KNOWLEDGE REPRESENTATION THROUGH SEMANTIC WEB – AN ARCHITECTURAL OVERVIEW

Mini U   Baby M D   Poulose Jacob K

Abstract

The World Wide Web is growing at an exponential rate. There are more and more technologies being developed to provide different ways of accessing this huge information resource, as well as representing the information stored. Because of the increase in information available and people or agents accessing it, the issue of securing this information has become paramount. Semantic Web is an extension of the current web in which the information is given well defined meaning, better enabling computers and people to work in co-operation.[1]. The web pages are designed to be read by humans and not machines. The Semantic Web is a project aimed to make web pages understandable by computers, so that they can search web sites and perform actions in a standardized way.

Keywords : Semantic Web, Ontology, URI,RFD

1. Introduction

Internet is a world wide collection of heterogeneous computers interconnected. The World Wide Web (or simply the Web) is a dynamic, cross-platform, global, distributed, interactive, graphical hypertext information system that runs over the Internet. The concept of Web was proposed by Tim Berners-Lee for information management at CERN to record, keep, and link the heterogeneous computers. Web is an enormous collection of pages linked one another written in HTML... Web is primarily on documents written in HTML with an emphasis on visual presentation, text interleaved with multimedia. Today's web is for Human consumption. It is Human centres- with no automatic processing of information. It requires searching and browsing which is laborious. It requires manual integration of information, you need to search and integrate information. Semantic Web is a project that intends to create a universal medium for information exchange by giving meaning (semantics), which is understood and interpretable by machines.

2. Why Semantic Web ?

From information based to knowledge based web, for Human and automatic consumption, Semantic Web is taking shape. The Semantic Web is a research project that will make available the content of Web pages and not just the keywords. The Semantic Web is an evolving collection of knowledge, built to allow anyone on the Internet to add what they know and find answers to their questions. Information on the Semantic web, rather than being in natural language text, is maintained in a structured form which is fairly easy for both computers and people to work with. Semantic Web will create an environment where the software agents roaming from page to page can readily carry out sophisticated tasks for users. Such agents will know the meaning of concepts used in the page. Creating a machine readable Internet, where computers can access and analyze online content as easily as human's is the goal of Semantic Web.

In short it can be said that by introducing semantics to the Web

- Machines can "understand" the content,   and process the information;

- Not only read it. A lot of information meaningfully linked together in order to be easily processed by machines on a global level

The Semantic Web is a Web-technology that lives on top of the existing Web by including machine-readable information in files without modifying the existing Web structure.

3.   Components of the Semantic Web

Semantic Web requires basically :

1.   Representation of meaning (Ontologies)   2.   Specialised Agents and web services

The two most important technologies for developing the Semantic Web are eXtensible Markup Language (XML) and the Resource Description Framework (RDF). XML allows content creators to label information in a meaningful way. Programs can make use of these tags in sophisticated ways, but the program has to know what the content creator uses each tag for. XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean. RDF expresses the meaning of XML. The W3C developed this new logical language to facilitate interoperability of applications which generate and process machine-understandable representations of data resources on the Web.

The basis for the augmented functionality of the Semantic Web is

- A global naming scheme (URIs);

- A standard syntax for describing data (RDF);

- A standard means of describing the properties of that data (rdf-schema);

- A standard means of describing relationships between data items (ontologies);

- The means to support trust and security.

The Semantic Web comprises of the standards and tools of XML, XML schema, RDF, RDF Schema and OWL.

Layered Model of Semantic Web :

source : Tim Berners-Lee0http://www.w3.org/2002/Talks/09-lcs-sweb-tbl
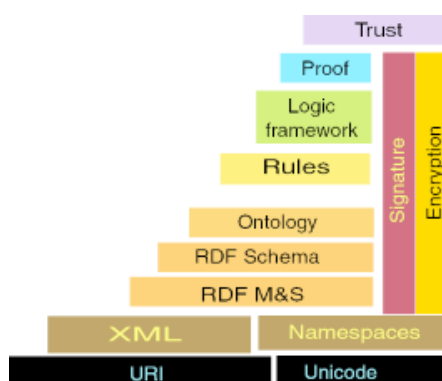
The Layered Stack can be explained :

UniCode	World-wide accepted character standard, which is a system for the interchange, processing, and display of the written texts of the diverse languages of the modern world.
URI	Uniform Resource Identifiers to talk about something, one has to first identify itResource: everything one can talks about, also things that do not exist physically.
XML	Defines the syntax for structured documents. A flexible document exchange format, including namespace, typing, etc.
RDF	Anyone can say anything about everything, A Data modelMake statements which are machine processable and understable
RDF-Schema	Typing system, simple ontology language
Ontology Layer	define meaning and relationships of terms
Logic	make complex logical expressions and the computer to reason on
Proof	Construct proofs by following semantic links
Trust	Digital Signature, Web of Trust

3.1 Ontologies

The current document-centric World Wide Web is evolving into a Semantic Web that communicates information associated with ontologies. Ontologies have been established as effective and efficient means of knowledge sharing. In philosophy, *ontology* is a theory about the nature of existence, of what types of things exists; ontology as a discipline studies such theories. Artificial-intelligence and Web researchers have co-opted the term for their own jargon, and for them the term ontology refers to a document or file that formally defines the relations among terms. A unified representation for Web data and resources is needed in today's large scale Internet data management systems. This unification through standards will allow machines to meaningfully process the available information and to exchange and integrate data coming from distributed databases and information management systems. For the semantic technologies to succeed in the field of information exchange and interoperability between cultural institutions there is a great need to gain interoperability. using standard ontologies. One of the crucial tasks towards the realization of the vision of the Semantic Web is the efficient encoding of human knowledge in ontologies. The proper maintenance of these, usually large and complex, structures and, in particular, their adaptation to new knowledge (*ontology evolution*) is one of the most challenging problems in the current Semantic Web research.

3.2 Uniform Resource Identifier (URI)

A Uniform Resource Identifier (URI) is a compact sequence of characters that identifies an abstract or physical resource. This specification defines the generic URI syntax and a process for resolving URI references that might be in relative form, along with guidelines and security considerations for the use of URIs on the Internet. The URI syntax defines a grammar that is a superset of all valid URIs, allowing an implementation to parse the common components of a URI reference without knowing the scheme-specific requirements of every possible identifier.

A URI is simply a Web identifier: like the strings starting with "http:" or "ftp:" that you often find on the World Wide Web. Anyone can create a URI, and the ownership of them is clearly delegated, so they form an ideal base technology with which to build a global Web on top of. A URI is not a set of directions telling your computer how to get to a specific file on the Web (though it may also do this). It is a name for a "resource" (a thing). This resource may **or may not be** accessible over the Internet. The URI may **or may not** provide a way for your computer to get more information about that resource. URL is a type of URI that does provide a way to get information about a resource, or perhaps to retrieve the resource itself

3.3   Resource Description Framework (RDF)

RDF gives you a way to make statements that are **machine-processable**. An RDF statement is a lot like a simple sentence, except that almost all the words are URIs. Each RDF statement has three parts: a subject, a predicate and an object. Subject is what the statement is about. Predicate is a property of the subject and Object is the value of the property A triple can simply be described as three URIs. A language which utilises three URIs in such a way is called RDF.

4.   Knowledge Representation

The complexity of the knowledge representation lies in the fact that it works well with the models, but the performance is not upto the expectation when scaled up. It makes sense with the limited use on small scale but not on a large scale. RDF model in Semantic web is a great basis for E-R modeling. In RDF models, the relationships are identified by URI so that anyone can say anything about anything. To quote an example, One may identify an automobile with wheels, other dimensions like weight, length etc., which doesn't stop another describing the same with colour vocabulary from elsewhere.

The semantic web data model is very directly connected with the model of relational databases. A relational database consists of tables, which consists of rows, or records. Each record consists of a set of fields. The record is nothing but the content of its fields, just as an RDF node is nothing but the connections: the property values. The mapping is very direct

- a record is an RDF node;

- the field (column) name is RDF property Type; and

- The record field (table cell) is a value.

Semantic Web is not designed to be just a new data model, but for linking data of different models. It allows adding information relating different databases on the web and operations across them.

5.   Semantic Web Service Systems

The Semantic Web Services (SWS) technology provides an environment in which new services can be added, discovered and composed continually. Semantic Web services combine Semantic Web technologies and Web services in order to overcome deficiencies of the current Web services technology stack built around SOAP, WSDL, and UDDI. Based on comprehensive semantic description frameworks, intelligent mechanisms are applied to enable dynamic discovery, composition, contracting, mediation, and execution of Web services. WSMO and OWL-S are two major initiatives for Semantic Web services. Both have been submitted to the W3C as input towards a standardization effort for Semantic Web services.

6.   Semantic Web Tools and Toolkits

There is a need for existence of high quality readily available Semantic Web tools   if the Semantic Web approach is to fulfill its promise. Such toolkits would enable the development of myriad example applications, helping to embody the Semantic Web vision and encouraging its adoption by the industry at large and by individuals all over the world. A large number of open source tools and toolkits have been developed (Parsers, Reasoners, APIs, etc.) in various languages (Java, Lisp, Python, PHP, Perl, C, C#, etc.). A typical powerful, flexible and easy Java based toolkit for developing applications within the Semantic Web is *Jena. RDFlib* is a python toolkit for manipulating RDF models. It is coupled with Redfoot to provide a common Semantic Web toolkit for driving a variety of information management applications.

RDF and OWL as general information models are applicable to many uses, the advantage being the data reuse, freeing the data from the application that created it. Breaking down the barriers of domain knowledge. Where "Anyone can say anything about anything", Trust plays an important role. Hence the Open systems are gaining more prominence. Future is moving towards an infrastructure that gives network effect of the data. Initiative is there for development of Ontology registers for Open vocabularies to help bootstrap application deployment in interoperable manners

7.   Issues for the Semantic Web

Semantic Web vision do not depend on technology alone . It depends on people and organization that makes data freely available in the web in such a way that it can be reused by others. Open Source Data is required. People have a bit of unrealistic expectation of what is Semantic Web achievement. The techniques that make the mapping between programs, data and the real world itself more explicit are an important area of research for computer science. Also there are standardization and research issues of bringing XML and RDF/XML standards closer together.

8.   Semantic Web Applications

There has been an evolution towards a Semantic Web. The emphasis is on methods for extracting existing data from documents, servers and databases. This is translated into common form using Semantic Web languages. The applications that range from social networking (FOAF), content description (Adobe creative suite), learning about licensing constraints of web content(Creative commons) and others shows that web users have began enjoying the benefits of the Semantic Web. A typical example can be quoted as Personal Information Environment Haystack, (http://haystack.lcs.mit.edu). The Haystack Project is investigating approaches designed to let people manage their information in ways that make the most sense to them. By removing arbitrary application-created barriers, which handle only certain information "types" and relationships as defined by the developer, this let users define their most effective arrangements and connections between views of information. Such personalization of information management isl dramatically improving everyone's ability to find what they need when they need it. This includes Piggybank as well as what they call the universal information client. Another example is the Tucana Enterprise Information Integration (http://www.tucantech.com) which exposes diverse data sources as RDF. W3c Semantic Web activity is facilitating   Semantic Web deployment and identify future areas of standardization. Now that with core specifications in place, more applications, toolkits and softwares are added everyday. New class of Semantic Web applications at individual, enterprise and web scale are being developed.

9. Conclusion

Although the Semantic Web as a whole is still very much at a grassroots kind of level, people are starting to take notice; they're starting to publish information using RDF, and thereby making it fit for the Semantic Web. The Semantic Web is dissimilar in many ways from the World Wide Web, including that you can't just point people to a Web site for them to realise how it's working, and what it is. One of the objectives of the advanced development component of the Semantic Web activity is to demonstrate how RDF and Semantic Web technologies can be applied to the W3C Process to increase efficiency, reliability, etc. According to Lee, the Web will continue to evolve and adapt and the Semantic Web is part of this evolution. As the Semantic Web becomes more pervasive,   new challenges will be addressed in terms of usability, accessibility along with the application of these technologies in a variety of new domains: mobile, scientific, cultural and so on. These may involve the creation of new standards.

10. References

1.   Berners-Lee T., Hendler J., Lassila O.: "The Semantic Web". Scientific American. 2001 available at http://www.scientificamerican.com/2001/0501issue/0501berners-lee.htm

2.   "Getting Into RDF & Semantic Web Using N3" available at http://www.w3.org/2000/10/swap/Primer

3.   Berners-Lee T "Semantic Web Roadmap", 1998 available at http://www.w3.org/DesignIssues/ Semantic

4.   "What Is The Semantic Web?" available at http://purl.org/swag/whatIsSW

5.   "Semantic Web Primer" available at   http://uwimp.com/eo.htm

6.   "Semantic Web Introduction – Long" http://logicerror.com/semanticWeb-long

7.   "Building The Semantic Web" http://www.xml.com/pub/a/2001/03/07/buildingsw.html

8.   "The Semantic Web, Taking Form" ,http://infomesh.net/2001/06/swform/ ,2001

9.   Berners-Lee T , "Putting the Web back into Semantic Web", Nov 2005 ISWC2005 Keynote available at   http://www.w3.org/2005/Talks/1110-iswc-tbl/

10.   "SW Activity Statement " available at http://www.w3.org/2001/sw/Activity

11.   "which Semantic Web?" available at http://www.csdl.tamu.edu/~marshall/ht03-sw-4.pdf

12.   Thomas Baker, Fraunhofer-Gesellschaft , "Introduction to Semantic Web", Nov 2004

About Authors

Mini Ulanat is a Professor, Rajagiri School of Computer Science.Cochin
E-mail : mini_u@cusat.ac.in

Dr. Baby M D is a University Librarian, Cochin University of Science & Technology, Kochi, Kerala.
E-mail : babymd@cusat.ac.in/ ul@cusat.ac.in

Dr. K.Poulose Jacob is a Director, School of Computer Science, Cochin University of Science & Technology, Kochi, Kerala.
E-mail : kpj @ cusat.ac.in