Gopal Pandey/ International Journal of Engineering Research and Applications
(IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 1,Jan-Feb 2012, pp.780-786

The Semantic Web: An Introduction and Issues

Gopal Pandey

(Department of Information Technology, Shantilal Shah Engineering College, Bhavnagar, Gujarat, India)

ABSTRACT

Current Web content has been designed for direct human processing and thus lack of computer processing elements. Semantic Web aims at computer process able information. Semantic Web is considered to be the next generation web, so lot of research and development is going on. This paper describes the Semantic Web-stack and working of its layers. Further Semantic Web is compared with Current Web using various factors. Finally focuses on major issues of Semantic Web.

Keywords - **Annotation, OWL, RDF, SPARQL, XML**

1. INTRODUCTION

The Semantic Web is "a mesh of information linked up in such a way as to be easily processable by machines, on a global scale ". You can think of it as being an efficient way of representing data on the World Wide Web, or as a globally linked database.

The Semantic Web was thought up by Tim Berners-Lee, inventor of the WWW, URIs, HTTP, and HTML. There is a dedicated team of people at the World Wide Web consortium (W3C) working to improve, extend and standardize the system, and many languages, publications, tools and so on have already been developed. However, Semantic Web technologies are still very much in their infancies, and although the future of the project in general appears to be bright, there seems to be little consensus about the likely direction and characteristics of the early Semantic Web.

There is no global system for publishing data in such a way as it can be easily processed by anyone. For example, just think of information about local sports events, weather information, airline timetable, educational institute data, news, television guides etc. All of this information is presented by numerous sites, but all in HTML. And due to that it is difficult to use this data in the ways that one might want to do so.

So the Semantic Web can be seen as a huge engineering solution we may find that a large number of Semantic Web applications can be used for a variety of different tasks, increasing the modularity of applications on the Web.

The Semantic Web is generally built on syntaxes which use URIs to represent data, usually in triples based structures: i.e. many triples of URI data that can be held in databases, or interchanged on the WWW using a set of particular syntaxes developed especially for the task. These syntaxes are called RDF (Resource Description Framework) syntaxes.

Semantic Web enables semantic interoperability, which involves understanding the meaning of data, not the syntactic structure. Syntactic interoperability is primarily concerned about parsing the data, while semantic interoperability refers to the definition of mapping between unknown terms and known terms.

2. SEMANTIC WEB – STACK

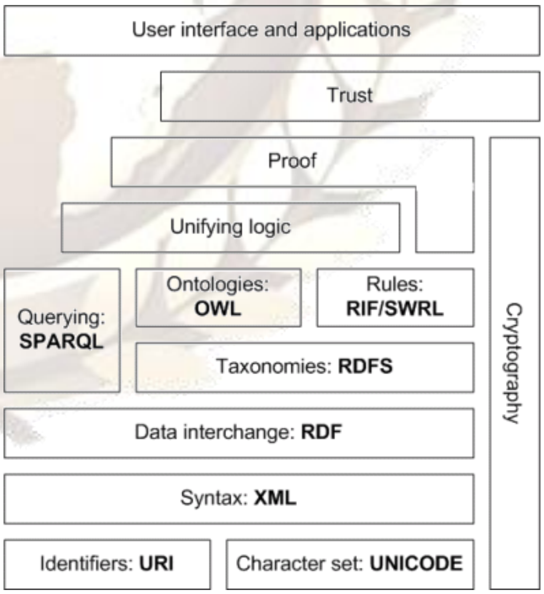

Fig. 1 Semantic Web Stack-A layered approach

Layer 1: URI and Unicode

The lowest layer is responsible for the encoding of any character written in any language and responsible for uniquely identifying different resources. This layer standardized the text using Unicode and URI (Uniform Resource Identifier).

Unicode serves to represent and manipulate text in many languages. Semantic Web helps to bridge documents in different human languages.

Layer 2: XML

XML is used to achieve interoperability on the internet. This layer is provides base to mix different elements from different vocabularies to a specific function. XML provides an elemental syntax for content structure within documents, yet associates no semantics with the meaning of the content contained within. XML is not at present a necessary component of Semantic Web technologies in most cases, as alternative syntaxes exists, such as Turtle. Turtle is a de facto standard, but has not been through a formal standardization process.

Layer 3: RDF

The first two layers consist of basic internet technologies. From this layer actual Semantic Web starts.

Resource Description Framework (RDF) is W3C standard for describing Web resources. RDF is written in XML. RDF is not design for presentation to human but for the purpose of read and understood by the computers.

RDF is a simple language for expressing data models, which refer to objects ("resources") and their relationships. An RDF-based model can be represented in a variety of syntaxes, e.g., RDF/XML, N3, Turtle, and RDFa. RDF is a fundamental standard of the Semantic Web and recommended by W3C.

RDF Schema (RDFS) extends RDF and is a vocabulary for describing properties and classes of RDF-based resources, with semantics for generalized-hierarchies of such properties and classes.

Layer 4: The Ontology layer

Ontology is about the exact description of things and their relationship. Sometime RDFS definition of particular Web resource is not sufficient then more extensive ontological vocabulary is needed.

OWL (Web Ontology Language) is a language for processing web information. It is a W3C standard.OWL has three sublanguages:

- OWL Lite
- OWL DL(Description Logic)
- OWL Full

OWL adds more vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties and characteristics of properties (e.g. symmetry), and enumerated classes.

However RDF and OWL are much of the same thing, but OWL is a stronger language with greater machine interpretability with a larger vocabulary and stronger syntax than RDF

SPARQL (Simple Protocol and RDF Query Language) is a protocol and query language for Semantic Web data sources. SPARQL is a RDF query language, used to query any RDF-based data (statements involving RDFS and OWL).

RIF (Rule Interchange Format) is a W3C recommendation. SWRL (Semantic Web Rule Language is a proposal for a Semantic Web rules language, combining sublanguages of OWL (OWL DL + OWL Lite)

RIF or SWRL will bring support of rules. This is important for example to allow describing relations that cannot be directly described using description logic used in OWL.

Cryptography is important to ensure and verify that Semantic Web statement is coming from trusted source. This can be achieved by appropriate digital signature of RDF statements.

Layer 5: The Logic layer

For expressing rules a logic layer is needed. This layer is used to enhance ontology language further.

Layer 6: The Proof layer

The Proof layer involves the actual deductive process as well as the representation of proofs in Web languages. In the vision of Tim Berner-Lee the production of proofs is not the part of the Semantic

Web. The reason is that the production of proofs is still a very active are of research and it is by no means possible to standardize this. A Semantic Web engine should only need to verify proofs. Someone sends to site X a proof that he is authorised to use the site, then site X must be able to verify that proof. This is done by a suitable inference engine.

Layer 7: The Trust layer

Without trust the Semantic Web is unthinkable. If company A sends information to company B but there is no way that B can be sure that this information really comes from A to that A can be trusted then there remains nothing else to do but throw away that information. Digital Signature can be used for this purpose. Trust to derived statements will be supported by verifying that the premises come from trusted source and relying on formal logic during deriving new information.

Layer 8: The UI and Application layer

UI (User Interface) is the final layer that enable human to access the Semantic Web application.

3. Comparison between Current Web and Semantic Web

Semantic Web can be compare with current web within several parameters such as contents, conceptual perception, scope, environment and resource utilization.

3.1 Content

Semantic Web encompasses actual content along with its formal semantics. Here the formal semantics are machine understandable contents, generated in logic-based languages such as OWL. In today's web there is no formal semantics of existing contents. These content are machine-readable but not machine understandable.

3.2 Conceptual Perception

Current web is look like a book having multiple hyperlinked documents. In book scenario, index of keywords are presented in each book but the contexts in which those keywords are used, are missing indexes. Means there are no of formal semantic of keywords in indexes. To check which one is relevant, we have to read the corresponding pages of that book. That is the case with current web. In Semantic Web this limitation is eliminated via ontologies where data is given with well-defined meanings, which can be understandable by machines.

3.3 Scope

It is estimated that there are billions pages of information available on the web, and only a few of them can be reached via traditional search engines. In Semantic Web formal semantics of data are via ontologies, which are essential components of Semantic Web accessible to semantic search engines.

Table 1 Current Web Vs Semantic Web

Sr No	Web / Factors	Current Web	Semantic Web
1.	Conceptual Perception	Large hyperlinked book	Large interlinked database
2.	Content	No formal meanings	Formally defines
3.	Scope	Limited	Boundless
4.	Environment	Web of documents	Web of ontologies, data and documents
5.	Resource utilization	Minimum-normal	Maximum
6.	Inference/ reasoning	No	Yes
7.	Knowledge management	No	Yes
8.	Information searching, accessing, extracting	Difficult, time consuming task	Easy and efficient
9.	Timeliness, accuracy, transparency of information	Less	More
10.	Semantic heterogeneity	More	Less
11.	Ingredient	Contents, presentation	Contents, presentation and formal semantics
12.	Text simplification, clarification	No	Yes
13.	Technology	AJAX	RDF/OWL

3.4 Environment

Semantic Web is web of ontologies having data with formal meanings. This is in contrast to current web which contains virtually boundless information in the form of documents. On the other hand Semantic Web is having data as well as documents that machines can process, transform, assemble and even act on data in useful ways.

3.5 Resource Utilization

There are many web resources that may be very useful in over everyday activities. In current web it is difficult to locate them, because they are not annotated properly by the metadata understandable by machines. In Semantic Web there will be network of related resources. It will be very easy to locate and to use them in Semantic Web world. Similarly there are some other criteria factors for comparison between current web and Semantic Web listed in the Table 1.

4. ISSUES OF SEMANTIC WEB

New technology has always introduced certain issues. Automated reasoning systems will have to deal with all of these issues in order to deliver on the promise of the Semantic Web. Major issues of Semantic Web are as follow.

4.1 Ontology development:

Development of Semantic Web depends a lot on the ontology. Attempts for exploring different aspects of ontology, such as ontology representation languages, ontology learning approaches and ontology library systems, which manage, adapt, and standardize ontology's.

4.1.1 Management

Ontology is purposed to enable knowledge sharing and re-use. Typical ontology library system supports identification and versioning with open storage and organization. The later shows how the ontology's are stored and organized to facilitate access and management of ontology's. Identification provides a unique identifier whilst versioning is an important feature on time basis as it ensures congruency amongst versions of ontology's.

4.1.2 Adaption

As time passes by, ontology's evolve and their extension and their updation becomes an issue to be resolved. Resulting in searching, editing and reasoning of ontology's in an library system ontology.

4.1.3 Standardization

For any open system like Semantic Web, integration and interoperability is always an area of concern. At internet level, expectation of being scalable is normal. For ongoing basis, multiple representation languages have been proposed and even library systems have been built. But still standardization is not done because; every entity had their own set of advantages and disadvantages. Since, Semantic Web is in its early gestation period enforcement of standard may not seem to be appropriate. Popularity of these representation languages will contribute in making them standards in future. XML would be the meta-language for facilitating integration and interoperation amongst these representation languages.

4.2 Formal semantics of the Semantic Web languages

Semantic Web is composed of three functional layers i.e. the metadata layer, the schema layer and the logical layer. For metadata layer, RDF is the most popular data model. Although RDF is considered efficient for metadata layer, the semantics of reification (statement about statement) is yet to be defined by that. Another model known as RDFS extends RDF as a popular schema layer language. RDFC deficits formal semantics and a proposal for metamodeling architecture is defined similar to UML (Universal Modeling Language) which further defines the formal semantics. Though semantic confusion persist, RDFS has shown that a formal semantic is possible. For further development of Semantic Web, RDFS needs to be resolved by defining a formal semantic for it.

4.3 Proof and trust

"Anybody can say anything on anybody" is the basic principle on which the Semantic Web carries out its working. Conflictions and contradictions are possible amongst views of people around the world. Hence, one needs to make sure that the original source does make a particular statement (proof) and that source is trustworthy (trust).

4.3.1 Proof

For proofing, digital signatures would play an important role. The source would be allowed to sign the statement he makes for making agents cross-check whether the information provided is proprietary to the claimant source. Encryption and access control are the other additional technologies that can be used.

4.3.2 Trust

Trust parameters should be defined individually. This means, any user on the Semantic Web can define the trust he possesses for any source present. Now, to define the extent of trust, "web of trust" is the solution. Here, when a user trusts the sources form XYZ user, he would automatically trust the sources that are in turn trusted by the XYZ user. This would create a huge hierarchical network which facilitates agents to infer information based on the trust knowledge.

At present, the notions for both the proof and trust are yet to be formalized and their overall integration into the Semantic Web is yet to be developed. However, these technologies are very important and are the foundation of building real commercial applications (e.g., B2B and B2C systems).

4.4 Availability of content

At present, the content which is compatible to the Semantic Web is very much limited in availability. Existing web content should be upgraded to Semantic Web content including static HTML pages, XML content, dynamic content, multimedia and web services.

Ontology's are used to annotate the Semantic Web content, which defines the meaning of words and the relationship existing between them. Without annotation, Semantic Web is a bizarre.

To annotate web pages with ontological information, HTML was used. HTML didn't turn out to be a successful attempt and resulted in a failure. Hence, creation of Semantic Web content is posed to be a serious challenge. As the architecture of Semantic Web i.e. RDFS, DAML+OIL, OIL, etc is till date under development, the amount of Semantic Web content available is less. However, besides architecture development, researchers are concentrating on building tools to support the annotation for the web content. These tools would primarily contribute in the success of Semantic Web. But, these tools do possess intrinsic limitations: 1. These tools would annotate only static pages, and 2. Major focus would be laid on creating new content.

As a solution to the above mentioned problems, a need does arise for creating a set of annotation services (as middleware) which would govern both the static and dynamic web documents possibly containing multimedia and other web services. The result form these annotation services will be craved as per the languages defined in the language pyramid for the Semantic Web enabling different agents understanding dissimilar languages of the Semantic Web. Innovation is guaranteed in this approach of annotation of resources, though ongoing efforts do propose annotation in a single language with this layered approach being taken into account in the development of languages.

4.5 Scalability

Noteworthy effort has been made for organizing the Semantic Web content, storing it and providing the mandatory functionalities for finding it. Each of the above tasks needs to be carried out and maintained in a scalable manner, as these solutions contribute to the growth of the Semantic Web.

As Semantic Web content is available, the issues are witnessed in managing the content in a scalable manner which means how its organization, storage, and appropriation of it would be carried out. Two major factors as issues are outlined as:

The storage and the organization of the Semantic Web pages is the first problem encountered. The "basic" Semantic Web consists of ontology based annotated pages whose linking structure reflects the WWW structure where the pages are linked by the means of hyperlinks. This hyperlinked configuration doesn't fully make use of the associated semantics of the Semantic Web pages. Programming indexes are realized to be helpful in grouping Semantic Web content based on related topics. This would be fruitful in making applications derive content in order to provide additional services. The generation of the semantic indexes would be carried out dynamically using ontological information and the annotated documents.

The other trouble is the discovery of information in the Semantic Web, also known as the co-ordination among the semantic indexes. A technology needs to be provided for easily finding the Semantic Web content taking into consideration the semantics of the web resources. As a solution, a P2P (peer-to-peer) architecture can be thought of similar to the current configuration of routers in WWW, to be known as a "semantic TCP/IP protocol" ", the new European Semantic Web (SWAP, Semantic Web and Peer-to-peer) project is dedicated this topics (Semantic Web and Peer-to-peer). Indexes would be treated as active agents that know their role of finding the content. Topics not found in the index are semantically routed to the neighboring indexes. Scope of use of the agents should be discovered for negotiation techniques for obtaining the semantic routing of the topics. In this manner, no central register for semantic content would be needed for making the whole infrastructure more scalable, making it hold continuous addition of the content of the Semantic Web. Note: Such arrangement would be intended of the original WWW.

4.6 Multilingualism

This problem does exist in both the Web and the Semantic web. Any of the Semantic Web approaches should provide facilities for accessing the information in multiple languages, thereby allowing the creation and Semantic Web content access independently of the native language of the content provider and the user.

Multilingualism plays an important role at the given scenarios: 1. At the level of ontology's of annotations, and 2. At user interface.

At the ontology level, native language might be used by the ontology builders for the development of the ontology's on which the annotations would be based. As all the users would not be of the ontology builder category, this leval has the lowest priority. Popular existing multilingual and linguistic resources such as the WordNet [wordnet], EuroWordNet [eurowordnet], etc can be used at this level to support Multilingualism.

At the other level, i.e. annotation level, the annotation of the content in various languages can be performed. As greater number of users especially content providers would rather annotate content than developing the ontology's, proper support is needed allowing the providers to annotate content in their native language. Semantic Web content generation effort is minimized because no requirement exists for a French to annotate content neither in the German language nor vice versa.

At last, at the user interface level, millions of people would like to access relevant content in which annotations are present in their native language without even considering the source language. As of now, most of the content is found in English language; more of the content in other languages is expected. Any possible Semantic Web approach should include facilities for accessing the information in multiple languages. Internationalization and localization techniques should be adapted and executed for availing personalized information access to the user in its native language.

4.7 Visualization

Non-rational visualization of the Semantic Web content will become increasingly important in solving the heighten amount of information overload, as easy recognition of relevant content is demanded by the user. New techniques must be developed which differ from the linking structure of the hypertext in the current Web architecture.

As users would demand easy recognition of the relevant content, information overloading does occur where the intuitive visualization of the web content would hold greater importance. Also, the use of semantic indexes and routers for the storage, organization and finding of the information will need enhancements in visualization. New visualization techniques for visualizing the Semantic Web content in any of the several Semantic Web languages such as RDFS, OIL and DAML+OIL.

4.8 Stability of Semantic Web languages

At the end of all, sincere efforts must be made for standardizing entities in this emerging field of the web which would then enhance the creation of the technology supporting the Semantic Web.

In the well-emerging field of the Semantic Web, W3C would provide with some effective recommendations on the languages and the technology that would be of use in this field. Necessity of standards brought into consideration has increased. Otherwise, a layered approach to ontology languages creation and the annotation has already been adopted by the community finding this field of worthy importance.

In the development of the Semantic Web, tool support also stands essential which are a bit dependent on the Semantic Web languages they are supposed to deal with. The W3C Semantic Web Activity comprise of the sources from Semantic Web language standard and the other necessary tool support.

5. Conclusion

In this paper I have introduced the Semantic Web and discussed different layers of Semantic Web. This is also known as Semantic Web Stack. I have compared Semantic Web with Current Web using several parameters like contents, conceptual perception, scope, environment and resource utilization. Hence have shown how semantic web is better than the current web.

Semantic web is still in development phase and thus contain few problems that need to be resolved. I have mentioned such critical issues or challenges along with their probable solutions.

Acknowledgements

I thank Prof. B.V.Buddhadev(MTech, IIT Delhi) of L.D.Engineering College, Ahmedabad for his valueable suggestions and reviewing this paper.

References

Journal Papers:

[1] Benjamins R., Decker S., Fensel D., Gómez-Pérez A.,(KA)2: Building Ontologies for the Internet: A Mid Term Report, International Journal of Human Computer Studies (IJHCS). 51(3), September 1999, 687-712.

[2] Gómez-Pérez, A.,Guidelines to Verify Completeness and Consistency in Ontologies, The third World Congress on Expert Systems. February. 1996, 901-908

[3] Alexander Maedche, Steffen Staab, Learning Ontologies for the Semantic Web, Semantic Web WorkShop 2001, Hongkong , China

[4] Knoblock, C.A., Minton, S., Ambite, J.L., Ashish, N., Muslea, I., Philpot, A.G., Tejada, S., The ariadne approach to web-based information integration. International the Journal on Cooperative Information Systems (IJCIS) Special Issue on Intelligent Information Agents: Theory and Applications, 10(1/2) , 2001, pp145-169.

[5] Gerber, Aurona; Van der Merwe, Alta; Barnard, Andries; (2008), A Functional Semantic Web architecture, European Semantic Web Conference 2008, ESWC’08, Tenerife, June 2008.

[6] Frank van Harmelen, Jeen Broekstra, Christiaan Fluit, Herko ter Horst, Arjohn Kampman, Jos van der Meer and Marta Sabou, Ontology-based Information Visualisation, Proceedings of the workshop on Visualisation in conjunction with the 5th International Conference on Information Visualisation, London 2001

Books:

[7] Thuraisingham B., XML, Databases and the Semantic Web(CRC Press, 2002)

[8] G. Antoniou and F. van Harmelan, A Semantic Web Primier (MIT Press, 2003)

[9] Liyang Yu, A Developer's Guide to the Semantic Web(Springer)

[10] Allemang, D., Hendler, J, RDF—The basis of the Semantic Web. In: Semantic Web for the Working Ontologist (2nd Ed.)

[11] Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph , Foundations of Semantic Web Technologies, (CRCPress)

[12] Thomas B. Passin , Explorer's Guide to the Semantic Web(Manning Publications)

[13] Jeffrey T. Pollock, Semantic Web For Dummies (For Dummies)

Web Resources:

[14] Web3.0,Available: http://webuser.hsfurtwangen.de /~heindl/ebte-08ss-web-20-Suphakorntanakit.pdf, viewed on 16 December 2011

[15] Klyne G, Carroll JJ, McBride B (eds) (2004) Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation. W3C. http://www.w3.org/TR/rdf-concepts/. Accessed 16 December 2011

[16] Brickley, D., Guha, R.V. Resource Description Framework (RDF) Schema Specification. W3C Proposed Recommendation. March, 2000. http://www.w3.org/TR/rdf-schema/. Accessed 16 December 2011.

[17] Frank van Harmelen, Peter F. Patel-Schneider and Ian Horrocks, Reference description of the DAML+OILlanguage.http://www.daml.org/2001/03/reference.html. Accessed 16 December 2011.