SEMANTIC WEB TECHGNOLOGIES IN DIGITAL LIBRARIES

Anshul Chugh[1] , Sanjay Bhatnagar[2] , Gurpreet Kaur [3]
[1]Assistant Professor,[2] Associate Professor,[3] Assistant Professor
Department of Computer Applications
Giani Zail Singh , PTU Campus
Bathinda
Anshu.chugh9@gmail.com

Abstract: *Digital libraries have been an important source of information throughout the history of mankind. Notably, semantic web has given a marvellous change to the traditional libraries and presently took the shape of semantic digital libraries which can be accessed at any time with more meaningful search. This paper explores the Semantic Web and the technologies that support the functioning of digital libraries including RDF, OWL and ontologies. Semantic Web technologies are considered from the perspective of digital libraries, and how Semantic Web technologies can enhance the functioning of digital libraries. A current research has been reviewed on the use of semantic web technologies in digital libraries and a vision on future trends of digital library have been proposed.*

I. Introduction

The internet has played an important role in the evolution of modern lifestyles, which subsequently became a valuable invention and which has made internet accessible and more convenient to its users. Such innovation formed the 'Web'. The Web was invented in 1989 by a graduate of Oxford University, England. His name was Tim-Berners Lee. He was responsible for writing the first web client and server in 1990 at CERN (which is the European Particle Physics Laboratory). His vision directed him toward the designing of an internet-based hypermedia system which would facilitate the uploading of information from all corners of the globe so that everyone could benefit from that knowledge [1]. The later developments in the web lead to the evolution of 'Semantic Web', which was similar, with the vision of

Tim-Berners Lee in the sense of making information global and universal for its users. Semantic Web elaborated the vision of information by enabling the web to meet the demands of the user and computers so that they can utilize the web for information. Semantic Web is based on a set of design principles and the concept of providing useful results to the user. Importantly, the web offered services and information which was comprehended by humans alone; however, in contrast, semantic web is able to produce information in a manner that can be understood by computers [4]. Some of the main technologies of Semantic Web are:

- **RDF**: **Resource Description Framework** initially implemented as a metadata data model. With the passage of time, it has proceeded into a mode of conceptual description and a mechanism to provide more meaning to information on the web.

- **Ontology**: can be described as the method of representing information in a formal and structured manner with the aid of a set of concepts and their respective relationships.

- **OWL**: represented as the set of information modeling languages which can be utilized in order to produce ontologies

ISSN: 2229-6913 (Print), ISSN: 2320-0332 (Online) -, Web Presence: http://www.ijoes.vidyapublications.com

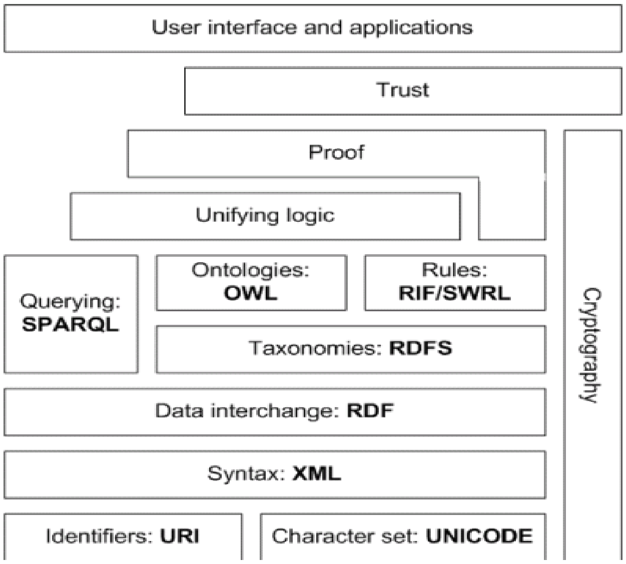


Figure 1: Semantic Web Stack

A. RDF SCHEMA

RDF Schema (RDFS) introduces the idea of a class. A class is a type of thing. For instance, you and I are members of the class Person. Computers are members of the class Machine. Person and Machine are classes. That is to say, they are themselves members of the type Class. The first higher-level predicate is the RDF: type predicate. (RDF is the usual namespace abbreviation).It relates an entity to another entity that denotes the class of the entity. The purpose of this predicate is to indicate what kind of thing a resource is. But, as with anything else in RDF, the choice of class is either by convention or arbitrary. To add class information into the vendor files from a few sections ago, a vendor would simply add this:

Adding Type Information

vendor1: product RDF: type general: Product

B. WEB ONTOLOGY LANGUAGE (OWL)

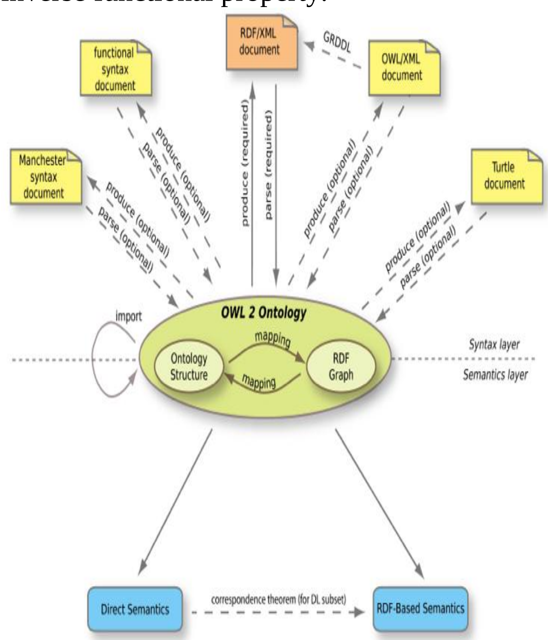
Web Ontology Language (OWL) defines more classes that let RDF authors define more of the meaning of their predicates within RDF. Four classes of predicates defined by OWL include:
owl:SymmetricProperty,
owl:TransitiveProperty,
owl:FunctionalProperty, and
owl:InverseFunctionalProperty[5].

OWL symmetric properties tell applications that the following inference is valid. If the application sees the statement S P O, and if P is typed as a symmetric property, then O P S is also true. For instance, we think of the has-friend relation are being symmetric. If you're my friend (ME HAS_FRIEND YOU), I'm your friend (YOU HAS_FRIEND ME).

OWL transitive properties work like this. If the application sees the statements X P Y and Y P Z, and if P is typed as a transitive property, then X P Z is also true. Rdfs: subClassOf is a transitive relation. If Mammal is a sub-class of Animal and Animal is a sub-class of Organism, then Mammal is a sub-class of Organism.

OWL functional and inverse-functional properties indicate how many times a property can be used for a given subject or object. A functional property is one that has at most one value for any particular subject. An example is the has Birthday relation between a person and his or her birthday. Everyone has just one birthday, so for any given subject (person), there can be just one object (birthday). But, the owns relation between an owner and ownee is not functional. People can own more than one thing[2].

Inverse functional properties do the same in reverse. For any object, there is only one subject for a particular inverse functional property.



Research Cell : An International Journal of Engineering Sciences, Issue December 2014, Vol. 3
ISSN: 2229-6913 (Print), ISSN: 2320-0332 (Online) -, Web Presence: http://www.ijoes.vidyapublications.com

II. BENEFITS OF SEMANTIC WEB TECHNOLOGY

1. With the help of multiple schemas, it enables distinct semantics for the various concepts in the domain.
2. By using the <s-p-o> structure of RDF, a tactful and simple mechanism provided to represent ontology.
3. Provides mechanisms to formulate generic queries (SPARQL) and instantiate them at runtime in order to answer the queries posed.
4. Provides mechanisms to create parts of the ontology and query on it seamlessly, using various technologies
5. To create derived facts rule evaluation and execution mechanism are also provided.
6. Provides mechanisms to link in *external concepts* with existing concepts of the domain through simple <s-p-o> structures[2]

III. PITFALLS

1. Reduced anonymity on the Web

As the active measures have already been taken to keep ones identity non-indexed. It can be located on the semantic web, information about ones identity, interest and other information are trivial to discover. Whenever we sign for an account on the various sites, the information about ones identity is fed during & after registration with your activities & contribution.

As the amount of available personal information increases we could begin seeing Websites that rely on querying the "Web as a database" for information about its visitors for mission critical functionality. If this change takes place having personal information on the Web may become the comfortable norm. One day we may see a shift in the importance of anonymity. Openness and transparency may become the "in thing."

2. Increased invasion of privacy

The problem raised out from the reduced anonymity on the semantic web. This drawback signifies that if vast amount of information about everyone is available that one can easily abuse that information .We want our self in the new era of personalization.

Contextual ads that examine a Website's content for hints of interests may be replaced with ads that target specific visitors based on their personal preferences, behaviours, lifestyle, friends, income, etc. In a similar way we will likely notice that e-commerce Websites will become better at figuring out just what it is we are going to want next.

Invasion of privacy brought about by the abuse of personal information — which would be more accessible than ever — will prove it quite annoying. We have privacy issues and have better solutions to avoid fake interference[5].

3. Intelligent content scraping

The content scrapers of today are really quite simple compared to what we will have to deal with in the Semantic Web. Essentially the scraper will access a Website or feed and extract and store the desired content. In most situations the content scraper must be customized or otherwise manually configured for the Website or feed (less so with feeds as they follow a standard format).

Content scrapper of the semantic web read the content within the web document and feed the data. Through natural language processing a semantic content scraper can read a blog entry or by surfing other blogs, a new blog can be prepared with entirely different wording.

The technology does not yet fully exist that would give us the ability to avoid content scrapping, however the bottom-up approach to semantic content scraping would be to scrape the content of metadata written in RDF / OWL. The "bottom-up" scraper would not have the ability to extract information from the content in the way that a top-down content scraper (using an NLP agent) could[2].

4. Value paradigm shifts

In The value of current datasets in the Semantic Web it is suggested that the ability to easily mine new and non-obvious types of data from the Semantic Web will turn information into more of a commodity than any past advancement in technology. No doubt with the help of semantic web it will become very simple to access any kind of information is no longer the bottleneck in our development. We have to differentiate between the commoditized information & information that could be better served as non-commoditized information. Most Websites earn money through visitors clicking on advertisements and generally attract those visitors with their content. Issues of commoditization are already springing up as we continue to explore the usage of feeds to deliver content to readers. Currently there are really only two solutions to the issue and those are embedding advertisements in feeds and publishing partial content to encourage readers to click into your Website and continue reading. It's possible that if we do not develop ways to generate revenue from commoditized content we will never see the Semantic Web come to fruition because it would receive little commercial backing.

5. Vocabulary incompatibilities

The vocabularies we use to classify information are the backbone of the new information frontier. One problem we're going to run into is when two different people are using two different vocabularies which happen to use the same terms to describe different meanings.

The problem with multiple vocabularies that contain the same terms but apply different meaning to them is that we destroy the author-intended meaning of the information if we attempt to merge the information. It is bad to assume binary compatibility between the meanings expressed in vocabularies. There will be a great need for an open, unified vocabulary in the Semantic Web.

IV. Future Trends

One of the primary considerations in an open environment like Internet where anyone and anywhere can access and provide information is the credibility and reliability of available information. We cannot have a total control. We also can presume that the reliability and credibility wanted could be provided by corporate status like government institutions. But even then information could be approximate or uncertain. It is well known that humans have a natural capacity to make a good discernment when they use web documents but computers don’t. That means we need a computer-discernment to realize an automated validation of online data.

This is the future vision about a web of trust.

Of course there are some problems to resolve, like how to establish the credibility of data or how to determine the reliability of data. This is the most important part because we cannot know the exact source of the information and also we cannot control which information should be part or not of the Web. In order to have a little control of what is on the Web may be a solution would be to create a system of validation, in the way the detecting email spasms system work, where every web resource to have a signature given by an authorized party. But the inconvenience of that solution is that it costs a lot of time and effort.

Other vision about Semantic Web is a strong and bigger collaboration between researchers and business corporations for industry solutions and products which use semantic web technology to increase profits and reduce costs; a web with better defined semantic languages with an increase expressivity and a wide area of covered domains used everywhere in the simple possible way in different corporations by non-expert users without even realizing what they use. We cannot miss out the online video/TV trend that has already becoming a daily possible activity for anyone especially because of the success of YouTube. For this trend we believe that in few years TV will be available on mobiles through the Internet everywhere with higher quality pictures than now, and in a personalized way for every user after the country, age, culture or other properties that matter for each individual in part[3].

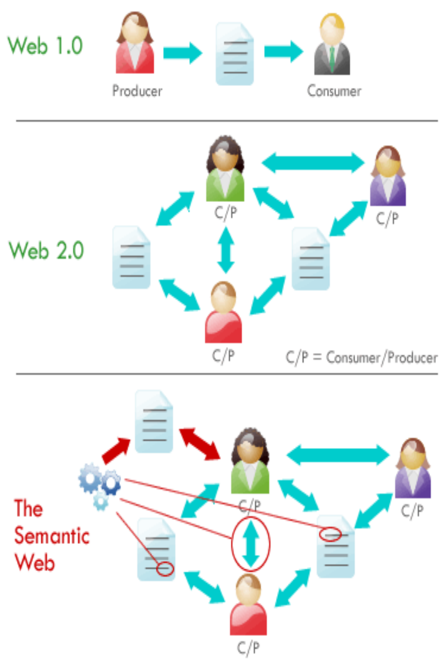


FIGURE 3.

6. CONCLUSIONS

Right now the semantic web techniques cannot replace a human. He still must validate all the results that a computer generates. Still the human is the one to formal define concepts, things, and events, real live and presented in a machine-understandable form. Even if the vision about the Web of trust can be still far way, we have to point out the important steps already achieved: RDF and OWL standards have been completed; many semantic web applications have been developed in the last years making collaboration with corporations much stronger, and given the right benefits to semantic web technologies.

There is much to be fulfilled but the opportunities are big because of the incredible capacity humans have, called knowledge.

REFERENCE

[1] http://en.wikipedia.org/wiki/Semantic_Web

[2]http://eprints.soton.ac.uk/262614/1/Semantic_Web_Revisted.pdf

[3]http://www.slideshare.net/skruk/semantic-digital-libraries

[4]http://www.l3s.de/kweb/kwepsy2006/FinalSubmissions/kwepsy2006_g uha.pdf

[5] J. Krupansky. Richness of semantic infrastructure, 2011. Online

[6] http://semanticabyss.blogspot.com/2011/06/ richness-of-semantic-infrastructure.html. Accessed on 2011-10-10

ISSN: 2229-6913 (Print), ISSN: 2320-0332 (Online) -, Web Presence: http://www.ijoes.vidyapublications.com