Semantic Web: An Organised Way for Web Resources

D. K. Pradhan*
Dr. B. K. Choudhury**
Y. Srinivas Rao***

Abstract

Semantically enabled technology is expected to bring a number of benefits to the users of digital libraries. This technology will help people find relevant information more efficiently and effectively, and give better access to that information and aid sharing of knowledge. This paper therefore put an overview of the basic concept of semantic web and its related technology.

Keywords : Semantic Technology, HTTP, XML , RDF ,RDFS, OWL(Web Ontology Language), URI, URL,URN.

Introduction

The digital library is a place where collection, service and human resources are grouped to format the complete cycle of the creation, diffusion, and preservation of the data with the aim to generate a new structure of the information [7]. Now the web is the most important resources of the digital library where the hyper textual document are present in linear fashion, and the World Wide Web is the biggest repository of information ever created, with growing contents in various languages and fields of knowledge. The web is virtually uncontrolled; any one can put any thing in the web by just accessing Information Service Provider (ISP). Search engines might help to find content containing specific words. But that content might not be exactly what you want. The search is based on the content of pages and not the *semantic meaning* of the page's contents or information about the page. The problem with the search engine is that, they use mostly statistical methods like frequency of occurrence of the words, co-occurrence of words, etc this results in the search queries irrelevant. This problem might be solved by attaching metadata to the concept of web, which makes easier to retrieve [4].

In a traditional library, library catalogue is used as a tool for easy access of information, so a catalogue is the centrally located tool where each documented by some entry. Each entry of the document contains descriptions about the document title, author, place, publisher, year pages, etc. Just like library catalogue, metadata provide descriptive elements on the web. But, unfortunately none of the metadata sets can describe all the documents of the world and often we use the more than one metadata sets to describe a documents.

In this context semantic web can provide the ability to tag all content on the Web, describe what each piece of information is about and give semantic meaning to the content item. Thus, search engines become more effective than they are now, and users can find the precise information they are hunting.

* Lib.Info.Asst (LIA), National Institute of Technology (NIT), Rourkela-769008,dibyakishor@gmail.com.
** Reader, UGC-DRS-SAP Dept.of Lib & Info.Sc, Sambalpur University-768019,bkc_123@rediffmail.com.
*** Asst.Librarian (AL), National Institute of Technology (NIT), Rourkela-769008,ysraoo@gmail.com.

1. Semantic Web: the Possible Way

'Semantics' means the study of meaning, the word "semantic" comes from the Greek word *semantikos,* or "significant meaning," derived from *sema,* or "sign". Semantic Web technologies help separate meanings from data, document content, or application code, using technologies based on open standards [3].

The Semantic Web is an idea of World Wide Web. HTML inventor Tim Berners Lee says that the Web as a whole can be made more intelligent and perhaps even intuitive about how to serve a user's needs. Berners-Lee observes that although search engine index much of the Web's content, they have little ability to select the pages that a user really wants or needs. He foresees a number of ways in which developers and authors, singly or in collaborations, can use self-descriptions and other techniques so that context-understanding programs can selectively find what users want [2].

According to World Wide Web consortium "The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries." Also W3C described "The Semantic Web is a Web that includes documents, or portions of documents, describing explicit relationships between things and containing semantic information intended for automated processing by machines" (http://www.w3.org/1999/04/WebData).

Hence, Semantic Web can be defined as a web of data and it is about two things:

- It is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents.
- It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing.

2. Why Semantic Web

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. It is based on the idea of having data on the Web defined and linked such that it can be used for more effective discovery, automation, integration, and reuse across various applications [6].

The Semantic Web will provide an infrastructure that enables not just web pages, but databases, services, programs, sensors, personal devices, and even household appliances to both consume and produce data on the web. Software agents can use this information to search, filter and prepare information in new and exciting ways to assist the web user [6].

In Internet retrieving and referencing information across remote systems was still an expert's game. The Internet existed, so it was quite possible that you could get access to remote systems. This semantic web system helps in accessing the relevant information [9].

Hence, the Semantic Web encompasses all technologies for sharing, reused, applied data between the groups of people. It is an efficient way to represent data on the World Wide Web, or as a database that is globally linked, in a manner understandable by machines, to the content of documents on the Web.

3. Technologies for Semantic Web

Semantic technologies represent meaning using *ontologies* and provide reasoning through the relationships, rules, logic, and conditions represented in those ontologies. For development of semantic web the most important part is preservation of relation among the concept of documents that can be done by ontologies, so ontology is the study of nature and relation of existence. To represent the Semantic Web, the following technologies are used: [1]

- A global naming scheme.
- A standard syntax for describing data.
- A standard means of describing the properties and relationship of that data.
- A standard means of describing relationships between data items.

(W3C Semantic Web Layer: Source: http://www.w3.org/)

3.1 Global Naming Scheme

3.1.1 ***Uniform Resource Identifier (URI)***: To identify items on the Web, a uniform system is to be needed, in which identified items are to be considered as a "resource" i.e. Uniform Resource Identifiers" (URI).A *URI* is simply a Web identifier, like the strings starting with http or ftp that you often see on the World Wide Web. URI can be given to anything, and anything that has a URI can be said to be "on the Web". "Anyone can create a URI, and the ownership of URIs is clearly delegated, so they form an ideal base technology on top of which to build a global Web [10]. So, the URI is the foundation of the Web. While nearly every other part of the Web can be replaced, the URI cannot, it holds the rest of the Web together. Every data object and every data schema/model in the Semantic Web must have a unique URI.

3.1.2 ***Uniform Resource Locator (URL)***: It is an address of the website that allows people to visit a webpage and notify the computer where to find a specific resource. URL is a URI that identifies and locates the resources by describing the primary access mechanism or network location[10] for example, the URL http://www.nirkl.ac.in, is a URI that identifies a resource of NIT Rourkela's home page. It implies that, a representation of the resource (such as the home page's current HTML code, as encoded characters) is obtainable through HTTP from a network host named http://www.nitrkl.ac.in.

3.1.3 ***Uniform Resource Name (URN)***: URI that identifies a resource by name in a particular namespace, for example, the URN urn: ISBN: 81-0-7666-X is a URI that, like an ISBN book number, allows to talk about a book, but doesn't suggest where and how to obtain an actual copy of it [1].

A URI is not a set of directions, saying the 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. A URL is a type of URI that provides a way to get information about a resource, or perhaps to retrieve the resource itself, and other methods for providing information about URIs and the resources they identify are under development. It is also true that the ability to say things about URIs is an important part of the Semantic Web.

3.2 Standard Syntax Describing Data

3.2.1 Extensible Markup Language (XML):

XML is the most popular technologies used in the development of Semantic Web. XML allows for extensible data formats unlike HTML which is inextensible and rigid. It was designed in a simple way to send documents across the Web. It allows anyone to design their own document format and then write a document in that format. These document formats can include markup to enhance the meaning of the document's content[4]. This markup is "machine-readable," i.e., programs can read and understand it. By including machine-readable meaning in documents, it makes them much more powerful.

3.2.2 Why XML?

XML has the following features [8]:

- XML is a markup language as like HTML.
- XML is designed to described data (means it define its own tag/ code).
- XML tags are not predefined. Anybody can define their own tags.
- XML can be used to create new languages.

For example, to define a book title 'Computer system Architecture' authored by 'M.Morris Mano', and published by 'Prentice Hall of India' can be represented as:

<book>

<title> Computer system Architecture</title>

<author> M. Morris Mano </author>

<publisher> Prentice Hall of India </publisher>

</book>

Hence, in the development of ontology, XML is most suitable because, one can define his/her own tag based on the structure of the ontology. But the basic problem is that syntax and structure in XML are non-standardized so that should be standardized. Therefore, a platform is required with a standard description. Metadata schema provided a standard set of terms to define a resources belonging to a particular domain of knowledge, e.g DC (Dublin Core), OIL (Ontology Interchange Language), DAML(DARPA Markup Language), RDF (resource Description Format), etc are serving in these areas.

3.2.3 Resource Description Framework (RDF)

An official W3C recommendation, RDF is an XML-based standard for describing resources that exist on the Web. RDF builds on existing XML and URI (Uniform Resource Identifier) technologies, using a URI to identify every resource, and using URIs to make statements about resources. RDF statements describe a resource (identified by a URI), the resource's properties, and the values of those properties. RDF provides a consistent, standardized way to describe and query Internet resources, from text pages and graphics to audio files and video clips. It offers syntactic interoperability, and provides the base layer for building a Semantic Web. RDF defines a directed graph of relationships. RDF statements are often referred to as "triples" that consist of a subject, predicate, and object, which correspond to a resource (subject) a property (predicate), and a property value (object) [5].

- ***Resource*** - A resource can be defined as any existing entity having some property or attributes. So, a resource can be a web page or entire web site, books, etc.
- ***Property*** - Any characteristics of resources or it attribute which is used for the description of the resource is known as property.
- ***Value*** - A property must have a value which is equivalent to object.
- ***Statement*** - A specific resource together with a named property plus the value of that property of that resource is a Statement.

For example, a statement Dspace@NITR is the institutional repository of http://www.nitrkl.ac.in. Here the resource is the http://www.nitrkl.ac.in (NIT Rourkela's web site), its property is its institutional repository which gives the value/object DSpace@NITR.

3.3 Standard Means of Describing the Properties and Relationship of That Data

3.3.1 RDF Schema (RDFS)

RDFS is used to create vocabularies that describe groups of related RDF resources and the relationships between those resources. An RDFS vocabulary defines the allowable properties that can be assigned to RDF resources within a given domain. RDFS also allows creating classes of resources that share common properties.

Using the same triples paradigm defined by RDF, RDFS triples consist of classes, class properties, and values that define the classes and relationships between the resources within a particular domain. In an RDFS vocabulary, resources are defined as instances of classes [6]. A class is a resource too, and any class can be a subclass of another. This hierarchical semantic information is what allows machines to determine the meanings of resources based on their properties and classes.

RDF XML Example

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:contact="http://www.w3.org/2000/05/contact#">
```

<contact:institute rdf:about="http://www.nitrkl.ac.in/Institute/contact#webmaster">

<contact: name>webmaster</contact: name>

<contact: mailbox rdf:resource="mailto:webmaster.nitrkl.ac.in"/>

<contact: phone>0661-2432169</contact: phone>

</contact: institute>

</rdf:RDF>

In this example the RDF statement about a resource says it is the institute http:// www.nitrkl.ac.in/. The institute can be identified by the URI http://_www.nitrkl.ac.in /Institute/ contact#; and its maintainer is the web-master, its e-mail address mailto:webmaster@nitrkl.ac.in, and its phone number is 0661-2432169. Overall, RDFS is a simple vocabulary language for expressing the relationships between resources.

3.4 Standard Means of Describing Relationships between Data Items

3.4.1 Web Ontology Language (OWL)

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. OWL is a third W3C specification for creating Semantic Web applications. It is basically a schema that formally defines the hierarchies and relationships between different resources. Semantic Web ontologies consist of taxonomy and a set of inference rules from which machines can make logical conclusions [3].

Taxonomy in this context is system of classification, such as the scientific kingdom/ phylum/class/order/family/genus/species. System for classifying plants and animals that groups resources into classes and sub-classes based on their relationships and shared properties. Since taxonomies (systems of classification) express the hierarchical relationships that exist between resources, we can use OWL to assign properties to classes of resources and allow their subclasses to inherit the same properties. OWL also utilizes the XML Schema data types and supports class axioms such as subclass of, disjoint with, etc., and class descriptions such as union of, intersection of, etc. Many other advanced concepts are included in OWL, making it the richest standard ontology description language available today.

The OWL Web Ontology Language is designed for use by applications that need to process the content of information rather than just presenting information. OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema by providing additional vocabulary along with formal semantics. OWL has three sublanguages: in order of decreasing expressiveness, they are *OWL Full, OWL DL,* and *OWL Lite* [1].

- **OWL Full**: The OWL Web Ontology Language taken as a whole is called **OWL Full.** OWL Full uses all OWL language primitives and allows combinations of those primitives in arbitrary ways with RDF and RDF Schema. OWL Full is fully upward-compatible with

RDF, both syntactically and semantically: any legal RDF document is also a legal OWL Full document.

- **OWL DL** supports those users who want maximum expressiveness without losing computational completeness. OWL DL is a sublanguage of OWL Full language constructs with restrictions such as *type separation* (for example, a class can not also be an individual or property, and a property cannot also be an individual or class).
- **OWL Lite** supports those users primarily needing a classification hierarchy and simple constraint features. The advantage of this language is that it is both easier to understand and easier to implement than the other two; however, it restricts expressivity.

Conclusion

The web is the ocean of information and the semantic web is the extension of the web (World Wide Web). The major philosophical difference is the Semantic Web is supposed to provide machine accessible meaning for its constructs where as in the WWW the meaning is provided by external mechanism. So the establishment of semantic web can provide precise and authentic information, reduces data redundancy, and provides uniform semantic meaning across applications. Hence, in the librarianship point of view Semantic Web can helps in providing right information in right manner in right time.

References

1. Balani,N.The Future of the Web is Semantic;2005http://www-128.ibm.com/developerwork/web/library/wa-semweb/
2. How the Semantic Web Works ?<http://www.w3.org/2002/03/semweb/>
3. Palmer,S.B. The Semantic Web:an Introduction< http://infomesh.net/2001/swintro/>
4. Prasad,A.R.D & Patel,D.An Overview of the Semantic Web;2003< http://www.dspace@drtc.isibang.in>
5. Semantic Web Roadmap http://www.w3.org/Disgin/Issue/Semantc/
6. Semantic Web in Breadth.<http://logicerror.com/semanticweb-low/
7. Thurlow,I, Duko,A & Davies,J. Applying Semantic Web Technology in Digital Libraryhttp://proton.semanticweb/org/
8. Tripathy,A. Resource Description Frame Work:A Tutorial for Developing Web Ontology;2003http://www.dspace@drtc.isibang.in
9. What is a semantic web?http://purl.org/swag/WhatIsSw/