Kazienko P., Litwin M.: *On Using Topic Maps for Knowledge Representation*. Information Systems Applications and Technology ISAT 2003 Seminar. Proceedings of the 24th International Scientific School. Wrocław University of Technology, 2003, pp. 100-107

topic maps, XTM, knowledge representation, semantic network, e-enzyme, visualization, RDF

Przemysław KAZIENKO*, Magdalena LITWIN*,

ON USING TOPIC MAPS FOR KNOWLEDGE REPRESENTATION

In a world of infoglut finding requested information is a difficult task. Main reason of that is the lack of semantic structure in most of web applications. Popular search engines and other services enabling information retrieval utilize a simple textual search without any semantic layer. Topic Maps ISO standard provide a solution to this problem. Its paradigm is based on representing knowledge in the shape of concepts, relation among them and resources.
The paper contains some detailed information on the topic maps paradigm. An application of enzymes knowledge database – the e-enzyme system is presented as an authors' example of the topic map. Some problems in developing of topic maps are pointed out: manual ontology building, visualization, lack of querying standard and methodology for topic maps design.

1. INTRODUCTION

In the electronic information world we do not suffer from lack of information but from huge amounts of it. In order to help to find relevant data a new model for knowledge representation is highly desirable. Such new model should allow not only to store data but also to organize them in a way that supports information retrieval. Topic Maps (TM) are a powerful mechanism which enables to organize knowledge so that the retrieval and sharing with other users would be easy. Topic maps provide not only storing data but also their meaning therefore they can be compared to semantic networks [14].

3. TOPIC MAPS AS A MODEL FOR KNOWLEDGE REPRESENTATION

Topic map is a model of knowledge representation which is based on three main issues:

- extraction of **topics** (subjects) which are concepts typical for modeled domain of knowledge,
- defining **associations** (relations) among topics,
- linking topics with a data layer (resources).

Each topic can have a name (none, one or more) and should have one or more topic types. A relation between topics and topic types is a simple class-instance association. The links between topics and their related information (e.g. web resources) are defined by objects called **occurrences**. The linked resource can be located in or outside the map. Occurrences like topics can be of a certain **type**. Types of occurrences are also defined as topics. There is a possibility to define relations between topics which are called **associations**. Each association can have an **association type** which is also a topic. There is no constraint about how many topics can be related by one association. Topics can play specific roles in association, described by **association role types** which are also topics. Using associations and topics from fig. 1 we can tell that "Vincent van Gogh *painted* Sunflowers", "Vincent van Gogh *was living* in Paris".

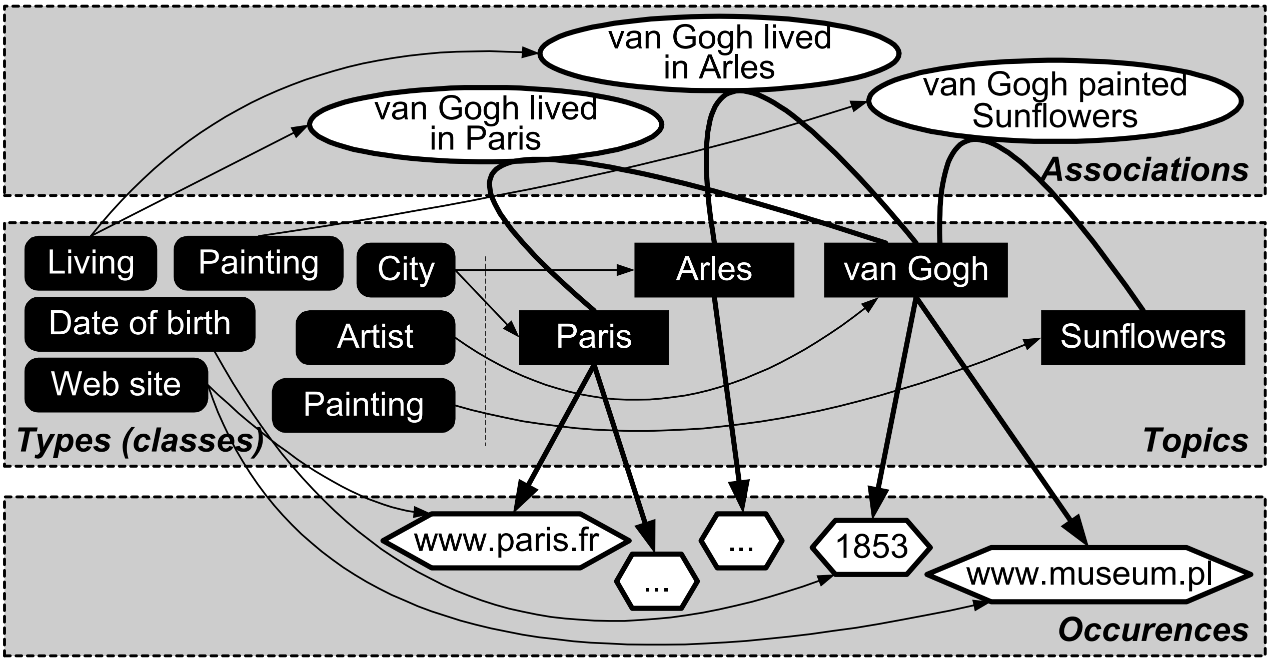

Fig.1. A topic map example

Sometimes there is a need to define constraints on topics in order to explain when they are valid. It can be reached using **scopes** assigned to topics, occurrences or associations. An example of scope for association between van Gogh and Paris can be the particular period the painter spend in Paris.

Topic maps provide also a mechanism which allows identifying seemingly disparate topics. Each topic can have a unique **subject identity** which describe topic in an unambiguous way. Subject identity is used for topic map merging when there is a need to recognize which topics describe the same subject.

3. RELATED WORKS

The idea of organizing data with some additional (semantic) information is not new. Currently there are a few standard, specifications, and techniques enabling organise knowledge e.g: semantic network [14] and RDF (Resource Description Framework) [10]. Semantic networks are used in the area of artificial intelligence for representing knowledge.

This concept mapping technique has a lot to do with topic maps. Both of these models are organised into a network of information, they enable to assign some semantic information to nodes and links between nodes. But the main difference is that topic maps focus on association between topics and on linking them with some resources while semantic networks concentrate more on concepts (nodes) [14].

Another mechanism for representing data and metadata is provided by RDF. Both RDF and topic maps standards are similar [9]. However RDF focus on network resources while topic maps starts from human being point of view. In particular in topic maps associations are always bi-directional, names are emphasized features of topic. There are scopes in topic maps in opposite to RDF. But the most important difference between both approaches is that in topic maps ontology is close to human perception while in RDF data are organized rather to facilitate machine understanding and processing.

Topic maps became the ISO standard in 2000 [7]. One year later an independent organization TopicMaps.org developed the XTM 1.0 (XML Topic Maps) [15] specification, in which it proposed using XML 1.0 for markup syntax and XLink for linking syntax [8]. XTM was created in order to simplify the ISO topic map specification and enable its usage for the Web.

Parallel some works have been carried out in particular topic map domains like querying and visualization. Initially the most common way of searching topic maps was to walk among topics, occurrences and associations. While such approach is suitable for not complicated and small maps, it does not turn out to be useful for large sets of topics. There is a need to create a query language specialized for topic maps. There are a few proposals of different query languages:

- TMQL (Topic Maps Query Language), in syntax similar to SQL provided for querying and manipulation of data (insert, delete, update) [17],
- AsTMa? - similar to SQL language with queries in the form of LET - IN - WHERE - RETURN expressions. However native data format for a map is AsTMa= [1].
- Tolog – based on logic programming language prolog. In consequence topic map data have to be stored in the form of facts and rules. [4]

Moreover some known visualization techniques such as hyperbolic trees, virtual worlds [11] and cone trees [5] can be used for representing topic maps in a more transparent and effective way.

4. E-ENZYME KNOWLEDGE DATABASE – TOPIC MAP IMPLEMENTATION

The e-enzyme system — worked out by authors — is the internet database of enzymes i.e. compounds that regulate some chemical reaction occurring in cells of organisms accessible via the web. This application is based on topic maps. It gathers:

- information about enzymes' names, identification numbers, etc.
- a division according to the reaction that enzymes catalyze,
- compounds that are related to particular enzymes,
- information about optimal environment for enzymes (temperature, pH),
- organisms in which enzymes are located,

- references to literature about enzymes

There were a few steps of designing the e-enzyme topic map database. The three main are listed below [11]:

- a domain analysis (the domain definition for the system and determining which fragment of the domain the system should cover),
- extraction of main topics (topic types/classes) – building declarative part of the map,
- implementation of the topic map.

4.1 ENZYMES IN A NUTSHELL - DOMAIN ANALYSIS

All enzymes are classified into six classes according to the type of reaction they catalyse. Every class consists of other subclasses. The whole class hierarchy has four levels. Every enzyme has a **recommended name**, **systematic name** and can have some additional names – **synonyms**. Some identification numbers: **EC number** and **CAS registry number** can be assigned to each enzyme. **Enzymes** are compounds that accelerate the rate of a chemical **reaction**. They **react with** other compounds named **substrates** forming transitional enzyme substrate complex. Some enzymes can attach to substrates only with the help of **metal ions**. When a complex fall apart, a **product** of the reaction **is released**. There are some factors which affect the activity of enzymes. Main factors are: **temperature** and **pH**. Enzymes can be found only in living **organisms**. Biologists worked out an official taxonomy of organisms. From the point of view of topic maps it is important that the classification of organisms is hierarchical.

4.2 BULIDING DECLARATIVE LAYER AND DATA LAYER OF THE E-ENZYME TOPIC MAP

A domain analysis (presented in 4.1) allows to create a declarative (conceptual) layer [16] of a topic map. This layer consist of types of topics, associations and occurrences. Most above emphasised topics can be utilized for topic types and occurrences types. Two topics: systematic name and synonym were defined to provide scope for topics.

In a e-enzyme topic map four associations types were designed:

- ‘enzyme location in organism’ – linking instances of topics: ‘organisms’ and ‘enzymes’,
- ‘enzyme reaction with substrate’ – linking instances of topics: ‘substrates’ and ‘enzymes’,
- ‘enzyme inhibition’ – linking instances of topics: ‘inhibitors’ and ‘enzyme’,
- ‘product release’ – linking instances of topics: ‘ products’ and ‘enzymes’

Having declarative layer (types) i.e. the model of the map, concrete data – instances (topics, occurrences) can be assigned to particular types. In this way the data layer is obtained.

A fragment of e-enzyme topic map describing information about ‘pinene synthase’ enzyme is presented on a fig. 2.

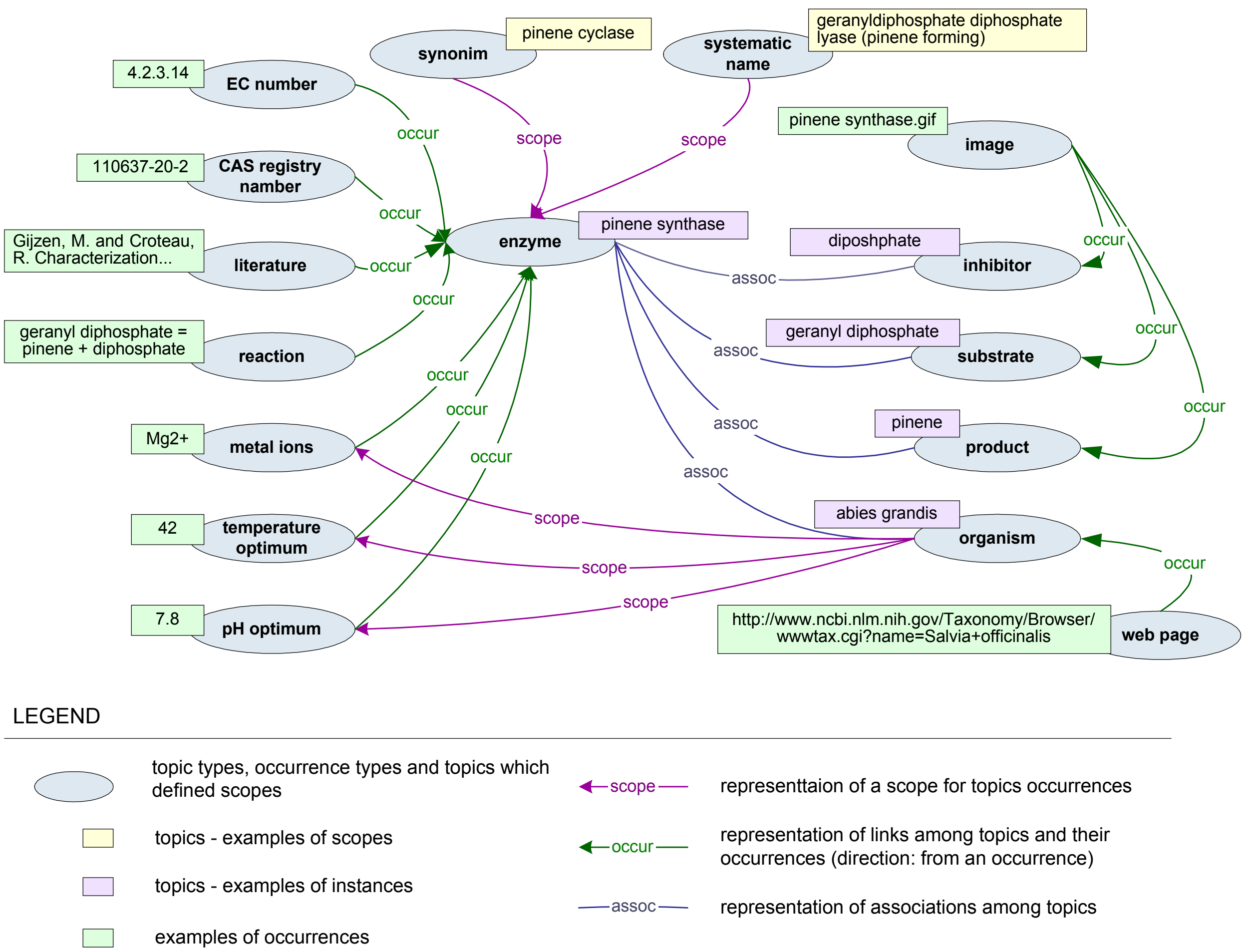

Fig.2. The e-enzyme topic map for ‘pinene synthase’ enzyme

4.3 TOPIC MAP IMPLEMENTATION

The e-enzyme implementation process was divided into two parts: creating topic maps using Empolis k42 knowledge server and developing user interface which enabled users to navigate among topics, occurrences and associations. The topic map was built according to the declarative layer.

System e-enzyme provides a few mechanisms for searching information about particular enzymes:

- simple and advanced search,
- enzymes tree,
- meta index.

Simple search require user to write down any phrase. The e-enzyme system returns list of enzymes which contain this phrase in any topic or occurrence related to them. Advanced search gives a possibility to determine features of enzymes that are to be found. User can specify name, EC number, CAS registry number, inhibitor, substrate, product, organism, reaction, metal ions, pH and temperature optimum. List of enzymes that meet user criteria is presented. After choosing a particular enzyme detailed information are displayed (fig. 3).

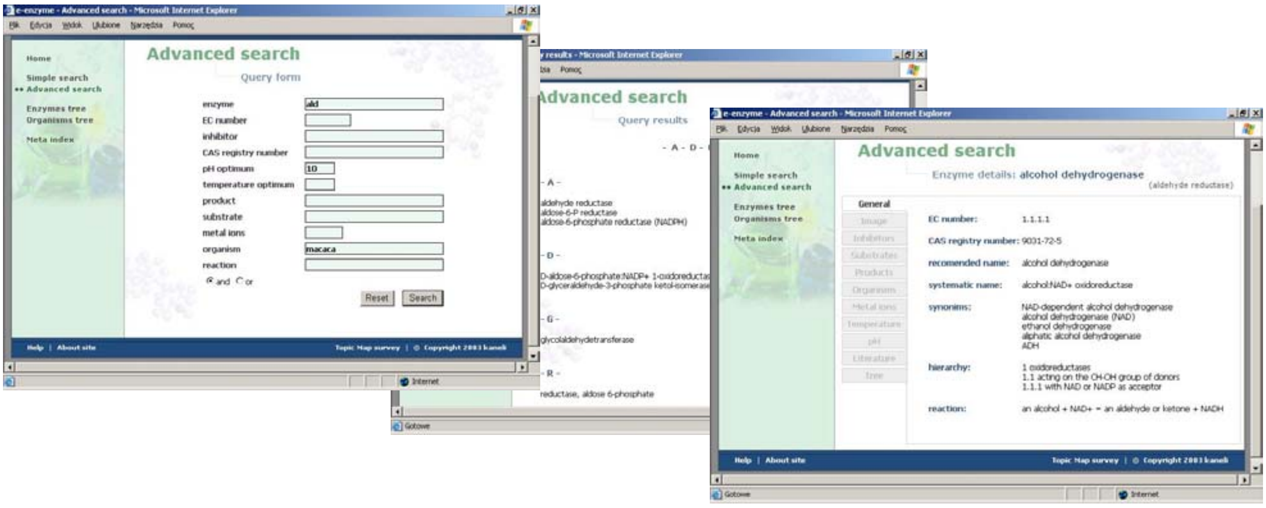

Fig.3. The advanced search in the e-enzyme system

The enzymes tree is a mechanism which enables to navigate enzymes hierarchy in the form of the hyperbolic tree (fig. 4). The hyperbolic tree is a way of visualization of topic maps implemented in a k42.

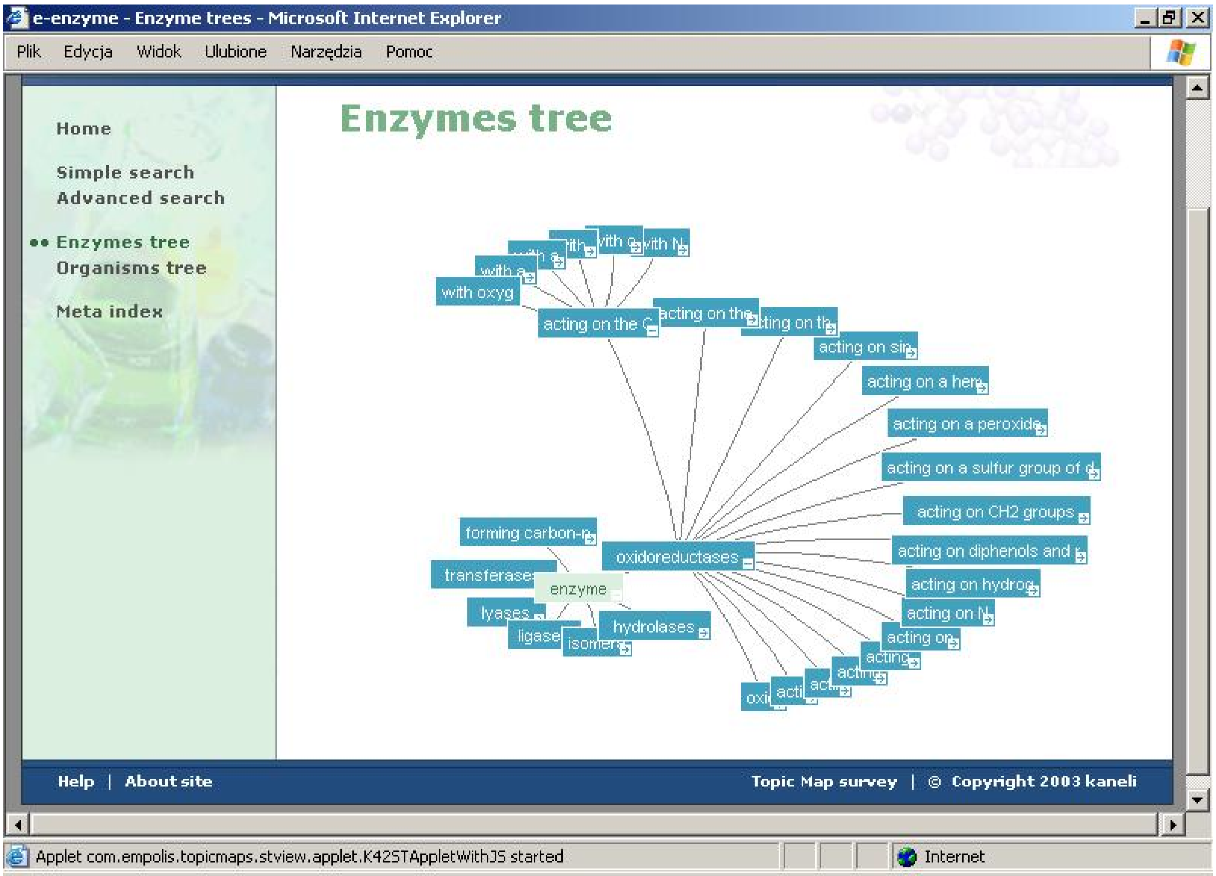

Fig.4. Tree view of enzymes' hierarchy in the e-enzyme system

Meta index allow reaching a particular enzyme through compounds (substrates, products, inhibitors) and organism associated with it.

5. TOPIC MAPS MAIN PROBLEMS

While developing the e-enzyme system many weaknesses of topic maps have been found. Let's focus on some of them:

- In most cases the ontology creation has to be a manual process. This requires in-depth analysis of a modelled domain. That is of course the significant problem at any knowledge modelling approaches.
- There is an immense need of a powerful and efficient tool for design, creation, storing and merging topic maps. This tool should enable to use all of mechanisms for modelling information supplied by topic maps. Existing systems, including k42, do not meet these requirements.
- There is no distinct border between declarative and data layer of topic maps. The next version of topic map ISO standard should provide Topic Maps Templates (TMT) which would enable to separate a topic map schema from data.
- Topic map ISO standard does not cover any validation mechanism. There is no way to ensure the quality of a map e.g. information that EC number is always in a form of *n.n.n.n* can not be included to topic maps
- A lack of a standard query language is also a serious problem. Topic Map Query Language (TMQL) has not been announced as a standard thus existing tools use their own languages.
- A powerful graphic interface for navigation is needed in order for fully utilization abilities of topic maps. There are some methods of visualization e.g. hyperbolic and cone trees, virtual words, etc. but many research has to be done to make them suitable for topic maps [11].

6. SUMMARY

ISO Topic Maps is a standard whose paradigm is based on giving structure to unstructured information. They represent part of the world in a way that human being perceives it – people are used to separating objects and relationships among them. Therefore topic maps are flexible tool for organizing data. In particular they can be utilized for:

- building independent knowledge database (like e-enzyme),
- creating semantic layer in Content Management Systems [3],
- integrating data from heterogeneous resources, databases [12, 13] and warehouses [2],
- enabling navigation among tasks in workflow processes [6].

Nowadays, works on a standard of a query language (TMQL) are in progress. Probably a language ensuring quality of topic maps (Topic Maps Constraint Language TMCL) [16] will be drown up next. In addition many people developing topic maps perceive also a need of existence of Topic Map Templates standard which would enable to separate declarative part of a map from its data [16]. All of these mechanisms would help topic maps to become more efficient standard independent from developing tools and suitable for web knowledge representation systems.

REFERENCES

[1] BARTA, R.: *AsTMa? Language Definition*. Bond University Technical Report, 2003, http://astma.it.bond.edu.au/astma%3F-spec.dbk?style=printable.

[2] BRUCKNER R.M., LING T.W., MANGISENGI O., TJOA A.M.: *A Framework for a Multidimensional OLAP Model using Topic Maps*. Proceedings of the 2nd International Conference on Web Information Systems Engineering (WISE'01), Vol. 2, IEEE Computer Society 2001, pp. 109-118.

[3] GARSHOL, L.M.: *Tolog: topic maps query language*. Proceedings of the XML Europe 2001 Conference, IDEAlliance, 2001. http://www.ontopia.net/topicmaps/materials/tolog.html

[4] GARSHOL, L.M.: *Topic maps in content management*, 2003. http://www.ontopia.net/topicmaps/materials/itms.html

[5] Le GRAND B., SOTO M.: *Visualisation of the Semantic Web: Topic Maps Visualisation*. International Conference on Information Visualisation, 10-12 July 2002, London, UK, IEEE Computer Society, pp. 344-351.

[6] HUTH C., SMOLNIK S., NASTANSKY L.: *Applying Topic Maps to Ad Hoc Workflows for Semantic Associative Navigation in Process Networks*. Seventh International Workshop on Groupware CRIWG, IEEE Computer Society, 2001, pp. 44-49

[7] *ISO/IEC 13250:2000 Document description and processing languages – Topic Maps*, International Organization for Standarization ISO, Geneva, 2000.

[8] KAZIENKO P., GWIAZDA K.: *XML na poważnie*. Helion, Gliwice 2002. (in Polish)

[9] LACHER M.S., DECKER S.: *On the Integration of Topic Map data and RDF data*. 1st International Semantic Web Working Symposium (SWWS `01), Stanford University, Stanford, CA, July 29-Aug 1, 2001, Jul. 2001.

[10] LASSILA O., SWICK R.R.: *Resource Description Framework (RDF). Model and Syntax Specification. W3C Recommendation 22 February 1999*. W3C, 1999.

[11] LITWIN M.: *Topic Maps. Master Thesis*. Wrocław University of Technology, Department of Information Systems, Wrocław, 2003 (in Polish).

[12] LUCKENEDER T., STEINER K, WÖß W.: *Integration of Topic Maps and Databases: Towards Efficient Knowledge Representation and Directory Services*. Database and Expert Systems Applications, 12th International Conference, DEXA 2001 Munich, Germany, 2001, LNCS 2113, pp. 744-753.

[13] OUZIRI, M.; VERDIER, C.; FLORY, A.: *Semantic Indexing for Intelligent Browsing of Distributed Data*. Proceedings of the International IIS: IIPWM´03 Conference, Zakopane, June 2-5, 2003, Advances in Soft Computing, Springer Verlag 2003, pp. 189-198.

[14] PARK, J.; HUNTING, S., eds: *XML Topic Maps - Creating and Using Topic Maps for the Web*. Boston, Addison Wesley, 2002.

[15] PEPPER S., MOORE G., eds: *XML Topic Maps (XTM) 1.0. TopicMaps.Org Specification*. TopicMaps.Org, 06 Aug 2001. http://www.topicmaps.org/xtm/1.0/.

[16] RATH, H.H.: *Technical Issues on Topic Maps.* Proceedings of the Metastructures 1999, Montreal, Quebec, Canada, 1999.

[17] WRIGHTSON A.: *TMQL Draft (Topic Map Query Language).* Ontopia, BSI, 2000. http://www.y12.doe.gov/sgml/sc34/document/0186.doc.