Semantic Web

Eshtaartha Basu

1. Introduction

Semantic Web is an extension of World Wide Web which enables computers to make sense of the documents or data that is being processed instead of just viewing them as raw data. The term "Semantic Web" was coined by Tim Berners-Lee with a vision of a web of data that can be processed by machines. Semantic web is a "set of standards and best practices for sharing data and the semantics of that data over the web for use by applications".

The main objective of this paper is to provide a detailed analysis of Semantic Web by comparing it with existing web technology, highlight the critical role semantic web could play in shaping our daily lives and future technologies and also understanding the challenges that need to be overcome in order to fully utilize the potential of semantic web.

2. Semantic Web

In this section, we try to understand the Semantic Web in further detail by going through its structure and other aspects.

2.1 Semantic Web: The Structure

Semantic Web follows the standards and practices given by World Wide Web Consortium(W3C). Some of the standards and practices are:

a. Resource Description Framework(RDF) data model.

b. The SPARQL query language

c. The use of URIs to name things.

Now, let's look at the structure of Semantic Web Stack

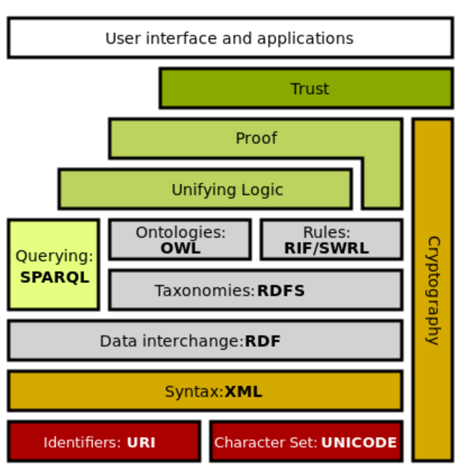

Fig.1 Semantic Web Stack

As shown in Fig.1, Semantic Web uses UNICODE character set to represent data. Uniform Resource Identifiers(URIs) are used to identify the data. XML provides elemental syntax and structure within documents. The Resource Description Framework(RDF) is used to model the data. The RDF is made up of data entities called triples. A triple defines the relationship between a subject, a predicate and an object. Modeling the data in this way and defining relationship with the data elements will help computers make sense of the data in a better way. Web Ontology Language(OWL) is used to enhance the vocabulary for describing properties. SPARQL serves as the query language. The web rules are defined by Rule Interchange Format(RIF).

2.2 Semantic Web Technology versus Web 2.0[1]

This section compares semantic web and Web 2.0

The Web 2.0 has been developed as a network of documents for humans whereas semantic web is a network of meaningful information for computers and humans.

Using Web 2.0 technology, machines(computers) merely display the data. They cannot make much sense of it. But, Semantic web technology will bring structure to the "meaningful content" of web pages.

Semantic will enable computers to perform sophisticated tasks without the need of high end artificial intelligence or extensive processing.

Let's consider an example to understand why semantic web can reduce processing time and improve performance by several folds. Joe is active on social media and has accounts on Facebook, Twitter and LinkedIn. If Joe updates his job information on LinkedIn, it won't be reflected on Twitter or Facebook because Joe's job status is not a unique global property linked to Joe which the can be understood by the browser/computer. As a result, Joe will have to update his job status in multiple sites repeatedly. Even if his job status were to be extracted automatically, it would require extensive processing using AI algorithms. Such problems can be easily avoided using Semantic Web.

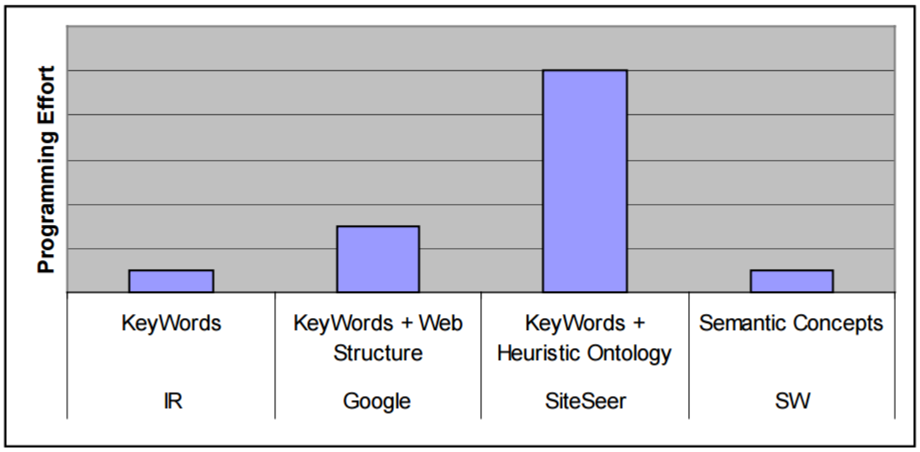

Fig. 2 Programming effort required for various web applications [4]

Assuming semantic web content already exists, the programming effort required for various web apps is shown in figure 2. Semantic web apps will require significantly less programming effort compared to other web 2.0 sites.

2.3 Why Semantic Web is critical for future technologies and how it can improve our lives

The Semantic Web's broad goal parallels that of many bioinformaticians and other scientists. There are vast quantities of biological data and associated annotations, or knowledge, now available on the Web. These resources are highly distributed and heterogeneous. This heterogeneity exists at many levels, the most pernicious of which are the semantic heterogeneities in the schema and the values placed in those schema. Semantic Web technologies and the vision itself offer a solution to this longstanding problem in creating an integrated view of bioinformatics.

With the advent of social media and big data there is an ever increasing demand for smooth and efficient information exchange and information retrieval while maintaining minimum costs. Semantic web will serve this purpose although initial setup costs may be a bit high during the transition from web 2.0 to semantic web.

Once we achieve complete transition to semantic web, user experience will be enhanced due to smoother and faster exchange of information, reduced cost of internet based services and improved automation.

3. Major challenges[1][3]

This section mainly deals with the major challenges which Semantic Web needs to overcome in order to become successful. They are:

1. *Lack of semantic content*: Currently there is very little semantic web content available. All the existing web content needs to be upgraded to semantic web content.

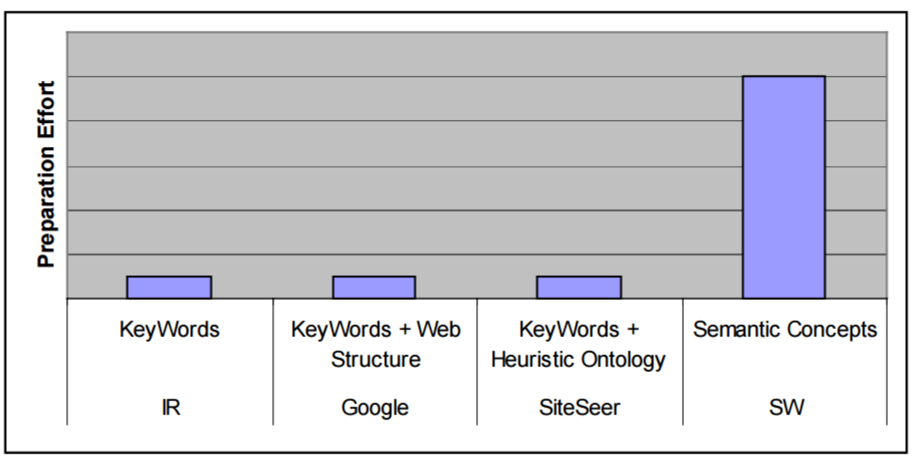

Fig 3. Web content preparation effort for various web applications [4]

As shown in figure 3, a simple keyword based IR system, Google and SiteSeer do not need any web preparation effort to work whereas a semantic web app needs significant annotation effort to put the knowledge into web and turn it into semantic web.

2. *Scalability of mechanisms*: Mechanisms and ontologies that can deal with semantic data needs to be developed and made available in a scalable and organized manner.

3. *Vagueness of information*: There are several imprecise properties/concepts associated with users like ‘tall’ or ‘short’, ‘slim’ or ‘fat’ etc which arise due to vagueness in user queries and vagueness is information provided by content providers.

4. *Multilinguality*: Data and content needs to be made available in several languages. This will involve development of software algorithms which can operate without being affected by the native language used to display the content.

5. *Inconsistency*: When large number of ontologies from several sources are combined, logical contradictions may arise.

6. *Need for better visualization*: Since Semantic Web has the potential to deliver lots of meaningful data (information overload), better visualization techniques need to be implemented to make sure the user can quickly and easily grasp the information being displayed.

7. *Need for maturity*:

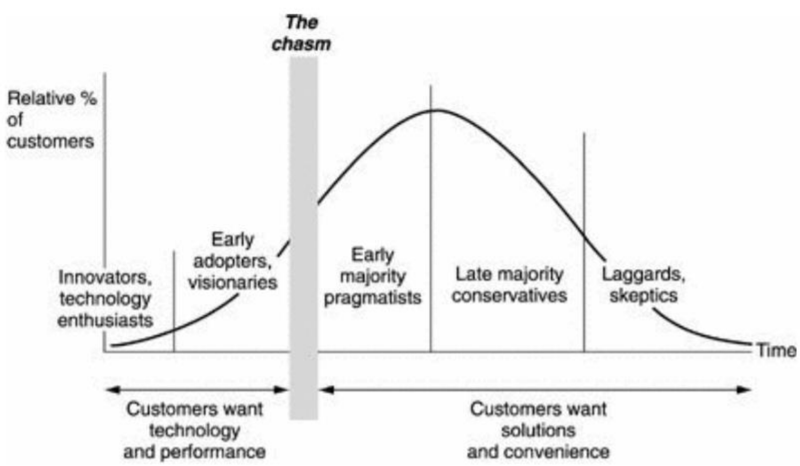

Fig 4. Variation of customer base and customer expectations over time for any new technology[2]

The above figure shows the variation of customer base and expectations over time for any new technology. The semantic web has to be pass the initial chasm in order to be successful. In its current state, semantic web is on the verge of passing the chasm.

4. Conclusion

In this paper, the concept of semantic web and its structure was introduced. A detailed analysis of semantic web was done by comparing it with existing web technologies and assessing its role in future technologies while keeping the challenges in mind.

5. References

[1] Web 2.0 vs. the Semantic Web: A Philosophical Assessment

http://blogs.oii.ox.ac.uk/floridi/wp-content/uploads/sites/67/2014/05/w2vsw.pdf

[2] Semantic Web in 10 years: Semantics with a purpose

http://stko.geog.ucsb.edu/sw2022/sw2022_paper9.pdf

[3] Semantic Web Technologies and Data Management

https://www.w3.org/2007/03/RdfRDB/papers/ma.pdf

[4] Six Challenges for the Semantic Web

https://wikis.gsic.uva.es/juaase/images/b/bb/Benjaminsetal.pdf