Chapter 4. Semantic Web Technology

4.1 Introduction

Semantic Web is one of the interesting developments in the WWW, in which semantics of web information is defined, so that it will be possible for an information seeker in retrieving the 'meaningful' contents from the web. In other words semantic web defines the meaning (semantics) of information and services on the web so as to make it possible for machines/agents to 'understand' and satisfy the requests of people and to use the web contents effectively. So, semantic web facilitates the data on the web defined in a way so that it can be used by the machines not just for display purpose, but for automation, integration and reuse of data across applications [1].

Semantic web concept was introduced by Tim Bernes-Lee, who also introduced Hyper Text Markup Language (HTML), Hyper Text Transfer Protocol (HTTP), Uniform Resource Identifiers (URI), and World Wide Web. His visualization of the Semantic Web is that intelligent software agents would analyze the situation and accordingly present the best possible alternative. This will make the information more meaningful for machine processing thereby providing efficiency in the information retrieval [2]. Semantic web relies on 'structured sets of information and inference rules that allow it to "understand" the relationship between different data resources' [3]. So, the Semantic Web is a mesh of information linked in a way so that it can be easily processable by machine [4].

According the World Wide Web Consortium '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'[5].

4.2 Elements of Semantic Web

4.2.1 Uniform Resource Identifier (URI)

For successful implementation of the semantic web concept, and its fulfillment, the most important requirement is the availability of unique identifiers to identify resources across the globe irrespective of its location or availability in different formats.

Uniform Resource Identifier (URI) is an identifier used to uniquely identify objects in a space [2]. URL is one of the most popular URI. URI specifies a generic syntax. The syntax of a URI is a scheme name followed by a colon which is again followed by a path as per the specifications of the given scheme.

4.2.2 XML

Extensible Markup Language (XML) is a simple, very flexible text format derived from SGML (ISO 8879). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere [6]. XML was designed to transport and export data [7]. XML is one of the most popular technologies used in the semantic web and it provides freedom for structuring the data in resources using our own preferred vocabularies to define various elements and the attributes.

'XML was the first language to separate the markup of web content from web presentation, facilitating the representation of task-specific and domain-specific data on the web' [8].

4.2.2.1 Extensible Stylesheet Language Transformations (XSLT)

XSLT (Extensible Stylesheet Language Transformations) is a XML based language used for transformation of XML documents into XHTML documents or to other XML documents. Generally XSLT do the transformation by transforming each XML element into an (X)HTML element. XSLT can be used for exchanging XML data between different XML schemas, or to make changes to documents within the scope of the same schema [9].

Transformation of the data using XSLT involves one or multiple XML documents which act as source, one or multiple XSLT stylesheet modules, the XSLT template processor; and one or more output documents. The XSLT stylesheet contains a collection of template rules i.e. instructions and other directives that guide the processor in the production of the output document [10].

4.2.3 Resource Description Framework (RDF)

Using Extensible Markup Language, resources can be described in many different ways. For example one can represent an element as 'Price' and same can be represented by other as 'cost'. Such non standard way of data representation creates problem. In order to provide a common framework which can bridge the gap between these different descriptions, RDF came in to picture. 'RDF is an XML-based standard for describing resources that exist on the Web, intranets, and extranets. 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 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)' [3]

Using triples of subjects, predicates, and objects, RDF allows machines to make logical assertions based on the associations between subjects and objects. Each resource is associated with a specific definition available on the web and RDF uses URIs to identify resources uniquely. Though RDF provides a model and syntax for describing resources, it does not specify the meaning i.e. semantics of the resources. To define semantics, RDF Schema (RDFS) and OWL are needed [3].

4.2.4 RDF Schema (RDFS)

RDFS is generally identified as a vocabulary language for expressing relationships between resources. RDF Schema is used to create vocabularies for describing groups of related RDF resources and the relationships between them. An RDFS vocabulary identifies the allowable properties that can be assigned to RDF resources within a given domain. RDFS allows creating classes of resources that share common properties. Similar to the triples concept used in RDF, triples in RDFS consist of classes, class properties, and values that define the classes and relationships between the resources within a particular domain. In RDFS vocabulary, resources are defined as instances of classes. A class itself is a resource, and one class can be a subclass of another. This type of hierarchical semantic information allows machines to identify the semantics of resources based on their properties and classes [3].

4.2.5 Ontology

Ontology is a 'formal representation of a conceptualization which provides definition of concepts relevant to a domain, system, or application [11] and ontology is considered as one of the pillars of the Semantic Web [12]. According to Gruber, [13] Ontology is a specification of a conceptualization.

In the context of semantic web, ontology is 'a schema that formally defines the hierarchies and relationships between different resources' [3]. Hence Ontology can be said to be the definition of the entities and their relationship with each other [2]. There are different types of ontologies existing such as general ontologies and subject specific ontologies. Ontologies can be developed with XML and RDF.

4.2.5.1 Examples of Ontology and Vocabularies

Examples of some of the Ontologies and vocabularies are given below for illustrative purpose [12].

Basic Geo Vocabulary	*RDF Schema vocabulary for representing latitude, longitude, and altitude information in the WGS84 geodetic reference datum*
BIO	*Vocabulary for biographical information*
DOAP	*Vocabulary which is used for describing open source software projects*
Finance Ontology	*Ontology for securities handling*
FOAF	*RDF based schema to describe persons and their social network*
GoodRelations	*A lightweight ontology for annotating offerings and other aspects of e-commerce on the Web*
MetaVocab	*RDF vocabulary for describing authors and generator tools of RDF documents*
OntoCAPE	*Ontology for chemical process engineering*

Table 4.1 Examples of Ontology and Vocabularies

4.2.5.2 Web Ontology Language (OWL)

During recent years different ontology languages have been developed, specifically Web Ontology Language (OWL).

OWL is a W3C specification for developing semantic web applications. Building upon RDF and RDFS, OWL defines the types of relationships that can be expressed in RDF using an XML vocabulary to indicate the hierarchies and relationships between different resources [2].

'When RDF resource descriptions are associated with an ontology defined somewhere on the Web, intranet, or extranet, it's possible for machines to retrieve the semantic information associated with each resource' [3].

4.3 Examples of Semantic Web Projects and Tools

Large number of projects and tools are being developed for application of the semantic web. Below examples are not exhaustive but it is given for illustration.

4.3.1 DBpedia

DBpedia [14] is a community effort to extract structured information from Wikipedia and to make the extracted information available on the Web. DBpedia facilitates querying over the Wikipedia-derived dataset and facilitating interlinking, re-use and extension in other data-sources.

4.3.2 GoodRelations

The GoodRelations [15] ontology is a popular vocabulary for expressing product information, prices, payment options, etc for electronic commerce. The GoodRelations ontology provides the vocabulary for annotating e-commerce offerings to sell, lease, repair, dispose, or maintain commodity products and to provide commodity services.

4.3.3 NextBio

NextBio [16] is the provider of platform which enables life science researchers to search, discover, and share knowledge locked within public and proprietary data. NextBio is consolidating high-throughput life sciences experimental data tagged and connected via biomedical ontologies

4.3.4 SIOC

The SIOC initiative [17] (Semantically-Interlinked Online Communities) aims to enable the integration of online community information. SIOC provides a semantic web ontology for representing rich data from the Social Web in RDF.

4.3.5 SIMILE

SIMILE [18] i.e. Semantic Interoperability of Metadata and Information in unLike Environments is a joint project of the MIT Libraries and MIT CSAIL. SIMILE enhances interoperability among digital assets, schemata, vocabularies, ontologies, metadata, and services.

4.3.6 OpenPSI

OpenPSI project [19] is a community effort to create UK government linked data service that supports research. It is collaboration between the UK government and University of Southampton. It uses semantic web standards to provide an integration point for UK governmental sourced information, and provide an interactivity space between the governmental information publishers, the research community who need access to governmental information and the new form of information intermediaries, the mashup creators.

4.4 Semantic Interoperability

Interoperability is the ability of two or more systems or components to exchange information and use the exchanged information without special effort on either system [20]. According to Institute of Electrical and Electronics Engineers (IEEE), interoperability is the ability of two or more systems or components to exchange information and to use this information [21].

According to Cechich, *et al.,* [22] interoperability encompasses three different aspects such as Technical, Semantic and Organizational. In technical interoperability issues like linking computer systems, defining open interface, protocols etc are involved. Organizational interoperability deals with modeling institutional process, aligning information architecture with organizational goals etc. Semantic interoperability deals with ensuring that precise meaning of exchanged information is understandable by any application involved.

One of the factors that affect interoperability is heterogeneity. There exist 2 different levels of heterogeneity such as information heterogeneity and system heterogeneity [23]. This study concentrates on the heterogeneity in the information side of electronic governance.

Semantic interoperability is the ability to meaningfully interchange information among different sources and systems. Semantic Interoperability is an important goal that should be addressed by using standard encoding and controlled vocabularies to identify semantic inconsistencies in their domains of interest [24]. Semantic Interoperability is the ability of computer systems to communicate information and have that information properly interpreted by the receiving system in the same sense as intended by the transmitting system [8].

According to Ojo and Estevez [11] Semantic Interoperability is the 'capacity of organizations in public, private, and other sectors, and their information systems to:

- Discover required information;
- Explicitly describe the meanings of the data they wish to share with other organizations;
- Process received information in a manner consistent with intended purpose of such information'.

Information or semantic interoperability is 'concerned with ensuring that the precise meaning of exchanged information is understandable by any person or application receiving the data' [25].

4.4.1 Elements of Semantic Interoperability

There are three technical elements that define semantic interoperability capability such as:[11]

Describe - Semantic description of the information source based certain level of conceptualization.

Mediate - For resolving the conceptual difference while searching for the information.

Discover - Semantic discovery of assets based on semantic description and, in required semantic mediation.

Semantic mediation is the least developed aspect of semantic interoperability. The objective of the semantic mediation is to resolve the semantic variation which might come across while exchanging information between participants in an interaction like service delivery [11].

4.5 Semantic Web and e-Governance

As explained in other chapters, e-governance provides online service to the users utilizing Internet technologies. With the tremendous growth of information and communication technologies numerous services and related information are made available over internet. Many of the e-governance information, is available on the web. Data that is generally hidden away in HTML files is often useful in some contexts, but not in others [26]. Such information embed in HTML is difficult to be used in a large scale, so that it can be processed by any agent/person. For example, data about agricultural offices, services offered by the government departments, population data, economic details of constituencies, tax details etc available in different websites created using HTML. Such data is generally difficult to process or use the way one wanted to use for various purposes. But, using semantic web, intelligent web applications and machines could collect information from various sources, combine them, and present it to the citizens in a meaningful way.

The present e-governance is suffering from lack of interoperability, resource sharing, operation integration and concept of collaborative work. Due to such problems citizen-services become delayed [26].

Let us discuss a scenario, to illustrate the application of semantic web in e-governance. 'FRIENDS' is a project of the Government of Kerala for providing common utility services over common service centres called 'FRIENDS'. A search on 'FRIENDS' on the web produce large volume of information including ‘Friends’ in various context like school friends, family friends, office friends, and several other unrelated context. After sifting through multiple listings and reading through the linked pages only, one is able to find information about the ‘FRIENDS’ common service centres. Due to the difference in the semantic associations of the search term ‘FRIENDS’ the results received will vary in relevance, and one still have to do a lot of work to find the information he/she looking for. But in case of a Semantic Web-enabled environment, one can use a Semantic Web agent to search the Web for ‘FRIENDS’ where ‘FRIENDS’ is a Common service centres used in Kerala. In such a search, based on the semantic web context, will produce more relevant results.

Based on the semantic information available for FRIENDS, semantic web agent also presents with a list of common service centres available in different places, or in other states, and along with other relevant information.

Similarly, semantic web also facilitates the finding, interpretation and acting on data for various e-governance information. For such services/functionalities semantic web relies on structured sets of information and inference rules that allow it to “understand” the relationship between different data resources.

‘Due to its vast challenge to achieve interoperability, given the manifold semantic differences of interpretation of, for example, law, regulations, citizen services, administrative processes, best-practices, and the many different languages to be taken into account within and across regions, nations and continents, the domain of e-governance is unique. These semantic differences are related to a great variety of IT solutions (on a local, regional, national, and international level), which will have to be networked (despite any effort of standardization)’ [27].

4.5.1 Advantages of Semantic Web in e-Governance

Application of the Semantic Web Technologies in e-governance has advantages like:

i) Reduction of cost & risk

ii) Providing better services to the e-government community such as stakeholders, administrations, end users.

iii) Semantic data integrator

iv) Semantic service discovery and choreography i.e. re-use of existing services and the dynamic automation of process

v) Precise and concept aware search using knowledge representations across multiple knowledge sources

vi) Virtual consultant such as understanding customer's goals and offering products and services which can help them meet those goals

viii) Aggregating services on the basis of user preferences better support GUI by allowing one -stop, customer focused, and multiple viewpoint access to services and shared information' [26].

4.5.2 Government Semantic Interoperability Frameworks

Today, several countries have understood the importance of the interoperability in e-Governance and accordingly they have developed interoperability frameworks mainly dealing with the technical and organizational interoperability. Attempts of North America and Europe in developing semantic interoperability solutions are also worth mentioning. European commission has been working on a European Interoperability Framework, United Kingdom has e-government interoperability framework called e-GIF and New Zealand has interoperability framework called e-GIF, based on standards mainly using XML [22]. Other major Government Semantic Interoperability Frameworks include: [11]

- US Data Reference Model (DRM), which is one of the five reference models of the Federal Enterprise Architecture (FEA);

- EU Content Interoperability Strategy, which is a part of the European Interoperability Framework;
- UK e-Government Metadata Standard, a part of the e-Government Interoperability Framework;
- Dutch Interoperability Framework (proposed); and
- Australian Information Interoperability Framework, part of the Australian Government Interoperability Framework

4.6 Conclusion

Semantic Interoperability is arguably the least developed aspects of Government Interoperability Frameworks [11]. Researchers are working across the world to develop Semantic Web based e-Governance services [28]. Application of Semantic web in e-governance will become a necessity, as the use of semantic web based applications is increasing. Considering the heterogeneous nature of the current e-Governance information and necessity for seamless interoperation, semantic web will be a promising solution, with the help of standard metadata.

4.7 References

1. Terziyan, Vagan., (2006), "Semantic Web Course Introduction". www.cs.jyu.fi/ai/vagan/SW_Introduction.ppt (accessed January 4, 2010).
2. Prasad, ARD., and Dimple Patel. (2003) "An Overview of the Semantic Web", Paper presented at *the DRTC Workshop on Semantic Web*, Bangalore, December 8-10, 2003: 1-10.
3. Altova, (2005), "What is the Semantic Web?" http://www.altova.com/semantic_web.html (accessed January 4, 2010).
4. Palmer, Sean B., "What is the Semantic Web?". http://infomesh.net/2001 /swintro/ (accessed January 6, 2010).
5. Berners-Lee, Tim., *et al.*, (2001), "The Semantic Web". *Scientific American*, 2001(May) http://www.scientificamerican.com/article.cfm?id =the-semantic-web (accessed January 4, 2010).

6. W3C. (2010), "Extensible Markup Language (XML)". http://www.w3.org/XML/ (accessed January 4, 2010).
7. Refsnes Data, "Introduction to XML". http://www.w3schools.com/xml/xml_whatis.asp (accessed January 5, 2010).
8. Wikimedia Foundation Inc. (2010), "Semantic Web" http://en.wikipedia .org/wiki/Semantic_Web (accessed January 5, 2010).
9. Refsnes Data. (2010), "XSLT Introduction". http://www.w3schools.com /xsl/xsl_intro.asp (accessed January 19, 2010).
10. Wikimedia Foundation, Inc, (2010) "XSLT". http://en.wikipedia.org/wiki/XSLT (accessed January 19, 2010).
11. Ojo,A., T Janowski., and Estevez E. (2009), "Semantic Interoperability Architecture for Electronic Government", In the *Proceedings of the 10th International Digital Government Research Conference*, Mexico, 17-21 May, 2009:63-72
12. Semanticweb.org.(2010),"Ontology", http://semanticweb.org/wiki /Ontology (accessed January 6, 2010).
13. Gruber, T.R. (1993), "A Translation Approach to Portable Ontologies". *Knowledge Acquisition*, 5(2) 1993:199-220
14. Bizer,Christian, *et al.* (2009), "DBpedia - A crystallization point for the Web of Data". *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154-165. Accessed February 22, 2010 doi: 10.1016/j .websem.2009.07.002.
15. Universität der Bundeswehr München. "GoodRelations: The Web Ontology for e-Commerce". http://www.heppnetz.de/projects /goodrelations/ (accessed February 12, 2010).
16. NextBio Website (2010), http://www.nextbio.com/ (accessed February 12, 2010).
17. National University of Ireland. "Semantically-Interlinked Online Communities". http://sioc-project.org/ (accessed February 12, 2010).
18. Massachusetts Institute of Technology. (2007), "The SIMILE Project". http:// simile .mit.edu/wiki/SIMILE:About (accessed February 13, 2010).

19. OpenPSI. (2009), "OpenPSI". http://www.openpsi.org (accessed February 13, 2010).
20. American Library Association, Committee on Cataloging: Description & Access, Task Force on Metadata. (2000), "Final Report". http://www.libraries.psu.edu /tas/jca/ccda/tf-meta6.html (accessed November 10, 2009).
21. IEEE. (1991), "IEEE-STD-610, ANSI/IEEE Std 610.12-1990. IEEE Standard Glossary of Software Engineering Terminology". IEEE, 1991
22. Cechich, Alejandra., (2008), " Interoperability Frameworks for Electronic Governance". Paper presented at the *ICEGOV2008*, Egypt, 2008: 490-491.
23. A. Sheth. (1998), "Changing Focus on Interoperability in Information Systems: From System, Syntax Structure to Semantics". In *Interoperating Geographic Information Systems*. Kluwer, 1998: 5–30.
24. Rothenberg, Jeff *et al.*, (2005), *Designing a National Standard for Discovery Metadata: Improving Access to Digital Information in the Dutch Government.* Pittsburgh: RAND Corporation, 2005. http://www.rand.org/pubs/technical _reports/2005/RAND_TR185.pdf (accessed February 23, 2010).
25. Lallana, Emmanuel C. (2009). *e-Government Interoperability,*Thailand: UNDP, 2008. : http://www.apdip.net/publications/iespprimers/eprimer-gif.pdf (accessed January 12, 2009)
26. Marjit, Ujjal., et al., (2009) "A semantic Web Service based Approach to E-Governance", In the Second International Conference on the Applications of Digital Information and Web Technologies, 2009: 232 – 237.
27. Pulido, J.R.G., *et al.*, (2006) "Ontology Languages for the Semantic Web: A Never Completely Updated Review". *Knowledge-Based Systems* 19 (7) 2006: 489-497.
28. Dutta, Biswanath. (2007), "Semantic Web Service Infrastructure for e-Governance". In *the Proceedings of the International Conference on Semantic Web and Digital Libraries, ICSD 2007*, Bangalore, February 21-23, 2007: 144-153.