

Comparative Study Of Semantic Search Engines

Vidhi Shah¹, Akshat Shah² and Asst. Prof. Khushali Deulkar³

¹Department of Computer Engineering, Dwarkadas J Sanghvi College of Engineering, Mumbai, India
vidhi240194@gmail.com

²Department of Computer Engineering, Dwarkadas J Sanghvi College of Engineering, Mumbai, India
akshatshah1717@gmail.com

³Department of Computer Engineering, Dwarkadas J Sanghvi College of Engineering, Mumbai, India
khushali.deulkar@djsce.ac.in

Abstract- The amount of information accumulated in the internet is massive. The information is searched in the internet using a search engine. The current search engines searches for the needed information based on the keywords which the users have typed. These search and retrieval are based on syntactic analysis of keyword instead of contextual analysis. In order to overcome this issue, the need of Semantic Search Engines is increasing. The Semantic Web is an extension of the current web in which Information retrieval is based on the contextual analysis of user search query which is a more meaningful search. In this paper we identify the different approaches and techniques used in different search engines as well as the analysis and comparison of various semantic web search engines based on various parameter to find out their features and limitations. Based on these analyses of different search engines, a comparative study is done to identify relative strengths in semantic web search engines.

Keywords: Semantic Search Engines, Semantic Web, Information Retrieval, Semantic search.

1. INTRODUCTION

With the continuous increase of information on the internet, the results provided by traditional search engines in response to user queries do no longer satisfy the needs of users[1,3]. The traditional search method provides answers that match the keywords contained in user queries. However, these answers are not precise and relevant enough according to the semantics of user queries. This problem has motivated a new era of search systems named as Semantic Search Engines.

[2] A Semantic Search Engine (SSE) is a semantic Web that can answer questions based on the meaning of users query, resources in the repositories and based on predefined domain semantics or a knowledge model. SSE can return relevant results on users topics based on its context and do not necessarily mention the word you searched for explicitly.

The goal of this paper is to study and discuss various semantic search engines, as well as identifying features and main approaches used in them and their limitations. It also provides a comparative analysis of different semantic search engines.

The rest of the paper is structured as follows: the semantic search engine is described in the second section the third section describes different semantic search engines, the comparative evaluation of various semantic search engines has been discussed in the fourth section. In the last section conclusion is stated.

2. SEMANTIC SEARCH ENGINE

A semantics search engine attempts to provide search results based on contextual search [2]. It automatically identifies the concepts structuring the texts. An important challenge of the semantic search is disambiguation of the queries as well as of

the content on the web. What this means is that the search engine through natural language processing will know whether you are looking for a fruit or a phone when you search for "apple" It finds the results based on the context of the user search query. Semantic search engines performs the following functionalities:

Interpretation of user query.

Extracting the relevant concepts from the sentence.

Building a user query using the predefined concepts.

Finally displaying the results to the user

The seven search engines below all use semantic analysis to sift through the present data and uses different techniques for semantic search.

SOME SEMANTIC SEARCH ENGINES

3.1 Hakia

Hakia, is a general purpose semantic search engine that search structured text. Hakia is a search engine designed to provide search results based on the meaning of their content rather than page's popularity. It concentrates on accurate results. It is build using three technologies [4]:

Ontosem is Hakia's repository which is a linguistic database in which words are categorized according to their different meanings.

QDEX is used for query detection and extracting inverted index. QDEX is the replacement for index which is used to

store the web content by more over all the engines. It obtains all possible problems relevant to the content of the page. Semantic rank algorithm used to rank contents on its own based on sentence analysis.

The information related to date can be obtained It is well organized by the tabs Web results, Credible sites, images and news. Credible site includes the sites which are vetted by the librarians and other information professionals. It produces resumes for some of the user queries. They are the portals which gives all information related to that subject. For each resumes there will be an index of links which refers to the corresponding pages, it helps for a quick reference. According to the query the content of the resume will vary. Resumes are one the most impressive feature of Hakia. Hakia will also provide the related queries, which help the user to reach the target or to get the thing very easy.

3.2 Swoogle

Swoogle [7] is a search engine for Semantic Web ontologies, documents and data on the internet. Swoogle uses crawlers to discover RDF and HTML documents with embedded RDF content. Swoogle is a crawler-based indexing and retrieval system for the Semantic Web documents i.e. RDF or OWL documents. It analyzes the documents 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 URIs as terms to find documents matching a user's query or to find the similarity among a set of documents. One of the interesting properties computed for each Semantic Web document is the measure of the document's importance on the Semantic Web i.e. rank. Swoogle is also a content based search engine that analyses, discovers, and indexes knowledge in the web. Swoogle uses web crawlers to discover RDF documents and then extracts metadata related to each discovered document. Then it computes the relationship between these documents. Semantic Web search engine has three main functions.

3.2.1 Finding appropriate ontologies

Finding appropriate ontologies is the main factor in this engine because if it does not return appropriate ontology then it creates new ontology. It then uses ontology rank algorithm to rank the returned ontologies.

3.2.2 Finding instant data

Swoogle collects unstructured appropriately by querying Semantic Web Data based on the classes and properties defined by them

3.2.3 Collecting metadata

One of the important characteristics of semantic web engine is collecting metadata such as the inter-document relations.

The main features of the Swoogle are as follows [6]:

- It efficiently queries a comprehensive database of SWD metadata
- It enables "distributed" knowledge sharing
- It makes knowledge visible and easy to access
- It finds appropriate ontologies

- It finds instance data
- It Characterizes the Semantic Web

3.3 Lexxe

Lexxe [8] is a Natural Language Processing based semantic search engine. It gives the results on the basis of meaning .It provides the documents most relevant in content to the user's query. Lexxe search engine uses clustering to provide results with subset cluster which helps the engine select the most relevant links to the context. In this engine most of the answers to the queries are taken from the unstructured texts floating on the internet. It uses a phrase recognition technique to extract relevant factual information for the user .It also helps the engine to understand whether the key words are formed as one or more phrases. Lexxe also has features such as Word Sense Disambiguation, Part-of-speech Tagging and Parsing.

3.4 Factbites

Factbites [10] is the semantic web search engine that is based on contextual search. It has the following features which makes it useful:

3.4.1 It focuses on beyond keyword matching

Factbites focuses on meaningful, relevant and accurate results rather than keyword matching.

3.4.2 It provides an informative search

It is a cross between an encyclopedia and a search engine. Factbites provides searchers with meaningful sentences from every site in the search results rather than links.

3.4.3 Filters spams

Factbites filters out spams efficiently. Since it analyses the meaning of the page that has been retrieved it filters the spam sites that lacks relevant information.

3.5 DuckDuckGo

DuckDuckGo (DDG) [11] is a privacy-focused Internet search engine. It focuses on maintaining user privacy and therefore does not store any personalized search results. It emphasizes on getting information from the best and the most relevant sources instead of the number of sources. It does not store IP addresses, does not log user information and uses cookies only when needed to maintain the privacy of the user. DuckDuckGo does not record or share any personal information. Along with the indexed search results, DuckDuckGo also displays relevant results which are called Instant Answers displayed in grey boxes above the results on top of the search page. These Instant Answers are collected from either 3rd party APIs or static data sources like text files. The Instant Answers are called "zeroclickinfo" because the aim behind these is to provide the summaries of the topic as well as the related topics on the search result page itself by just a single click so that the user doesn't have to click any results to find what they are looking for. This is one of the distinct features in DuckDuckGo.

3.6 Sensebot

SenseBot [12] is a semantic search engine that has a unique feature of summarizing the top results for a query. It uses Natural Language processing. Sensebot tries to understand the

concept of the query and gives an appropriate result. Sensebot prepares the text summary according to the user's search query. It identifies key semantic concepts by using text mining algorithms that parse the Web Pages. The retrieved multiple documents are then used to perform a coherent summary and accurate search results. This coherent summary becomes the final result for user's query. The summary gives a good overview of the topic of the query. Hence, SenseBot will save time by providing an overview of the topic, and providing information from the right sources. The user need not go through many web pages to get the appropriate results [13]. Generally, a user has to visit various links returned by the normal search engine to acquire the required information but in SenseBot provides a summary of the user query is presented at the top of the search result.

3.7 Kngine

Knowledge Engine abbreviated as Kngine [14], is a hybrid implementation of Knowledge-Based approach and statistical approach. It is a web 3.0 semantic search engine and focuses on providing high degree of relevance to the user's search query. It analyzes the keywords from the user's query to form a relationship between them, links the different kind of data together and then the results are returned. The widely available unstructured data on the internet is collected and organized by Kngine's Extraction Engine by crawling the web. With the help of natural language processing, machine language and data mining information relations among each keyword is learnt. Knowledge graph is made up on this engine. Kngine is a question answering engine. It understands the question and proceed across the knowledge graph to solve the answer with the help of natural language processing and machine learning.

COMPARISON OF VARIOUS SEMANTIC SEARCH ENGINES

In this part, we have delineated the various semantic search engines with its features and limitations and based on different approaches and techniques we have compared the above semantic web search engines. Various techniques, approaches and pros and cons of each the above mentioned semantic engines are shown in the table I.

Table 1: Comparison of various semantic search engines

Sr. No	Semantic Web Engine	Approach	Technique	Type of Result	Features	Limitations
1	Hakia[4,5]	Related searches, NLP	1.OntoSem(sense repository) 2.QDEX(Query indexing method) 3.Semantic Rank algorithm	Link & Free text	1.Excellent resumes 2.Easily Identifies Information from Credible Sites Searches the data with Focus 3.Saves Time	It does not index everything .It needs other search engines too.
2	Swoogle[6]	Content based	A crawler based Indexing semantic search 2004 engine that searches ontologies and instance data	OWL, RDF	1.Finds appropriate ontologies . 2.Finds appropriate instance data structures of the semantic web.	Extending Swoogle to index and effectively query large amounts of instance data is still a challenge.
3	Lexxe[8]	NLP	It uses Semantic key technology which enable users to query with a conceptual keyword.	Selects link from Subset cluster	1.Part-of-speech tagging 2.Parsing 3.Word sense disambiguation	It does not work well with long queries
4	Factbites[9,10]	Contextual search	Ai and computational linguistics with Rapid intelligence	Summary	1.It has the ability to Filters out spam websites in the search results 2.It searches based on	It works better with general questions rather than specified topics

					the topic rather than the keywords	
5	Duckduckgo[1]	Clustered search, NLP	Instant Answers are collected from either 3rd party APIs or static data sources like text files	Summary	1.Zero click info 2.Emphasizes privacy and does not record user information. 2.Produces result based on many sources and its own web crawler.	It lacks feature for image and video searching
6	SenseBot[12,13]	Concept Search	Identifies key semantic concept from user's query by using text mining algorithm that parse the Web Pages which are then used to perform coherent summary.	Summary	Multi-document summarization	It only works with Firefox as a browser extension and Google search engine to display results.
7	Kngine [14]	Knowledge -Based approach and statistical approach	Knowledge base determines synonyms, relations between concepts, meaning document analysis and context-based fuzzy search.	The answer is placed in the first line with image and information	1.It displays search results in the form of images 2.It is multi-lingual 3.Kngine has the ability to allow the users to search in Parallel manner.	It does not provide silent mode option.

4 CONCLUSION

It is concluded that searching the internet today is a challenge and Semantic search engine enhances the traditional web search technique. This paper gives a brief overview of some of the best semantic search engines that uses various approaches indifferent ways to yield unique search experience for users and provides the comparative study of the same. 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] G. Madhu and Dr. A. Govardhan, Dr.T.V.Rajinikanth, "Intelligent Semantic Web Search Engines: A Brief Survey," International journal of Web Semantic Technology (IJWesT) Vol.2, No.1, January 2011.
- [2] Anusree.ramachandran, R.Sujatha," Semantic search engine: A survey",International Journal of Computer Technology and Applications(IJCTA),2011
- [3] [3] Faizan Shaikh, Usman A. Siddiqui, IramShahzadi, Syed I. Jami, Zubair A. sheikh,"SWISE: Semantic Web based Intelligent Search Engine," 978-1-4244-8003-6/10 IEEE,2010
- [4] [4] ArunRadhakrishnan, "Hakia's Semantic Search: The Answer to Poor Keyword Based Relevancy", Search Engine Journal,2007
- [5] [5] Jennifer Zaino "Hakia Semantic Search Engine Powers New StockIntelligence Service" <http://semanticweb.com/hakia-semantic-searchengine-powers-new-stock-intelligence-service>
- [6] [6] Li Ding, Tim Finin, Anupam Joshi, Rong Pan, R.Scott Cost, Yun Peng, PavanReddivari, Vishal Doshi, Joel Sachs, "Swoogle: A Semantic Web Search and Metadata Engine".
- [7] ArunRadhakrishnan, "Semantic Search Swoogle: An Engine for the Semantic Web" Search Engine Journal August 12, 2007.
- [8] ArunRadhakrishnan, "Semantic SearchLexxe : Search Engine that Answers Exact Queries", Search Engine Journal ,2007.
- [9] .Loren Baker," Factbites Search Engine Encyclopedia Hybrid "Search Engine Journal,2005
- [10] Rapid Intelligence "Introduction to Factbites" <http://www.factbites.com/>
- [11] Carlos Escalera,"Duckduckgo Search engine is actually good"<http://www.ohow.co/search-without-tracking/> , 2015
- [12] "The Search Engine that finds sense in a heap of Web pages"<http://www.sensebot.net/about.htm>
- [13] Arun Radhakrishnan,"Summarization, the Answer to Web Search: Interview with Dmitri Soubbotin of SenseBot" Search Engine Journal, 2007.
- [14] Why,What,HowKngine "We're Kngineers" <http://www.kngine.com/company.html>