LNCS 5616, pp. 83
9. Nazemi, Kawa; Burkhardt, Dirk; Praetorius, Alexander; Breyer, Matthias; Kuijper, Arjan: Adapting User Interfaces by Analyzing Data Characteristics for Determining Adequate Visualizations. In: Kurosu, Masaaki (Ed.): Human Centered Design: Second International Conference: HCD 2011. Berlin, Heidelberg, New York: Springer, 2011, LNCS 6776, pp. 566-575.
10. Nazemi, Kawa; Stab, Christian; Fellner, Dieter W.: Interaction Analysis for Adaptive User Interfaces. In: Huang, De-Shuang (Ed.) et al.: Advanced Intelligent Computing Theories

and Applications: 6th International Conference on Intelligent Computing. Berlin, Heidelberg, New York: Springer, 2010, pp. 362-371.

11. Stab, Christian; Nazemi, Kawa; Fellner, DieterW.: SemaTime - Timeline Visualization of Time-Dependent Relations and Semantics. In: Bebis, George (Ed.) et al.: Advances in Visual Computing. 6th International Symposium, ISVC 2010: Proceedings, Part III. Berlin, Heidelberg, New York: Springer, 2010, pp. 514-523.
12. Treisman A. M.: Preattentive processing in vision. Computer Vision, Graphics and Image Processing 31:2 (1985), 156177.
13. Treisman A. M.: Features and objects in visual processing. Scientific American. 255 (1986), 106115
14. Tukey, J. W. and Tukey, P. A.: Computer graphics and explaoratory data analysis: An introduction. In Proceedings of the Sixth Annual Conference and Exposition: Computer Graphics'85 3 773-785. National Computer Graphics Association, Fairfax, VA (1985).
15. Wolfe J. M.: Guided search 4.0: Current progress with a model of visual search. W. Gray (Ed.), Integrated Models of Cognitive Systems (2007), 99119.