An Eagle Eyed View over the Search Implementation of Semantic Web

K. Valli Priya Dharshini
Assistant Professor,
Computer Engineering,
Pimpri Chinchwad College of Engineering & Research,
Ravet, Pune, Maharashtra, India

ABSTRACT
To know more about a particular domain, the basic requirement of an user has become a search engine now – a – days. Web technology which is playing a major role in providing the acquaintance for humanity, process the technical background followed by knowledge representation based technologies. Here the semantic web makes use of the benefits and implementation that combine together to satisfy the user needs. It implements the idea of searching technique with enhanced capability of doing the search in a meaningful manner, hence the name "*Semantic web*". The basic idea of the semantic web analyzes the fact about user needs and develops the queries to start the search. This paper explores the applicability of semantic web, its search implementation over the existing search engine techniques.

General Terms
Information retrieval, Ontology, UDDI.

Keywords
Ontology, Metadata, OWL, RDF, XML, Web crawling

1. INTRODUCTION
According to W3C, semantic web provides a common framework that allows data to be shared and re used across application, enterprise and community boundaries. The term was coined by Tin – Berners – Lee, the inventor of WWW and the director of W3C. He defines web of data, mesh of information linked up so that it can be processed directly and indirectly by machines. The main purpose is to find, shareto improve, extend the standardized system, and many languages, publications, tools and combine information easily. Humans are capable of using the web to carry out tasks such as finding "Irish" word for folder, to search lowest price DVD, reserving a library book, machines cannot perform without human direction, which need some interpretation by a third party. Normally a search engine acts as that third party. Here the semantic web is a vision of information that can be readily interpreted by machines, so that machine can perform more of the tedious work involved in finding, combining and acting upon that information on the web.

2. SEMANTIC WEB – A FRAMEWORK
2.1 URI
Every search starts with an URL. Here semantic web starts with URI i.e., Uniform Resource Identification. An URL is an URI, but every URL is not an URI. i.e., a URL is a specialization of URI that defines network location of a specific representation of given resource. An URI identifies a resource either by a location or by a name. A URI is simple a Web Identifier: like the strings starting with "http:" or "ftp:" that you often find on the World Wide Web. Anyone can create a URI, and the ownership of them is clearly delegated, so they form an ideal base technology with which to build a global Web on top of. In fact, the World Wide Web is such a thing: anything that has a URI is considered to be "on the Web". The Semantic Web is generally built on syntaxes which use URIs to represent data, usually in triples based structures: i.e. many triples of URI data that can be held in databases, or interchanged on the World Wide Web using a set of particular syntaxes developed especially for the task. These syntaxes are called "Resource Description Framework" syntaxes.

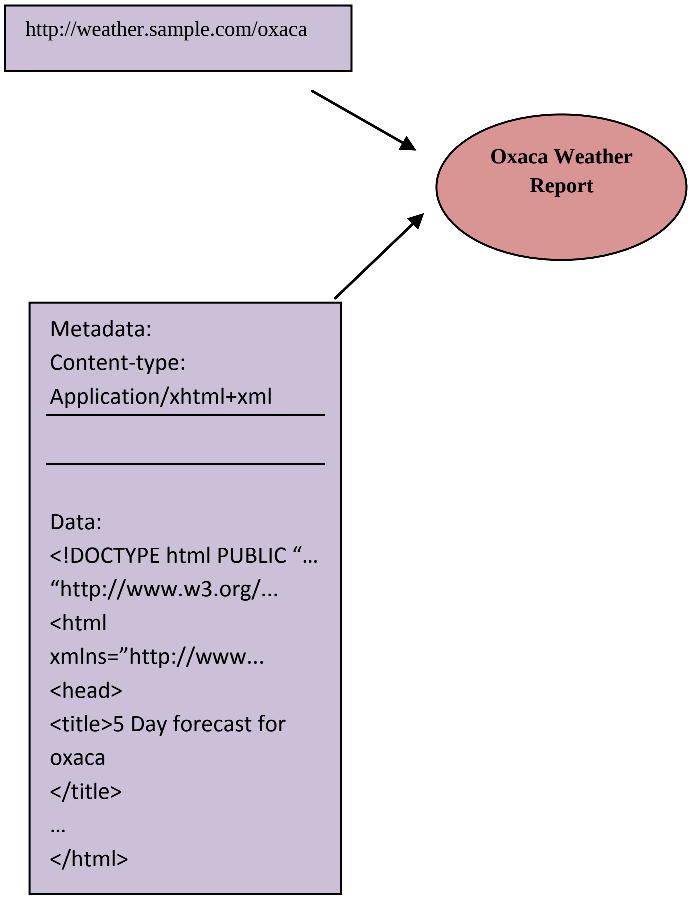

Figure 1 .Locating the metadata.

When a user enters a query into a search engine (typically by using key words), the engine examines its index and provides a listing of best-matching web pages according to its criteria, usually with a short summary containing the document's title and sometimes parts of the text. The index is built from the information stored with the data and the method by which the information is indexed. Here URI process the metadata as shown above.

2.2 Metadata

Information about information in real world for searching .Suppose in a hunt of treasure, the need is a to find the exact location of treasure. If that key contains another key to find the location, that indirect key can be visualized as metadata and such it works for accessing the resource using RDFs.

2.3 RDF

A triple can simply be described as three URIs[4]. A language which utilises three URIs in such a way is called RDF: the W3C have developed an XML serialization of RDF, the "Syntax" in the RDF Model and Syntax recommendation. RDF XML is considered to be the standard interchange format for RDF on the Semantic Web, although it is not the only format. For example, Notation3 (which we shall be going through later on in this article) is an excellent plain text alternative serialization. Once information is in RDF form, it becomes easy to process it, since RDF is a generic format, which already has many parsers. XML RDF is quite a verbose specification, to learn XML RDF properly, and to understand a little about XML, a quick look at an example [4], of XML RDF is illustrated here below:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-yntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:foaf="http://xmlns.com/0.1/foaf/">

<rdf:Description rdf:about="">

<dc:creator rdf:parseType="Resource">-

<foaf:name>XXXX</foaf:name>

</dc:creator>

<dc:title>The Semantic Web </dc:title>

</rdf:Description>

</rdf:RDF>
```

This piece of RDF basically says that this article has the title "The Semantic Web: An Introduction", and was written by someone whose name is "XXXX". Here are the triples that this RDF produces:-

```
<><http://purl.org/dc/elements/1.1/creator>

_:x0 .this <http://purl.org/dc/elements/1.1/title>"The Semantic Web" .

_:x0 <http://xmlns.com/0.1/foaf/name>"XXXX" .
```

This format is actually a plain text serialization of RDF called "Notation3". Note that Even XML RDF is preferable for Notation3, it is generally accepted that Notation3 is easier to use, and is of course convertible to XML RDF anyway. The benefit that one gets from drafting a language in RDF is that the information maps directly and unambiguously to a model, a model which is decentralized, and for which there are many generic parsers already available. This means that if an user knows the details about syntactic fluff and semantic bits of data of an RDF application, without often implicit specification, RDF can resolve those syntactic fluff and semantic bits of data.

RDF being a part of the Semantic Web, it is beneficial drafting the data in RDF now draws parallels with drafting your information in HTML in the early days of the Web. XML Schema is a language for restricting the syntax of XML applications.

However, using XML Schema in conjunction with RDF may be useful for creating data types and so on. Therefore the answer is "possibly", with a caveat that it is not really used to control the syntax of RDF. This is a common misunderstanding, perpetuated for too long now.

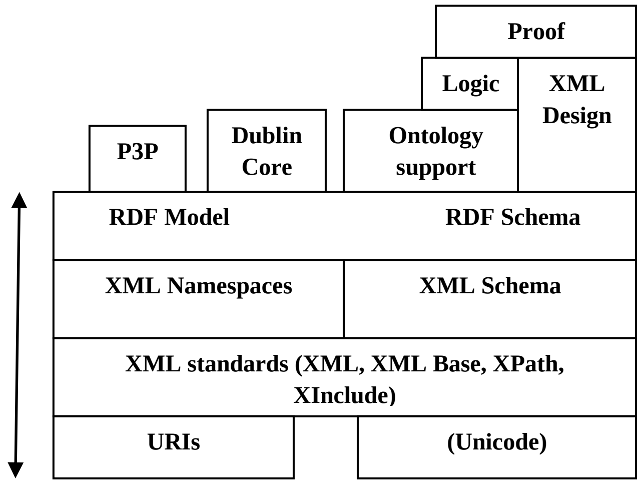

Figure 2. Building blocks of RDF

2.4 XML

At the W3C Query Language workshop, there was a clear difference of view between those who wanted to query documents and those who wanted to extract the "meaning" in some form and query that. This is typical with RDF model, even though those things can be expressed in terms of the XML model. Since some of the XML features has to ce paid attention and was discussed in this part.This note assumes that the XML data model in all its complexity, and the RDF syntax as in RDF Model and Syntax, in all its complexity. It doesn't try to map one directly onto the other -- it expresses the RDF model using XML as discussed by Tim Berners Lee [4]. In the XML Schema method, the meaning for data is provided by separate means.XML Schema uses schema documents for this purpose. These separate documents are given special meaning and are processed differently from documents that carry data.

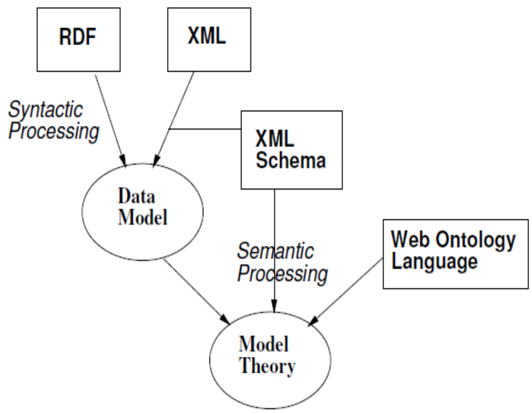

Figure 3.Role of XML in RDF

2.5 Ontology

W3C Web Ontology Working Group Releases Working Drafts for OWL Semantic Mark up Language. Three initial working draft documents on 'OWL' have been published by the W3C's Web-Ontology Working Group (WebOnt). OWL is a semantic mark up language for publishing and sharing ontology on the World Wide Web. OWL is derived from the DAML+OIL Web Ontology Language and builds upon the Resource Description Framework [5].

The designers expect that OWL will support the use of automated tools which "can use common sets of terms called ontology to power services such as more accurate Web search, intelligent software agents, and knowledge management." The OWL Web Ontology Language is being designed "in order to provide a language that can be used for applications that need to understand the content of information instead of just understanding the human-readable presentation of content. OWL facilitates greater machine readability of web content than XML, RDF, and RDF-S support by providing an additional vocabulary for term descriptions."

The *Feature Synopsis for OWL Lite and OWL* introduces the OWL language. The *OWL Web Ontology Language 1.0 Reference* provides a systematic, compact and informal description of all the modelling primitives of OWL. An OWL knowledge base is a collection of RDF triples as defined in the RDF/XML Syntax Specification; OWL prescribes a specific meaning for triples that use the OWL vocabulary. The Language Reference document specifies which collections of RDF triples constitute the OWL vocabulary and what the prescribed meaning of such triples is. The *OWL Web Ontology Language 1.0 Abstract Syntax* document describes a high-level, abstract syntax for both OWL and OWL Lite, a subset of OWL; it also provides a mapping from the abstract syntax to the OWL exchange syntax.

The W3C Web Ontology Working Group has published an initial working draft document outlining requirements for the Ontology Web Language (OWL) 1.0 specification. The draft document "specifies usage scenarios, goals and requirements for a web ontology language. Automated tools can use common sets of terms called ontology to power services such as more accurateWeb search, intelligent software agents, and knowledge management."An'ontology' in terms of the WG charter "defines the terms used to describe and represent an area of knowledge.

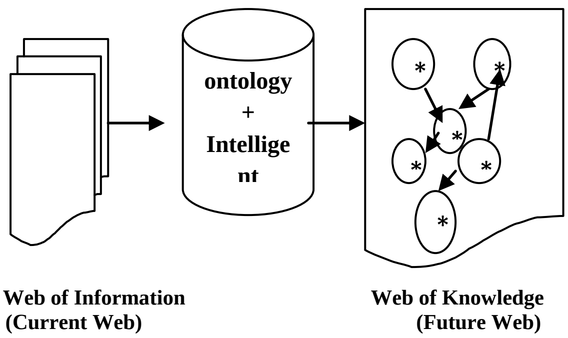

Figure 4. Ontology – A backbone of Semantic Web

Ontology is used by people, databases, and applications that need to share domain information, where a domain is just a specific subject area or area of knowledge, like medicine, tool manufacturing, real estate, automobile repair, financial management, etc. Ontology includes computer-usable definitions of basic concepts in the domain and the relationships among them. Ontology formally defines a common set of terms that are used to describe and represent a domain. The WD specification motivates the need for a Web ontology language by describing six use cases. Some of these use cases are based on efforts currently underway in industry and academia, others demonstrate more long-term possibilities. The use cases are followed by design goals that describe high-level objectives and guidelines for the development of the language. These design goals will be considered when evaluating proposed features

3. SIGNIFICANCE OF SEMANTIC WEB

3.1 Functionalities

1- Handling morphological variations
2- Handling synonyms with correct senses

3- Handling generalizations

4- Handling concept matching

5- Handling knowledge matching

6- Handling natural language queries and questions

7- Ability to point to uninterrupted paragraph and the most relevant sentence

8- Ability to enter queries freely, no special formats like quotes, or Boolean operators.
9- Ability to operate without relying on statistics, user behaviour, or other artificial means.
10- Ability to detect its own performance.

This also describes the things that the semantic web how deviates from the normal search engines. These are considered as the most important functionalities of semantic web with which it performs the search in a meaningful manner as discussed by T.Berners Lee and his group of researchers[3].When there is no semantic content analysis in a search algorithm, relevancy scores refer to artificial measurements, like how popular the page is. A semantic search engine is expected to produce a relevancy score that reflects the degree of meaning match.
This capability provides flexibility for the developers to apply meaning thresholds. Accordingly, the search engine can understand its poor performance to automatically flag areas of improvement that is needed.

3.2 Search Implementation

3.2.1. Working of a normal search engine

To make a study over the search implementation, first working of a normal search engine must be considered. A web search engine is designed to search for information on the World Wide Web servers. The search results are generally presented in a list of results often referred to as SERPS, or "search engine results pages". The information may consist of web pages, images, information and other types of files.

To initiate web access, when web service application is compiled, a WSDL file is generated which specifies how web service can be accomplished. A client application finds WSDL pages from UDDI (Universal Description and Definition Language) which provides an authorized web access, i.e., a server which was supposed to provide service and that must get registered in the UDDI registry. All these can be done by a web reference on any .Net project.

3.2.1. Working of a normal search engine

A semantic web collects information based on semantic contents relevant to both the search query and the information on web pages. Supplementary technologies such as Ontology, Modelling languages, Mark up language must be incorporated for the extra added search engine capabilities [7]. Ontology incorporates tools that convert the syntactic content into semantic content despite of the explicit specification of metadata in the XML coding since RDF act as a tool that abstracts the end level metadata in its triples. The remarkable variation in the step processes on working of semantic web search engines are:

i. It collects available semantic content on the Web,
ii. It analyses it to mine the useful metadata and its relevant files and
iii. It implements efficient query amenities for admittance of the data.

4. CONCLUSION

Even then, semantic web techniques found to have some pinpoints that cannot replace a normal search engine. Few of them are implementation infrastructure might require fine concepts, things, and events, that may incorporate extra hardware, recent technology exposure etc. If the vision about the Web of trust can be still far way, we have to point out the important steps already achieved: RDF and OWL standards have been completed; many semantic web applications have been developed in the last years making collaboration with corporations much stronger, and given the right benefits to semantic web technologies.

There is much to be fulfilled such as finding the unauthorized registries from UDDI, balancing the load of web server using a Man In the Middle Server, etc which all these things lead to the future scope and will make the next generation search engine a successful one.

5. REFERENCES

[1] Bijan Parsia and Peter F. PatelSchneider 2004 Meaning and the Semantic Web Technical Report. University of Maryland at College Park.

[2] Simona Elena VARLAN, Alexandru Ioan Cuza 2010 University of Iasi, Faculty of Economics and Business Administration Iasi, Romania – advantages of semantic web in the knowledge based society

[3] T. Berners-Lee, J. Hendler, and O. Lassila. May 2001. Feature article. The semantic web - *Scientific American*.

[4] G. Klyne and J. J. Carroll. 10 Feb. 2004 Resource description framework (RDF): Concepts and abstract syntax. W3C Rec.,

[5] Prof. Latifur Khan 24 August 2009 Technical Report. Ontology Alignment and RDF Query Processing Intelligence Advanced Research Projects Activity and National Geospatial Intelligence Agency.

[6] Oscar Corcho, Mariano Fernandez-Lopez, Asuncion Gomez-Perez 21 August 2002 Methodologies, tools and languages for building ontologism. Where is their meeting point? Universidad Polite cnica de Madrid, Campus de Montegancedo s/n, Boadilla del Monte, Madrid 28660, Spain.

[7] Fakorede Oluwatoyin Johnson September 2007 An Investigation into the implementation issues and challenges of service oriented architecture. A research project. Implementation Issues and Challenges of Service Oriented Architecture