THE RISE, SHINE AND DECAY OF TOPIC MAPS

Daniel Dragu[1]; Ioan Emeric Köles[2], Elena Muncut[3]

1 A University "Aurel Vlaicu" of Arad, B-dul Revoluţiei Nr. 77, România. P. O. BOX 2/158 AR, dragudaniel@yahoo.com
2 A University "Aurel Vlaicu" of Arad, B-dul Revoluţiei Nr. 77, România. P. O. BOX 2/158 AR, koles54@yahoo.com
3 A University "Aurel Vlaicu" of Arad, B-dul Revoluţiei Nr. 77, România. P. O. BOX 2/158 AR, muncutstela@yahoo.com

ABSTRACT: The paper presents a brief review of the current state of Topic Maps (TM) and an analysis of its evolution on the market, trying to understand why a technology with brilliant perspectives is no longer in the spotlight. In the end, the paper points out some of the reasons that led to a decline of the visibility of TM in the world of Computer and Information Technology.
KEYWORDS: Topic Maps, semantic technology, knowledge representation.

1. INTRODUCTION

Topic Maps is a standard (ISO/IEC 13250) designed to represent, manage and interchange information in a subject centric way (fig. 1). Its creation process began in the 90's and has been going on for several years. The efforts of its contributors were initially focused on the development of a data model able to represent indexes, glossaries, thesauri and table of contents, but they realized that TM could increase the semantics of the traditional databases by connecting concepts to the various places where they occur. They could create and manage knowledge bases. There were so much hope and confidence about the new technology and its widely adoption that some placed TM and Resource Description Framework (RDF) together, forming the foundation of the Semantic Web [1].

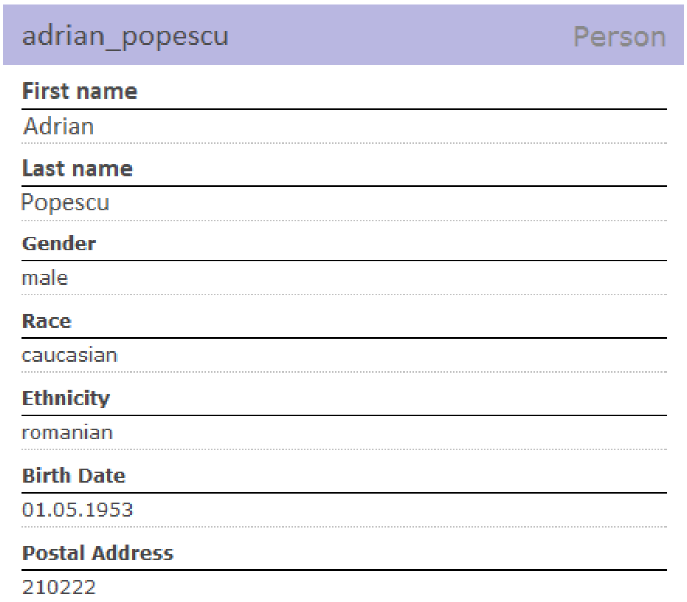

Figure 1. Subject-centric way to visualize information

After about two decades from the start, TM seems to be neglected more and more [2], and this happens as there is a growing demand for knowledge bases as support for the systems based on artificial intelligence.

2. WHAT A TOPIC MAP IS

According to the specifications within Clause 3 of the TMDM standard, Topic Maps is a technology used to encode knowledge. The term "topic map" refers an instance of TM. A topic map consists of a set of topics and associations [3].

TMDM is a 2-layer data model (fig. 2) consisting of:

- a layer of knowledge, including topics and associations;
- a layer of resources that may exist in various forms: digital or not, encoded as text in any format or notation, graphics, audio, video, etc. [4].

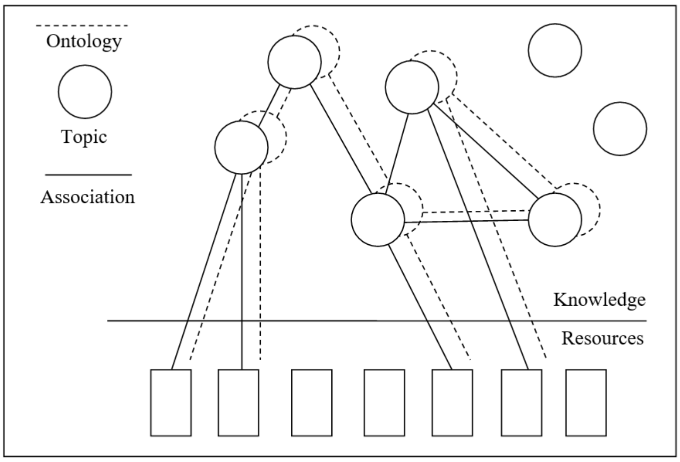

Figure 2. Topic Maps

A topic is a representation of a subject of interest. A subject is similar to the term "resource" from RDF and is defined in TMDM as "anything whatsoever, regardless of whether it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever" [3].

TMDM supports n-ary associations, meaning that two or more topics may play different roles in a relationship that connects them.

Contexts can be modelled through the scope, which determines the validity of TM artefacts [3].

The two layers of the TMDM are linked together through the occurrences and links. An occurrence connects a topic with a relevant resource, and a link points to a location where pertinent resources for that topic can be found.

Topics are representations of subjects. A subject should be uniquely identified, and there was a proposal stating that a subject is identified using a Published Subject Identifier (PSI) and "A Published Subject Identifier must be a URI" [5]. At the time of the development of XML Topic Maps (XTM) 1.0, three lists of PSIs was published on the Internet, identifying subjects within XTM 1.0 core, countries and languages of the world [6]. The real value of unique identification can be uncovered during the merging process, provided by TMDM to avoid redundant TM artefacts in a topic map, and, by extension, to join together two or more topic maps.

A topic map is a way to organize and represent knowledge so that the result is easily understood by both humans and computers.

Only a few of the TMDM specifications have been reviewed, but they are enough to make a picture of a data model able to represent "anything whatsoever"[3].

3. THE RISE, SHINE AND DECAY

Even from the beginning, there were lots of initiatives and resources involved in the process of TM development, so today the standard includes a data model [3], an XML syntax for interchanging instances of TM (fig. 3), a canonical XML format to control the serialization processes of TM instances, a reference model specifying how to constrain and interpret TM artefacts, a text-based notation named compact syntax, and as part of another standard, a constraint language for the development of TM schemas "in a precise and machine-readable form" [7]. Many researchers and enthusiasts have created various tools in order to make TM stronger and stronger, and the limits seemed not to exist if we refer to its ability to represent information. There were proposals for:

- simpler syntaxes, like JSON Topic Maps (JTM) [8],
- graphical notation and visualization tools [9],
- TM ontologies and query languages [10],
- interoperability with other data models,
- data modelling technics and semantic integration [11],
- merging tools: MaJorToM – Merging Topic Maps Engine,
- automatic acquisition of knowledge from other various sources, like relational databases, documents, etc.,
- programming interfaces for various programming languages, in order to facilitate the rapid development of applications: TMAPI, phptmapi, etc.;
- desktop tools to create, collect, store, manage, retrieve and publish information and knowledge;

online information management systems that implement, more or less, the specifications of the data model, like Topincs [12], Ontopia, Ruby Topic Maps etc., to name only a few of them.

```xml
<topic id="dunarea">
    <baseName>
        <baseNameString>The Danube</baseNameString>
    </baseName>
    <instanceOf>
        <topicRef xlink:href="#river">
    </instanceOf>
    <occurrence>
        <instanceOf>
            <topicRef xlink:href="#webpage" />
        </instanceOf>
        <resourceRef>
            xlink:href="https://ro.wikipedia.org/wiki/Por%
        </resourceRef>
    </occurrence>
</topic>
<topic id="hpf1">
    <instanceOf>
        <topicRef xlink:href="#hydropower_plant">
    </instanceOf>
    <baseName>
        <baseNameString>Hidrocentrala Portile de Fier I</b
    </baseName>
</topic>
<association>
    <instanceOf>
        <topicRef xlink:href="#built_on" />
    </instanceOf>
    <member>
        <roleSpec>
            <topicRef xlink:href="#hydro_power_plant" />
        </roleSpec>
        <topicRef xlink:href="#hpf1">
    </member>
    <member>
        <roleSpec>
            <topicRef xlink:href="#source_of_water" />
        </roleSpec>
        <topicRef xlink:href="#dunarea">
    </member>
</association>
```

Figure 3. A topic map in XML format

Besides these, the series of six scientific conferences on TM, between 2005 and 2010, known as Topic Maps Research and Applications (TMRA) [13], aimed to disseminate the achievements in the field.

4. THE REASONS

There were plenty of opportunities for development and widespread integration at a time when a paradigm shift to the subject-centric computation led to a new programming approach - object-oriented programming. Unfortunately, TM was not received with the same enthusiasm on the market, and users still preferred relational databases to store their data. This maybe happened because the data model is so versatile that its instances can be stored as well in RDBMS (Relational Database Management System), and users might think that a topic map is nothing else but a relational database. They're wrong! A topic map is a way to organize data and information, enriching them with semantics for an easier interchange and integration into artificial intelligence systems. An accurate implementation of the TMDM is a knowledge base.

So, what were the reasons that impeded the widespread use of TM?

In our opinion these are some of the most important reasons that led to a limited usability of TM technology:

- creating, managing and maintaining knowledge bases are more difficult than relational databases;
- knowledge bases require a higher level of abstraction than relational databases, so not all developers will migrate from a well-known traditional way of storing data to a new more abstract data model;
- TM requires highly specialized human resources in the field of represented domain and TM specialists working together, so, we can say that TM implementations are more appropriate for companies or enthusiastic collaborative initiatives that can cope with these costs;
- there were too few and too specialized implementations, forcing potential users to dig deep in order to use them;
- there are applications based on TM paradigm, but their owners avoid to mention that;
- decreasing interest of direct contributors for the further development of technology and maintenance of tools [2].

5. HOW TO USE TOPIC MAPS

A topic maps is not just a simple database, even if it can be stored in one. The difference is given by the semantics inside them. Relational databases consist of tabular organized data, while topic maps are based on ontologies that organize the data using associations between subjects, in a manner similar to the human brain.

Assuming that there are two related topics, *hpf1* and *dunarea*, according to the TMDM specifications, the relationship between those two topics can be represented as in figure 4.

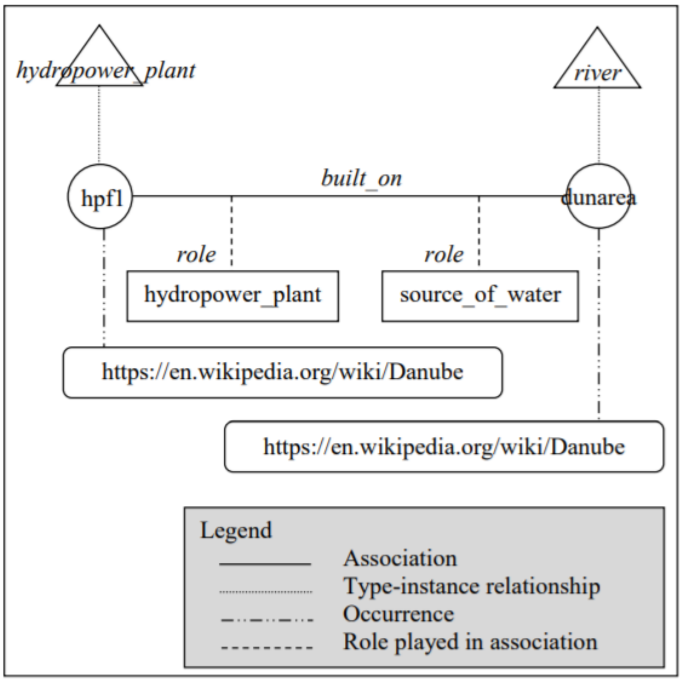

Figure 4. Graphic representation of an association

The following example is a topic map containing information about the previous introduced subject, "Porțile de Fier I", the largest hydropower plant on the Danube river and part of the national energy system of Romania (fig. 5).

```
/*topics*/

        [hpf1:hydropower_plant="Hidrocentrala Portile de Fier I"]
        [dunarea:river="Dunarea"]
        [sen:"Sistemul Energetic National"]
        [electrical_power:energy]
        [romania:country="Romania"]

/*associations*/
        built_on(hpf1:hydropower, dunarea:source_of_water)
        part_of(hpf1:component, sen:system)
        produces(hpf1:source, electrical_power:result)
        owned_by(sen:property, romania:owner)

/*internal_occurrences*/
        {hpf1, description,
                "the largest hydropower plant on the Danube River"}
        {hpf1, installed_power, 1080 MW}
        {hpf1, location, 44.671453, 22.528600}

        {dunarea, description, "Europe's second longest river"}
        {dunarea, length, 2860 km}

        {sen, description, "all installations for the production,
                transport, distribution and use of electricity
                in Romania"}

        {romania, description, "a sovereign state located at the
                crossroads of Central, Eastern, and Southeastern Europe"}

/*external_occurrences*/
        {romania, website, https://ro.wikipedia.org/wiki/Rom%C3%A2nia}
        {dunarea, website, https://en.wikipedia.org/wiki/Danube}
        {sen, website,
                http://www.transelectrica.ro/web/tel/sistemul-energetic-national}
```

Figure 5. An example of a topic map

Topic maps can be represented in various formats, but a simple text notation, LTM - Linear Topic Map Notation, facilitates the human readability.

In previous examples, the topic *hpf1* stands for the subject of interest "Hidrocentrala Porțile de Fier I", and the topic *dunarea* represents the Danube River. The two topics play roles of *hydro_power_plant*, and *source_of_water*, respectively, in a relationship referred here as *built_on*.

The constructs *hydropower_plant* and *river* are topic types, and they are involved in a type-instance relationship with the topics *hpf1*, and *dunarea*, respectively.

Occurrences in a topic map give relevant information in order to create an accurate view of the subjects represented by topics.

Querying topic maps can be done over any topic map construct (topics, topic names, topic types, occurrences, associations, roles and so on).

Merging topics is another important facility that should be considered in the development process of TMDM instances, in order to eliminate redundancy of TM constructs [3]. The subjects should be uniquely identified [14] using subject identifiers, subject indicators and subject locators, allowing to put together, in a single topic, the information from all representations of a specific subject of interest (fig. 6). Assuming that, in a topic map, there are two topics representing the same subject, the result of the merging process is a new topic (fig. 7).

```
/*topics*/
        ...
        [hpf1:hydropower_plant="Hidrocentrala Portile de Fier I"]
        ...
        [djerdap:hydropower_plant="HE Djerdap"]

/*associations*/
        built_on(hpf1:hydropower, dunarea:source_of_water)
        part_of(hpf1:component, sen:system)
        produces(hpf1:source, electrical_power:result)
        ...
        used_by(djerdap:hydropower_plant,
                serbia:country, romania:country)

/*internal_occurrences*/
        {hpf1, description,
                "the largest hydropower plant on the Danube River"}
        {hpf1, installed_power, 1080 MW}
        {hpf1, location, 44.671453, 22.528600}
        ...
        {djerdap, serbian, "Хидроелектрана Ђердап"}
        {djerdap, location, 44.671453, 22.528600}

/*external_occurrences*/
        {hpf1, subject_indicator,
                @https://ro.wikipedia.org/wiki/Por%C8%9Bile_de_Fier_I}
        ...
        {djerdap, subject_indicator,
                @http://www.djerdap.rs/}
        {djerdap
        , subject_indicator,
                @https://ro.wikipedia.org/wiki/Por%C8%9Bile_de_Fier_I}
```

Figure 6. Two topics representing the same subject

The names and occurrences of the new topic are given by the reunion of names and occurrences of the two merged topics. The new topic replaces the two merged topics in any association every of them plays a role(fig. 5).

```
/*topics*/
        [hpf1_djerdap:hydropower_plant=
                "Hidrocentrala Portile de Fier I",
                "HE Djerdap"]

/*associations*/
        built_on(hpf1_djerdap:hydropower, dunarea:source_of_water)
        part_of(hpf1_djerdap:component, sen:system)
        produces(hpf1_djerdap:source, electrical_power:result)
        ...
        used_by(hpf1_djerdap:hydropower_plant,
                serbia:country, romania:country)

/*internal_occurrences*/
        {hpf1_djerdap, description,
                "the largest hydropower plant on the Danube River"}
        {hpf1_djerdap, installed_power, 1080 MW}
        {hpf1_djerdap, location, 44.671453, 22.528600}
        {hpf1_djerdap, serbian, "Хидроелектрана Ђердап"}
        {hpf1_djerdap, location, 44.671453, 22.528600}

/*external_occurrences*/
        {hpf1_djerdap, subject_indicator,
                @https://ro.wikipedia.org/wiki/Por%C8%9Bile_de_Fier_I}
        {hpf1_djerdap, subject_indicator,
                @http://www.djerdap.rs/}
```

Figure 7. The result of merging topics

The merging process can be applied to many topic maps, but precautions are required to avoid damaging existing representations. Constructs according to other data models can be integrated in topic maps due to the representing neutrality of the TMDM.

These are only a few examples of how topic maps can be used, but its range of implementation is much wider than that.

Beyond these, TM exists and there are still hopes for that data model to be adopted in different areas of work. Novel unconventional technologies provide a field for new implementation opportunities to facilitate knowledge discovery and sharing, information integration, data mining, intelligent control, signal processing and so on. The specialists in these branches could benefit from large knowledge bases in order to adapt more easily to current technologies and, implicitly, to make their products more and more competitive on the market. The same knowledge bases could be further used as a cornerstone for controlling optimization and prediction processes as well as for the implementing reasoning and decision support systems.

6. CONCLUSION

For simple stand-alone data collections, it is more appropriate to use relational databases, while TM are a better solution for systems where there is a need for information interchange, automated knowledge acquisition, reasoning, natural language processing, in learning machines or in other artificial intelligence systems.

The resolution and the correctness of the topic maps depend of many aspects, like the application requirements, the knowledge of the topic map builder, or the boundaries of the represented domain, to name only a few. Users have to be advised that information within topic maps may be altered in time, so there is a need for a permanent and specialized maintenance of TM constructs.

The standard ISO/IEC 13250 specifies that a TM abstract structure may be represented in a wide variety of forms, as simple text files, relational databases, internal data structures of running programs, or even the mind of human being [3].

Syntaxes are just ways to formalize the representations and they are not a must when someone wants to create a topic map. Therefore, why not to enlarge the view and see that the idea of TM resides in almost any type of information representation, accomplishing its original purpose: topic maps have to be "neutral envelopes, hospitable to any existing or future schema for knowledge representation" [1].

7. REFERENCES

1. M. Biezunski, "Introduction to the Topic Maps Paradigm", J. Park, S. Hunting (Eds.), XML Topic Maps: Creating and Using Topic Maps for the Web, Addison-Wesley Professional, 2002, pp 17.
2. M. Biezunski, "Topic Maps Now", June 23, 2017 https://www.xml.com/articles/2017/06/23/topic-maps-now/
3. ISO/IEC 13250-2:2006 "Information technology -- Topic Maps -- Part 2: Data model".
4. S. Pepper, "The TAO of Topic Maps: Finding the Way in The Age of Infoglut" Conference Proceedings XML Europe 2000, Paris, France, 2000, pp. 167-180.
5. S. Pepper, "Published Subjects: Introduction and Basic Requirements", Working Draft 9 Mar 2003, https://www.oasis-open.org/committees/download.php/1217/wd-pubsubj-introduction-01.htm
6. S. Pepper, G. Moore, "XML Topic Maps (XTM) 1.0,TopicMaps.Org Specification", http://www.topicmaps.org/xtm/1.0/
7. ISO/IEC 19756:2011 Information technology -- Topic Maps -- Constraint Language (TMCL)
8. JSON Topic Maps (JTM), http://www.cerny-online.com/jtm/
9. H. Thomas, T. Redmann, M. Pressler, B. Markscheffel, "GTMalpha – Towards a Graphical Notation for Topic Maps". L. Maicher, L. M. Garshol, (eds.): Subject-centric Computing. Fourth International Conference on Topic Maps Research and Applications, TMRA 2008, Leipzig, Germany, pp. 310-316.
10. L. M. Garshol, "Tolog – a Topic Maps Query Language." Charting the Topic Maps Research and Applications Landscape. Springer Berlin Heidelberg, 2006, pp. 183-196.
11. T. Neidhart, P. Pinchuk, and B. Valentin, "Semantic Integration of Relational Data Sources With Topic Maps", Linked Topic Maps - Fifth International Conference on Topic Maps Research and Applications, TMRA 2009, 2009, pp. 185–192.
12. R. Cerny, "Topincs: A Software for Rapid Development of Web Databases", Proceedings of the Internetional Conference on Knowledge Management and Information Sharing, KMIS 2011, SciTePress, 2011, pp. 187-194.
13. Topic Maps Research and Applications, http://tmra.de/
14. L. M. Garshol, "What Are Topic Maps", September 11, 2002, https://www.xml.com/pub/a/2002/09/11/topicmaps.html