

## Semantic Knowledge at User side through Web Browser Plug in: A Review

Hitesh Kumar

Pursuing Master of Technology (Computer  
 Science and Engineering)  
 GJUS&T  
 Hissar, Haryana

Dr. Dharmender Kumar (Assist. Proff.)

Computer Science and Engineering  
 GJUS&T  
 Hisar, Haryana

**Abstract** –The semantic web is the web in which searching is based on the meaning of words. It provides the semantic knowledge to the user in a structured format. Semantic web is not the link between web pages it is the link between things. In this paper we will discuss the related terms to the semantic web and how the information becomes the knowledge. This article presents the role of ontology, browser, database to extract the semantic knowledge and how the relevant terms are related to each other.

**Keywords:** Semantic web, ontology and semantic knowledge.

### 1. Introduction

The word semantic stands for the meaning of. So semantic web is the web in which searching is based on the meaning of words. The semantic web is the web of data which describes the thing in a way that computer or machine can understand. Tim Berners-Lee is the inventor of the term semantic web. There is a lot of data on web pages. Due to this large amount of data it is more difficult to find desired data by user. So there is a need to have a technique which provide useful information to user to make an efficient searching.. This useful information is called knowledge. Ex. Suppose you give me a book How much knowledge you pass me?

It depends on the amount of knowledge means command on the language that is used in the book In semantic web information is analyzed. After performing semantic analysis the information becomes knowledge. So it can be said that network of knowledge is the semantic web. It extends the network by inserting machine readable metadata about web pages. It also tells how this metadata or things are related to each other. So semantic web is “web of data that can be processed directly or indirectly by machines”[1].

We know statements are built by following the syntax. Syntax of a language defines the rules which

built the statement of a language. Now question arises that how can syntax become semantic?

In simple web (www) searching is based on the syntax rules. In this links are between web pages. The Semantic Web links are not between the web pages. The Semantic Web describes the relationships between things (like A is a part of B) and properties of things like (size, shape, color). So it goes in a semantic way from syntax.

The semantic web consists of XML, XML Schema, RDF, RDF Schema and OWL(web ontology language) that are organized in the Semantic Web Stack It allow users to look for words or meaning of words found in that page[2].

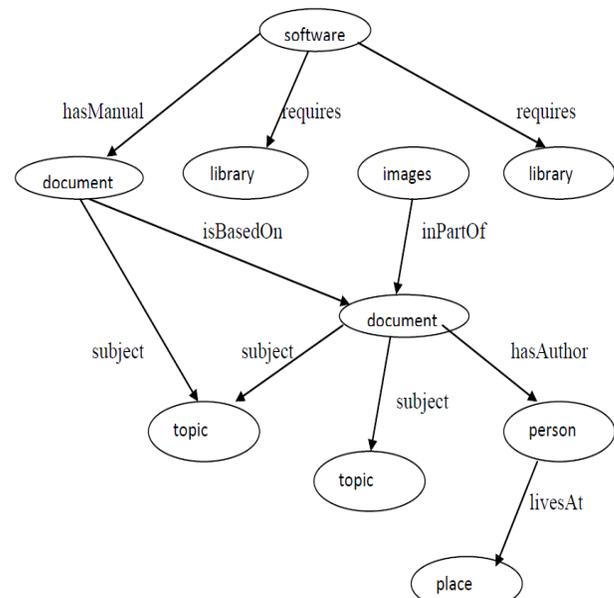


Figure1: semantic web example.

The diagram is describing whole about semantic web. The software is consists of document and library. Document can be manual and based on other document. It also consists of the topic that can be related to any subject. There is a author of document

and it is a person. This person can live at a place .These all are related to each other by their part or properties and in this way it goes in semantic way.

## 2. Data base distribution and information filtering:

There is a huge amount of unstructured data on web. To make it in the structured format a database is created. Content of database is published on web pages by a server called d2r server .In the semantic web there are many resources which are distributed in a proper way.

The resource description frame (RDF) is the simple language for expressing data models which refers the object (resource) and their relationship. Resources are the title, author, and content and modification date in a book. Information goes to the rdf files and it is possible to search, discover, collect and analyze the information from web. If there is different information in RDF file like music, car etc. then intelligent web application collect it from other resource, combine these and present to the user in a meaningful way. RDF schema is the extension of RDF. It is used for describing properties and classes of RDF base resources. RDF schema specification describes the structure of metadata. Metadata is used to describe the www resources [3].

RDF and RDF schema are enabled by the D2r server.D2r server mapping specifies how resources are identified and how values are generated. It supports hyperlink navigation by providing links on RDF and X html.

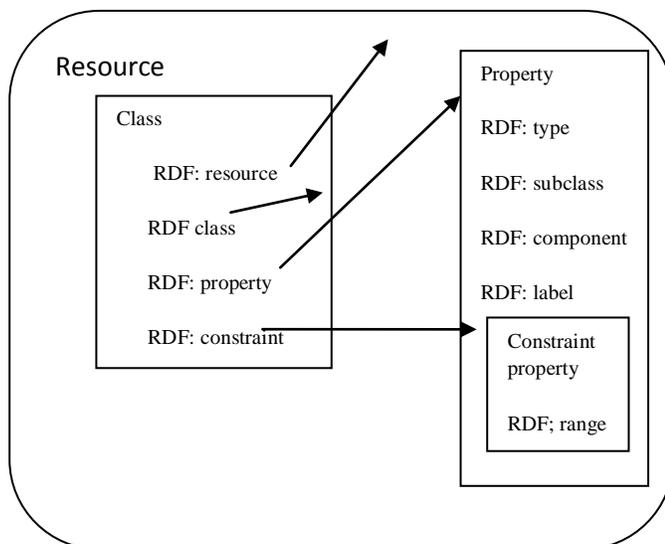


Diagram: Resource description frame (RDF) Classes and Resources [3].

Properties are defined by the RDF: Property class. It can be seen as attributes to describe resources by assigning values to them. Resources themselves are properties. The RDF Schema Specification defines four specific properties. (Rdfs: subclass of, RDF: type, rdfs: range, rdfs: domain)

There is a large amount of content available on the World-Wide Web. There are many pages which are written by many authors .These are in unstructured form. Now the question is how to use this effectively. There can be user preference in which they visit the web pages most of the time. So there is a need of filtering of web-based information according to user preferences [4]

Since information extracting or filtering is not only searches for exact keywords that is given by user. This will search the evidence phrases such as synonyms, hyponyms and hyponyms of keywords generated from WORDNET.

URL rating is based on 3 techniques:

1. Evidence phrase (EP). IFA(information filtering agent) searches for evidence phrases (EP) (such as exact keywords, synonyms, hyponyms and hyponyms) in a webpage after removing all function words(e.g., “and” and “of”. The IFA filters the context of web pages that contain keywords in users’ queries by considering part relation or whole relations of words such as metonymy (part relation) and homonym (whole relation).[11]
2. Evidence frequency (EF): It measure the frequency of EP. It determines the each non function keyword in the query .Non function keywords have variable frequency in collection of text It is measured by the equation  $EFL$  in  $UL$  is computed by taking the average of all no keywords function.

3. Evidence nearness  $N= 1/D$ : information retrieved can be increased by nearness. If the query is “London symphony orchestra”. Then in a web page if these 3 words are adjacently occur then it will be more relevant as compare to individually come.

Users’ rating is based in part on model of information seeking behaviors. Each user is asked to perform both actions browsing and filtering the URL [5].

In searching the web Users were instructed to scan the content of a website for keywords and/or related information. Information retrieval is a field of study that helps the user to find required information from a large collection of database .The goal of information retrieval system is to help a user to locate the most similar document that is required to satisfy the user

needs. If keywords and/or related terms are found, users are instructed to determine whether the information is of the correct sense. If the information is of the correct sense, the rating in 1 is retained. Otherwise, a rating of 0 is assigned [5]. The IRS (information retrieval system) is used to retrieve the most similar documents that satisfy the user requirement. Users make a query and send it to the IRS. The information retrieval system searches for proper matches in the document database and retrieves the relevant information using background knowledge drawn from controlled vocabularies or ontological information sources [6].

### 3. Ontology in semantic web:

There are the questions which are raised in the mind of every user are:

What is ontology? What principles should I follow to build ontology? What are the components of ontology? What types of ontology exist? How are ontologies organized in libraries? What methods should I use to build my own ontology?

Ontology is the description of the whole world. It is about the domain and identifies the concept and organizes them. It is used in other technologies such as AI. Ontology is the representation vocabulary specified to some domain. It is the data model that represents the set of concepts and relationships between them [7]. Ontologies are used for both purposes to structure the web and to provide a platform for understanding between the systems. There can be any preference of user in so many unstructured web pages. In the acquisition of user preference, ontology is needed. It investigates how knowledge domain can help to identify user preferences. In knowledge-based systems, ontological representations are also seen which use relationships between web entities (bookmarks, web pages, page authors etc.) [4].

Ontology creation is an important work in semantic web and knowledge management. Concepts are represented in a proper manner and relationships between those concepts is designed by ontology in a specific domain. There are many domains of ontology like medical domain, electronic domain etc. There are many concepts in the domain. These concepts are represented in a proper way in an ontology and these concepts are also related to each other. Ontology is used to increase the efficiency and consistency by description of resources for the purpose of sharing common descriptions and relations

with in a particular community there are RDF, RDF schema, web ontology language for the context of Foa. The RDF format has been chosen in ontology as it is best suited for its needs and purpose [9].

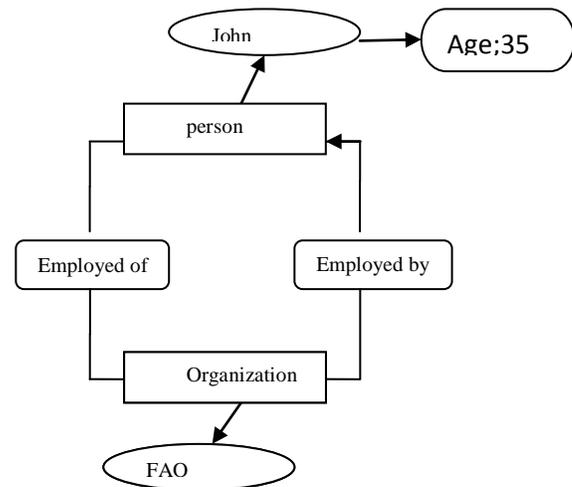


Figure 2.1 Ontology diagram of the relationship between person and organization.

In ontology, there is a relation between the object and concepts that is organized in a proper way. In the diagram, person and organization are the classes, employed by and employed of are the properties or relations, John and FAO are instances, and age is an attribute. As shown, person is linked to organization through a property employed by/of, and instances are the actual data.

There are 2 types of ontology learning algorithms for the creation of ontology. One is human-guided learning algorithm and the second is guided machine learning algorithm. In most cases, ontology is highly task-specific and subjective. There are two major questions for constructing ontology by guided machine learning:

1. Can ontology provide the same quality made by guided machine learning as human-guided learning?

2. Can a guided machine learning approach learn from individual users and capture the distinction among their personal preferences?

It can be said that guided machine learning is able to produce an ontology of the same quality as human do and it also learns from individual

preferences [8]. To build, test, and maintain services are very expensive in Knowledge-based systems. so there is a need of methodology based on formal specifications of shared resources, reusable components, and standard services. We believe that specifications of shared vocabulary can play an important role in such a methodology. For such knowledge level communication we need conventions at three levels: Representation language format level, agent communication protocol level, specification of the content of shared knowledge [10]

To make ontology we need a language that is called web ontology language. America and Europe have already identified the more powerful ontology modeling language. This is OWL (Web ontology language). It is the combination of 2 languages DAML (American) and OIL (a European language) [11]. Main requirements of an ontology language it should be well-defined syntax and well-defined semantics. An efficient reasoning support should be in the language it should provide sufficient expressive power.

#### 4. Role of browser for semantic knowledge:

Web browser retrieve the content and display from remote web pages using protocols HTTP and other technology like CSS and java script. Some of the subsystem in reference architecture may be treated as a single subsystem for simplicity. On the other hand it may be of multiple subsystems in the web browsers.

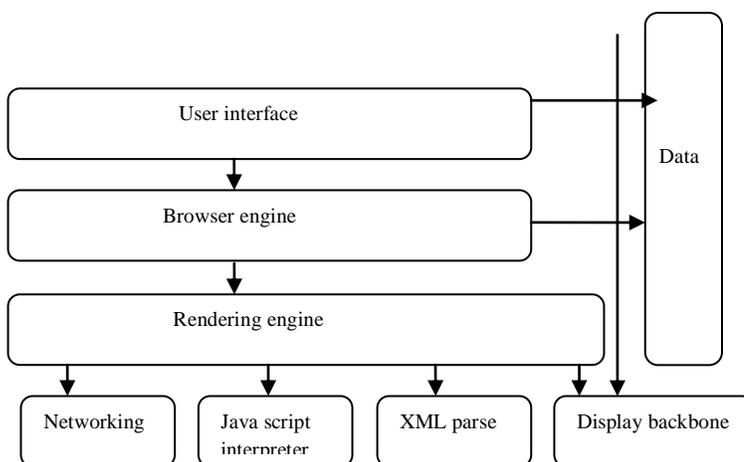


Figure: web browser and layers [12].

Web browser has many parts:

1. User interface: It is the interface between user and the browser. It provides features such as toolbars, visual page-load progress, smart download handling, preferences, and printing.
2. Browser engine: It provides the interface for querying. It loads a given URI and supports primitive browsing actions such as forward, back, and reloads.
3. Rendering engine: Performs parsing and layout for html documents. The Rendering Engine subsystem produces a visual representation for a given URL It calculates the exact page layout and may use reflow algorithms to incrementally adjust the position of elements on the page.
4. The Networking subsystem implements file transfer protocols such as HTTP and FTP. It translates between different character sets, and resolves MIME media types for files.,
5. JavaScript Interpreter evaluates JavaScript (also known as ECMA Script) code, which may be embedded in web pages,
6. Xml parser, The XML Parser subsystem parses XML documents into a Document Object Model (DOM) tree.
7. The Display Backend subsystem provides drawing and windowing primitives, a set of user interface widgets, and a set of fonts. It may be tied closely with the operating system..
8. The Data Persistence subsystem stores various data associated with the browsing session on disk. This may be high-level data such as bookmarks or toolbar settings, or it may be low-level data such as cookies, security certificates, or cache. [12]

Semantic web browser http offers a simple standard by which information can be fetched from any web server. Url gives a way for searching of any web page that is required.. Through its use of http and http server publisher is able to give the right way to present its information to end users. User can gain direct access to the underlying information and control .for managing connection to the rdf repository and various approaches for federating metadata [13]. The frequent use of Web pages to aggregate related items and building on the notion of customization is introduced by bookmark managers in the Web browser. Bookmark and history panes are like an evidence of web pages that was searched in Internet

Explorer and Mozilla. Browser should be maintaining the original architecture of the web page when any plug in is attached with it. Mozilla fire fox have this capability to maintain the original structure of web page .It was introduced in1998, it renders very fast. Now many features are added to it to increase its power. Linux is the only text searching browser today. due to its only text searching it is very simple.

### **5. Role of plug in:**

An application or software that usually cannot be run independently and enhances another software called a plug in . It modifies the interface and behavior of web browser. It is used to add new features to browser. By a plug in power is enhanced in a specific manner. When hardware is installed then plug in is used as a common standard.

Mozilla firefox also supports plug-in. Since there is generally a clear separation between the browser and the plug-in. When the browser encounters references to content a plug-in the results are discrete objects embedded within a webpage. The same distinction between plug-in and extensions is in use by other web browsers, such as Microsoft Internet Explorer, where a typical extension might be a new toolbar, and a plug-in might embed a video player on the page. Since plug-ins and extensions both increase the utility of the original application, Mozilla uses the term "add-on" as an inclusive category of augmentation modules that consists of plug-ins, themes, and search engines.

Compiler plug-in called annotation processors processed the annotation algorithm when Java source code is compiled. Processors can produce informational messages or create additional Java source files or resources, these may be compiled and processed, and also modify the annotated code itself. The Java compiler conditionally stores annotation metadata in the class files, if the annotation has a Retention Policy of CLASS or RUNTIME. Later, the JVM or other programs can look for the metadata to determine how to interact with the program elements or change their behavior [14].

### **Works on semantic web:**

[1] Describes the meaning of semantic web. In this searching related to the meaning of words. [2] Tells that in this all the extends markup languages are used like xml and java script. In today era there is a need of this type of web. [3] Describes that anybody can be author of the web page. So there is so much unstructured data on web pages. So there is

requirement to make it in structured format . How it can be in the structured format is discussed by [4]. There is so much relevant terms in the web pages. Identification of the relevant terms is not an easy work. These can be according user preferences. is described by [4] Techniques for this like annotated algo comes in light by [5]. Information is filtered by information filtering agent. But it have some difficulties which are removed by some tools or some other algo. Ontology is a stone of mile in semantic web. Ontology is not a short terms or not of a short domain whole world can be in the ontology term is discussed by [6, 7]. Ontology can be constructed. Manually or by machine .Ontology can be created by machine guided learning best as human guided [8]. For construction of ontology there is a language is required. Languages are owl for construction of ontology allows restriction classes. A restriction class is a class that is characterized by certain limitations on its instances in owl[11] Semantic web works by selection of ontology. By selecting the specific ontology we can get the semantic knowledge .Now for browsing and to extract the information there is requirement of browser. There are many layers in the web browser [14] plug in is also used for browsing semantic web at these layers. Web browser is compatible with ontology in a specific manner [13]. [12] Browser is compatible with web pages in a structured format for rendering the web pages. It is software to enhance the other software capability. Web pages can be interpreted by technology like annotation algo. IT allows users to associate meta-information with web resources, which can then be used to facilitate their interpretation. In this an algo is implemented for semantic knowledge [15]. Other is complementary knowledge source in which a user can call service to gain instantaneous access to the background knowledge which is relevant to a web resource. In this a semantic layer is created and then knowledge is extracted. The work will be on this type of technique [10]. So by the semantic web useful information is extracted from large data.

### **Conclusion:**

There is a large amount of data on the web .it is difficult task to retrieve the relevant information in a semantic way due to web world changing day by day.in this paper we discussed whole about semantic web and knowledge. There are many techniques which provide semantic information to user in their own way to help them to find specific information. To fulfill the users need to search in a semantic way for the knowledge at user by plug in is an emerging research field in semantic web.

## References

- [1] Tim Berners-Lee ,Yuhsin Chen ,Lydia Chilton , Dan Connolly ,Ruth Dhanaraj, James Hollenbach , Adam Lerer ,David Sheets , Tabulator "Exploring and analyzing linked data on the semantic web" In Proceedings of the 3rd International Semantic Web User Interaction Workshop [52 citations — 1 self] 2006.
- [2] Stefano Mazzocchi "Toward the semantic web a view of XML from outer space" Apache CON EU, 2007.
- [3] Brickley, D. and Guha, R., "Resource Description Framework (RDF) Schema Specification". 2000, WorldwideWebConsortium.
- [4] Middleton, S., De Roure, D., and Shadbolt, N. "Capturing knowledge of user preferences: Ontologies in recommender systems". In Proc. Of the ACM K-CAP'01 Conference. Victoria, Canada: ACM Press 2001.
- [5] Sini, M., Salokhe, G., Pardy, C. "Towards an Ontology-Enhanced Information Filtering Agent" journal ACM SIGMOD Record volume33, pp95-10 2004.
- [6] D. L. McGuinness. "Ontological Issues for Knowledge-Enhanced Search". In Proc. of Formal Ontology in Information Systems, pp. 302-316, 1998.
- [7] B Orgun, M Dras, A Nayak, G James "Approaches for semantic interoperability between domain ontologies", proceeding of the second Australian workshop on advances in ontology Volume: 72, pp 41-50, 2006.
- [8] Hui Yang and Jamie Callan, "Human-Guided Ontology Learning". Second Workshop on Human Computer Interaction and Information Retrieval (HCIR2008).Microsoft Research, Redmond, WA, USA. Oct, 2008.
- [9] Sini, M., Salokhe, G., Pardy, C., et al. "Ontology based navigation of bibliographic metadata example of the Food, Nutrition and Agriculture" Journal, 2007. In International Conference on Solid Dielectrics, Winchester, Hampshire UK, 8-13 July 2007, pp.1-13. (Published) [Conference paper].
- [10] Gruber, T.R., "A Translation approach to Portable Ontology specifications. Knowledge Acquisition", 1993 5(2): pp. 199-221 year 1993.
- [11] World Wide Web Consortium, 2004, Web Ontology Language, [Online],
- [12] Alan Grosskurth, Michael W. Godfrey "A Reference Architecture for Web Browsers", Proceedings of the 21<sup>st</sup> IEEE International Conference on Software Maintenance pp 661 – 66, 2005.
- [13] Dennis Quan, et al, May 2004," How to Make a Semantic Web Browser", World Wide Web Conference New York, USA, pp. 2-3 2004.
- [14] Plug-in en.wikipedia.org/wiki/Plug-in may 05,2011.
- [15] Ovsianikov, I.A., Arbib, M.A., and McNeill, T.H., "Annotation Technology". International Journal of Human-Computer Studies, 1999. 50(4): p. 329-362.