****

**Wollo University**

**Kombolcha Institute of Technology**

**College of Informatics**

**Department of:-computer science**

**Selecting Topic Assignment Given By:-**

**Instructor Hassan**

**Done by:-**

**Name id no**

***1. Tesfaye Admasu……….…..WOUR/1374/07***

***2. Sirag Mohammed ………….WOUR/2024/08***

***3. Tadese Adugna…………….WOUR/2028/08***

***4. Yordanose Kassahun….…….WOUR/2065/08***

***5. Zekiy Seid………………….WOUR/2068/08***

***Submitted date 19/10/2011 E.C***

**Title: SEMANTIC WEB**

1. ***INTRODUCTION***

Today we are living in the information age and society can be labeled as information society. Knowledge, a predominant element, pervades and dominates the day to-day activities in information society. There is a sea change in information generation, distribution, and access. In recent years, finding information from the Web is becoming more and more complex. Searching on the Internet can be compared to dragging a net across the surface of the ocean.

The traditional search engines are unable to provide satisfactory solutions to this. Research is in progress to build a Web, which is semantically richer than the current one. The ability to translate knowledge from different languages is an important ingredient for building powerful artificial intelligent (AI) systems, by easing the difficult and time-consuming task of knowledge base construction.

The vision of semantic Web proposes an environment where the data and services on the Web can be semantically interpreted and processed by machines to facilitate human consumption. The semantic Web relies heavily on the formal ontologies that structure underlying data for the purpose of comprehensive and transportable

-The idea of the semantic Web was conceived by Tim Berners-Lee, the founder of the WWW. He envisions that in future, the vast amount of information on the Web will bear machine readable metadata, resulting in computers being able to manipulate the contents automatically, without human intervention.

-Therefore, the semantic Web is imagined as an extension of the Web, in which information is given a well defined meaning. It is the application of advanced knowledge technologies to the Web and distributed systems in general. It describes methods and technologies to allow machines to understand the meaning or “semantics” of information on the WWW. To accomplish this, the provided information should be structured, accompanied by sets of inference rules that can be used by computers to conduct automated reasoning.

-According to Tim Berners Lee, “The Web has developed most rapidly as a medium of documents for the people rather than of information that can be manipulated automatically. By augmenting web pages with data targeted at computers and by adding documents solely for computers, we will transform into semantic Web.

***2. Expressing Meaning***

Pete and Lucy could use their agents to carry out all these tasks thanks not to the World Wide Web of today but rather the Semantic Web that it will evolve into tomorrow. Most of the Web’s content today is designed for humans to read, not for computer programs to manipulate meaningfully. Computers can adeptly parse Web pages for layout and routine processing— here a header, there a link to another page—but in general, computers have no reliable way to process the semantics: this is the home page of the Hartman and Strauss Physio Clinic, this link goes to Dr. Hartman’s curriculum vitae.

The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users. Such an agent coming to the clinic’s Web page will know not just that the page has keywords such as "treatment, medicine, physical, therapy" (as might be encoded today) but also that Dr. Hartman works at this clinic on Mondays, Wednesdays and Fridays and that the script takes a date range in yyyy-mm-dd format and returns appointment times. And it will "know" all this without needing artificial intelligence on the scale of 2001’s Hal or Star Wars’s C-3PO. Instead these semantics were encoded into the Web page when the clinic’s office manager (who never took Comp Sci 101) massaged it into shape using off-the-shelf software for writing Semantic Web pages along with resources listed on the Physical Therapy Association’s site.

-The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The first steps in weaving the Semantic Web into the structure of the existing Web are already under way. In the near future, these developments will usher in significant new functionality as machines become much better able to process and "understand" the data that they merely display at present.

-The essential property of the World Wide Web is its universality. The power of a hypertext link is that "anything can link to anything." Web technology, therefore, must not discriminate between the scribbled draft and the polished performance, between commercial and academic information, or among cultures, languages, media and so on. Information varies along many axes. One of these is the difference between information produced primarily for human consumption and that produced mainly for machines. At one end of the scale we have everything from the five-second TV commercial to poetry. At the other end we have databases, programs and sensor output.

-The Semantic Web will enable machines to COMPREHEND semantic documents and data, not human speech and writings.

Documents for people rather than for data and information that can be processed automatically. The Semantic Web aims to make up for this.

-Like the Internet, the Semantic Web will be as decentralized as possible. Such Web-like systems generate a lot of excitement at every level, from major corporation to individual user, and provide benefits that are hard or impossible to predict in advance. Decentralization requires compromises: the Web had to throw away the ideal of total consistency of all of its interconnections, ushering in the infamous message "Error 404: Not Found" but allowing unchecked exponential growth.

1. ***Knowledge Representation***

Knowledge Representation is a branch of Artiﬁcial Intelligence that provides access to a structured collection of information and a set of inference rules. The information and rules can then be used for automated reasoning, e.g. with the help of software agents. In this section we will have an introduction to the markup languages used to represent knowledge.

For the semantic web to functions, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning. Artificial-intelligence researchers have studied such systems since long before the Web was developed. Knowledge representation, as this technology is often called, is currently in a state comparable to that of hypertext before the advent of the Web: it is clearly a good idea, and some very nice demonstrations exist, but it has not yet changed the world. It contains the seeds of important applications, but to realize its full potential it must be linked into a single global system.

Traditional knowledge-representation systems typically have been centralized, requiring everyone to share exactly the same definition of common concepts such as "parent" or "vehicle." But central control is stifling, and increasing the size and scope of such a system rapidly becomes unmanageable.

Moreover, these systems usually carefully limit the questions that can be asked so that the computer can answer reliably— or answer at all. The problem is reminiscent of Gödel’s theorem from mathematics: any system that is complex enough to be useful also encompasses unanswerable questions, much like sophisticated versions of the basic paradox "This sentence is false." To avoid such problems, traditional knowledge-representation systems generally each had their own narrow and idiosyncratic set of rules for making inferences about their data. For example, a genealogy system, acting on a database of family trees, might include the rule "a wife of an uncle is an aunt." Even if the data could be transferred from one system to another, the rules, existing in a completely different form, usually could not.

Semantic Web researchers, in contrast, accept that paradoxes and unanswerable questions are a price that must be paid to achieve versatility. We make the language for the rules as expressive as needed to allow the Web to reason as widely as desired. This philosophy is similar to that of the conventional Web: early in the Web’s development, detractors pointed out that it could never be a well-organized library; without a central database and tree structure, one would never be sure of finding everything. They were right. But the expressive power of the system made vast amounts of information available, and search engines (which would have seemed quite impractical a decade ago) now produce remarkably complete indices of a lot of the material out there. The challenge of the Semantic Web, therefore, is to provide a language that expresses both data and rules for reasoning about the data and that allows rules from any existing knowledge-representation system to be exported onto the Web.

Adding logic to the Web— the means to use rules to make inferences, choose courses of action and answer questions— is the task before the Semantic Web community at the moment.

A. mixture of mathematical and engineering decisions complicate this task. The logic must be powerful enough to describe complex properties of objects but not so powerful that agents can be tricked by being asked to consider a paradox. Fortunately, a large majority of the information we want to express is along the lines of "a hex-head bolt is a type of machine bolt," which is readily written in existing languages with a little extra vocabulary.

Two important technologies for developing the Semantic Web are already in place: eXtensible Markup Language (XML) and the Resource Description Framework (RDF). XML lets everyone create their own tags—hidden labels such as or that annotate Web pages or sections of text on a page. Scripts, or programs, can make use of these tags in sophisticated ways, but the script writer has to know what the page writer uses each tag for. In short, XML allows users to add arbitrary structure to their documents but says nothing about what the structures mean.

Meaning is expressed by RDF, which encodes it 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 sister of," "is the author 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.

Human language thrives when using the same term to mean somewhat different things, but automation does not. Imagine that I hire a clown messenger service to deliver balloons to my customers on their birthdays. Unfortunately, the service transfers the addresses from my database to its database, not knowing that the "addresses" in mine are where bills are sent and that many of them are post office boxes. My hired clowns end up entertaining a number of postal workers— not necessarily a bad thing but certainly not the intended effect. Using a different URI for each specific concept solves that problem. An address that is a mailing address can be distinguished from one that is a street address, and both can be distinguished from an address that is a speech.

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 zip code, we need to know which fields in each database represent names and which represent zip codes. RDF can specify that "(field 5 in database A) (is a field of type) (zip code)," using URIs rather than phrases for each term.

1. **Knowledge Representation Languages XML**

-The eXtensible Markup Language (XML) and the XML validation document XML Schema (XMLS) are used to add arbitrary structure to the documents. XML is a well-formed markup language which lets everyone create own tags that may surround a portion of content, but says

-nothing about what the structure means. On the Web XML is often used to store meta data, because it is application independent and has a human readable form, but exchanging XML documents between systems is only reasonable when both systems know what the tags denote.

-XML is not a knowledge representation language, but the syntax is used in many knowledge representation languages, e.g. RDF and OWL. Hence, it is important to know how an XML ﬁle –

is built. XMLS is a template and validation documents that deﬁnes the valid elements and attributes of an XML ﬁle.

Xml example

<?xml version=”1.0”?>

<rdf:RDF xmlns:rdf=”http://www.w3. org/1999/02/22−rdf−

syntax−ns#” xmlns:ab=”http://www. about .com/”

xml:base=”http://www. henrys page .com/”>

<rdf:Description rdf:ID=”Henry”

ab:work=http://www. job .com/

ab:age=”23”>

<ab:friend rdf:nodeID=”s3fo” />

</rdf:Description>

<rdf:Description rdf:nodeID=”s3fo”

ab:age=”23”>

</rdf:Description>

</rdf:RDF>

1. ***Ontologies***

-The general idea behind ontologies is to make knowledge explicit by expressing concepts and their relationships.

-In other words ontologies deﬁne the common words and concepts used to describe and represent an area of knowledge or collection of information about data and how the data is related. Thus, ontologies provide the means for establishing a semantic structure. Referring to the theory of communication, ontologies would represent the context of the terms. In the Semantic Web ontologies are semi-structured and represent an open world, which means that the model can grow with the data and that ontology does not contain every existing real world entity. An ontology model can be merged with another ontology model. Thus, ontologies in the Semantic Web are partial and modular. When we talk about ontologies we distinguish between the semantically richness of the various types.

**5.**1**Taxonomies**

A. taxonomy is a method to classify or categorize a set of terms in a hierarchical structure. In general, it is the study of the general principles of scientiﬁ classiﬁcation. When we adapt this dentition to the domain of information technology we can say that taxonomy is the classiﬁcation of information entities in the form of a hierarchy, according to the presumed relationship of the real-world objects that they represent.

B.Taxonomy, in general, is semantically weak, because it does not express rich meaning and does not distinguish between aggregation and generalization/specialization relations.

**5.2 Thesauri**

-A thesaurus deﬁnes relationships between words and phrases structured in a taxonomy. Some examples of term relations would be synonyms, homonyms, there is narrower than and the is broader than relations.

Homonyms describe two or more equal words with deferent meanings, where as synonyms are deferent words with the same meaning. The relations is narrower than and is broader than are relations between a parent and a child sub classiﬁcation, where is narrower than declares that the subject is more speciﬁc than the object and vice versa for is broader

5.3 **Conceptual Models**

-Conceptual Models are common in modeling databases or applications. The Uniﬁed Modeling Language (UML) is a widespread Conceptual Model in software engineering

* 1. **Logic Theories**

Logic Theories are built on axioms or statements deﬁned in a knowledge base and inference rules, which together are used to prove theorems about the domain. With these evidence it is possible to create new knowledge.

1. ***Incremental Ontology creation***

The vision of extending the current human-focused Web with machine process able descriptions of Web content was first formulated in 1996 by Tim Berners-Lee, the original inventor of the Web. The semantic Web has been actively promoted since then by the WWW Consortium, the organization that is chiefly responsible for setting technical standards on the Web. At this point, the field of knowledge representation and reasoning took center stage, but outcomes from other fields of AI have also been put to use to support the move towards the semantic Web. For example, natural language processing and information retrieval have been applied to acquire knowledge from the WWW4.

As semantic Web is a relatively new and dynamic field of investigation, it is difficult to precisely delineate the boundaries amongst these. Semantic Web communities have defined their community by including those researchers who have submitted publications or held an organising role at any of the past International Semantic Web Conferences, or the Semantic Web Working Symposium. Today semantic Web has created its own importance by developing meaningful results.

The following steps are involved in the construction/creation of Ontology:

- Acquiring the domain knowledge: This step consist of identifying and collecting appropriate expertise and information resources to present in a common language, all descriptive terms with consensus and consistency

-Design the conceptual structure: Identify the key concepts within the domain along with their associated properties. Further identifying the relationships among the concepts

-Develop the suitable details: Include concepts, relationships and various instances to achieve the level of detail to satisfy the given purpose of the ontology

-Verify: Ensure the structure for uniformity. If required, revise any logical, semantic or syntactic errors among the elements

-Commit/submit: Once domain expert ensures, the ontology is committed by publishing it within its planned deployment location. Then the ontology can be referred and used within the environment

There are many ways in which one can contribute to creating the Semantic Web. Here's a few of them:-

- Publish some globally useful data in RDF.

-Write an inference engine in the language of your choice.

- Spread the word: do some education and outreach.

-Help in the development of RDF Schema and/or DAML.

- Contribute in representing state in RDF, a rather neglected field of research.

- Apply your own development backgrounds to the Semantic Web, give us all a new angle to consider it from.

-Instead of using some proprietary system for your next application, consider making it a Semantic Web project instead.

1. ***Evolution of the Semantic Web***

The internet was conceived way back in the 1960’s as an antidote to any sabotage or external attack threatening the communication system of any country. For more than two decades, the internet was a closely held technology that was being used by the academia, military and scientists. It was later on in the mid-1990’s that the internet was released for public use. And the vast amount of information that was available in the public domain needed to be made meaningful so that information can be searched in such a manner that automated search engine robots can pick it up and serve them up to users.

The semantic web was now more about ubiquity, click streams and personalization.  Meta data or the information about where users go, which places they visit and what they click on when they browse the web. By using this data, the first baby steps towards creating a more intelligent web were taken.

Search engines began to tag web pages with some kind of universal machine readable code that informed search engine bots what these web pages are all about, what information they contain and how they might change over time.

With such an initiative the web became more like a structured database akin to lexis-nexus, hovers etc. This made it easier to find information and even easier to aggregate web content from various sources and deliver it in a digestible format.

How we reached where we are today

It took years of hard work on behalf of mathematicians, scientists and internet experts to create a reasoning engine that blended with a search engine. Such an engine not only indexed the complete list of all occurrences of a particular word but also weeded out superfluous and meaningless results. Developing such an engine offered a sound commercial incentive as companies could offer targeted ads at users who  were aggressively searching for their products. The best way to understand the semantic web is to describe things in such a manner that the computer can understand. The following equation will explain how search engine spiders interpret search queries.

A is related to b

B is related to c

then a is related to c

This is the way search spiders go about finding all the relevant information about a term that is queried in the search engine.

Birth of Folknomy

With such a strong commercial incentive in place to get search right, other portals started offering users the freedom to tag anything they liked and share it with others. This led to the birth of social networking sites that gave search engines some kind of fuzzy idea about the relevance of a particular search term and what kind of results will best highlight the attributes of the product, service or search term being queried.

Syndicated feeds followed and search engines could now offer a better user experience as they had figured it out as to what exactly were users searching for when they typed a search query. They began supporting their results with images, videos, book results and took the semantic web to the next level of growth.

1. ***Advantages***

The main purpose of the Semantic Web is driving the evolution of the current Web by enabling users to find, share, and combine information more easily. The semantic web is a vision of information that can be readily interpreted by machines, so machines can perform more of the tedious work involved in finding, combining, and acting upon information on the web.

**Capability**

The technology must be capable of:

- Retrieving large amounts of textual data quickly.

- Allowing users to add annotations so that a reasoning capability exists.

-Making text retrieval more specific.

-Allowing conclusions to be drawn by data on the Web and across organizations.

**Challenges**

Some of the challenges for the Semantic Web include vastness, vagueness, uncertainty, inconsistency, and deceit. Automated reasoning systems will have to deal with all of these issues in order to deliver on the promise of the Semantic Web.

**BENEFITS OF THE SEMANTIC WEB**

• Information is captured in a language agnostic format.

• A central repository for knowledge is created.

• More precise, relevant information is captured.

• Processes and procedures are mapped to data sources.

• One collective view of knowledge across enterprise applications is created.

As a result:

• Point-to-point integration becomes obsolete.

• Application integration is easy and efficient.

• Superfluous data decreases.

• Knowledge across applications becomes consistent.

• Upgrades and maintenance are simplified.

8. ***COMPONENTS OF THE SEMANTIC***

**URI – UNIFORM RESOURCE IDENTIFIER URI’s**

are simple web identifiers that are often found on the World Wide Web (i.e. http, ftp).

**RDF – RESOURCE DESCRIPTION FRAMEWORK**

RDF is used by The Semantic Web to describe data, thus allowing it to be shared more conveniently. It enables software developers to design products that will deploy better search engines by utilizing the metadata. As a result, users have more control over what they are viewing. Additionally, RDF is vocabulary agnostic which creates an interoperable environment capable of supporting a diverse range of ontologies.

**RDF SCHEMA**

The RDF Schema is a language used by The Semantic Web to describe the data properties used in RDF. display it. Semantic Web uses OWL to add reason to data by identifying and describing relationships between data items.

**OWL ontologies**

are capable of processing the content of information, rather than just presenting the data to users.