3 Introduction To Semantic Web

Objective:

This chapter introduces different technologies that are perceived to be the building blocks of the Semantic Web. The chapter focuses on defining the meaning of each and every term and terminology that plays a key role in constructing the Semantic Web.

3.1 Defining Semantic Web

Semantic Web has many well-known definitions. Listed below are a few of them:

Tim Berner Lee's definition for Semantic Web:

"People keep asking what Web 3.0 is. I think maybe when you've got an overlay of scalable vector graphics – everything rippling and folding and looking misty-on Web 2.0 and access to a Semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource."

Google's CEO, Eric Schmidt stated:

"Web3.0 is a series of combined applications. The core software technology of Web 3.0 is artificial intelligence, which can intelligently learn and understand the semantics. Therefore, the application of Web3.0 technology enables the Internet to be more personalized, accurate and intelligent."

Netflix founder, Reed Hastings thinks that Web 3.0 would be a full video Web as stated below:

"Web 1.0 was dial-up, 50K average bandwidth; Web 2.0 is an average 1 megabit of bandwidth and Web 3.0 will be 10 megabits of bandwidth all the time, which will be the full video Web, and that will feel like Web 3.0."

Yahoo founder Jerry Yang stated at the TechNet Summit in November 2006:

"Web 2.0 is well documented and talked about. The power of the Net reached a critical mass, with capabilities that can be done on a network level. We are also seeing richer devices over last four years and richer ways of interacting with the network, not only in hardware like game consoles and mobile devices, but also in the software layer. You don't have to be a computer scientist to create a program. We are seeing that manifest in Web 2.0 and Web 3.0 will be a great extension of that, a true communal medium…the distinction between professional, semi-professional and consumers will get blurred, creating a network effect of business and applications."

Explanation:

The Semantic Web is an evolution and extension of the existing web that allows computers to manipulate data and information. The idea of the Semantic Web is still undergoing research and development. There is no reliable way to process semantics which questions the purpose of developing such a large-scale project. Yet there is a strong appeal behind the concept of having such an extension to pre-existing documentation of information that is presently distributed on the World Wide Web. There is a great appeal in a web that has the potential ability to 'know' and 'understand' data. This adds a more humanistic quality to standard data processing because the Semantic Web seeks to close the gap between merely providing documents to people and automatic data and information processing. However, in order to reach to this point, developers are challenged by providing a language that can express both data and rules for reasoning.

3.2 Characteristics of Semantic Web

Intelligence:

Experts believe that one of the most promising features of Web 3.0 will be the web with intelligence, i.e., an intelligent web. Applications will work intelligently with the use of Human-Computer interaction and intelligence. Different Artificial Intelligence (AI) based tools & techniques (such as, rough sets, fuzzy sets, neural networks, machine learning, etc.) will be incorporated with the applications to work intelligently. This means, an application based on Web 3.0 can directly do intelligent analysis, and then optimal output would be possible without much user intervention. Documents in different languages can be intelligently translated into other languages in Web 3.0. It should enable us to work through natural language. Therefore, users can use their native language for communication with the others around the world.

Personalization

Another characteristic of Web 3.0 is Personalization. Personal or individual preferences would be considered during different activities such as information processing, searching, forming a personalized portal on the web, etc. Semantic Web would be the core technology for Personalization in Web 3.0.

Reasoning

Semantic Web allows search, integration, answering complex queries, connections and analysis (paths, sub graphs), pattern finding, mining, hypothesis validation, discovery, visualization etc.

Interoperability

Interoperability refers to the aspects such as the seamless integration of data from heterogeneous sources, dynamic composition, interoperation of web services, and the next-generation search engines. Web 3.0 applications would be easy to customize and they can independently work on different kinds of devices. An application based on Web 3.0 would be able to run on many types of Computers, microwave devices, handheld devices, mobiles, TVs, automobiles and many others.

Usability

Usability encompasses new information retrieval paradigms, user interfaces, interaction and visualization techniques, which in turn require methods for dealing with context dependency, personalization, trust and provenance, amongst others, while hiding the underlying computational issues from the user.

Applicability

Applicability refers to the rapidly growing application areas of Semantic Web technologies and methods, the issue of bringing state-of-the-art research results to bear real-world applications, and to the development of new methods and foundations driven by real-application needs, from various domains.

Note 1:

- Semantics with metadata and ontologies for heterogeneous documents and multiple repositories of data, including the web was discussed in 1990s
- Tim Berners-Lee used the term "Semantic Web" in his 1999 book.
- Initial 5 years of Semantic Web research, saw too much of AI/DL, but more practical/applied work has dominated in recent years.

Note 2:

Pervasive web –

Pervasive Web is the term used to describe the phenomenon where the web is operable to a wide range of electronic devices.

3.3 Semantic Web Vs Artificial Intelligence (AI)

In reality, Semantic Web technologies are as much about the data as they are about reasoning and logic. RDF, the foundational technology in the Semantic Web stack, is a flexible graph data model that does not involve logic or reasoning in any way. The realization of the Semantic Web vision does not rely on human-level intelligence. In fact, the challenges are approached in a different way. The full problem of AI is a deep scientific one, perhaps comparable to the central problems of physics (explain the physical world) or biology (explain the living world). In AI, partial solutions may not work.

But on the Semantic Web partial solutions will work. Even if an intelligent agent is not able to come to all the conclusions that a human user might draw, the agent will still contribute to a web much better than the current Web. If the ultimate goal of AI is to build an intelligent agent, exhibiting human-level intelligence (and higher), the goal of the Semantic Web is to assist human users in their day-to-day online activities. It is clear that the Semantic Web will make extensive use of current AI technology and that advances in that technology will lead to a better Semantic Web. But there is no need to wait until AI reaches a higher level of achievement; current AI technology is sufficient to go a long way toward realizing the Semantic Web vision.

3.4 SDLC – An Overview

SDLC stands for software development life cycle. SDLC consists of following activities:

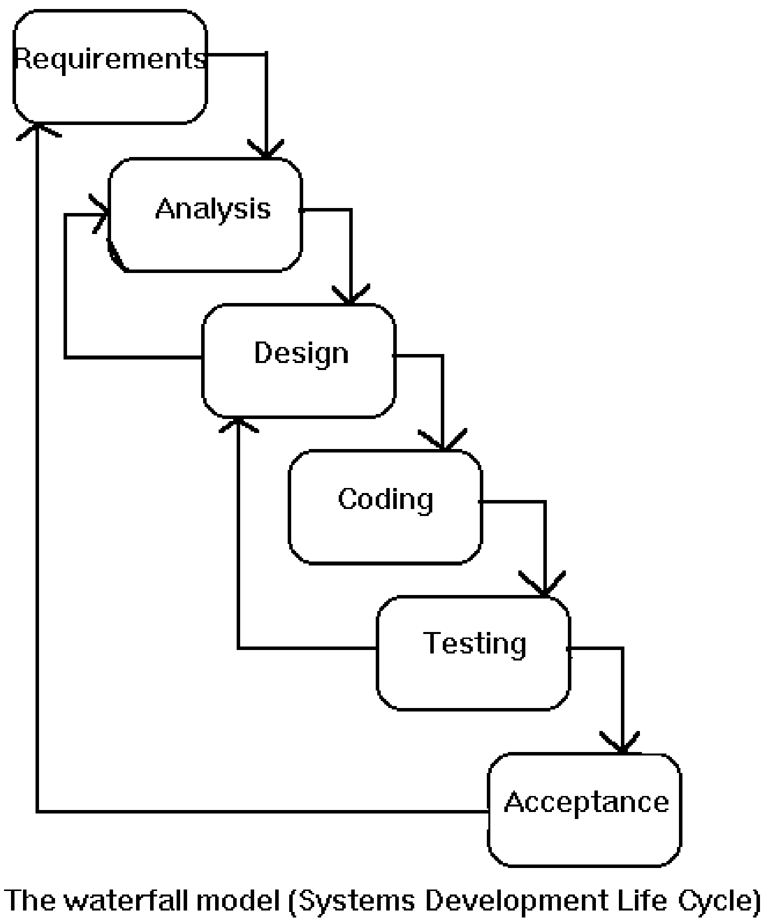

Figure 3.1: SDLC Waterfall model

- **Planning:**
  The most important part of software development, i.e. requirement gathering or requirement analysis is done by the most skilled and experienced software engineers in the organization. After the requirements are gathered from the client, a scope document is created in which the scope of the project is determined and documented.

- **Implementation:**
  Implementation refers to the coding done by the software engineers to implement the client's requirements.

- **Testing:**
  This is a process of finding defects or bugs in the created software. There are different types of testing, performed at different stages of software development. For example, unit testing and regression testing are usually done by developers, whereas smoke-testing and white-box testing are performed by testers.

- **Documentation:**
  Every step in the project is documented for future reference and for the improvement of the software in the development process. The design documentation may include writing the application programming interface (API), the business requirements, intended audience, etc.

- **Deployment and maintenance:**
  At this stage, the software is deployed after it has been approved for release.

- **Maintaining:**
  This is the final stage of the SDLC. Software improvement and new requirements (change requests) can take longer time than the time needed to create the initial development of the software.

3.4.1 There are several software development models followed by various organizations:

Waterfall Model:

This model involves finishing the first phase completely before commencing the next one. When each phase is completed successfully, it is reviewed to see if the project is on track and whether it is feasible to continue.

V-shaped Model:

This model focuses on execution of processes in a sequential manner, similar to the waterfall model, but with more importance placed on testing. Testing procedures are written even before the commencement of coding. A systematic plan is generated before starting the development phase.

Incremental Model:

This life cycle model involves multiple development cycles. The cycles are divided into smaller iterations. These iterations go through a set of phases including requirements, design, implementation and testing. A working version of the software is produced during the first iteration, so working software is created early in the development process.

3.5 Building-blocks of Semantic Web

3.5.1 Ontology

Now, after a deep introduction to Semantic Web, let us try to understand its major building blocks. Let's start with the most important of its kind i.e. Ontology. With ontology, computers can sometimes act as if they 'understand' the information they are carrying. This is where the term "semantic" comes in. In this web, we try to make the meanings so clear that even a computer can understand them. To have truly intelligent systems, knowledge needs to be captured, processed, reused, and communicated. Ontologies support all these tasks. The term 'ontology' can be defined as an explicit specification of conceptualization. The exact meaning depends on the understanding of the terms 'specification' and 'conceptualization'. Explicit specification of conceptualization means that ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This definition is consistent with the usage of ontology as a set of concept definitions.

The backbone of ontology is often taxonomy. Taxonomy is a classification of things in a hierarchical form. It is usually a tree or a lattice that expresses sub-assumption relation – i.e., A subsumes B, means that everything that is in A is also in B. An example is classification of living organisms. The taxonomy usually restricts the intended usage of classes, where classes are subsets of the set of all possible individuals in the domain. Ontologies are considered as one of the pillars of the Semantic Web, although they do not have a universally accepted definition. A (Semantic Web) vocabulary can be considered as a special form of (usually light-weight) ontology.

In order to share the knowledge among the agents, an agreement must exist on the topics which are being communicated. This raises the issue of ontological commitment. Ontological commitments allow a number of agents to meaningfully communicate about a domain without necessarily operating on a globally shared theory individually. In the context of multiple agents, a common ontology serves as a knowledge-level specification of the ontological commitments of a set of participating agents. A common ontology defines the vocabulary with which queries and assertions are exchanged among the agents, thereby providing the means to bridge the semantic gap that exists between the lexical representations of information and its non-lexical conceptualization.

3.5.2 RDF/OWL

RDF is a specification that defines a model for representing the world and syntax for serializing and exchanging that model. The W3C has developed an XML serialization for RDF. RDF XML is the standard interchange format for RDF on the Semantic Web, although it is not the only format. For example, Notation3 is an excellent plain text alternative serialization of RDF XML. 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. These are represented by object-attribute-value triples, that is, an object O has an attribute A with the value V.

3.5.3 SPARQL

SPARQL is a RDF query language which is capable of retrieving and manipulating the data stored in Resource Description Framework format. It was standardized by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium, and is recognized as one of the key technologies of the Semantic Web. SPARQL allows a query to consist of triple patterns, conjunctions, disjunctions and optional patterns.