Semantic Web and Electronic Information Resources
Danica Radovanović

Abstract

The usage of electronic resources depends on good possibilities of searching and the Semantic Web concept can be convenient solution for information retrieval (IR). WWW (World Wide Web) enables, with the help of search engines and huge number of available (meta)information, data that can satisfy user's need for information, but only at some extent. At the same time, there are more and more research efforts to increase the efficiency for IR until one gets as much as possible relevant information on the Web. As one of the latest results of this W3C efforts, Semantic Web presents a group of organized technological standards, IT products, and information linked in such a way that can be easily indexed and semantically filtered through the process of classification on the global scale. Semantic Web and its principles make IR easier because it can be also observed as very useful and successful way of representing data on the WWW or as a group of globally linked databases. The architecture of Semantic Web consists of three important IT standards: XML (eXtensible MarkUp Language), RDF (Resource Description Framework) and the ontologies. Semantic web is still under development and is not in common usage but it promises that it will radically improve the possibility of searching, sorting and classification of information.

Key words: Semantic Web, electronic information resources, information retrieval, information representation, Internet, standards

1. INFORMATION NEEDS, WORLD WIDE WEB, AND THE SEMANTIC WEB

As in many areas of global activity, information and communication technologies have found it's use in librarianship a long time ago. They are mostly used for searching the information and accessing the existing databases over the Internet. Usage of electronic information resources depends on good search capabilities. Because of that the Semantic Web concept was developed in pursue of more efficient solutions for finding information.

Since the World Wide Web exists for more than ten years, search engines can help in finding information within the vast amount of available metadata, but that fulfills the information needs only in a certain degree.

At the same time, more and more research is done towards creating more efficient Web search techniques so the more relevant information can be found. The need for such efforts remains because the search results are often completely irrelevant and the users of information systems and databases are forced to refine their search queries and to, often, manually filter the data to get to the more precise results.

---

1 The paper is presented at the 8th ZBUS conference- Usage of electronic resourses of information, Belgrade, 26.9.2002, later published at peer reviewed Infotheca journal 4(2003)2, p. 157-163

Tim Berners-Lee, the professor at Massachusetts Institute of Technology (MIT), who in 1989 developed the World Wide Web by creating the language for linking and sharing information, in 1994 also founded World Wide Web Consortium (W3C) at the MIT (www.w3.org). To further develop the Internet service he invented Berners-Lee, together with his colleagues, works on it's semantics so that the information resources on the Internet will not just be located, but they will also, with the help of computers, be semantically processed. As one of the latest results in these W3C efforts, the Semantic Web represents a set of organized technical standards, products, and information connected in such a way that they can be easily reached and semantically filtered by processing it on the global scale.

2. DEFINITIONS

The Semantic Web is an abstract representation of the World Wide Web data based on RDF (Resource Description Framework) specification and other standards. In literature there are several definition of this concept.

"The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation." - Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001.

"The Semantic Web is the Web of data that can be processed directly or indirectly using the machines." - Tim Berners-Lee, Weaving the Web.

The Semantic Web as a concept makes possible to organize available Web information resources, and to use them not only by syntax and structural methods, but also by the semantic ones. It represents the synergy of programs that collect the contents from the Web, from different resources, process the information, and exchange the results with other programs.

The Semantic Web contributes to efficient searching by making it possible to present the information, that can be perceived as the set of globally linked Internet databases, in a special way. For the Semantic Web to function, the computers need to have access to the structured data collections, and rules for automated management have to be established. Internet search engines use keywords and search terms that coincide with search words in user queries to solve the problem of lack of meaning on the Web and to analyze the presentations. In the beginning search engines used keywords, and the new generation analyzes the context through automated or manual review of the links between Web pages and content. In that way keywords and abstracts are being interpreted as terms. Automated results classification and categorization is common nowadays; modern search technologies include technology that finds search terms in the Web pages and links them to the profiles that describe the results' characteristics to the user.

3. SEMANTIC WEB ARCHITECTURE

While HTML (Hypertext MarkUp Language) is used to present the data and to describe, using the format tags, the way it will look on the Web page, the Semantic Web architecture is comprised of two basic IT standards, and the third element which has the key role:

<RDF primer u XML-u>

```
<?xml version="1.0"?> <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-
ns#"          xmlns:dc="http://purl.org/dc/elements/1.1/">          <rdf:Description
rdf:about="http://www.ilrt.org/people/cmdjb/">    <dc:title>Dave    Beckett's    Home
Page</dc:title> <dc:creator>Dave Beckett</dc:creator> <dc:publisher>ILRT, University
of Bristol</dc:publisher> </rdf:Description> </rdf:RDF>
```

- XML (eXtensible Markup Language) - a language that defines the data structure;

- RDF (Resource Description Framework) - a basic Web model conceived as a W3C standard for semantic links; RDF actually describes the semantic links between electronic resources; and

- the ontology, the most important factor of the Semantic Web, which is also the hardest one to achieve.

3.1 XML (eXtensible Markup Language)

XML is a well known W3C standard that belongs to the SGML group (Standard Generalized Markup Language). It is created in 1996 and serves to describe and exchange data on the Web. XML allows the authors to create their own markup (ie. <Author>), a syntax that holds a part of semantics itself. Tags, such as <Author>, hold more semantic meaning that tags like <H1>. The computer simply does not know what "author" means and what is the relation between the term "author" and the term "person". XML can help the creator, the user, to anticipate which information can be put "between the tags". XML allows the website designers to specify their own tags for the Web documents. Tags put the information in the structural hierarchy which users can understand. Existing HTML tags are mostly used to encode simple things such as the document layout (underline, italic, bold, etc.) or hypertextual links. But to create the Semantic Web a vast amount of tags is needed. They have to be used to describe the text subject and the author, so that the search engine can differentiate between, for instance, the works written by Kafka, and an essay or an article about him. For (inter)links to be complete, all elements must be specified to the last detail and because of that XML alone is not enough to achieve the semantics.

3.2 RDF (Resource Description Framework)

RDF (Resource Description Framework) is the basic Web model and W3C standard for semantic links - www.w3.org/RDF. It actually describes the semantic links between the electronic resources. RDF integrates different applications - from library catalogues and world guides and portals, to consolidated and collected data, software and contents, to multimedia collections (music, photographs, event records) and catalogues - using the XML as a syntax for mutual exchange. RDF specifications make it possible to create an ontological system that supports information on the Web by describing the structural semantics in a subject-predicate-object triplet. These triplets can be written using the XML tags (see the example schematics of an RDF written using the XML). Using the RDF standard, document is classified in such way that certain objects (people, web content, etc.) have properties that can be assigned certain values. RDF triplets stand between the information on the Semantic Web and allow the computers to find them, collect the requested information, and

have the insight in their movement.

In this way the presented model contains the data about the resources. And resources are anything connected to a URI (Universal Resource Identifier). Subject and object are each identified by URIs that are used to connect Web pages; URL (Uniform Resource Locator) is the most well know URI type. Verbs are also specified by URIs, and that allows the constitution of a new concept, a new verb, by defining an URI for each of them, somewhere on the Web. Besides Universal Resource Identifiers, which are globally recognizable, important concepts also are: unique element definitions, systems with certain processing rules, and the usual Internet infrastructure protocols. All types of metadata have to use established keywords or the search engine would not function when the user puts in a query. The following example shows how the elements of Dublin Core metadata standards are used with the RDF. (A lot of Internet databases and mutual library catalogues in the world use this concept.)

Also, the relation between the facts, data, and their interlink, are important, as well as the data that is relevant to a certain query or request. What follows is the example that shows a link that redirects from one URL (one location) to another (Web page). The identity of RDF resource is the ID of one of many concerned points on the map nods.

The goal of the Semantic Web is to allow the software to find the requested data on the Web, to semantically understand it, to create a connection-reference, to transfer it, and to apply all that to a certain request, query. That implies a combination of XML, RDF and an ontological concept.

3.3 Ontologies as a specification method

The Semantic Web can not be created without the metadata. Metadata make possible for middle search level to exist, so the user does not browse through the massive amounts of irrelevant full HTML text that can be found on the Web. Metadata from Web pages and catalogues have to be connected to special documents that define the metadata terms and the relationship between them. These collections of terms and their interlinks are called ontologies.

```
<rdf:RDF xmlns:rdf="http://www.w3c.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdf="http://www.w3c.org/2000/01/rdf-schema#"
   xmlns:tms="http://www-db.stanford.edu/rdftmmapping/tm-schema#"
   xmlns="http://www-db.stanford.edu/rdftmmapping/tm-schema#">
   <rdfs:Class ID="t"/>
   <rdfs:Class ID="a"/>
   <rdfs:Class ID="s"/>
   <rdf:Property ID="associationMember"/>
   <rdf:Property ID="associationScope"/>
   <rdf:Property ID="associationTemplate"/>
   <rdf:Property ID="scopeComponent"/>
   <rdf:Property ID="roleLabel"/>
   <rdf:Property ID="scr"/>
   <rdf:Property ID="sir"/>
</rdf:RDF>
```

To record the relationship between the objects and their characteristics, the Semantic Web creators use the ontologies, the third basic component of the Semantic Web. The term ontology is taken from the philosophy. It is a part of metaphysics' investigation of the nature of being. In the information context of using the electronic resources and knowledge, the ontology is defined as a specification

of ideas, conceptualization. Conceptualization is an abstract presentation of the world, or a part of it, from a certain point of view or with a specific reason. Every information or knowledge based database is, explicitly or implicitly, connected to a certain conceptualization.

The ontology is an explicit specification of a conceptualization that defines the terms and a connections between them. The ontology is a description (as a formal program specification) of terms and links that can exist inside a hierarchical term system. Obviously it is a tremendous task to create ontologies that will describe the term systems related to everything that exists on the Web. Ontologies' origins are in taxonomies: parts of ontologies that represent the classification, the structured layout of the data inside the classes, and categorize certain field and links between the elements. They are the result of generic scheme data, record, and object naming studies, as well as the rules that constitute the relationships between the data. The example of a digital taxonomy can be found on http://digitaltaxonomy.infobio.net/[2].

The Semantic Web describes the ontologies within the RDF specifications using the XML as the instrument. The ontologies allow the computers to gather information from different Web pages and presentations, and to give the answers to sophisticated user queries. In short, the Semantic Web turns the computer into a personal search assistant capable of diving into the whole structure of information and finding the necessary information in the speed of light. Libraries and information centers could support this project because they have the knowledge in building this sort of classification mechanisms.

4. THE FUTURE OF THE SEMANTIC WEB

The Semantic Web is a vision that is starting to get realized: to have the requested data on the Web, defined and connected, linked in a way that can be used by the machines, not only for the purpose of presenting it but also for using it in different search applications and for data availability. The technology to create the Semantic Web exists and it is known how to build the terminology and how to use the metadata. The whole idea depends on the usage of common standards.

Tim Berners-Lee is convinced that in the near future the Semantic Web will be used because RDF implementation in document compatibility projects is rising. He sees the Semantic Web as a possibility that will break barriers (intellectual and cultural) that exist on the Web today. It is expected that the full force and application of the Semantic Web will be achieved by 2010.

4.1 The Semantic Web and Applications

The Semantic Web should define the structure of Web page content and create the conditions for programs (agents) to move from page to page looking for the answers to user queries.

Some of the institutions that were among the first ones to start using Semantic Web principles are DARPA (U.S. Defense Advanced Research Projects Agency) and a Manchester (UK) based commercial organization Network Inference. They, for a long time now, use the available instruments for developing the Semantic Web infrastructure and user applications. Some companies and institutes work on those issues: Nokia (Espoo, Finland) - markup, Hewlett-Packard - Java tools for creating and managing the metadata, MIT and Free University of Amsterdam. And there are several languages for creating the ontologies.

4.2 Semantic Web Application Examples

---

[2] In 2002, when the article was written, this example was hosted at http://www.geocities.com/RainForest/Vines/8695

Besides Internet bot software that shows certain intelligent software agent characteristic when, in a very fast way and on a daily basis, visiting and indexing websites, and the most successful online indexing robot, the google.com second generation, there are also other examples of the Semantic Web structure usage.

• Math-net (http://www.math-net.de) uses the Semantic Web structure for their math science portal.

• An example of a RiboWeb database that uses a RDF scheme http://www.medwebplus.com/obj/25678

5. CONCLUSION

The Semantic Web is still a vision that implies that the requested data exists on the Web, that it is defined and linked in a way that can be used by machines, not only for the purpose of presenting it, but also to use it in different applications.

The Semantic Web can be created together with further development of the standards and information technologies (RDF, Z39.50, DC, FRBR, XML). Z39.50, for instance, is nearly an international standard that allows the information and metadata exchange between heterogeneous computer systems. At the moment, however, it is still a question if the Semantic Web concept will be successful. Computer science experts work with librarians on solving the information search, as well as knowledge and information exchange problems. Cost for creating ontologies is very high: it takes a lot of time and requires permanent development. According to some, the whole Semantic Web idea is a contradiction.

Technology that can make the Semantic Web a reality exists, and the ways to create the terminology and use the metadata are known. The idea depends on using the mutually accepted standards. The Semantic Web is still in development and is not generally used, but there is a chance that it will radically improve the capabilities of finding, sorting and classifying the information.

6. LITERATURE

URL References
1. The World Wide Web Consortium Semantic Web, http://w3.org/2001/sw
2. Getting Into RDF & Semantic Web Using N3, http://www.w3.org/2000/10/swap/Primer
3. Semantic Web Roadmap, http://www.w3.org/DesignIssues/Semantic
4. What Is The Semantic Web?, http://purl.org/swag/whatIsSW
5. Semantic Web Primer, http://uwimp.com/eo.htm
6. Semantic Web Introduction - Long, http://logicerror.com/semanticWeb-long
7. SciAm: The Semantic Web, http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html
8. Building The Semantic Web, http://www.xml.com/pub/a/2001/03/07/buildingsw.html
9. The Semantic Web, Taking Form, http://infomesh.net/2001/06/swform/
10. SW Activity Statement, http://www.w3.org/2001/sw/Activity
11. (SWAD) Semantic web road map, http://www.w3.org/2000/01/sw/
12. Web design issues: semantics, Tim Berners-Lee, http://www.w3.org/DesignIssues/Semantic.html
13. Web design issues; what semantics can represent, Tim Berners-Lee, http://www.w3.org/Designissues/RDFnot.html

14. Realizing the full potential of the web, Tim Berners-Lee,
    http://www.w3.org/1998/02/Potential.html
15. HP's metadata standards website,
    http://h10014.www1.hp.com/buildguide/meta_data_standards.htm
16. Berkeley electronic issue: Finding Information on Internet,
    http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/FindInfo.html

    Ontology References
1. Guarino N. Formal Ontology and Information Systems. In N. Guarino (ed.), Formal
   Ontology in Information Systems. Proc. of the 1st International Conference, Trento, Italy, 6-
   8 June 1998. IOS Press http://www.ladseb.pd.cnr.it/infor/Ontology/Papers/FOIS98.pdf
2. Ontology, Conceptual Modeling, and Knowledge Engineering Group at: Institute for
   Systems Science and Biomedical Engineering of the Italian National Research Council
   (CNR), Padova, Italy http://www.ladseb.pd.cnr.it/infor/ontology/ontology.html
3. SENSUS - 70,000-node terminology taxonomy. SENSUS is the continuance and the
   reorganization of WordNet (including the other ontologies connected to the ISI project)
   http://mozart.isi.edu:8003/sensus2/
4. Other ontology projects can be found on
   http://www.isi.edu/natural-language/projects/ONTOLOGIES.html