A Systematic Analysis of Semantic Web Search based on Ontology Modeling and It's Search Engines

1. Ms. Soniya Sharma, M tech Scholar (IV sem), Dept. of CSE, Greater Noida Institute of Technology, Greater Noida (UP), India
2. Mr. Arun Mittal , Asst. Professor, Dept. of CSE , Greater Noida Institute of Technology, Greater Noida (UP), India

Abstract: Semantic Web search is currently one of the hottest research topics in both Web search and the Semantic Web. The Semantic Web is an extension of the current web in which information is given well-defined meaning. Semantic web technologies are playing a crucial role in enhancing traditional web search, as it is working to create machine readable data. But it will not replace traditional search engine. The Semantic Web works on the existing Web which presents the meaning of information as well-defined vocabularies understood by the people. Semantic Search, at the same time, works on improving the accuracy of a search by understanding the intent of the search and providing contextually relevant results. The World Wide Web data is growing rapidly in the data repositories because of various factors such as users, systems, sensors and applications. For example, millions of transactions that occur daily, and the social media tools such as Facebook, Twitter, LinkedIn, Google+, and Tumblr, add vast of information. These large data create several challenges that called V attributes : Velocity, Volume and Variety. Clearly, the velocity means the data comes at high speed, while volume focus on large and growing files and the variety means the files come in various formats (e.g. text, sound and video). These issues enable a competition among the developers to search about a technique that help to extract the accurate data and overcome the current problems in order to reach a semantic search.The paper describes a semantic approach towards web search and semantic web search based on ontology and search engine. The ontology, created from the browsing history, was then parsed for the entered search query and the corresponding results were returned to the user providing a semantically organized and relevant output.

Keywords: Semantic Web, Semantic Search, Ontology, RDF, OWL, Search Engine

I. INTRODUCATION

Semantic web was introduced by Tim Berner lee in 2000.Semantic web is an extension of the current web in which information is given well defined meaning, enabling computers and people to work in cooperation efficiently. It requires strong language and set of rules that can express the data and link the data, reason for the data in a most efficient way. It was discovered because web 2.0 was not machine understandable. World Wide Web is universal that means "anything can link to anything". Over the past years, the ever-increasing growth in the World Wide Web has brought into light the insufficiency of the currently existing techniques used for searching information on the web [2]. For a given query entered by the user, search for a target web page in most search engines is based on keyword-based searches and popularity based ranking. Although the results might be good enough, not all the search results turn out to be relevant to the given query. Web has developed for people but not for automatic processing by machines. It was not context aware or adaptive. There was no formal semantics of the data. Contents were machine readable but not machine understandable. Web 2.0 is like a book with multiple hyperlinked documents. Index of the keywords is present but the contexts in which those keywords used are missing. To check which one is relevant we have to read the corresponding pages. In Semantic Web this limitation is eliminated via ontologies where data is given with well-defined meanings, which can be understandable by machines. Traditional search engines are unable to reach and give accurate precise results due to their lack of context aware. Semantic web on the other hand is adaptive and context aware. [11].

According to Merriam-Webster, semantic means "of or related to meaning". The Semantic Web is, necessarily, a vision for the future of the World Wide Web where the available information is given a *meaning*, providing *logical connections of terms* and making it easier of the machine to integrate data and process the information available on the Web [8]. They play a key role in the vision of the Semantic Web where they provide the semantic vocabulary used to annotate websites in a way meaningful for machine interpretation. As studied in the context of information systems, ontologies borrow from the fields of symbolic knowledge representation in Artificial Intelligence, from formal logic and automated reasoning and from conceptual modeling in Software Engineering, while also building on Web-enabling features and standards.Semantics is the study of meaning. For the search query entered by the user, a Semantic Web Search ensures contextually relevant results by understanding *intent* and *meaning* of the query provided.

With the need for Semantic Web, W3C defined the first Resource Description Framework (RDF) specification for semantic interoperability in 1997. RDF required triple-based representations for Universal Resource Identifiers (URIs). Expression of structured vocabularies was then introduced in RDF Schema (RDFS). Web Ontology Language (OWL) provided greater expressivity in the objects and relations of the RDFS [4]. These ontologies provided a strong semantic structure to the data.

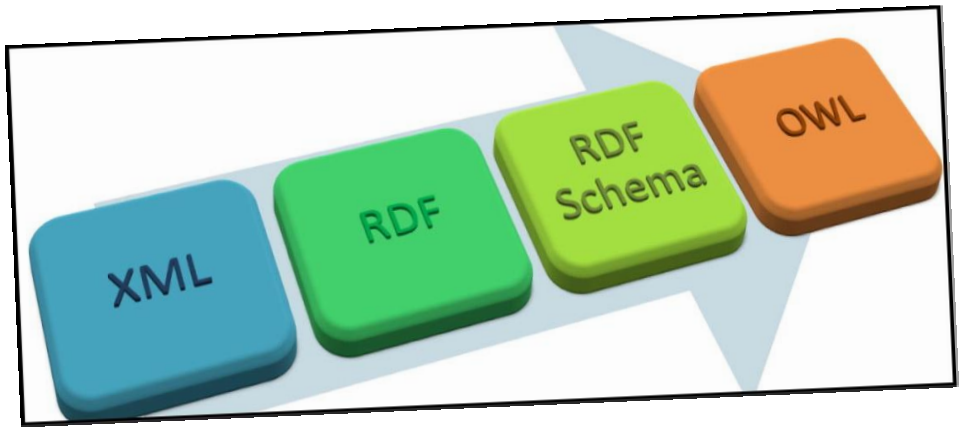

Figure 1. Hierarchy of Semantic Search

In a semantic, the data is stored in different levels as it is illustrated in Fig. 1, the hierarchy of layers to reach a proposed semantic search. It start from XML (Extensible Markup Language), RDF, RDFs (RDF Schema) and OWL (Ontology). Each concept is a complementary for the next and the last two concepts are the crucial to get semantic search. While the RDFS [3], [4], [5] suffers from many weaknesses, that leads to create a movement and extend it to the Ontology upper layer. For instance, RDFS has a weakness to describe resources in sufficient details because there is no localized range and domain constraints. In addition, it is difficult to provide reasoning support and has no existence/cardinality constraints and no transitive, inverse or symmetrical properties.Ontology gets over from the issues of RDFs that makes this concept the nearest one to the semantic search. Actually, the term Ontology has been used for several years ago by the artificial intelligence and knowledge representation community.

II. BACKGROUND

The idea of search engine and info retrieval from search engine is not a new concept. The interesting thing about traditional search engine is that different search engine will provide different result for the same query. While information was available in web, we have some fields of problem in search engines. Information retrieval by searching information on the web is not a fresh idea but has different challenges when it is compared to general information retrieval. Different search engines return different search results due to the variation in indexing and search process. Google, Yahoo, and Bing have been out there which handles the queries after processing the keywords. They only search information given on the web page, recently, some research group's start delivering results from their semantics based search engines, and however most of them are in their initial stages. Till none of the search engines come to close indexing the entire web content, much less the entire Internet.

1. Many times this happened that the particular result is available on the web but due to not availability of intelligent retrieval system.

2. The another main program with search engine is result that contain information will scattered in different pages, so there is need of hyperlinking of these pages.

Semantic is the process of communicating enough meaning to result in an action. A sequence of symbols can be used to communicate meaning, and this communication can then affect behaviour. Semantics has been driving the next generation of the Web as the Semantic Web, where the focus is on the role of semantics for automated approaches to exploiting Web resources. 'Semantic' also indicates that the meaning of data on the web can be discovered not just by people, but also by computers. Then the Semantic Web was created to extend the web and make data easy to reuse everywhere.

III. SEMANTIC WEB LANGUAGE

The semantic web languages used in this application are RDF, RDFS and OWL [8]-[10]. The Resource Description Language is a general-purpose triple-based language used for representing information in the Web. The triples in RDF are represented as *subject-predicate-object*. RDF Schema is a semantic extension over RDF, providing vocabulary descriptions over the triples-based RDF. Web Ontology Language is used to make information available to be processed by applications, where the meaning of each term and their inter-relationships are explicitly represented. The representation of the terms and the relationship between thee terms is called an *ontology*. Due to the greater expressiveness of OWL, this language has the ability to represent machine interpretable content on the Web [8].

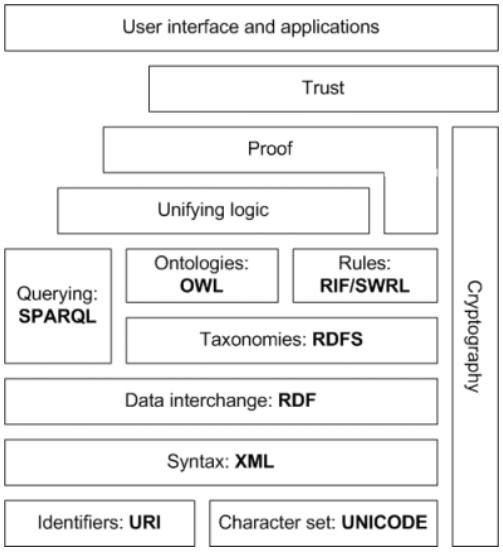

Figure2.Architecture of Semantic Web; Source-W3C

The layered tower of Semantic Web Languages shown in Fig. 2 is Semantic is the vehicle dreamed of to bring the Semantic Web to its full potential. The recognition of the importance of ontologies for the Semantic Web has led to the revolution and extension of the current web markup languages surveyed here. The various constructs in the OWL language, mostly used in the ontology created in the application, are the RDF Schema features (rdf:subClassOf, rdf:Property, etc.) for defining the classes, subclasses, properties, sub-properties, etc. , (In)Equality (differentFrom, distinctMembers, etc.) for specifying the inequalities between the various individuals, Property Restrictions (Restriction, onProperty, allValuesFrom, someValuesFrom, etc.) for defining restrictions on the individuals and Annotation Properties (rdfs:label, rdfs:comment, AnnotationProperty, etc.) for specifying the details for each individual [9].

IV. SCOPE OF SEMANTIC WEB

Semantic web is slowly gaining power and collaborating with other areas of research like bioinformatics, eCommerce, eGovernment and social web. It's most significant use is seen in the field of Bioinformatics A fast developing trend in biomedical network analysis is about combining multiple biomedical associated data, which can be highly heterogeneous into coherent bio-molecular interaction networks to enable integrated network analysis. This is possible due to progress in semantic web.

Applications like genomic ontologies, semantic web services, automated catalogue alignment, ontology matching, blogs and social networks are constantly increasing, driven by companies like Google, Amazon, YouTube, Face book and LinkedIn. The need for combining information in a meaningful way creates the potential and demand for research in Semantic web.

V. SEMANTIC SEARCH

Semantic search denotes search with meaning, as distinguished from lexical search where the search engine looks for literal matches of the query words or variants of them, without understanding the overall meaning of the query.[1] Semantic search seeks to improve search accuracy by understanding the searcher's intent and the contextual meaning of terms as they appear in the searchable dataspace, whether on the Web or within a closed system, to generate more relevant results. Semantic search systems consider various points including context of search, location, intent, variation of words, synonyms, generalized and specialized queries, concept matching and natural language queries to provide relevant search results.

The development of a new semantic search technology for the Web, called semantic search on the Web, is currently a very hot topic, both in Web-related companies and in academic research. There is a fastly growing number of commercial and academic semantic search engines for the Web. Connect the information on existing Web pages with background ontological knowledge. Mapping Web pages/objects to a knowledge base relative to an ontology; vertical vs. general search.Make current search engines more "semantic" / "intelligent"(adds meaning and structure to Web pages and queries).Semantic search on the Web on top of standard Web search: can immediately be applied to the existing Web (and not only to the future Semantic Web), and it can be done with existing Web search technology (and so does not require completely new technologies).

More complex search queries and more precise answers; reasoning over the contents of Web pages.

Examples

- A search for "president of the USA" should also return Web pages that contain "George W. Bush" (who was one of the presidents of the USA according to some background ontology).
- A search for "the president of the USA on September 11, 2001" should return Web pages mentioning "George W. Bush" (who was the president of the USA on September 11, 2001, according to some background ontology).
- When searching for Web pages about the first president of the USA, "Washington", semantic annotations and background knowledge allow us to restrict our search to Web pages that are actually about Washington as the name of the president, and so to ignore, e.g., Web pages about the state or town.

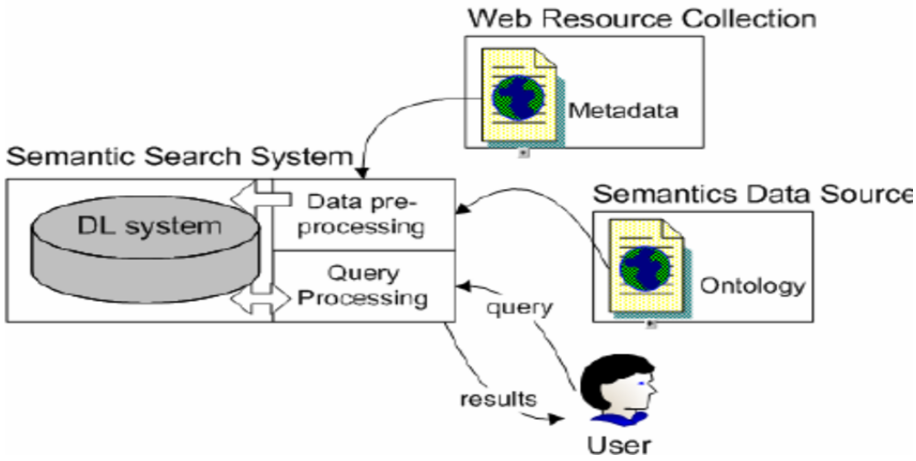

Figure3. Architecture of Semantic Search System

VI. SEMANTIC WEB SEARCH ENGINE

Currently many of semantic search engines are developed and implemented in different working environments, and these mechanisms can be put into use to realize present search engines.

Semantic Search Engines That Will Change the World of Search

1. HAKIA

Hakia is widely used semantic search engine that work like Wikipedia. Hakia calls itself a meaning based search engine. The main goal of these search engines is that they provide search results based on meaning match rather than by popularity of search query. The current news, blogs are processed by hakia's proprietary are semantic technology called QDEXing.
The ideal search engine would be able to match the search queries to the exact context and return results within that context. While Google, Yahoo and Live continue to hold sway in search, here are the engines that take a semantics (meaning) based approach, the end result being more relevant search results which are based on the semantics and meaning of the query, and not dependent upon preset keyword groupings or inbound link measurement algorithms, which make the more traditional search engines easier to game, thus including more spam oriented results.

2. EXALEAD

The image search engine was unique for its host of options to narrow down search based on image size, color, and content. Many of these features have since appeared across other image search engines. Exalead is a must try for image search. The company has been focusing on the enterprise search market, essentially attempting to solve the problem of search for content where link analysis is of little help.

3. SenseBot

The technology powering this engine creates a summary of the top results that are returned for a user query, often negating the need to drill down into the URLs to get the information that one is seeking. Semantic Engines LLC, the company behind the engine provides a variety of products around this technology.
There is Link Sensor, a tool that can be used on major blogging platforms (WordPress, Blogger, etc.)for automatically picking up key concepts from the post and linking them to related articles from the same blog or publisher. It is possible to point to other venues as well, e.g. to another blog from the same publisher – perhaps with a higher CPM. The tool increases user engagement. The company has also started providing APIs for returning summaries of results for a query from a set of URLs that are also passed in as parameters to the APIs. This is one interesting approach that helps save time when an exact answer is what one is looking for.

4. COGNITION SEARCH

The Cognition Search NLP Product is a solution companies can use to extract relevant results from their content. The application of this technology could range from better search across the enterprise to fetching more relevant ads. The company provides APIs for access to these technologies. I could not locate the free search, but definitely with the showdown featured on GigaOM, the product has its utility. And it also provides a definite business model.

5. SWOOGLE

Swoogle is strictly for the semantic web. The engine indexes documents developed on the concepts and standards for semantics (such as the RDF Format). The engine performs crawling of semantic documents like most web search engines and the search is available as web service too. The engine is primarily written in Java with the PHP used for the front-end and MySQL for database. Swoogle is capable of searching over 10,000 ontologies and indexes more that 1.3 million web documents. It also computes the importance of a Semantic Web document. The techniques used for indexing are the more google-type page ranking and also mining the documents for inter-relationships that are the basis for the semantic web.

Figure 4. Swoogle Semantic search engine

VII. ONTOLOGY

Currently, Ontology is becoming very important because we have a lack of standards (shared knowledge) which are rich in semantics that represented in machine understandable form. Moreover, it has been proposed as a solution for the problems that arise from using different terminology to refer to the same concept or using the same term to refer to different concepts [6]. Ontology is built to develop the required conceptualizations and knowledge representation in order to meet various challenges**.**

Ontology is considered as a portal to make the engines more intelligent and powerful. It is a respectful mission for the current generation of the web which known as Web 3.0 and the future mission for Web 4.0. Ontology is powerful and has a correct and reliable data that stores in its repositories that called the ontological graphs. It enables user to get and retrieve a direct answer without any complexities.

Why develop an Ontology?

- To enable a machine to use the knowledge in some application.
- To enable multiple machines to share their knowledge.
- To help yourself understand some area of knowledge better.
- To help other people understand some area of knowledge.
- To help people reach a consensus in their understanding of some area of knowledge.

Table 1. Ontological search engines

Search engine	Specialty	Repository	Search approaches	Results	Voice recognition	Portability	Language support
Kngine	Knowledge Engine	Wikipedia and other sites	Knowledge-Based approach and the statistical approach	Direct answer or link to web pages	Yes	Yes	Multi-language (supports Arabic)
Google	Search Engine	Wikipedia	Hummingbird approach	Direct answer	Yes	Yes	Multi-language (support Arabic)
Wolfram| Alpha	Computational Knowledge Engine	Curated data of other sites	It is own computational approaches	Direct computational Answer	Yes	Yes	Multi-language (doesn't support Arabic)

Currently, there are some engines that based on the concept of semantic such as Kngine [10], Wolfram Alpha [10] and the most popular engine nowadays Google.

1. Kngine

Kngine is the first multi-language question answering engine which supports around four languages and English, Arabic with them. Kngine stands for Knowledge Engine that is Web 3.0 Knowledge Engine. It is designed to provide customized and exact meaningful search results. For instance, semantic information about the keywords, user's queries, list things, find out the relations between the keywords. The exciting characteristics of this search engine, it gives precise results which links different kinds of related information together to present them to the user such as: movies, photos, and prices and the users reviews.

2. Wolfram Alpha

Wolfram Alpha is a computational knowledge engine or answer engine which developed by Wolfram Research. It is an online website that answers factual queries directly by computing the answer from externally sourced "curated data" or structured data, rather than providing a list of documents or web pages.

VIII. PROPOSED SYSTEM

The research had been carried out in three modules:

- Web Content Extraction (Module 1)
- Semantic Knowledge Base (Module 2)
- Reasoner (Module 3)

The architecture of the system, in accordance to the modules, is shown below.

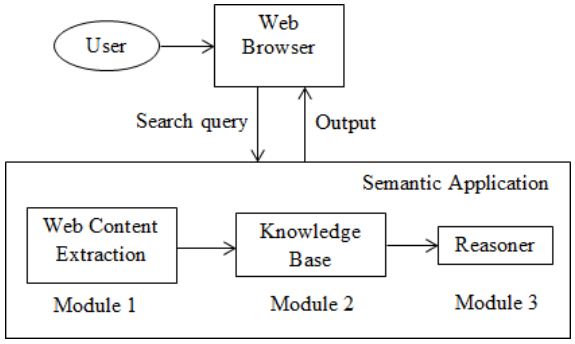

Figure 5 Architecture of the system search based on ontology

Here, the user enters the search query into the semantic application. The semantic application in itself consists of the three modules, wherein, the browsing history of the user was contained in module 1, the ontology created in accordance to the browsing history was contained in module 2 and module 3 parsed through the created ontology to return the result. This result was then displayed to user through the semantic application on the browser.

IX. RESULT AND FURTHER WORK

The application, hence, aides the search of the user by providing a list of useful and relevant web pages earlier visited by him, in accordance to the query, thereby providing him with results that would be relevant and helpful. In this paper we implemented proposed system with intelligent semantic search engine.

Figure 6. Homepage of Proposed System [GUI]

The major area of further work in the system would be to make the application work for web browsers . Accessing databases of Google Chrome, Safari, Internet Explorer, etc. and storing them in the MySQL database along with the existing database would make the application work for the other web browsers as well. To make the application more accessible to the user, the system can be made available as a plugin to be installed into the web browser, providing semantically organized results to the user simultaneously with the current searches being made.

Also, further work can be done in increasing the knowledge base by providing automation in the process of adding information to the ontology.

X. CONCLUSION

In this paper, we make a brief survey of some the semantic web search and search engine that uses various methods to search experience for users. With the increasing amount of data being stored on the Web every day, Semantic Search would ensure the provision of useful and relevant information to the user. This application works on making the procedure more user-friendly by providing contextually relevant visited pages to the user along with the searches available to the user otherwise through the various search engines. Understanding the user's query and providing just the content required is the objective of the application.

REFERENCES

[1] G.Sudeepthi1, G. Anuradha, Prof. M.Surendra Prasad Babu, "A Survey on Semantic Web Search Engine", International Journal of Computer Science Issues, Vol. 9, Issue 2, No 1, March 2012.

[2] www.searchenginejournal.com

[3] A.H. Doan, J Madhavan, Pedro Domingos and Alon Halevy, "Learning to map between Ontologies on the Semantic Web," in *Proc. WWW'02*, 2002.

[4] Doms A. and Schroeder M., "GoPubMed: exploring PubMed with Gene Ontology," *Nucleic Acids Res.*, 33(Web Server issue):W783-6, Jul. 2005.

[5] www.google.com

[6] G. Madhu, Dr. A. Govardhan and Dr. T. V. Rajinikanth, "Intelligent Semantic Web Search Engines: A Brief Survey," in *Proc. ICIET'10,* 2010.

[7] Dinesh Jagtap, Nilesh Argade, Shivaji Date, Sainath Hole, Mahendra Salunke, " Implementation of Intelligent Semantic Web Search Engine" International Journal of Engineering & Technology, vol. 4 issue 4, April 2015.

[8] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, R. Scott Cost, Yun Peng, Pavan Reddivari, Vishal Doshi and Joel Sachs, "Swoogle: A Search and Metadata Engine for the Semantic Web," in *Proc. CIKM'04,* 2004.

[9] Maedche A. and Staab S., "Ontology Learning for the Semantic Web," *IEEE Intelligent Systems*, vol. 16 issue 2, pp 72-79, Mar.-Apr. 2001.

[10] OWL Web Ontology Language Reference: *http://www.w3.org/TR/owl-ref/*

[11] Sampson D. G., Lytras M. D., Wagner G. and Diaz P., "Ontologies and the Semantic Web for E-learning", Educational Technology and Society, vol 7 issue 4, pp 26-28, 2004.

[12] Steffen Staab, "Semantic Web – Multimedia Ontology," in *ISWeb'01*, 2001.

ABOUT AUTHOR

SONIYA SHARMA received the B.C.A degree from chaudhary charan singh university, Meerut, India, in 2015 and the M.C.A from HIMT College (Greater Noida) which is affiliated to Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India, in 2017. She is currently pursuing the M.Tech in Computer Science from GNIOT College (Greater Noida) which is affiliated to Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India. Her areas of interest include Semantic Web ,Soft Computing, Artificial Intelligence, Database technologies, .Net Programming, Operating Systems and Software Engineering.

ARUN MITTAL was born in Mathura on 29th March 1987. He is Professional Graduated in 2007 in Computer Science and Completed M.TECH in 2011.His Professional Experience included 1.6 years in Siemens Technologies as a System Engineer and 2.6 years in IVSIT, Mathura as a Lecturer and 4 year experience in AIET (IET group of Institutions),Alwar as Assistant Professor. Currently working in GNIOT group of institutions, Greater Noida. During his Professional Period He Published 3 National and 3 International Research Paper and 2 journals. His area of interest includes Microprocessor, computer Organization, Digital Logic Design and DBMS.