RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS
No. 85 — 2009
Advanced Information Technologies for Management – AITM 2009

Helena Dudycz

Wrocław University of Economics, Wrocław, Poland
e-mail: helena.dudycz@ue.wroc.pl

PROBLEM OF VISUALIZING SEMANTIC LAYER OF A LARGE NUMBER OF TOPICS IN TOPIC MAPS

Abstract: In the last few years an increasing interest of visualization's applications to the presentation of business information has occurred. Visualization should support the user in finding relevant information. One of the solutions which can help managerial staff to obtain essential information is that of a topic map. This is the standard uniform introduced by the International Organization for Standardization used to describe knowledge structures and associate them with information resources. Topic map is the visualization concept which can support exploration of large data sets. The advantage of visual data exploration is that the user is directly involved in the data mining process. Purpose of this article is to solve problem of large number of topics in usage topic map as a tool enabling visual exploration of huge, complex and multidimensional data set. Relating to that in the first part standard of topic map was characterized and possibilities of the use of a topic map for creation application enabling visual exploration of data were pointed out. The next part describes the solutions of the problem of a large number of topics.

1. Introduction

Now collecting data and information is not a problem. However, while information becomes richer and more complex, extracting value from information collections becomes progressively more difficult (see [Zhu, Chen 2005, p. 139; Keim 2002, p. 105]). The analysis of such massive data sets is a challenging task. Nowadays researchers and analysts are interested in patterns in the data including associations, correlations or exceptions (see [Keim, Schneidewind 2005, p. 1767]). Therefore "alternative information-organization methods are needed to more effectively and efficiently retrieve information from various system" (see [Yi 2008, p. 1898]).

Information visualization techniques may help to solve the problem, because "visualization offers a link between the human eye and the computer, helping to identify patterns and to extract insights from large amounts of information" [Zhu, Chen 2005, p. 139]. The basic idea of visual data exploration is "to present the data in some visual form, allowing the human to get insight into the data, draw conclusions, and directly interact with the data" [Keim 2002, p. 100]. In that way the user is

directly involved in the data mining process; it is possible to include human expertise or feedback in data mining, leading to more effective data exploration (see [Zhu, Chen 2005, p. 146; Keim, Schneidewind 2005, p. 1767]). Visual mining methods enable to overcome major problems of automatic data mining methods, e.g., the presentation of uninteresting results, lack of acceptance of the discovered findings, or limited confidence in these (see [Atzmueller, Puppe 2005, p. 1752]).

One of methods of visualization which can be used in the data mining process is topic map. This solution is a relatively new form of presentation of knowledge, which put emphasis on data semantics and ease of finding desired information (see [Ahmed, Moore 2006; Freese 2001; Newcomb 2002]). Usually a topic map is a semantic graph, that contains definitions of a set of topics and a set of association between topics called ontology of a domain (see [Korczak, Dudycz 2009, p 86]).

As topic maps may contain a lot of topics and associations, a critical issue is to visualize semantic layer efficiently (see [Grand, Soto 2000]). The aim of researches is to identify solutions which allow a better, faster and more intuitive exploration of very large data resources (including database management and data warehouse systems).

In this article solutions of the problem of large number of topics in a topic maps application are discussed. The elaboration is structured as follows. In the next section ideas, benefits and limitations of topic map are briefly described. There is also description of visual data exploration in the application of a topic map. Section 3 illustrates the solutions of the problem of a large number of topics. A number of examples of visual interpretation are presented and discussed. Finally in the last section a summary of this work is given and a future research projects are indicated.

2. Characteristics of a topic map

2.1. Idea of a topic map

Topic map is an International Organization for Standardization (ISO) standard [ISO/IEC 13250] used to describe knowledge structures and associate them with information resources (see [Pepper 2000; Grand, Soto 2000]) and allowing to solve problems related to information retrieval and interpretation through the structuring of information resources according to user requirements (see [Korczak, Dudycz 2009, p. 87]). A topic map is a kind of data structure.

An organization which adopts topic maps will be able to share the data and knowledge within those maps with other groups or business partners (see [Ahmed, Moore 2006].

The topic map model contains three elements (called characteristics) to any given topic:

– its name, which is useful to people in dealing with topics, and a topic does not actually need a name;

– its associations with other topics, which are the conceptual essence of topic maps, indicating how one topic relates to another;
– its occurrences (also called resources) which are the actual references – pointers to relevant information resources.

Topics and their associations build a semantic structure linked to information resources (see [Grand, Soto 2000]) that comprise topic maps and allows them to describe ontologies. In that way it is possible to build semantic network above information resources (see [Grand, Soto 2000; Rath 2001]), which enables easy navigation on scattered sources of data[1]. On Figure 1 the visualization of demonstration topic map is shown, where topics are represented as rectangles, whereas lines indicate associations between them.

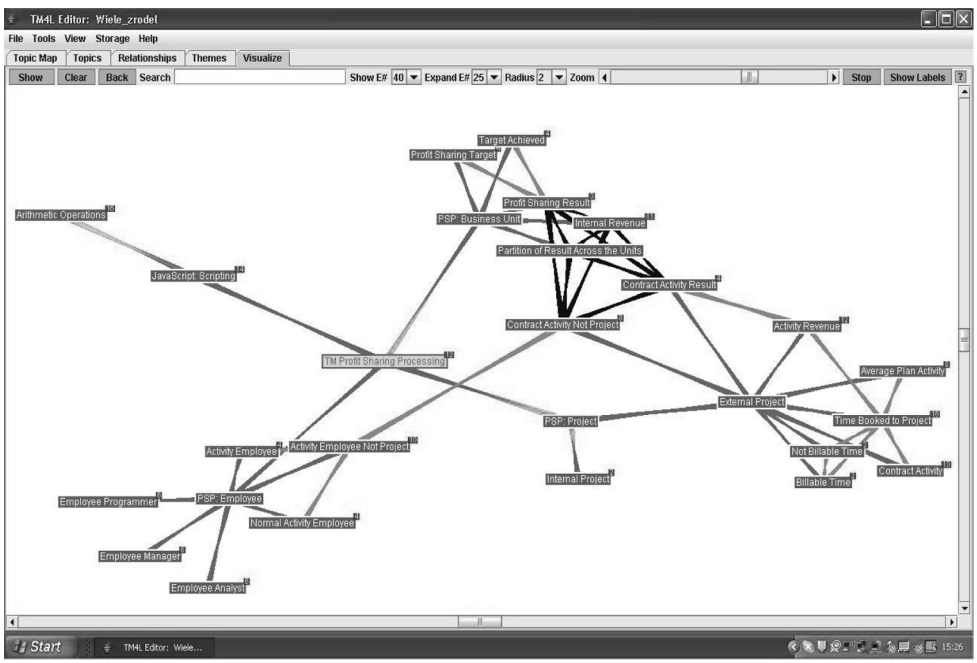

Figure 1. Visualisation of demonstration topic map

Source: own presentation based on TM4L Editor.

Topic maps are essentially interconnected graphs with (potentially) many dimensions of meta-data. There already are a number of approaches to the visualization of such data in the commercial domain (see [Ahmed, Moore 2006]):

[1] Analysis of potential possibilities of use of standard topic map was described among others in: [Ahmed, Moore 2006; Korczak, Dudycz 2009; Sanin et al. 2007].

– hyperlinked-trees – a graph can be interpreted as a hierarchical tree relationship with additional hyper-links between nodes,
– graphs – graph visualization displays the topic map as a set of interconnected nodes,
– landscapes – an interesting data visualization technique is displaying interconnected information as a landscape, assigning coordinates to topics according to their interconnections and assigning height to coordinates according to the degree of relevance or the degree of convergence of multiple topics,
– worlds – the data model of topic maps seems to lend itself well to the construction of three-dimensional spaces.

Topic maps as the foundation for integrating all sorts of data sources (see [Ahmed, Moore 2006]) can support the extraction of information from all relevant subsystems.

2.2. Visual data exploration in the application of a topic map

The recent integration of visualization with various data mining techniques has attracted attention, as huge volumes of data are routinely being generated and stored in databases (see [Zhu, Chen 2005, p. 145]). Visual analysis and visualization techniques have been proven – as said Keim and Schneidewind – "to be of great value in analyzing and exploring such large data sets, since presenting data in an interactive, graphical form often fosters new insights, encouraging the formation and validation of new hypotheses to the end of better problem solving and gaming deeper domain knowledge" (see [Keim, Schneidewind 2005, p. 1767]).

Visual data exploration usually follows a three steps process (which has been called by B. Shneidermann's Visual Information Seeking Mantra) (see [Keim 2002, pp. 100-101; Keim, Schneidewind 2005, p. 1768]):
– interactively by overview – the user needs to get an overview of the data and identifies interesting patterns,
– zoom and filter – the user focuses on one or more of interesting patterns,
– details-on-demand – the user needs to drill-down and access details of the data for analyzing the patterns.

In this interactive process, the user is able to subsequently concentrate on the interesting data elements by filtering uninteresting data, and focusing (zooming in) on the interesting elements, until finally details are available for an interesting subset of the analysed elements (see also [Atzmueller, Puppe 2005, p. 1756]). Important stage in this process is the use of appropriate solutions which allow to filter and zoom in (zoom out).

One of these visualization methods enabling visual data exploration is topic map application. It allows to display the whole semantic network (topics and associations) efficiently, as it is essential to select the relevant information. Fundamental

factors for good visualization interface of application of topic map are (see [Grand, Soto 2000]):
– an overview of the structure for a global understanding of the structure and of the relationships within the hierarchy,
– the ability to zoom and to select some nodes,
– dynamic requests in order to filter data in real time.

Filtering techniques are needed in order to select and display only relevant information. For example in the software TM4L Viewer there are three possible manners of interrogations (see [Korczak, Dudycz 2009, p. 95]):
– searching by topic name and/or type,
– selecting topics from a hierarchy (tree view),
– using a custom query to display a list of valid choices (this option can make use of the current context as a part of the query).

On Figure 2 there are shown screenshots visualizing the semantic network and visual data exploration in topic map application.

In next points benefits of a topic map and its limitation are indicated.

2.3. Benefits of a topic map

There can be shown seven basic benefits of creating topic map application for organization of information in existing information in the company.

Firstly, topic maps are saved as computer files in the open standard, so they are not limited to the particular form of data presentation shown here.

Secondly, topic map is characterized by flexibility, i.e. it can be freely modified by adding new topics and connections between them.

Thirdly, once the topic maps have been built, a user can easily merge data from one topic map to another (see [Korczak, Dudycz 2009]).

Fourthly, topic maps offer the alternative of indexing and searching for topic names, and then using topic occurrences to present links to all content related to the topics found by the search (see [Ahmed, Moore 2006]). Such searching is more efficient than that based on basic hierarchic structure (see [Garshol 2004; Yi 2008, p. 1899]). The results of Yi's study show that relationships-based query searches using this topic maps-based information retrieval system resulted in better recall and shorter search times than did those for fact-based query searches (see [Yi 2008, p. 1910]).

Fifthly, the layer of topics is separated from the layer of resources. Building semantic layers there is no need to modify subsystems or to duplicate content and logic of subsystems[2].

[2] A wide review of the architecture of information system with topic map application is presented in [Korczak, Dudycz 2009, pp. 88-89].

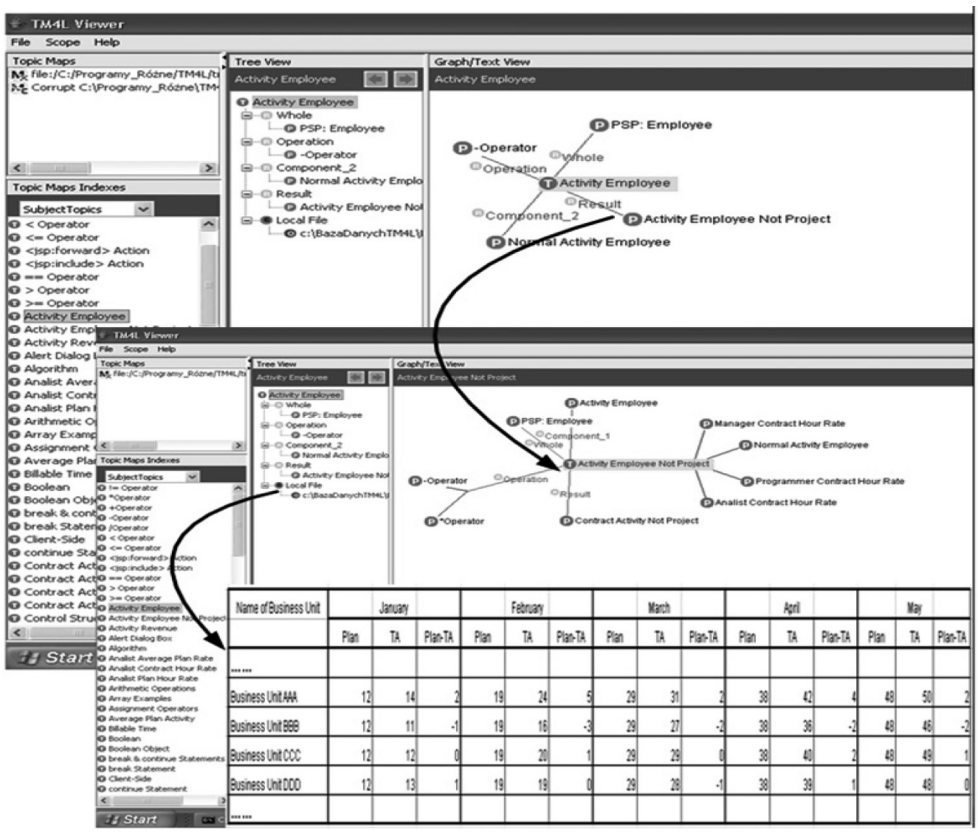

Figure 2. Example of visual exploration of data in topic map application

Source: [Dudycz, Korczak, 2009].

Sixthly, topic maps can be used to combine topic information received from multiple separate sources into a single functioning topic map (see [Ahmed, Moore 2006]).

Seventhly, for topic map's standard a large set of tools from various vendors becomes available, like query languages, graphic visualization, portal integration, content management, workflow, natural language querying (see [Garshol 2004]).

2.4. Limitations of a topic map

Having verified possibilities and properties of the topic maps application, the following limits and drawbacks were identified[3]:

[3] A wide review of the problem is presented in [Dudycz 2009].

– the potentially greater difficulties and longer time required to design the semantic net concerning the ontology of the field that is supposed to be written down in the topic maps,
– the difficulties concerning navigation in the topic maps when on the screen there is projected a big amount of topics and connections between them,
– the difficulties with identification at the right moment necessary sources when they were attributed to one topic in the great number,
– the wrong choice of software to create and activate the topic maps application thinking about needs and abilities of users,
– the usage of the topics' wrong nomenclature,
– the modification and expansion of the topic maps application to the next topics, associations and resources.

One of major limitations of topic maps application in visual data exploration is the problem of visualizing semantic layer of a large number of topics.

3. The solutions of the problem of a large number of topics

One of the main troubles with the effective exploitation of visual data exploration in application of topic map is the navigation when a lot of topics and associations are projected on the screen.

Nowadays the existing software to create the topic maps application permits visualization of the great number of topics and connections. There is no technological limitation of their range. However, if topic maps are very large, users may have problems with understanding them and finding relevant information (see [Grand, Soto 2000]). There is often a problem of legibility of the topic maps as well. This restriction can be minimized with the usage of the following solutions in the topic maps application:

– the creation of the ontology of the field, containing also the hierarchical connections,
– making various visual filtering techniques permitting zooms, translations and rotations available,
– the use of different shapes and lines symbolizing topics and associations.

The first solution consists in the analysis of the ontology of the field in order to isolate the potential groups and subgroups of topics. It allows to identify the semantic connections of Superclass-Subclass type or Whole-Part type. The efficiency of that way will be analyzed on the example of two applications built for the same ontology. In the first case the hierarchical structure was not separated (Figure 3), so all topics were put on the same level (window TreeView). Among this topics there are various semantic connections (window Visualize). Although there are not so many topics, the navigation between them with the aim of finding only relevant information is significantly hard. The user is forced to accurately analyze the whole image in order to look for interesting for him topics and connections. In the second application the

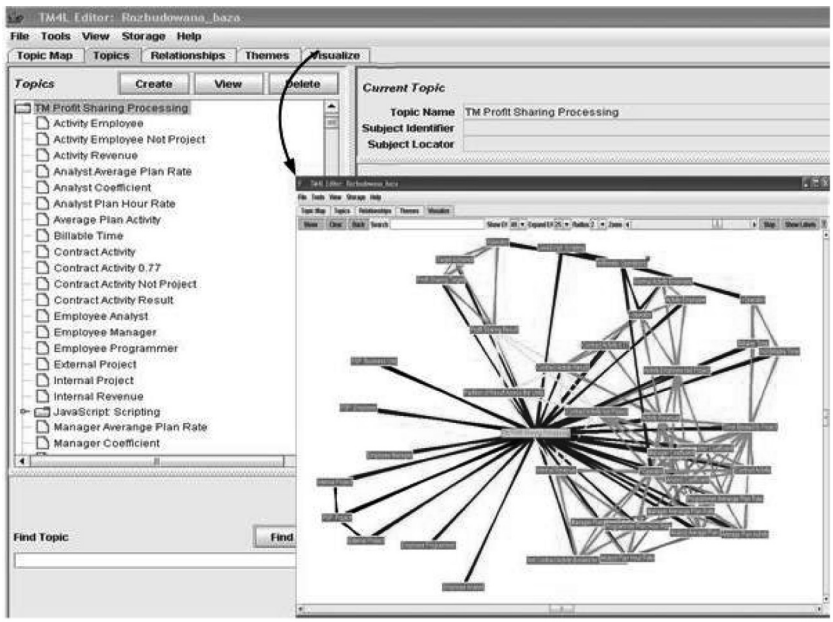

Figure 3. Example of topic map application without creating groups and subgroups of topics

Source: own presentation based on TM4L Editor.

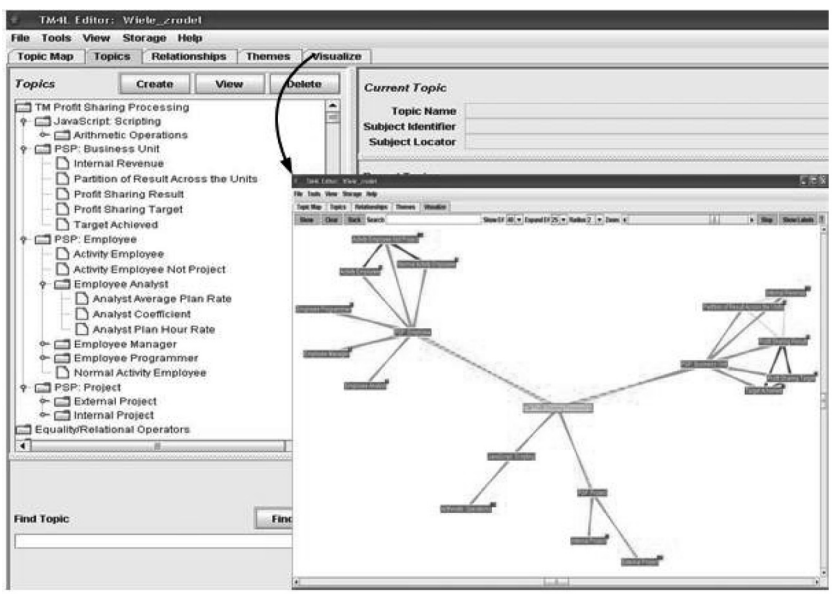

Figure 4. Example of topic map application with created groups and subgroups of topics

Source: own presentation based on TM4L Editor.

class hierarchies functionalities were defined (Figure 4). That allows to reduce the number of types to represent. In this way redundant overloading image with information can be minimized. The user has a view on the general view of semantic net and by use the right filtering techniques he can display only relevant information.

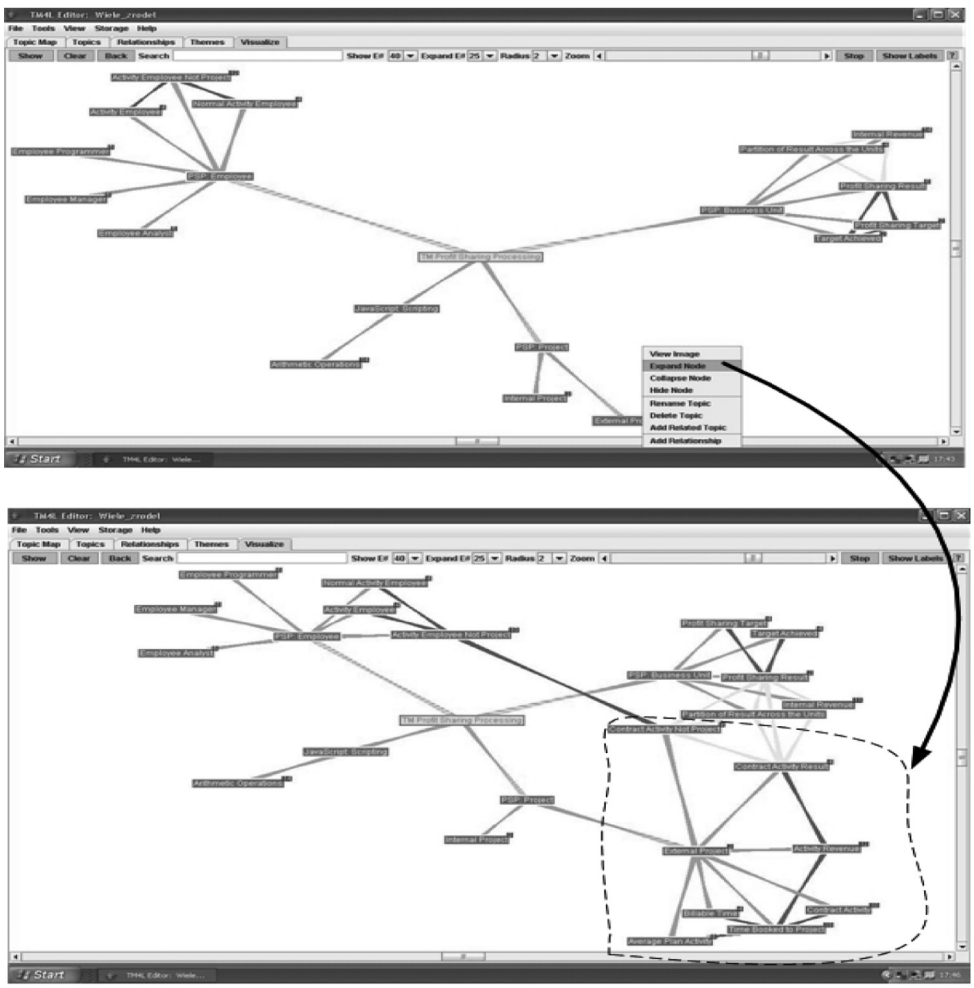

Figure 5. Example of expanding chosen topic in topic map application

Source: own presentation based on TM4L Editor.

The second solution requires use of the right visual filtering techniques in order to select and display important topics and associations. Nowadays the software for creating the topic maps application contains the most often functions allowing to show or hide parts of graph. It enables to modify projected image's contents using

interactive manipulation of views of node attributes. The user can expand interesting branches or delete irrelevant ones of the topic map (Figure 5). In this way the user can analyse at the right moment the essential part of semantic net. Therefore it is still necessary to make researches that point the most effective visual filtering techniques which are: highly interactive letting zoom, translate and rotate.

The third solution concerns use of graphic solutions making human's perception of view easier. Topic types and associations can be distinguished with different shapes, lines (full line, dotted line, etc.) or colours. The rule that definite colour means always the same topic or association should be obeyed[4].

4. Conclusion

The issues shown in the article are related to the use of a topic map as a tool of visual exploration of data. In this study emphasis was put on pointing the three solutions of the problem of large number of topics and associations in topic map application.

As the conception of a topic map is relatively new, information technology solution researches connected with creation of the topic map application for description diverse sources of data should be carried out. Future works will involve studying how large existing graphs visualization techniques may be applicable to topic maps.

References

Ahmed K., Moore G. (2006), Applying topic maps to applications, *The Architecture Journal*, January, http://msdn.microsoft.com/en-us/library/bb245661.aspx.

Atzmueller M., Puppe F. (2005), Semi-automatic visual subgroup mining using VIKAMINE, *Journal of Universal Computer Science*, Vol. 11, No. 11, pp. 1752-1765.

Dudycz H. (2009), Identyfikacja ograniczeń zastosowania mapy pojęć w wizualnej eksploracji danych, *Roczniki Kolegium Analiz Ekonomicznych*, SGH, Warszawa (forthcoming).

Dudycz H., Korczak J. (2009), Wizualna eksploracja danych finansowych za pomocą mapy pojęć, [in:] *Systemy wspomagania organizacji SWO'2009*, Vol. 2, Eds. T. Porębska-Miąc, H. Sroka, Prace Naukowe Akademii Ekonomicznej, AE, Katowice, pp. 253-263.

Freese E. (2001), Using topic maps for the representation, *Management & Discovery of Knowledge*, ISOGEN International, http://www.isogen.com.

Garshol L.M. (2004), *Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all*, http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html.

Grand B.L., Soto M. (2000), *Information Management – Topic Maps Visualization*, http://www.gca.org/papers/xmleurope2000/pdf/s29-03.pdf.

ISO/IEC 13250:2000 (2000), *Information Technology – SGML Applications – Topic Maps*, International Organization for Standardization, Geneva, Corrected, 2nd edition, http://www.y12.doe.gov/sgml/sc34/document/0322_files.

[4] For example in programme TM4L view generated in window: Visualize always contains topics of grey colour, whereas colours of lines are automatically changed – each time the same association is symbolised by different colour.

Keim D.A. (2002), Information visualization and visual data mining, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 7, No. 1, pp. 110-107.

Keim D.A., Schneidewind J. (2005), Scalable visual data exploration of large data sets via MultiResolution, *Journal of Universal Computer Science*, Vol. 11, No. 11, pp. 1766-1779.

Korczak J., Dudycz H. (2009), Approach to visualisation of financial information using topic maps, [in:] *Information Management*, Eds. B.F. Kubiak, A. Korowicki, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk, pp. 86-97.

Newcomb S.R. (2002), *The Economics of the Topic Maps Reference Model*, http://www.RenderX.com.

Pepper S. (2000), *The TAO of Topic Maps: Finding the Way in the Age of Information*, Proceedings of XML Europe, GCA, http://www.ontopia.net/topicmaps/materials/tao.html.

Rath H.H. (2001), *Making Topic Maps More Colourful*, http://www.gca.org/papers/xmleurope2000/pdf/s29-01.pdf.

Sanin C., Szczerbicki E., Toro C. (2007), An OWL ontology of set of experience knowledge structure, *Journal of Universal Computer Science*, Vol. 13, No. 2, pp. 209-223.

Yi M. (2008), Information organization and retrieval using a topic maps-based ontology: Results of a task-based evaluation, *Journal of the American Society for Information Science and Technology*, Vol. 59, No. 12, pp. 1898-1911.

Zhu B., Chen H. (2005), Information visualization, *Annual Review of Information Science and Technology*, Vol. 39, pp. 139-177.