5. Semantic Web

Background:

Suppose you wish to find the Mr. Jain you met at Unicode Conference last year. You don t remember his first name, but you remember that he is working for Government of India and that his wife was a student of famous dance school in your town. An intelligent search program can search through all the pages of people whose name is Jain (sidestepping all the pages relating to Jains, Jainism, the Jain holy places and so forth), find the ones that mention working for Government of India and follow links to Web pages of their wives to find if any was in the same dance school in your town.

A web with such capability is known as Semantic Web. The Semantic Web is an extension of the current Web, in which information is given well-defined meaning in a common format to be shared by multilingual people & multiplatform programmes.

eXtensible Markup Language (XML), the Resource Description Framework (RDF) and Ontologies are the three important components of the Semantic Web. With the use of XML developers may create their own tags. Scripts, or programs, can make use of these tags in useful ways, but the script writer has to know what the each tags mean for. In other words, XML allows developers to add arbitrary structure to their documents but do not explain the meaning of these structures. The meaning of these structures is expressed by RDF. The RDF encodes the structure meaning in sets of triples, each triple being rather like the subject, verb and object of an elementary sentence. These triples can be written using XML tags. In RDF, a document makes assertions that particular things (people, Web pages or whatever) have properties (such as is a husband of, is the creator of) with certain values (another person, another Web page). This structure turns out to be a natural way to describe the vast majority of the data processed by machines. Subject and object are each identified by a Universal Resource Identifier (URI), just as used in a link on a Web page. (URLs, Uniform Resource Locators, are the most common type of URI.) The verbs are also identified by URIs, which enables anyone to define a new concept, a new verb, just by defining a URI for it somewhere on the Web.

The triples of RDF form webs of information about related things. Because RDF uses URIs to encode this information in a document, the URIs ensure that concepts are not just words in a document but are tied to a unique definition that everyone can find on the Web. For example, imagine that we have access to a variety of databases with information about people, including their addresses. If we want to find people living in a specific PIN code, we need to know which fields in each database represent names and which represent PIN codes. RDF can specify that (field 5 in database A) (is a field of type) (PIN code), using URIs rather than phrases for each term.

A program that wants to compare or combine information across the two or more databases must know that particular terms being used in the database are to mean the same thing. Hence, the program must have some way to identify such common meanings or relations between these terms and this information should be available to the programme. The collection of such information is called Ontology. In other words, Ontology is a document or file that formally defines the relations among terms.

Taxonomy and a set of inference rules are the two important components for the Web ontology. The taxonomy defines classes of objects and relations among them. For example, an address may be defined as a type of location, and city codes may be defined to apply only to locations, and so on. Classes, subclasses and relations among entities are a very powerful tool for Web use. A large number of relations may be expressed, among entities by assigning properties to classes and allowing subclasses to inherit such properties. If city codes must be of type city and cities generally have Web sites, we can discuss the Web site associated with a city code even if no database links a city code directly to a Web site.

Inference rules in ontology supply further power. An ontology may express the rule If a city code is associated with a state code, and an address uses that city code, then that address has the associated state code. A program could then readily deduce, for instance, that the Kurukshetra University address, being in Kurukshetra, must be in Haryana State, which is in the India, and therefore should be formatted to Indian standards. The computer doesn t

truly understand any of this information, but it can now manipulate the terms effectively in ways that are useful and meaningful to the end user.

Ontologies can enhance the functioning of the Web in many ways. They can be used in a simple fashion to improve the accuracy of Web searches the search program can look for only those pages that refer to a precise concept instead of all the ones using ambiguous keywords. More advanced applications will use ontologies to relate the information on a page to the associated knowledge structures and inference rules.

In addition, this markup makes it much easier to develop programs that can tackle complicated questions whose answers do not reside on a single Web page.

The Semantic Web, in contrast, is more flexible. The consumer and producer agents can reach a shared understanding by exchanging ontologies, which provide the vocabulary needed for discussion. Agents can even bootstrap new reasoning capabilities when they discover new ontologies. Semantics also makes it easier to take advantage of a service that only partially matches a request.

The Semantic Web will break out of the virtual realm and extend into our physical world. URIs can point to anything, including physical entities, which means we can use the RDF language to describe devices such as cell phones and TVs. The Semantic Web can assist the evolution of human knowledge as a whole.

The Semantic Web Activity:

The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.

Goals:

The goal of the Semantic Web activity is to create a universal medium for the exchange of data. Facilities to put machine-understandable data on the Web are also one of the objectives of the Semantic Web activity. Goals may be summarized as:

- Web of data provides common data representation framework to facilitate integrating multiple sources to draw new conclusions
- Increase the utility of information by connecting it to its definitions and to its context
- More efficient information access and analysis

As discussed above there are three major components of the Semantic Web: these are the Resource Description Framework (RDF) Core Model, the RDF Schema language and the Web Ontology language (OWL). For RDF enabling the joining of decentralized collections of RDF data a standardized query language, SPARQL has been developed using these components. These languages all build on the foundation of URIs, XML, and XML namespaces.

The following groups are contributing/ contributed in the Semantic Web activity:

Semantic Web Coordination Group

The Semantic Web Coordination Group is tasked to provide a forum for managing the interrelationships and interdependencies among groups focusing on standards and technologies that relate to this goals of the Semantic Web Activity. This group is designed to coordinate, facilitate and (where possible) help shape the efforts of other related groups to avoid duplication of effort and fragmentation of the Semantic Web by way of incompatible standards and technologies.

Semantic Web Interest Group

The Semantic Web Interest Group is a forum for W3C Members and non-Members to discuss innovative applications of the Semantic Web. The Interest Group also initiates discussion on potential future work items related to enabling technologies that support the Semantic Web, and the relationship of that work to other activities of W3C and to the broader social and legal context in which the Web is situated.

This group has released RDF Calendar - an application of the Resource Description Framework to iCalendar Data. This Interest Group Note is a report on the state of the art for integrating calendar data with other Semantic Web data used in social networking, syndicated content, and multimedia.

Semantic Web Best Practices and Deployment Working Group

The focus of this Semantic Web Best Practices and Deployment (SWBPD) Working Group is to provide hands-on support for developers of Semantic Web applications.

This group has published three Working Drafts for the Simple Knowledge Organization System (SKOS): the Core Guide, Core Vocabulary Specification and the Quick Guide to Publishing a Thesaurus on the Semantic Web. The drafts explain how to express classification schemes, thesauruses, subject heading lists, taxonomies, terminologies, glossaries and other types of controlled vocabulary in RDF.

RDF Data Access Working Group

The focus of the RDF Data Access Working Group will be to evaluate the requirements for an query language and network protocol for RDF and defined formal specifications and test cases for supporting such requirements.

This group has released Last Call Working Drafts of the SPARQL (pronounced sparkle) Protocol, SPARQL Query Language for RDF, and the SPARQL Query Results XML Format. Using SPARQL, developers and end users will be able to integrate data over disparate sources, and write and consume search results.

The Rule Interchange Format (RIF) Working Group

The Rule Interchange Format (RIF) Working Group was launched to produce a language for the exchange of rules and their transfer between rule systems.

RDF Core Working Group

The RDF Core Working Group is chartered to consider update to the RDF Model and Syntax Recommendation, and to a few revisions to the RDF Schema specification. A further objective of this group is to respond to the Candidate Recommendation feedback and produced a revised RDF Schema document.

Web Ontology Working Group

The Web Ontology Working Group is chartered to build upon the RDF Core work a language for defining structured web based ontologies which will provide richer integration and interoperability of data among descriptive communities.

References:

www.w3.org

The Scientific America.