A Survey on Semantic Web Search Engine

G.Sudeepthi[1] , G. Anuradha [2] ,Prof. M.Surendra Prasad Babu [3]

[1] Department of CSE, CMRIT
Hyderabad, A.P, India

[2] Department of CSE, GMRIT
Rajam, Srikakulam dist, A.P. India

[3] Department of CS & SE, Andhra University, AU college of Engineering
Visakhapatnam, A.P, India

Abstract

The tremendous growth in the volume of data and with the terrific growth of number of web pages, traditional search engines now a days are not appropriate and not suitable anymore. Search engine is the most important tool to discover any information in World Wide Web. Semantic Search Engine is born of traditional search engine to overcome the above problem. 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. In this paper we made a brief survey on various promising features of some of the best semantic search engines developed so far and we have discussed the various approaches to semantic search. We have summarized the techniques, advantages of some important semantic web search engines that are developed so far.The most prominent part is that how the semantic search engines differ from the traditional searches and their results are shown by giving a sample query as input.

Keywords: *RDF,OIL, DAML+OIL, OWL, QDEX.*

1. Introduction

The rapid growth of Internet has given user an easy way of accessing information and services. Web is a huge semi structured database that provides with vast amount of information. With ever-increasing information overload, we are facing new challenges for not only locating relevant information precisely but also accessing variety of information from different resources automatically. Efficient searching is required to get high quality results which are based on pertinent matching between well-defined resources and user queries. When users use search engines to search for specific information, the quality of the search results will be improved significantly if they make use of advanced Techniques. Most of the traditional search engines get the answers syntactically correct but larger in amount. The Semantic allows the information to be precisely described in terms of well defined vocabularies. Semantic Web is gaining momentum. A semantic search engine gives selected results which the user is searching for. The main objective of Semantic Web is to make Web content understandable not only by humans, but also machine understandable. We need to ensure that semantics are not lost during the whole life cycle of information retrival.Various semantic search engines developed so far differ from each other through the results obtained & technologies involved which can be discussed in detailed in later sections.

2. Traditional search engine & its limitations

Present World Wide Web is the global database that lacks the existence of a semantic structure and hence it becomes difficult for the machine to understand the information provided by the user in the form of search strings. As for results, the search engines return the ambiguous or partially ambiguous result data set. Semantic web is being developed to overcome the following main limitations of the current Web.

2.1 Limitations

1. The web content lacks a proper structure regarding the representation of information.
2. Ambiguity of information resulting poor interconnection of information.
3. Unable to deal with enormous number of users and content ensuring trust at all levels.

4. Incapability of machines to understand the provided information due to lack of a universal format.
5. Automatic information transfer is lacking.

3. Semantic web search engine

The semantic search highly improves search accuracy of the query related data and the search engine delivers the exact content, the user intent to know. There's no denying the power and popularity of the Google search engine. But there are other ways to search the web, using semantic search engines. By using semantic search engine we will ensure that it results in more relevant and smart results. The search engines are able to compare or extract the data and gives very relevant results for the queries. The semantic web aims to provide an extra-machine understandable layer, which will considerably simplify programming and maintenance effort for knowledge-based web service[1].Many semantic web languages have been developed like RDF,OIL, DAML+OIL and OWL.

3.1 Approaches to semantic web

There are four approaches to semantic search. Different semantic search engines may use one or more of these approaches. The point of semantic search is to use meaning to improve the user's search experience. For example, first approach uses contextual analysis to help to disambiguate queries. for example, the word "strike," refers to baseball or labor or something else entirely.

Second approach focuses on reasoning. Given set of facts that are represented in the system, additional facts can be inferred from them. If the system knows who is Bach's children were, and it knows who each of their children were, then a reasoning system can infer who is Bach's grand children were.

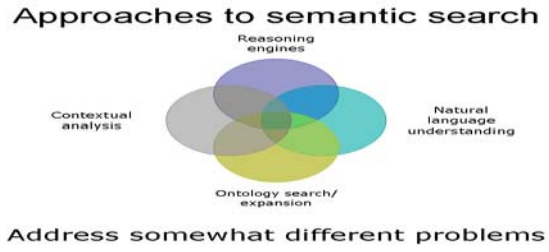

Fig 1: Approaches to semantic web

Third approach emphasize on natural language understanding. These engines process the content they index and the queries people submit try to identify the intent of the information. They use the syntax of the sentence and rules to identify people, places, organizations, and so forth. Powerset makes extensive use of natural language understanding.

The fourth approach uses ontology to represent knowledge about a domain and expand queries. In this approach, when a user enters a query for a word like "truck," the system adds terms from its ontology (e.g., "vehicle" because a truck is a kind of vehicle) to make the search more focused as well as more broad This approach is used by a large number of semantic search systems. There is not just one approach to semantic search. Most semantic search engines mix and match them in various ways to yield a unique search experience for their users. Each approach has much to contribute. Different kinds of search are intended to fulfill different kinds of functions. One size does not fit all. There is room for variety.

4. Results of some of the semantic search engines by giving sample query as input

Hakia

Hakia is a semantic search engine that brings relevant results based on concept match rather than keyword match or popularity ranking, but the engine prompts to enter not just keywords - but a question, a phrase, or a sentence. They provide search results based on meaning match, rather than by the popularity of search terms. A very important capability of hakia engine is that it gets results using equivalent terms such as "treat=cure" or "kill=murder". These associations are deployed with proper disambiguation methods. The search results are divided into Web, News, Blogs, video and can be re-listed according to relevance or date. Hakia's semantic search is essentially built around three evolving technologies. OntoSem (sense repository) it is a linguistic database where words are categorized into the various "senses" they convey.QDEX (Query indexing technique) which extracts all possible queries relating to the content[2]. Semantic Rank algorithm independently ranks content. It benefits user by offering quality results there by saving time. It Easily Identifies Information from Credible Sites.

Fig 2: Hakia

Kngine

In kngine search results are divided into either web results, or image results. They are preceded by information about the search term, known as 'Concepts.' for example, searching for the "iPhone 3GS" will be preceded by the device's specifications. Searching for a film will be preceded by information about the film, links to trailers, reviews and quotes. Searching for a city will be preceded by information about the city, local attractions, events, weather and hotels. Kngine currently contains more than eight million Concepts this is where the site's strength lies.

Fig 3: Kngine

Kosmix

Kosmix lies at the intersection of two important trends topic exploration and Deep Web[3].It is the first general-purpose topic exploration, information rich and elaborative search engine. It uses semantics in an attempt to link data from all over the web, giving relevant search results. Search results themselves, they are divided into Video, Web, News & Blogs, Images, Forums, Twitter, Amazon and Facebook.

Powerset

The Microsoft-acquired search engine Powerset focuses on doing only one thing and doing it really well by using natural language processing to understand the nature of the question and return pages containing the answer. All search results on Powerset come from Wikipedia, making it the ultimate way to search Wikipedia, using semantics Search terms can be formulated as questions, which will be answered, or as simple terms, and results will be aggregated from all the relevant pages on Wikipedia. It helps to give comprehensive view of the thing we search for. It aggregates the information provided by the different resources. It provides a set of suggestions about the query given and also the related queries.

DuckDuckGo

It is a feature-rich semantic search engine which gives countless reasons to leave Google. If we search for a term that has more than one meaning, it will give you the chance to choose what you were originally looking for, with its disambiguation results. For example, searching for the term Apple will give you a long list of possible meanings including fruit, computer company, bank.

Fig 4: Duck Duck go

Sensebot

Sensebot uses text mining to parse Web pages and identify their key semantic concepts. It then performs multidocument summarization of content to produce a coherent summary it gives a summarized accurate search result according to the query given. The summary gives a good idea of the topic of the query. The summary is readable and coherent. SenseBot will save time by providing an overview of the topic, and pointing to the right sources. The search engine itself tries to understand the concept of the query, actually what it contains and gives an appropriate result. The user need not go through many web pages to get the results.

Fig 5: SenseBot

6. Summary of various semantic web search engines

By going through the literature analysis of some of the important semantic web search engines, it is concluded that each search engine has some relative strengths. A summary is given in the below which summarizes the techniques, advantages of some of the important semantic web search engines that are developed so far.

XSEarch involves a simple query language, suitable for a naïve user. It returns semantically related document fragments that satisfy the user's query. Query answers ranked using extended information-retrieval techniques and are generated in an order similar to the ranking. Advanced indexing techniques were developed to facilitate efficient implementation of XSEarch[8].The performance of the different techniques as well.

XCDSearch is a context-driven search engine. It uses keyword-based queries as well as loosely structured queries, using a stack-based sort-merge algorithm. It employs Object-Oriented techniques for answering queries. The keyword query is answered by returning a sub graph that satisfies the search keywords[5] It builds the relations between data elements based solely on their labels and proximity to one another, while overlooking the contexts of the elements, which may lead to erroneous results. It employs stack-based sort-merge algorithm employing context driven search techniques for determining the relationships between the different unified entities.

Hakia is a general purpose semantic search engine that search structured text like Wikipedia. Hakia calls itself a "meaning-based (semantic) search engine". They provide search results based on meaning match, rather than by the popularity of search terms. The presented news, Blogs, Credible, and galleries are processed by hakia's proprietary core semantic technology called QDEXing.

Swoogle is a crawler-based indexing and retrieval system for the Semantic Web documents. It analyzes the documents it discovered to compute useful metadata properties and relationships between them. The documents are also indexed by using an information retrieval system which can use either character N-Gram or Uri's as terms to find documents matching a user's query or to compute the similarity among a set of documents. One of the interesting properties computed for each Semantic Web document is its rank -a measure of the document's importance on the Semantic Web

Kosmix is the first general purpose topic exploration engine to harness the Deep Web using a federated search Approach.Kosmix lies at the intersection of two important trends topic exploration and the Deep Web. Topic exploration is a new approach to information discovery on the web that satisfies certain use cases not served well by conventional web search. The Deep Web, an inhospitable region for web crawlers, is emerging as a significant information resource. Focus on the Kosmix approach to query transformations and caching, which is essential to ensure reasonable performance.

Falcons Concept Search, a novel keyword-based ontology search engine is developed to facilitate concept sharing and ontology reusing[6]. It integrates concept level search and ontology-level search by recommending ontology and allowing filtering concepts with ontology Through this users can quickly find ontology that satisfy their needs and present several supportive techniques including a new method of constructing virtual documents of concepts for keyword search, a popularity-based scheme to rank concepts and ontology.

Watson is a gateway for the Semantic Web, which has been guided by the requirements of Semantic Web applications and by lessons learnt from previous systems. It uses Ontology crawling exploration technique[7]. It provides explicit and implicit relations between ontology, providing rich, semantic access to data, focusing on semantic quality. It exploits the strengths of semantic technologies to provide fundamental functionalities for a more suitable access to online knowledge.

7. Conclusion

This paper gives a brief overview of some of the best semantic search engines that uses various approaches in different ways to yield unique search experience for users. It is concluded that searching the internet today is a challenge and it is estimated that nearly half of the complex questions go unanswered[4].Semantic search has the power to enhance the traditional web search.Wheather a search engine can meet all these criteria continues to remain a question. Future enhancements include developing an efficient semantic web search engine technology that should meet the challenges efficiently and compatibility with global standards of web technology.

References

[1] Lavanya Rajendran, "A comparative study on internet application development tools", International Journal of Engineering Science and Technology, Vol. 2(10), 2010, ISSN: 0975-5462.
[2] "Hakia" Semantic search technology, January 2010.
[3] Anand Rajaraman ," Kosmix: High Performance Topic Exploration using the Deep Web",ACM,VLDB'09, August 24-28,2009, Lyon, France.
[4] "Searching the Internet today is a challenge, and it is estimated that nearly half of customers' complex questions go unanswered."Bill Gates, CEO, Microsoft. (Source: Bill Gates officially launches MSN Search on MSN.com's homepage, February 2005)
[5] Kamal Taha,"XCDSearch: an XML Context Driven Search engine"IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, vil.22, No.12, DECEMBER 2010
[6] Yazhong qu and Gong Chen, Falcons Concepts Search: A practical search engine for web ontologies, IEEE Transactions on systems and Cybernetics.
[7] Mathieu d'Aquin, Marta Sabou, Martin Dzbor, Claudio Baldassarre, Laurian, "WATSON: A Gateway for the Semantic Web".
[8] Sara Cohen Jonathan Mamou Yaron Kanza Yehoshua sagiv, "XSEarch: A Semantic Search Engine
For XML".

G.Sudeepthi did her M.Tech (CSE) from Koneru Lakshmaiah College of Engineering, Andhra Pradesh, India. She has got one paper published in International journal. She is pursuing her Ph.D from Andhra University,visakhapatnam. Presently She is working as Assistant Professor in CSE department at CMR institute of technology,Hyderabad,AP,India.Her research area include semantic web on Intelligent systems.

G.Anuradha did her M.Tech from GITAM college of engineering, Visakhapatnam, Andhra Pradesh. She is working as an Associate professor in the department of CSE in GMRIT, Rajam, and Srikakulam district. She is pursuing Ph.D. in Andhra University on semantic web and intelligent systems. She has got one paper published in International journal, and more than five paper published in various National & international conferences.

Prof. M.Surendra Prasad Babu did his M.Phil and Ph.D. degrees from Andhra University. He is presently working as Professor in Dept. of Computer science & systems engineering in Andhra University college of Engineering in Visakhapatnam. He has got more than 30 years of teaching experience. He has more than 120 papers published in various National & International journals and conferences. His research areas include Artificial Intelligence & Expert Systems Machine Learning, Computer Networks, Neural Networks, and Object Oriented Technologies, web Technologies. Pervasive computing and similarity transformations.