Proceedings of the Pakistan Academy of Sciences 51 (2): 101–114 (2014)

ISSN: 0377 - 2969 (print), 2306 - 1448 (online)

Pakistan Academy of Sciences

Research Article

The Jigsaw of Resource Description Framework (RDF) and Topic Maps Serialization Formats: A Survey

Fakhre Alam, Muhammad Abid Khan, Shaukat Ali* and Shah Khusro

Department of Computer Science,
University of Peshawar, Peshawar, Pakistan

Abstract: The dawn of the 21st century saw the emergence of a number of technologies which not only revolutionized human lives but also opened new avenues of research. Semantic Web technology is one of the technologies which attracted tremendous research attention around the globe. To successfully implement Semantic Web, technologies have been developed to effectively represent, create, and navigate metadata relationship among the Web resources. Resource Description Framework (RDF) and Topic Maps (TMs) emerged as the leading technologies of the newly envisioned Semantic Web. RDF and TMs are independent technologies using different mechanisms to represent semantic relationships between the Web resources. A number of serialization formats have been investigated by the researchers in the past few years for successful storage and transmission of RDF and TMs data. However, different representation mechanisms and serialization formats result into the problem of interoperability, potentially resulting in the division of the Semantic Web into two islands. This paper presents a comprehensive survey of the available serialization formats of RDF and TMs by categorizing, analysing, and evaluating them using a set of criteria. For evaluation purpose, simple book ontology is developed in both RDF and TMs formats. It has been observed that serialization formats vary in different aspects but are powerful enough to encode their respective data models and the interoperability among them is possible, subject to some additional effort. This investigation provides a compact platform for researchers to solve the interoperability problem among serialization formats to accelerate the growth of Semantic Web.

Keywords: Serialization, Semantic Web, Resource Description Framework (RDF), Topic Maps (TM), Extensible Markup Language (XML)

1. INTRODUCTION

The World Wide Web (WWW) [1] is an Internet system in which billions of web pages containing text, images, audio and videos are interlinked with each other through hypermedia links. Despite of all of the efforts for the perfection, advancement and widespread use, the information in the current Web is still presented in human readable format with lack of any semantic and annotation capabilities, enough for machines to deduce any new information from the existing by its own [2]. Today huge amount of information is available on the Web in different structural formats imposing the burden of retrieving and interpreting of relevant information on the users. Thus, making users overwhelmed while searching for specific information.

To overcome the limitations of the current Web, Tim-Berner-Lee[1] (inventor of the WWW) coined the idea of Semantic Web, also known as the modern Web. Semantic Web is an extension of the current Web, insisting the creation of implicit and meaningful relationships among the Web resource in a manner to be directly understandable by the machines [3]. In Semantic Web, machines will integrate information, apply computation on the information, and organize them into a format to be reusable across wide array of domains. Thus, will establish an environment where people and

Received, May 2013; Accepted, May 2014
* Corresponding author: Shaukat Ali; Email: shoonikhan@upesh.edu.pk

[1] http://www.w3.org/People/Berners-Lee/

machines will function in cooperation and will make the availability of right information at the right time and the right place possible.

The vision of Semantic Web revolutionized Web by taking it into new avenues of research. To light the lamp, researchers around the globe put forwarded their contributions for turning the Semantic Web vision into reality. To achieve the goals of Semantic Web, a list of technologies were proposed by the research communities which will not only help in understanding Semantic Web but will also accelerate the development rate of Semantic Web applications. Among the other technologies, Resource Description Framework (RDF) and Topic Maps (TMs) technologies developed respectively by W3C[2] and ISO[3] served as the backbone of Semantic Web and gained high level of popularity. An RDF model represents metadata enabled relationships among web resource in an explicit and precise manner to be easily interpretable by the machines [4]. Each RDF model is composed of statements where each statement relates two Web resources by using the analogy of subject, predicate, and object. The idea of Topic Maps was originally developed for the representation of back of the book index construction [5]. The original idea of Topic Maps was further extended by the researchers and used it for wider applications like to represent exchange and convey knowledge on the Semantic Web. Information in Topic Maps are represented in the form of topic, associations between topics and their occurrences.

Serialization is the process in which data in one format is semantically converted into another format for the storage and transmission purpose[6]. The convertible data can be reused later in another computing environment, regardless of any change in the original data format. It is due to the serialization that different types of data formats can be stored and run on different hardware platform regardless of their underlying architecture, thus, promoting interoperability. Serialization also provides simplicity and common I/O interface to store and transform data from one format to another. XML provides persistent method for the interchange of data on the web to be stored or communicated regardless of the programming languages in a human readable format [7]. A number of serialization formats are proposed by the research communities, academia, and organizations for RDF and TMs Semantic Web technologies with the view of expressibility, easy to create, easy to understand by both humans and machines, and take less network bandwidth while in transmission.

While being overwhelmed with a number of serialization formats, it is rather cumbersome for researchers and users to select an appropriate one and find solutions to problem of their inter-conversion. This research paper is aimed to provide a comprehensive overview and analysis of the available RDF and Topic Maps serializations formats, covering all of their possible aspects, and pros and cons. Main contributions of the paper are:

- The topic is almost unique in its integrity and opens new area of research.
- To provide a compact platform for the users and researchers to grab almost all of the relevant possible information about the topic in a single document.
- To organize and classify the available literature about the topic in an attractive manner to catch and boost interest of the new researchers in the area and take them into new avenues of research.
- To briefly explain and analyze the available serialization formats which are currently available in various domains and with their success stories and common reasons of failure.
- Gaining practical experience of inter-converting serialization formats belonging to the two Semantic Web islands.

2. EVALUATION FRAMEWORK AND METHODOLOGY

A serialization format is a way of encoding information that can be stored (e.g., in a file, or memory buffer) so that, when passed between machines, it can be parsed and understand, and reconstructed in the same or another computing environment. Resource Description Framework (RDF) and Topic Maps (TMs) technologies, fulfilling the vision of Semantic Web, enrich the scattered resources in the Web with metadata to solve the problem of knowledge sharing

[2] http://www.w3.org/
[3] http://www.iso.org/iso/home.html

and integration. Although RDF and TMs are independent technologies from different standard organizations but their goal is the same, which is to represent the semantic relationship between web resources. RDF and Topic Maps data needs to be stored in a data structure (i.e., file format) which can be transmitted through a network to be used in another computing environment without changing in the original data format. The notions used to record RDF and Topic Maps data in a data structure is called serialization format. Practically, serialization format and RDF/Topic Maps are two different things. A serialization format describes an encoding technique of information, whereas, a RDF/Topic Map represents a mental conceptual data model of information in a domain. But, to make a RDF/Topic Map data model understandable to both humans and computers (especially), it needs to be represented in a serialization format. In other words, serialization is like the grammar of a language, while the data model is the informational content behind words. For example, the word "Green" spoken aloud in English is the serialization, while a data model is a way to define the concept of

Table 1. Detailed analysis and evaluation of Resource Description Framework serialization formats using set of criteria.

Category	Serialization	Exp/Comp	Simplicity	Compactness	Tools Supp.	Applicability	Remarks
XML	**RDF/XML**	Very High	Simple and user friendlyz	Less Compact	Available	Small scale and large scale	Merge and interchange information from multiple sources, Used to create ontologies, Can be used with relational databases and RSS, Hard format, Difficult to implement, Not processable with some standard XML tools.
	TriX	High	Simple syntax	Compact	Available	Small scale	Represent RDF graph more consistently using XSLT, and Xquery, due to which processing speed is high and concise, Syntax can be extensible to represent any type of RDF graph, Verbose and poor human readable un-extended syntax, Inefficient transformation, Lack of hierarchy template match
Non-XML	**N3**	Medium	Simple and can easily parsed	Highly Compact	Available	Small scale and large scale	De facto standard, Optimize the logic and inference mechanism, Compact, persistent and readable syntax, Shorten absolute URI, Reduce verbose nature as compared to RDF/XML, Unknown and long term stability, Clear requirements for correlations of different features of N3 in RDF/XML format
	N-Triple	High	Simple and easier to read and write	Less Compact	Available	Large scale	Automated processing during the comparison of the actual results with our desired results, Store and transform RDF graph(Triples) data in plain text and lines, Easily processable with N3 and Turtle tools, Line by line Parsing, Redundancy in code and Lack of support for nested resources
	Turtle	High	Very Simple	Highly Compact	Available	Large scale	Contains all the capabilities of N-Triple, Compatible with N3, Encode any type of RDF graph, Greater conciseness and human readability, Used with the popular tools and API like Jena and OWL API, Reification of complex expressions and translations of OWL axioms into triples is difficult

Table 2. Detailed analysis and evaluation of Topic Maps serialization formats using set of criteria.

Category	Serialization	Exp/Comp	Simplicity	Compactness	Tools Supp.	Applicability	Remarks
XML	**XTM**	Very High	Simple	Not so Compact	Available	Small and large scale	Advance features than SGML, Flexible, user friendly, easily represent Knowledge Representation and navigation, Implemented on web browser, Cannot fully represent the whole Topic Maps paradigm, Occupy large space, Less effective in semantic integration and interoperability of Topic Maps constructs.
	CXTM	High	Complex	Less Compact	Available	Small scale	Provide test suite for testing different TM technology, Directly compare the two data structures, Provide more reliability then other formats, It is only the variation of XTM used for the judgment of two TMDM
Non XML	**HyTime**	Medium	Simple	Not Compact	Available	To any type of Hypermedia	Provides an abstract facility for the TM paradigm to be expressed in SGML, Link and organized related topics which are time and space critical, Fulfill the needs of addressing in a convenient way, General purpose and abstract standard
	CTM	Very High	Simple	Highly Compact	Available	Small and large scale	Short and simple text based notation, Represent the syntax of TMQL and TMCL precisely, Efficient parsing mechanism, Need special parser for TM constructs, Lack of support for item identifiers in TM constructs that are not topics
	GTM	High	Simple	Highly Compact	Not Available	Small and large scale	Easily modeled, presented and communicated between users Due to visual nature, Create similar view on TM for programmers, Can be used with UML in a single model yielding high level performance, Weak due to the separation of two levels one for TMDM and another for TMCL
	LTM	Low	Simple	Highly Compact	Available	Small scale	Simple and text based, Can be converted into XTM format, Efficient and take less space, User friendly and compact syntax, Not yet standardized, Superseded by CTM, Unsuccessful when come to the large TM

"Green" such that it is unambiguous whether you say "Green" or "Verde" or think about the color of a leaf [8]. A data model is an abstract model – it does not matter how one represent it as long as one stay true to its abstract properties. A serialization format typically encodes the information provided by the associated data modeling technology. To win the race, both of the competitors RDF and Topic Maps supports a variety of serialization formats varying in different aspects such as readability, expressibility, and compactness etc. for their data representation. However, a serialization format providing excellent overall results will not only help to boot the associated technology in the race but eventually the Semantic Web technology. In this paper, to give insight knowledge and understanding of the on hand RDF and Topic Maps serialization formats, they are evaluated against a set of criteria including xml and non-xml, expressibility and computability, simplicity, applicability, compactness, tools and APIs support, and remarks. In order to compare the different serialization formats, simple book ontology is developed and serialized into different RDF and Topic Maps serialization formats to derive conclusions for the criteria (Table. 1 and Table. 2). Our book ontology covers most of the basic information related to a book as well as properties (.i.e., data properties and object properties), and objects. The overall ontological graphs in RDF and Topic Maps are shown in Fig. 1 and Fig. 2, respectively.

- ***XML-base and non-XML-based***: RDF and Topic Maps serialization formats can be classified into XML-based and non-XML-based. XML was the first serialization format used to produce text-based encoding which could not only be readable by humans and machines but could also be communicated between disjoint systems regardless of programming languages. XML-based encoding, however, lost more compact byte-stream-based encoding but the recent advancements in storage and communication capabilities made file size less of a concern. Binary XML was proposed more compact than regular XML but was not readable by plain-text editors and could not be exchanged between disjoint systems. Non-XML serializations were designed to be easier to write by hand and easier to follow because of using notions (.i.e., tabular notion etc), which encodes the underlying information in a document to be easily recognizable as compared to XML-based sterilizations. JavaScript syntax based serialization: JSON is another more compact and lightweight plain-text serialization proposed as an alternative to XML-based serialization.
- ***Expressibility***: The expressiveness power of a serialization format determines the level of constructs and features (.i.e., syntactic sugars) provided by the format to encode complex information. A serialization format with high expressive power is obviously a liked one but often result in low computational performance. Similarly, a highly expressive format can serialize any type of data model either valid or invalid. For example, Notion 3 and RDF/XML can serialize

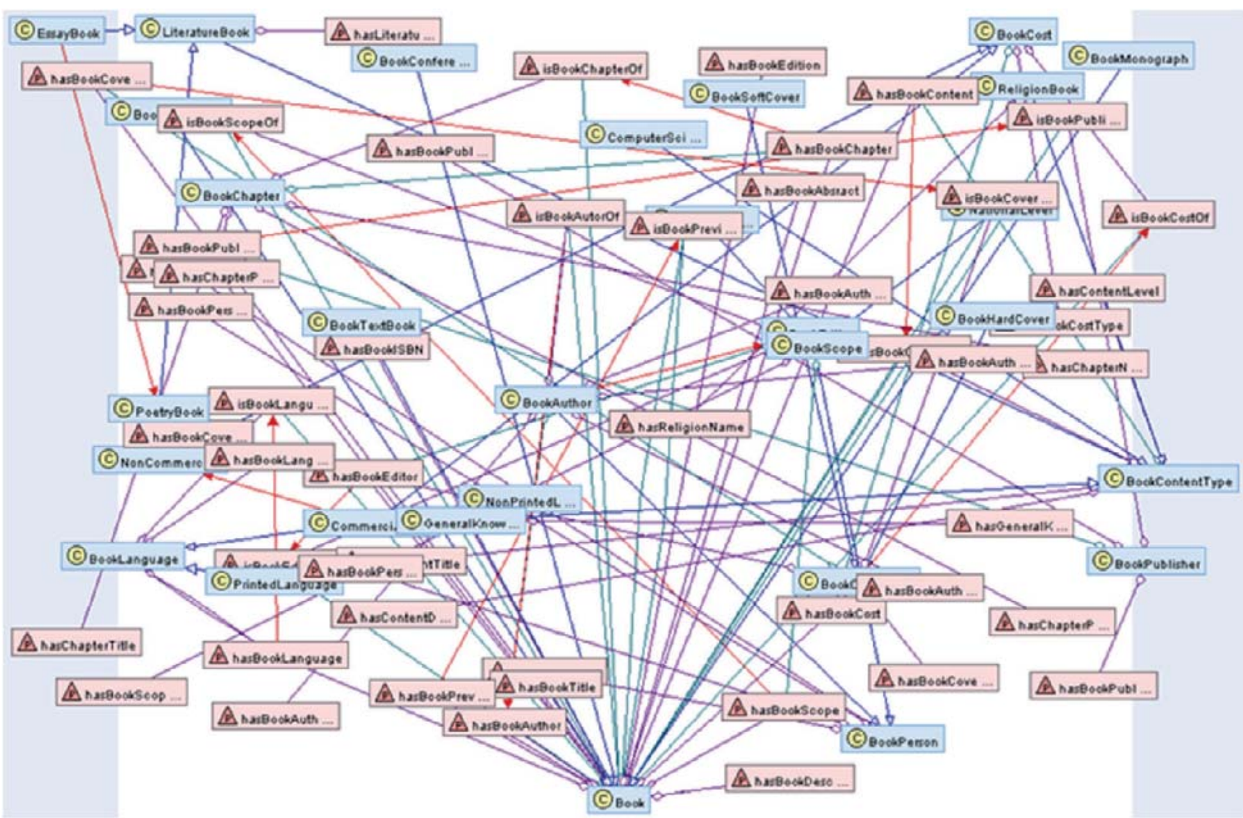

Fig. 1. Minimized RDF graph of book ontology.

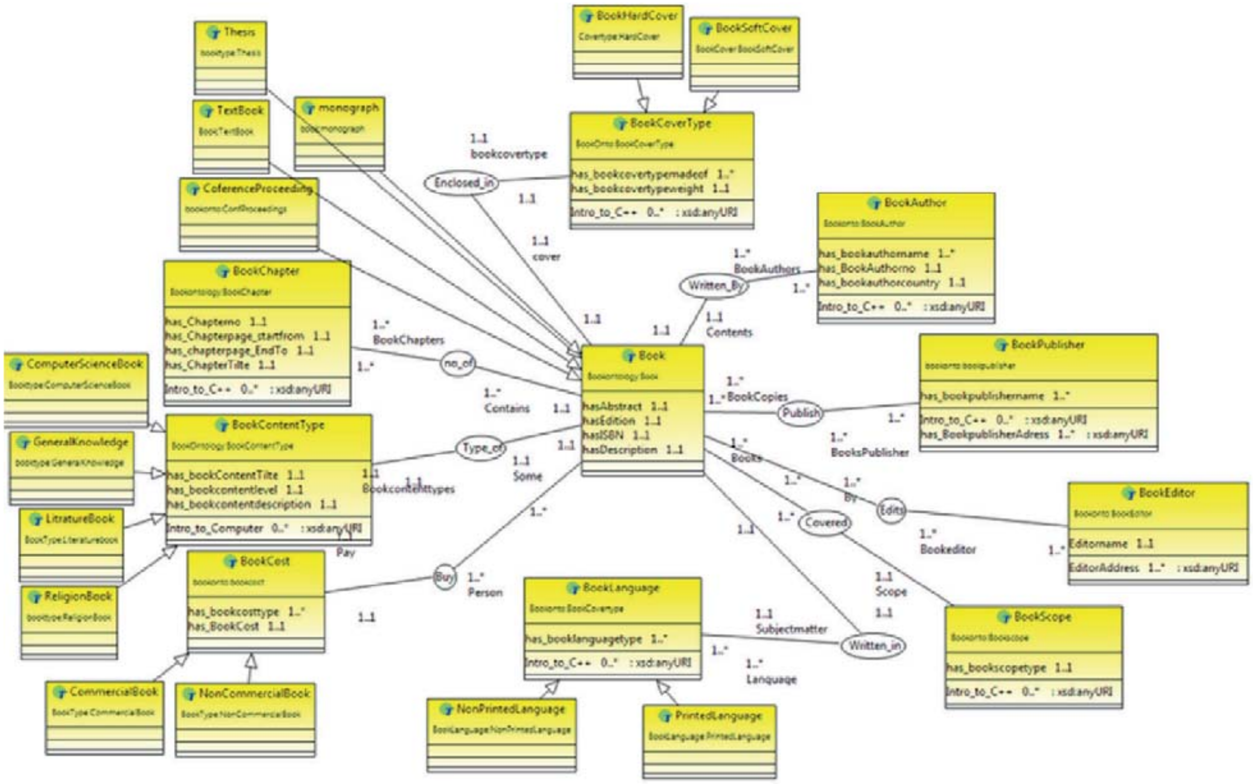

Fig. 2. Topic Maps schema of book ontology.

both valid and invalid RDF graphs but Turtle can serialize only valid RDF graphs [9]. For a useful serialization format to be rich enough to satisfy user requirements comprehensively whilst retaining good computational properties.

- ***Simplicity***: The simplicity of a serialization format defines how humane the format is: easy to understand by both humans and machines, quite readable/editable in raw form by humans, and relatively compact. However, compactness is not the critical requirement for simplicity because some serialization formats are quite simple but verbose such as N-Triples is a very simple but verbose serialization for RDF graphs [10]. Similarly, simplicity should not come at the cost of expressibility such as RDF/XML is very simple user friendly and highly expressive serialization format for RDF graphs.
- ***Compactness***: The compactness of a serialization format determines strength of the format to encode maximum information in small number of statements. The compactness of a serialization format typically depend on the encoding type (.i.e., XML-base and non-XML-based). XML based encodings are typically found more verbose (e.g., RDF/XML) and non-XML-based encodings are typically found more compact (e.g., Turtle [9]) but not necessarily (e.g., N-Triples [10]). XML-based serializations were plagued with non-compactness but current technological advancements in the storage and communication facilities made it of less concern. Compactness should not come at the cost of expressiveness. A Serialization format should be compact enough while retaining the desired expressiveness level. Similarly, simplicity should not be effect by compactness, a serialization format should use various methods to compact the data while still leaving it readable [10]. However, majorly simplicity affects compactness because compactness increases complexity while decreasing simplicity (e.g., N-Triples, and RDF/XML). Furthermore, using heavy compact encoding can impact encoders and decoders performances negatively.
- ***Tools Support***: For a serialization format to be accepted by users needs strong support from software tools and APIs paradigms. A tool or an API provides an environment for users to create,

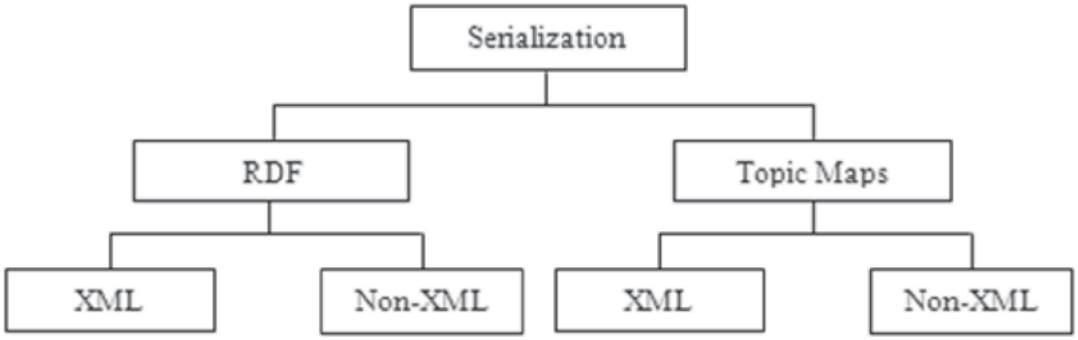

Fig. 3. Classification of RDF and Topic Maps serializations formats.

store, edit, maintain, and visualize a data model in a serialization format. However, hard coding in serialization format directly is a cumbersome task, therefore, tools and APIs should provide graphical user interface for users to interact with the data models in a click and play environment while producing the resultant serialization code automatically. For example, Protégé for RDF data model, and Ontoa for Topic Maps. An editor should support various serialization formats enabling users to select the required one. The more a serialization format is supported by the tools and APIs, the higher will be its market value.

- ***Applicability***: The applicability of a serialization format determines to address which scale of data models: small scale or large scale. The scale of a RDF model is measured by the number and types of objects and relations, whereas, of Topic Maps by the number and types of concepts, occurrences, and associations. The higher the number of ingredients the complex will be the model and the higher the will be the scale. However, large scale models are difficult to create and maintain, therefore, designers typically recommend a serialization format application for a specific scale level.
- ***Remarks***: In addition to the above criteria, several other features found during serialization formats investigation are highlighted in the remarks.

3. TYPES OF SERIALIZATION

A number of RDF and Topic Maps serialization formats are introduced by the Semantic Web researchers in the past several years. The available serialization formats of RDF and Topic Maps can be broadly divided into two categories: XML based and Non-XML based (as shown in Fig. 3) using the structure of serialized documents. Serialization formats belonging to both of the categories provides flexibility, user friendless and web centric application syntax while keeping the expressive power of both RDF and Topic Maps. Knowledge representation and navigation can also be done with simplicity and flexibility due to its support for different types of software and implementation on multiple platforms.

2.1. Resource Description Framework Serialization Formats

Resource Description Framework (RDF) represents data in the form of graph model which is not suitable for storage and manipulation. To efficiently store, transmit, manipulate, and reuse RDF data on the Internet or in local computer several types of serialization formats were researched by the Semantic Web researchers with the aims of providing flexibility and compactness. Using the underlying serialization structures, they are classified into XML based and Non-XML based serialization formats, as shown in Fig. 4.

2.1.1. XML Based Serialization

XML is a meta language standardized by the W3C, which stores, structures, and relates the contents of Web resources in a way to be easily accessible and interpretable by the machines [11]. To easily and efficiently interchange RDF documents on the Semantic Web, XML based Serialization like RDF/XML and Triples in XML (TriX) etc have been introduced by the Semantic Web researchers.

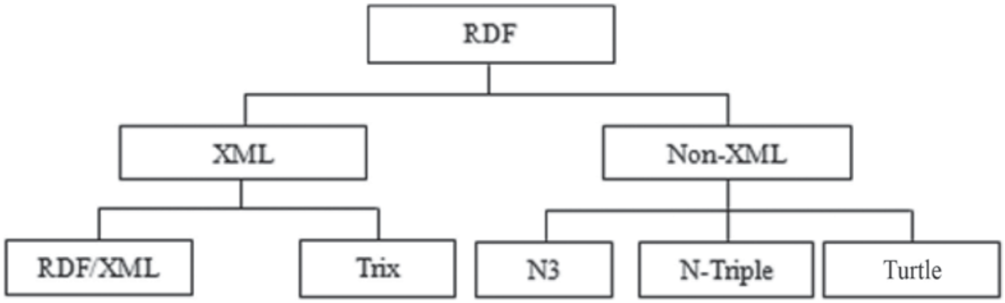

Fig. 4. Classification of RDF serialization formats.

***RDF/XML**:* RDF/XML is W3C standardized serialization format for the merging and interchangeability of information from multiple sources. Web resources can be expressed more meaningfully by RDF/XML Serialization. The expressive power of RDF/XML is much high, and the syntax is computable, simple, and user friendly. The parsing mechanism is quick and produces the output files in the same or another format by serializing the RDF triples [12]. One of the reasons of its popularity is the availability of numerous visual editing tools written in Java, like IsaViz etc [13]. As compared to other RDF serialization formats (e.g., N3, and Turtle, etc), RDF/XML based documents are less compact due to being lengthy, and larger in sizes [14]. But, on the other are more readable, easy to understand, and easy to implement as compared to them. Although RDF/XML is more flexible providing opportunities for representing RDF resource in multiple of ways, but results into making the syntax and format hard to read and implement [15]. Therefore, sometimes it is not processable with standard XML tools and creates difficulty while implementing some general types of RDF Graphs.

***Triples in XML (TriX)**:* TriX is a XML based serialization format which represents RDF graph (subject, predicate, and object) in a highly consistent and normalized way [16]. The reason of its high consistency and normalization is due to the use of XML tools like XSLT[4], and Xquery[5] etc for the manipulation of RDF graphs. TriX can express and serialize RDF named graphs as well, which are the combination of various RDF graphs in a single document represented by URIs. The syntax of TriX is more user friendly due to its unique features: relative URIs, and qnames. These features establish an environment where RDF documents can be easily read and write by eliminating redundant information and easing URI abbreviation. Supporting XSLT, TriX provides extensible syntax to express, represent and manipulate RDF graphs [16]. On the negative side, the syntax of TriX is incoherent and un-extended, therefore, decreases human readability. Despite of the fact that TriX documents are translatable using XSLT, but due to its lack of efficiency for hierarchal representation, each pattern match must query the entire document.

2.1.2. Non-XML Based Serialization

Although XML is well enough to provide interoperability, and platform independency but it suffers from numerous problems majoring that XML documents are normally larger in size by containing a number of words, which will not only create problems while transmission but will also severely effect human readability and understandability. To eliminate these problems researchers tried to find another way of representing RDF documents which is SGML. In the past several years, a wide array of non-XML based (SGML based) serialization formats for RDF documents including Notation-3 (N3), N-Triple, Terse RDF Triple Language (Turtle) are introduced by the research communities. It is argued that they have the potential to overcome the limitations and, in specific, the verbose nature of XML based serialization.

***Notation-3 (N3)**:* N3 is a de-facto standard developed by Tim Berners Lee and published by W3C in 2008 [17]. The basic idea was to optimize the logic and inference mechanism as compared to the other notations. The main reason behind the popularity of

[4] http://www.w3schools.com/xsl/
[5] http://www.w3schools.com/xquery/default.asp

N3 is because of its compactness, readability, and following many key features of Semantic Web not available in RDF/XML [18]. The distinguishable features of N3 include its simple and persistent syntax, expressive power, and ability to abbreviate URIs using prefixes. The expressive power of N3 in open web based environment is high and can express rules and logic by combining RDF properties with N3's extensions of RDF by including variables and nested graphs. The development of N3 notation for RDF has optimized the computability of data and logic due to its short notation and abbreviations used for representing RDF graph. N3 can parse RDF data quickly while its syntax is compact and can also be extended for much wider applications. N3 format has low redundancy, describe URI prefix at the beginning of file, shorthand notations are available in it and provides an abbreviation for both URI prefixes and base URI. Several types of visual and text editors and convertors are available for N3 which can create, edit and convert N3 document into another format both online and locally inside the system. The unknown and long term stability of N3 and its clear requirements for correlations of different features of N3 in RDF/XML format is the main limitation of this format [18]. The power of N3 is also limited in a situation when a user wants to make a statement about statement (reification).

N-Triple*:* Dave Beckett at the University of Bristol and Art Barstow at the W3C developed serialization format called N-Triple to solve the problem of automated processing while comparing the actual results with desired results [19]. RDF graph data can be stored and transmitted in N-Triple serialization format in the form of plain text and line. N-Triple syntax is easier to read and write and represents a single triple (subject, predicate and object) in a single line. N-triple is suitable in a situation when user want to store millions of triples. Although this serialization format is the subset of N3 and Turtle, however, it is easier to generate, merge, and parse in memory as compared to its predecessors. N-Triple serialization format can express RDF test cases and describe the association between RDF/XML serialization and the RDF abstract syntax. Parsing is also done in N-Triple file, line by line and a large file which cannot fit into main memory at once can be loaded one line at a time [20]. However, N-Triple is not as compact as RDF/XML and Turtle due to its lack of support for nested resources and elimination of repetitions of the resources URIs.

Terse RDF Triple Language (Turtle)*:* Turtle is a W3C's recommendation for representing almost every type of RDF graph data in a plain and compact text format [21]. Turtle is the extension of N-Triple and subset of N3, therefore, supporting any system which N3 and N-Triple supports. Turtle represents triples in the form of <subject >, <relationship>, and <object>. Some of the most promising features of Turtle includes high expressive power, user friendly than RDF/XML, and the availability of open source online editors and conversion tools. Apart from them popularity of Turtle serialization can be attributed to its wider use, greater conciseness, human readability, ease of generation, and use by the popular tools and APIs including Jena and OWL API[22]. Turtle suffers from the same limitations as RDF/XML having triple based syntax, where reification becomes necessary while translating complex expressions and OWL axioms into triples. Thus, makes the resulting RDF document more disruptive and unpleasant.

2.2. Topic Maps Serialization Formats

In Topic Maps serialization (TMs), data is semantically converted into a suitable format to be stored and transmitted over Internet and reused in a different computing environment, without requiring any change in the original data format. Topic Maps can be represented in a number of serialization formats including XTM, CTM, TM/XML, and LTM etc (as shown in Fig. 5), which could be exported into another computing environment conveniently. A number of applications (e.g., TM4J[6], and Wandora[7] etc) have been invented by the research communities which not only facilitate representing Topic Maps data in afore mentioned serialization formats but also provides consistent methods for interchanging data from one format into another. Using the underlying serialization structures, Topic Maps serialization formats are classified into two broad categories: XML based and non-XML based.

---

[6] http://tm4j.org/

[7] http://www.wandora.org/www/

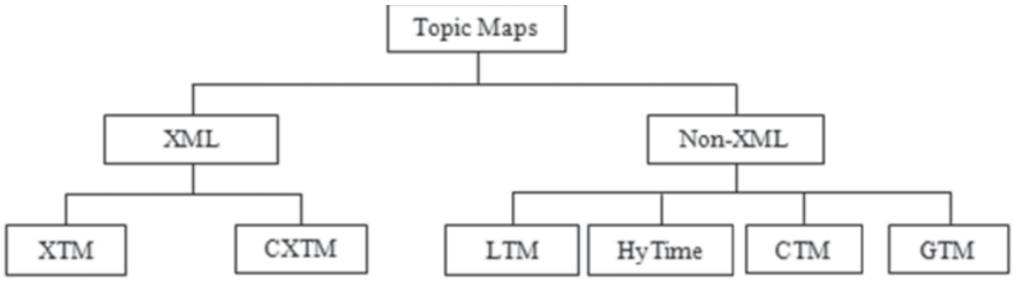

Fig. 5. Classification of Topic Maps serialization formats.

2.2.1. XML Based Serialization

XML based Serialization provides an efficient, easily implementable, and easily interchangeable syntax for Topic Maps documents, due to fact of its working on the lower layer of Topic Maps and direct insertion of any XML tag into Topic Maps information source. A number of XML oriented serialization formats are available, with the capability of smooth interchange of Topic Maps data into another format.

XML Topic Maps (XTM): Knowledge Representation in SGML Topic Maps was suffered for numerous problems including difficult to understand and create, application dependent, and lack of interchangeability. Therefore, a group of researchers in topicmaps.org developed a new syntax with more advanced features than SGML, called XTM 1.0 (XML Topic Maps version 1.0), which uses XML as a storage method and URL for linking and references [23]. Due to its flexibility, software support, and accessible in diverse domains and platform, XTM was standardized by ISO and published as XTM DTD[8] in October 2001. The prominent features of XTM includes simplicity, open architecture, flexibility, user friendly web-centric applications syntax, high expressive power, high level of computability, and the availability of editing and conversions tools [24]. XTM is inherently implemented in the web browser, therefore, knowledge representation, accessibility, and navigation on the Web across applications can be easily done irrespective of scale. The automatic translation process of XTM is made simple by the availability of CAT[9] tool, Translation Management System (TMS)[10] tool, and its support for all common file types. Beside several advantages provided by XTM for its users to create Topic Maps for the World Wide Web, XTM suffers from a number of problems including cannot fully represent the whole Topic Maps paradigm, less compact due to the verbose nature of XML, occupy large space as compared with other Topic Maps serialization formats, and not so much effective in semantic integration and interoperability of Topic Maps constructs [25].

Canonicalization of XML Topic Maps (CXTM): CXTM was developed to overcome the lack of canonicalization in XTM serialization. Canonicalization can be used for immediate comparison of two data model instances to find out correspondence and equality by comparing their canonical serialization [26]. CXTM is XML based serialization format standardized by the ISO/IEC-4 in 2009. The unique feature which distinguishes it from XTM is that the instances of two corresponding Topic Maps data models will generate byte by byte identical serializations at every time and the two Topic Maps data models dissimilar instances will generate different serialization [27]. Test suite for testing different Topic Maps technologies to show the existence of some particular set of behaviours can be effectively created using CXTM. The availability of editing and conversion tools makes the reading, writing, and parsing of CXTM documents pretty much easy. Despite of its good aspects, CXTM is very complex to understand, and is not user friendly. CXTM is only a variation of XTM useful for comparing two Topic Maps data models, which could be done with XTM, indeed.

2.2.2. Non XML Based Serialization

To overcome the limitations and the verbose nature of XML based serialization, several types of non

[8] http://www.topicmaps.org/xtm/1.0/xtm1.dtd
[9] http://www.metatexis.com/cat.htm
[10] http://cmswithtms.net/what-is-a-translation-management-system/

XML serialization formats have been researched for the interchange of Topic Maps data into another format.

Linear Topic Map (LTM): LTM serialization format for Topic Maps was developed by Ontopia[11] in 2001 and refined by Lars Marius Garshol [28]. LTM provide a unique Topic Maps textual editor, which isolates users from the syntactical details of the interchange format unlike that of XTM. Representation of LTM is simple, efficient and takes less space as compared to XTM. This interchange syntax is suitable for small scale Topic Maps constructs especially for personal use like email and chat discussions etc. The computability of LTM is ideal, the syntax is compact, and simple, having user friendly interface as compared to XTM, easily convertible to XTM format, and has efficient parsing mechanism. The expressive power of LTM is low than XTM because of its lack of representing the whole Topic Maps data model [29]. LTM did not succeeded in obtaining standard from any official body (i.e., ISO, etc), therefore, superseded by another Topic Maps serialization format called CTM[12].

Hypermedia/Time-based Structuring Language (HyTime): HyTime is another ISO standard interchange format for Topic Maps, based on SGML (Standard Generalize Mark-up Language, a language which describe the contents of documents rather than its look) [30]. HyTime provides an abstract facility for the Topic Maps paradigm to be expressed in SGML format. Originally HyTime was used to link and organize related topics which are time and space critical, but later on used for any types of Topic Maps representation as well. The expressive power of HyTime is much better, therefore, any sort of virtual notation can be used to express addressing [31]. The architecture of HyTime is flexible, providing simple and user friendly ways for defining new applications, having efficient parsing mechanism, provides convenient way to fulfill the addressing needs, and applicable to any type of hypermedia concepts [32]. The main limitation of HyTime is the lack of support for real SGML because currently there are several SGML systems having limited capability for the support of HyTime standard [33].

Compact Topic Maps (CTM)*:* Compact Topic Maps (CTM) is a text based notation standardized by ISO for representing Topic Maps data. CTM replaces LTM and AsTMa because of its compact syntax and unique feature of supporting TMQL's insert operation [34]. CTM provides several features including its short and simple notation, easy and maximum readability, and providing a common background for TMCL and TMQL. Due to these features CTM can represent any type of Topic Maps resources and their relationships as a simple test pattern. The CTM notation uses CTM processor for the precise computation, enabling users to interact with Topic Maps and perform several data operations such as evaluation, editing, and changing etc. according to user's rights on data. The editing facility provided by CTM is efficient, the transformation and parsing mechanism is fast, and supports all types of character encoding. The limitations of CTM includes the need of special parser for transforming Topic Maps constructs, and the lack of support for item identifiers in the constructs which are not topics.

Graphical Notation for Topic Maps (GTM)*:* Standardized by the ISO, GTM was developed for visual representation and creating similar views of Topic Maps instances and ontologies. The graphical modeling features of GTM are similar to UML, represents topics and associations using rectangles and edges respectively [35]. Names and occurrences of topic's classes are indicated in the lower part of each rectangle. In GTM compositional shapes are available for the representation of topics, associations, labels, occurrences and reification [36]. Due to its visual notation and compact syntax ontologies, TMDM and TMCL can be expressed and understood easily while designing, explaining, and instructing peoples. GTM is simple, user friendly, easy to read in paper form, and can parsed into other formats like XTM and LTM without any change in its original information or addition of new elements. GTM is applicable to both small scale and large scale Topic Maps instance data and ontologies. The unavailability of editors and authoring system, due to its early stages of development is major problem faced by the GTM users.

---

[11] http://www.ontopia.net/

[12] http://www.isotopicmaps.org/ctm/

4. COMPARISON OF RDF AND TOPIC MAPS SERIALIZATIONS

After thorough analysis it is not difficult to decide that both Resource Description Framework (RDF) and Topic Maps (TMs) are the finest examples of human's efforts for representing objects/concepts and their relationships/associations on the Web not only in human readable form but for the machines equally. To further refine the vision, researchers contributed their massive efforts in developing unique and powerful serializations formats not only to ensure interoperability but to provide flexibility, and strength at extreme levels. In the preceding sections of this paper, the available serialization formats for both of the technologies are highlighted in extreme. To further refine the idea and increase the strength of understanding, the discussed RDF and Topic Maps serializations are evaluated using the evaluation criteria describe in section 2.

A detail comparison of the main features of the RDF serializations is presented in Table 1. It is obvious from the table that every serialization format has their own capability to describe RDF graphs in their own formats. However, as a conclusion of analysis, the most appropriate serialization format to represent RDF graphs is Notation-3 (N3). The superiority of Notation-3 over its companion is due to a number of its reasons including: (1) logic and inference mechanism, (2) compact, persistent and readable syntax, (3) advanced parsing mechanism, (4) availability of open source visual and text based editors, (5) and its expressive power to represent any RDF graph.

Likewise, detailed comparisons of the main features of Topic Maps serialization formats are presented in Table 2. It is find out that every serialization format provides their own facilities, and potentials to represent Topic Maps constructs in different formats effectively. However, as a conclusion of analysis, it is deduced that the most appropriate serialization format to represent Topic Maps documents is XTM. The superiority of XTM over its companions is attributed due to some of its prominent features including: (1) having advanced features than SGML, (2) user friendly and flexible architecture, (3) high expressibility due to XML based syntax, (4) simple interface, (5) automatic translation facility, (6) availability of simple and efficient visual editors, (7) and its applicability for both small scale and large scale projects. Knowledge representation and navigation on the Web can also be done easily by XTM as compared to other serialization formats.

After analysis and comparison of the facts about the RDF and Topic Maps serializations formats presented expressively in the above two tables, it not difficult to reach at a solid conclusion. It is found that both RDF and Topic Maps serialization formats either XML based and non-XML based can be compared using evaluation criteria. Although the two sets of serialization formats have some degree of differences but have majority ratio of similarities, therefore, can be deemed as inter-convertible with the introduction of some technology. The list of similarities among both of the technologies includes it XML based syntax, its representation and navigation on the web, the computation of internal and external resources, advanced automatic translation facility, availability of visual and text based editors, online open source conversion software, and their flexible nature to be represented in any type of languages and their underlying formats.

The two non- XML based serialization formats (i.e., N3 and LTM) for RDF and Topic Maps respectively have also several commonalities and have the possibilities to be used together. Both of the formats can write their own technologies in a simple text editor, availability of prefixes for URIs, having high speed of computability, and represents information in simple and efficient documents requiring less space. Therefore, using the underlying facts it is easy to conclude that we can use RDF serialization formats for encoding Topic Maps data and vice versa with the provision of requiring development of advanced tools and technologies supporting both of the technologies.

5. CONCLUSIONS

The freedom of World Wide Web has lead into exponential growth of web pages containing enormous contents such as text, images, videos, and hyperlinks etc, resulting into the increase of human's cognitive overload. To elevate the problem and convert the web contents into machine readable and interpretable format Semantic Web is deemed as the ideal solution. To achieve the goals of Semantic Web, a number of technologies are invented by the research communities in the past several years having their foundations in RDF and Topic Maps. To further refine the ideas and achieve high level of interoperability and machine readability, a number of serialization formats are put forwarded by the researchers.

In this paper we presented a comprehensive overview of the taxonomy of Resource Description Framework (RDF) and Topic Maps (TMs) serialization formats by classifying them into XML based and non XML based. RDF and Topic Maps data can be serialized into XML, non-XML, or any other customized format. Thus, providing flexibility, and user friendliness for creating web centric applications while pertaining the expressive power of RDF and Topic Maps paradigms. It was augmented by the critics that the leading factor effecting RDF and Topic Maps interoperability is the underlying serialization formats used by the technologies. Therefore, to turn the vision of their interoperability into reality, a common XML based serialization syntax is needed which should be understandable to both of the technologies. But, in our investigation we came up with finding a number of commonalities which could be exploited effectively with the addition of novel advanced technologies to turn the idea of interoperability into possible. However, the introduction of such advanced technologies is not pretty easy and still needs a massive amount of research contributions from the research communities.

6. REFERENCES

1. World Wide Web. world wide web. [Online]. Internet Available: http://en.wikipedia.org/wiki/World_Wide_Web (Accessed: November 20, 2012).
2. Singh, D.R.K. Mishra & Chandrashekhar. Exploring semantic web using ontologies. *Stem Cell* 2: 11-15 (2012).
3. Berners-Lee, T.J. Handler & O. Lassila. The semantic web. *Scientific American* 284*:* 34-43 (2001).
4. Kaminski, P. *Integrating Information on the Semantic Web Using Partially Ordered Multi Hyperset.* MS thesis, Department of Computer Science, University of Victoria, Canada (2002).
5. Pepper, S. *Topic Maps: Encyclopedia of Library and Information Sciences 3rd ed*. Taylor & Francis Group, New York, USA (2010).
6. Serialization. Serialization [Online]. Internet Available: http://en.wikipedia.org/wiki/Serialization (Accessed: November 22, 2012).
7. Manoj, K.V. What is serialization in .NET, types of serialization and why we need it while developing an application? [Online]. Internet Available: http://www.c-sharpcorner.com/interviews/answer/499/ (27 July, 2006) (Accessed: October 24, 2012).
8. Gonzalez, R. RDF vs. XML [Online]. Inernet Available: http://www.cambridgesemantics.com/semantic-university/rdf-vs-xml (2007) (Accessed: May 17, 2012).
9. Benson, E. A Quick Tutorial on the Turtle RDF Serialization. Internet Available: http://haystack.csail.mit.edu /blog/2008/11/06/a-quick-tutorial-on-the-tutrle-rdf-serialization/ (2008) (Accessed: May 20, 2012).
10. Segaran*,* T., C. Evans, J. Taylor, S. Toby, E. Colin, & T. Jamie. *Programming the Semantic Web*. O'Reilly Media (2009).
11. W3school. XML tutorial. [Online]. Internet Available: http://www.w3schools.com/xml/default.asp (Accessed: October 14, 2012).
12. Lapalme, G. XML: Looking at the forest instead of the trees [Online]. Internet: http://www.igt.net/~ngrenon/ UdeM/cours/IFT3220/ForestInsteadOfTheTrees.pdf (18 November, 2005) (Accessed: September 13, 2012).
13. Powers, S. *Practical Resource Description Framework (RDF*) *(First ed.)*. O'Reilly & Associates, North Sebastopol, Canada (2003).
14. Yu, L. *A Developer's Guide to the Semantic Web*. Springer Publishing Company, Berlin Heidelberg, Germany (2010).
15. Jeni. RDF chimera [Online]. Internet Available: http://www.jenitennison.com/blog/node/171 (30 June, 2012) (Accessed: October 24, 2012).
16. Carroll, J.J. & P. Stickler. TriX: RDF triples in XML. *Proceedings of the 13th international World Wide Web Conference on Alternate Track Papers & Posters*: p. 412- 413 (2004).
17. Berners-Lee, T. & D. Connolly. Notation3 (N3): A readable RDF syntax [Online]. Internet: http://www.w3.org/ TeamSubmission/2011/SUBM-n3-20110328/ (28 March, 2011) (Accessed: October 28, 2012).
18. Grlicky, V. An overview of RDF model representation formats. *M. Bieliková (Ed.), IIT.SRC 2005*: 242-248 (2005).
19. N-Triples. N-Triples [online]. Internet: http://en.wikipedia.org/wiki/N-Triples (Accessed: December 11, 2012).

20. Heath, T. & C. Bizer. *Linked Data, Evolving the Web into a Global Data Space (First Edition)*, Morgan & Clay- pool publishers, USA (2011).
21. Beckett, D. & T. Berners-Lee. Turtle - terse RDF triple language [Online]. Internet Available: http://www.w3. org/TeamSubmission/turtle/ (28 March, 2011) (Accessed: January 12, 2013).
22. Horridge, M. OWL syntaxes [Online]. January 22, 2010. Internet Available: http://ontogenesis. knowledgeblog. org/88 (22 January, 2010) (Accessed: December 24, 2012).
23. Tolksdorf, R. C. Bizer. R. Eckstein. & R. Heese. Business to consumer markets on the semantic web. *Proceedings of the On The Move to Meaningful Internet Systems (OTM 2003) Workshops, Catania, Sicily, Italy*: p. 816-828 (2003).
24. Marshall, M. The semantic web & its implications on search marketing [Online]. July 27, 2007. Internet Available: http://www.searchenginejournal. com/the-semantic-web-its-implications-on-search-marketing/ 5390/ (27 July, 2007) (Accessed: December 26, 2012).
25, Pepper, S. & G. Moore. XML Topic Maps (XTM) 1.0 [Online]. Internet Available: http://www. topicmaps.org/ xtm/ (08 June, 2001) (Accessed: November 13, 2012).
26. Ahmed, K. Topic maps — canonicalization [Online]. Internet Available: http://www.isotopicmaps.org/ cxtm/ 2004-11-01/ (11 January, 2004) (Accessed: October 17, 2012).
27. Lee, J. L.M. Garshol. & M. Naito. Topic maps — canonicalization [Online]. Internet Available: http://www.isotopicmaps.org/cxtm/cxtm.html (27 January, 2009) (Accessed: October 17, 2012).
28. Garshol, L.M. The linear topic map notation: definition and introduction, Version 1.3 [Online]. Internet Available: http://www.ontopia.net/download/ ltm.html (17 June 2006) (Accessed: December 10, 2012).
29. LTM. LTM [Online]. Internet Available: http:// xsiteable.sourceforge.net/standard-ltm.html (Accessed: October 17, 2012).
30. Fierz, W. & R. Grutter. The SGML standardization framework and the introduction of XML *Journal of Medical Internet Research* 2: e12 (2000).
31. Biezunski, M. & S.R. Newcomb. Differences between XTM 1.0 and the HyTime-based meta-dtd [Online]. Internet Available: http://www.topicmapslab.de/ publications/differences_between_xtm_10_and_ hytime-based_meta-dtd (18 December, 2001) (Accessed: October 18, 2012).
32. Rutledge, L. J. V. Ossenburgen. L. Hardman. & D.C. A. Bulterman. A framework for generating adaptable hypermedia documents. *Proceedings of the Fifth ACM international conference on Multimedia, Seattle, Washington, USA*, p. 121-130 (1997).
33. Newcomb, S. R. Using the information addressing model of HyTime (ISO 10744) to add hypermedia functionality to legacy data and systems. *Proceedings of the Second International Workshop on Incorporating Hypertext Functionality into Software Systems held in conjunction with the ACM Hypertext '96 conference, Washington, USA* (1996).
34. Hopmans, H. & O. Pepper. WD 13250-6: Topic maps – compact syntax [Online]. Internet Available: http://lingo.uib.no/trond/TopicMaps/CTM Working Draft.pdf (15 September, 2006) (Accessed: January 22, 2013).
35. Witt, T. Conception and Implementation of a Visual Editor for Topic Maps [Online]. Diploma thesis, Department of Institut für Informatik Ludwig-Maximilians-Universität München, Germany (2009).
36. Thomas, H., T. Redmann, M. Pressler & B. Marksscheffel. GTMalpha – Towards a graphical notation for topic maps. *Proceedings of the Fourth International Conference on Topic Maps Research and Applications (TMRA 2008) - Subject-centric computing, Leipzig, Germany,* p. 56-66 (2008).