EXPLORING THE FULL POTENTIALS OF CONTEMPORARY SEMANTIC WEB

Senduru Srinivasulu[1]

Research Scholar,
Dept of IT Sathyabama University, Chennai
sendurusrinivas@gmail.com

P.Sakthivel[2]

Associate Professor,
Dept of ECE Anna University, Chennai
psv@annauniv.edu

Abstract

Technologies are evolving at an astounding rate. Due to this, platforms and services provided over the past few years have become outdated in the current world. The very evolution of the internet and web platform has been extremely fast paced and we are currently in the web 2.0 iteration. One of the primary contenders for the next iteration of the web platform is the semantic web technology (Web 3.0 as it is popularly known) and this has gained acceptance among various researchers and web experts. This research is not based on the effort to make the machines understand human language but to provide a better user experience. Many researchers believe that Semantic web will have a big impact on human life but there is still a lot to achieve before this vision becomes a reality. Implementation of semantic web faces challenges on multiple fronts. Even though web 3.0 is more theory than reality that hasn t stopped people from guessing what will come next. This paper explores the evolution of web and various factors that influenced the conceptualization of Semantic Web

Keywords: semantic web, web 3.0, web 2.0, web 1.0, RDF ,XML, ontology, OWL, contemporary semantic web

I. Introduction

The current day world-wide web is a large collection of HTML documents inter-connected by hyperlinks. Its main purpose is to enable human information seekers to browse, assimilate the contents and make use of it in appropriate ways. The popular mode of accessing the huge amount of information present on the web is through the use of search engines. However, search engines find it very difficult to keep the index that maps keywords to the set of relevant documents up-to-date. Thus, our ability to access information in contingent on the capabilities of the search engine. In addition, search engines find it difficult to extract the semantics behind the search keywords, the context of the search. Making semantic sense of the documents is also a tall order at present. In this context, a new set of standards termed as Semantic Web Technologies have emerged recently and it is believed that adopting these would take us forward towards making better sense of the information present of the web, interoperate information sources, and represent knowledge from diverse domains of interest.

The information available on the web is vast and it is increasing with each passing day but the technology that can process this information is not (Priddy).The reason to the increase is information is because the web can be accessed by millions of resources irrespective of their language and physical location (Benjamins, Contreras, Corcho, & Gómez-Pérez, 2002). The users of internet have grown from 16 million at the end of 1995 to 2280 million in the month of March, 2012 (Internetworldstats, 2012). This has resulted in an exponential growth in the amount of data from 18 exabytes in 2002 to 295 exabytes in 2007 (Jon Stewart, 2011) (Lyman & Varian, 2003). Just to describe the quantity of this data, an Exabyte is the equivalent of 1 billion gigabytes (Margaret Rouse, 2005). A significant challenge posed by this large amount of data is the available processing power to successfully analyze this data, negate the duplications and map them successfully to respective contexts. The computers have not yet advanced to perform complex task that can support the exchange of different structures and data formats at such an extensive level. A need to exploit other information rises to deal with the size, language and the formats issues (Benjamins, Contreras, Corcho, & Gómez-Pérez, 2002).That is when the Semantic web comes in.

Semantic web was an initiative of W3C (World Wide Web consortium) which was envisioned by its founder Tim Berners Lee. Since, the introduction of this technology several research and developments have been conducted which have resulted in a set of standards and tools to help support this vision (Matthews).

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation "Berners-Lee et al).

According to Tim Berners Lee the goal is not to make the computers understand human language but it is to define a universal model to express the information and a set of rules that machines can understand and use easily in order to process the information in a way as if they really understood it. It has the potential to overcome the drawbacks of current web by allowing the applications to integrate the content, access various resources and produce additional value for the user. (Alesso & F.Smith, 2009). Standards like the RDF (Resource description framework) and OWL (Web Ontology language were developed to describe the web semantically. The accomplishment of such a technology is considered to be a great challenge not because of the complexity involved in implementing it but also because it can be applied in vast areas (Pereira & Freire, 2009). Many researchers, companies and institutes believe that semantic web will have the same impact on life as the current World Wide Web has (Benjamins, Contreras, Corcho, & Gómez-Pérez, 2002).

The use of format s such as RDF have helped decouple the conventionally tightly intertwined logical model and the physical database (relational databases, flat files) (Feigenbaum, The What and Why of Semantic Web Technologies, 2012).This flexibility means that companies need not spend hours and millions to keep a track of changes in their information capture and upgrade their information storage accordingly. This flexibility helps various system software enables bypass multiple levels of ensuring coordination with the other information owners with whom the data is being shared. The key advantage of defining such formats is the ability to use information across various systems, devices and platforms. For example, the information on hard files such as excel sheets, across websites such as social networks can be easily integrated into keeping the internal software and systems up to date with the latest trends and forecasts. (Feigenbaum, 2012).

The Semantic Web has attracted a lot of attention from people from world over and it still manages to be in the news. The impact of semantic web on the public sector organizations has been stamped by the intention of the U.S. government to integrate semantic web features into its Data.gov website (BUSKIRK, 2010) and an investment of £30 Million in semantic web by the U.K. government into an Institute of Web Science run by Tim Berners-Lee (Jackson, 2010). In parallel, many areas of use for semantic web technology have been explored in the practical integration of social network information (The FINANCIAL, 2012) into bringing various systems up to date for disaster response, understand user responses and create information pages in a rapid cycle (PRWeb, 2011), utilization of widespread information in improving analytics for defence decisions (PRWeb, 2011) and other related areas of application.

II. Need for semantic web

Even though the existing web is in itself a big success, it lacks user support in finding, combining and extracting information (Fensel, Facca, Simperl, & Toma, 2011). The semantic web emerged as a result of lack of added functionality in the previous versions of the web. Web 1.0 lacked scalability, context and stability (Strickland, 2008).The next version Web 2.0 which is also known as the social web had some drawbacks. It lacked interoperability, personalization and true portability (Strickland, 2008).

Today most of the content on the Web is designed in such a way that it can be read by the humans and not by the computers programs to interpret or manipulate its meaning (Berners-Lee, Hendler, & Lassila, The Semantic Web, 2001).Searching for a relevant piece of information on the web is in itself a big task. Finding an information on the Web be it version 1.0 or 2.0 is mostly on the basis of keyword search which has restricted accuracy and recall (Fensel, Facca, Simperl, & Toma, 2011) and this is because the written language is ambiguous. For instance, the search results retrieved from a search engine for a keyword doesn t include the synonyms of the query keyword even though it would be relevant. To get to the right information one has to go through a huge amount of irrelevant information. Often the searches are not precise and mostly return pointers to thousands of pages (Fensel, Hendler, Lieberman, & Wahlster, 2005) and this becomes worse as the web grows.

In fact, as the web grows maintaining sources on the web also becomes difficult. Maintaining the information correct was not entirely supported by the web tools and therefore it was difficult to maintain any consistency (Fensel, Hendler, Lieberman, & Wahlster, 2005).As a result there are plenty of sites with unreliable information.

Another limitation of the keyword search is the usage of homonyms (Fensel, Facca, Simperl, & Toma, 2011). For example, searching for information on "Bear" would return pages containing information of both

"Bear" which is "the animal" and "Bear" which also means "Tolerate" even if the user wants information on only one of them. As mentioned in the beginning finding information is not the only limitation, extracting the information is also another limitation of the current web. Extracting information from the same web page is a difficult task because the web pages have different syntaxes and formats (Fensel, Facca, Simperl, & Toma, 2011).

Another limitation of the current web is in combining the information on the web. For instance, planning an inter-city travel would require transport which could be either a bus, train or even a flight. But the information when searched on the web would return different alternatives from different websites. Most often it is the case when a combination the solution is required to complete the request. (Fensel, Facca, Simperl, & Toma, 2011).

Due to the problems mentioned above there was an emerging need to provide a solution to these problems. To fix these problems there emerged a need to develop a machine that understands semantics for all the information that is presented in the web or the WWW (Fensel, Hendler, Lieberman, & Wahlster, 2005)."The solution, called the Semantic Web, enables machines to understand and potentially satisfy user requests, by processing the meaning or "semantics" of information" (Fensel, Facca, Simperl, & Toma, 2011).

If computers could understand the meaning behind the information provided by the user then, it would be able to learn and interpret what the user is interested in and can help the user find exactly what he wants. It can easily understand the relationship between things and recognize places, products, people, events etc. "Very broadly, things on the Internet will be described with descriptor languages so that computers can "understand" what they are" (Wong, 2011). This technology is referred to or called as the Semantic web technology. Semantic web or web 3.0 is an emerging technology

III. Semantic Web and its concepts

The web pages are all written in HTML. The structure of the information is described by HTML. "Standards must be defined not only for the syntactic form of documents, but also for their semantic content" (Decker, et al., 2000). HTML only describes the syntax and not the semantics HTML s simplicity hampered advanced applications for many tasks and domains (Fensel, Hendler, Lieberman, & Wahlster, 2005).Therefore, another language was defined which was XML. XML is a complementary technology and not a replacement of HTML. (Alesso & F.Smith, 2009).

The semantic web is layer which is made on top of the existing web. That is why it is called the extension of the current web. The semantic web has the architecture that can make the web a global database. The structure of the semantic web is in the form of stacks which is growing.(Refer Figure:1) semantic web architecture is designed in such a way , it supports scalability ,heterogeneity and can support any new advanced technology to sit on web stacks.

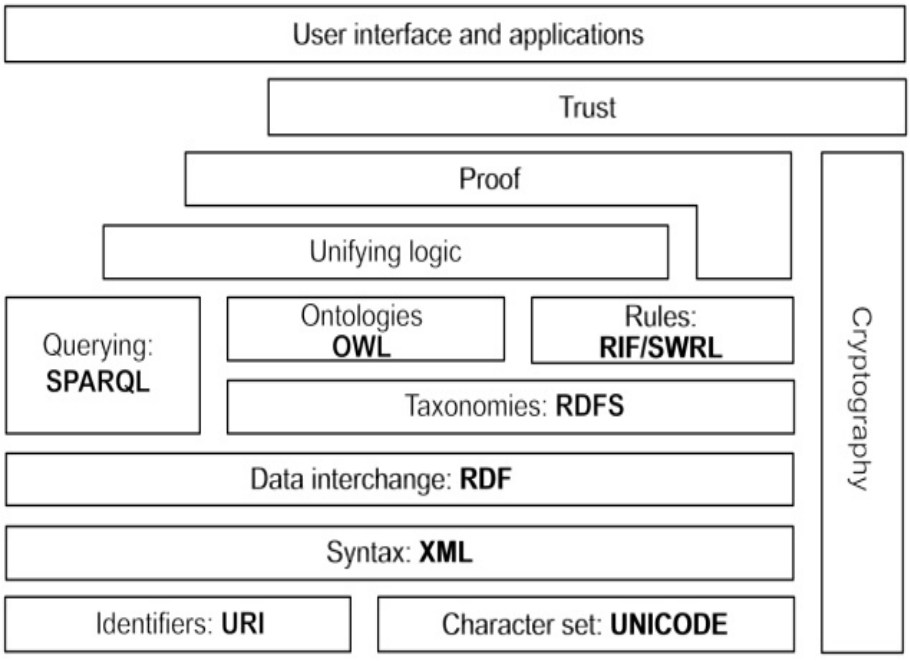

Figure:1The Semantic Web stack

Set of tools, standards and technologies form the building blocks to support the vision of semantic web. The architecture of semantic web starts with the URI and Unicode. URI is the Universal resource identifier which identifies the physical resource and Unicode provides a code or a number to each character independent of the language, program or the platform (Cardoso, 2007).The basic layer for data representation is the Resource Description Framework. It is based on the XML standards. The Ontology layer is built on top of the RDF (Fensel, Facca, Simperl, & Toma, 2011)which is also called the OWL language. The functionalities of XML, RDF, XML and RDF Schema and OWL are described below.(Figure 2)

IV. Construction of Semantic web

To develop the semantic web two main standards are important (Berners-Lee, Hendler, & Lassila, The Semantic Web, 2001) .

XML(Extensible Markup language)

RDF(Resource Description Framework)

a. XML and XML Schema

It is a mark-up-language which has brought about some great prospects to the development of semantic web (Lu, Dong, & Fotouhi, 2002).It is very useful for the exchange of metadata. Extensible in XML refers to the ability to exchange information regardless of the applications (Culley, 2006).It provides the syntax and structure so that it can be read by the whole web.XML schema describes the application of XML.

b. RDF and RDF Schema

RDF was developed by W3C.It assigns a URI which is specific to its individual fields. It is a data representation language about Web resources (Moldovan, 2012).RDF encodes and interprets the resources in such a way that any machine can access it in spite of the vocabulary it uses (Culley, 2006).RDF builds standards for XML application in order to intercommunicate and interoperate easily. It also facilitates system and data integration and interoperability (Cardoso, 2007).The RDF allows the developers to utilize metadata and design products that will create a better search engine that will enable the users to have control over what they want to view. It also supports a wide range of ontologies (Priddy).

RDF schema is a type of vocabulary that defines the properties and attributes of RDF resources.RDF Schema is a semantic extension of RDF (Brickley & Guha, 2004).

The XML and RDF do not provide any properties of their own and hence we require OWL (Web Ontology Language)

Table 1: Codes categorized into Technical and social issues.

THEME	TECHNICAL	SOCIAL
ISSUES WITH RDF	1. Improper planning 2. Hidden complexities 3. RDF stack slower than other systems 4. RDF format is complicated 5. Complicated documentation 6. Flawed concepts 7. Difficult to implement	1. Insufficient popularity 2. RDF expensive to run 3. High learning curve 4. Lack of awareness 5. Lack of common initiative
STANDARDIZATION	1. Complex standards 2. Errors due to high complexities 3. Standard technology issue	1. Multilingualism
LACK OF KNOWLEDGE	1. Unable to explain propositions	1. Fading popularity 2. Yet to deliver results 3. Setting incorrect expectations 4. Fading confidence in the potential 5. Lack of business focus 6. Lack of awareness of semantic web 7. Over-hyped by start-up firms. 8. Lack of common language

PRIVACY AND TRUST	1. Accuracy of information	1. Excessive trust in the data available on the internet 2. Misuse of information 3. Can keep track of consumer behaviour 4. High risk of fraud 5. Lack of privacy
HUMANS ARE IRREPLACEABLE	1. Machines improve productivity productive 2. Machines cannot perform unless told 3. Semantic web cannot match humans	1. Human memory 2. Semantic web v/s human thoughts

c. OWL: Web Ontology Language

It is vocabulary extension of RDF or Resource description framework. It can be used to represent the meaning of the terms explicitly in vocabularies and also the relationships between those terms (McGuinness & Harmelen, 2004).It uses the linking that is provided by RDF which allows the ontologies to be distributed across various systems (Shadbolt, Hall, & Berners-Lee, 2006)

Representing the terms and their inter-relationships is called Ontology.

d. **Ontology**

Ontology is concerned with the study of existence. In terms of computer science it is technically defined term to represent an object that enables the knowledge modeling of a domain which can either be real or imaginary. It allows domain communication between application systems and people (Cardoso, 2007).Ontology is defined by a set of representational primitives with which a domain of knowledge can be modeled (Gruber, 2009).The primitives are basically the classes; properties or relationship between the classes. Ontology presents the vocabulary required and also provides the relations among them (Ding, Fensel, & Stork, 2003).It can be used to develop communications between machines and humans. These can be grouped into three areas (Jasper & Uschold, 1999):

- To attain interoperability
- To help agents and humans communicate
- To improve the quality and process of software engineering systems.

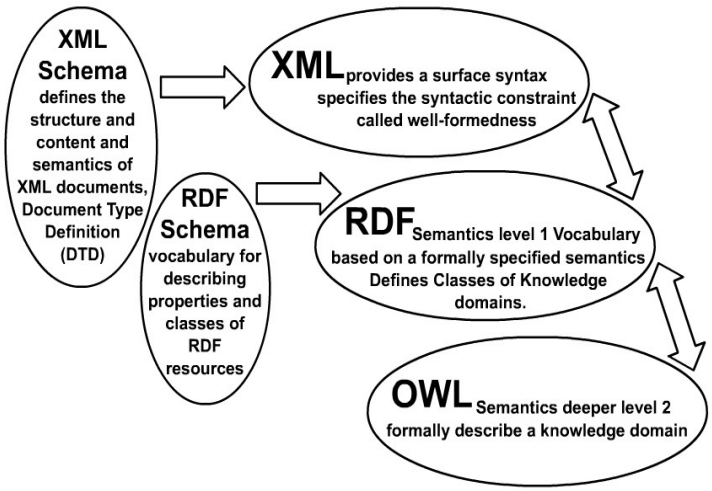

Figure 2: Semantic web technologies

Further, we will discuss the possible benefits of semantic web technology. It is to be noted that these benefits are part of a vision which is incomplete and yet to be realized and are in the process of development.

V. Key Benefits of Semantic web

a. Enhanced search

Semantic Web provides enhanced search in a complex web site (Roberta, Alexandre, Vincent, & Carlo, 2007). For example: An employee is searching for patient s information in a certain health organisation. The data for the search criteria is very less i.e. the gender and the birth year. This search would give a huge amount of information which would need to be isolated by the employee for the desired information. If the information is already streamlined in the organisation then the search would have produced very few results which would be closer to the desired information. (Priddy).The Semantic web identifies the web based data which generates effective and productive searches.

Semantic based search engine may prove more useful for the users compared to the search engines that exist today. Resources on the web will be characterized semantically in the form of ontologies. The ontologies will increase the precision of the result (Lu, Dong, & Fotouhi, 2002). For example: If you search for a query say "return all the versions of the book " Web 3.0" on a semantic search engine then the search result would return only the versions of the book and not the keyword" version" and/ or the word "Web 3.0".

b. Reaping benefits from own Data

Whenever a person access a site for any activity such as searching for information, shopping online etc. they leave a trail of information. All the information is recorded. Therefore, there is no control over the data that you share on the internet and this happens without the person s knowledge. The Semantic web claims to change this by providing the power to the users to control the information that is shared on the internet by using a concept called „personal data locker (Wong, 2011).This concept will grant selective access of the information that a person wants to share.

c. Software Engineering

There are lot of steps required to make software. It is a hideous task to design and implement them; hence it will be fruitful to be able to reuse software. Currently, it is possible with only projects that are similar. In Semantic web technology this can be possible with the use of RDF and ontologies.

There are number of software that are used and programmed world over. The idea behind introducing common ontologies in all the software was to make it easy to understand the terminology in other software. Therefore, it makes it easy to use global software with any platform (Harvinder, 2011).

VI Conclusion

The vision of the Semantic web is huge and very inspiring but it has not been able to make a mark even after over a decade of research and development. The World Wide Web has exploded in terms of the number of users and the data being stored and shared. This has created huge complexities in managing data and utilizing it across multiple platforms. Due to these reasons, semantic web has not taken off when compared to the expectations. The success of semantic web rides on the ability of experts and corporate organizations set common format and platforms of operating which can be gradually built to adapt to this burgeoning information levels on the web. There is still a lack of focus from the common public in general and also the developers themselves simply because there are very few people who are aware of this ongoing effort or because they are too comfortable with the current web and there is nothing they need to complain about.

VII. References

[1] Richard Benjamins , Jesús Contreras , Oscar Corcho , Asunción Gómez-pérezSix Challenges for the Semantic Web (2002) In KR2002 Semantic Web Workshop
[2] Lyman, Peter and Hal R. Varian, "How Much Information", 2003. Retrieved from http://www.sims.berkeley.edu/how-much-info-2003
[3] h. peter alesso, craig f. smith thinking on the web: berners-lee, godel and turing (wiley interscience series 2009.
[4] The what and Why of semantic Web Technologies By Lee Feigenbaum (@leefeigenbaum) Sep 4, 2012 D.
[5] Fensel et al., Semantic Web Services Descriptions ,DOI 10.1007/978-3-642-19193- 0_12, © Springer-Verlag Berlin Heidelberg2011
[6] Tim berners-lee, james hendler and ora lassila Scientific American: Feature Article: The Semantic Web: May 2007.
[7] Fensel, Hendler, Lieberman, & Wahlster Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential 2005
[8] Dieter Fensel, Ioan Toma, Federico Michele Facca, Elena Simperl , Semantic Web Services Springer 2011
[9] The semantic web: the roles of XML and RDF - Decker, Melnik, et al. – 2000
[10] McGuinness & Harmelen, OWL web Ontotlogy Language W3C Recommendation 10 February 2004
[11] Nigel shadbolt ,Tim Berners-Lee,Wendy Hall The Semantic Web Revisited IEEE Intelligent Systems Volume 21 Issue 3, May 2006