Semantic Search's Impacts on Searching & Ranking of WebPages

[1]Ashwani Sethi, [2]Suneel Dubey
[1,2]Maharishi University of Information Technology, Lucknow, India

Abstract
Searching for information on the Internet is much easier with the help of Search Engine, but relevant search is another complex task. Traditional search compare the user query keywords with web pages. Without meaning of query search will continue and produces non-relevant results. User will search again with different keywords for relevant results. Semantic Search improves all the shortcoming of traditional search and provides a meaning to query and can handle a large query from users. This paper presents architectureof Semantic Search, how it impacts on web page searching and ranking as compare to traditional search.

Keywords
SERP, Knowledgebase, Term, WWW, URL.

I. Introduction

Internet is today's wildly used tool for information manager. Every information system directly or indirectly deals with internet or with its applications. Internet can be expressed by a tool which used its all resource and produce according to requirements for users. Search Engines are well represented tool of Internet, which use all the resources of Internet and presents to users' world wild. With all these, Search Engines are used to determine the intention of search. This view of Search Engine is used to extract the information about users and helps to determine the real meaning of user query. Many Search Engines used Semantic Search for determines the real meaning of query and to extend it for more relevant search. Semantic Search also called search with meanings. Traditional information retrieval systems are based on purely occurrence of keywords in a webpage [5]. These traditional information Retrieval Systems are not capable to handle double meaning or Semantic queries and produces non-relevant results to users. Third generation of web includes the machine learning concepts in Information Retrieval Systems [1]. Where machine its self is capable to handle all Semantic Queries.Semantic Searchintroduces a way in which computer can understand to human being exactly what they means for particular query.Semantic Search can understand as an exercise to discover knowledge. Semantic Search manages relation between both Human being and Computer with knowledge. Semantic Search uses knowledgebase or Ontology to describe query meaning and to extend query for further search.Keyword Based search uses as a traditional search to extract information from WWW. Keyword Based Search targets the highlighted contents of Web pages like Meta, title, alt, bold, italic, underline and URL. Quality measurement of web page depends on these contents. This approach of search wildly bypass by spammers to bypass the Search Engine rules.

Semantic Search uses knowledgebase to extract the real meaning of query. After extracted the meaning form query knowledgebase extends query with more related knowledge keywords. As fig. 1 show when user puts a query into Semantic Search Engine, Search Engine search query related knowledge into knowledgebase to extend the query. Finally, extended query goes to search. Semantic Search Engines can handle a paragraph query.

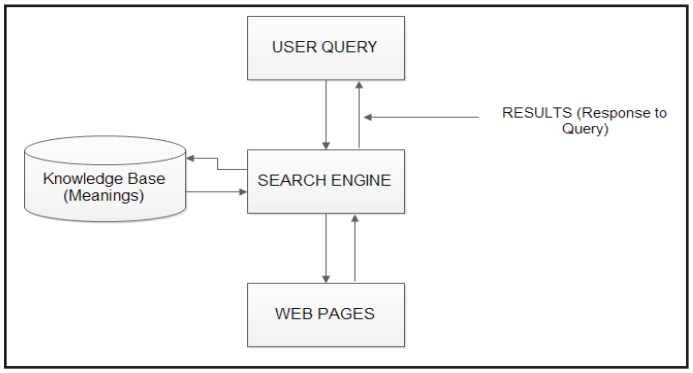

Fig. 1: Semantic Search Engine Parts

When user don't want to change their query and wants relevant results to current query the Semantic Search Engines are capable to do this.

II. Literature Survey

Junaidah Mohamed Kassim describes the Semantic search engine design and use as well as traditional search engine. The first generation of web 'web1.0' form 1990 – 2000 refers to internet at its emerging stage and produces a Producer – Consumer relationship. Web2.0 transforms the Web into a space that allows anyone to create and share information online. Web 3.0 shows more intelligence like the 'web machine' learns, suggests that what people like and would like to get. Semantic search integrates the technologies off Semantic Web and search engine to improve the search resulted gained by current search engines [1].

Yi Jin highlights the possibilities to improve a traditional search engine to create a semantic search engine. This paper describes the semantic web as well as the architecture of semantic search engine. The semantic web uses the tag notes to enrich the text documentation which machine can understand. These tag notes provides the meta-data of these texts, and will capture the meanings of these texts, which can be understand by machine. Meta-data are the fundamental building blocks of the semantic web. The researching purpose of Semantic Web is to describe the Web documents clearly. The foundation of the Semantic Web is the markup languages. The Architecture of Semantic web search engine focused on Retrieval and Reasoning driven processes. The system should be able to run some reasoning engine to identify the necessary facts and rules so as to achieve the desire conclusion. The semantic web tags can be used to get semantic notes. The major premise of Semantic notes is getting key words, terms and other elements from text documentation [2].

DuyguTumer analyzes the semantic search performance of search engines. This paper took three keyword based search engines Google, Yahoo, Msn and a Semantic search engine Hakia. Different queries of different topics analyze the performance of these search engines. Web search engines are computer programs which allow users to search their desired information from websites. The most popular search engines are Google, Yahoo, and Msn with 71.9, 71.7% and 4.2 volume of search ratio respectively. Hakia is

the publicly available semantic search engine. This paper has a table with ten different types of queries. These queries were run on the both keyword-based search engine as well as semantic search engine. Keywords were used to replace phrase. Beside the keywords phrase were used in Hakia for the main feature of Hakia Semantic search. Paper also highlights the concepts of relevant and non-relevant documents. A document which matches with the query keywords is called relevant and which don't called non-relevant. Google, Yahoo, MSN retrieved at least one relevant document for all queries [3].

Robin Sharma highlights the architecture of Semantic information retrieval to enhance the search result. And an algorithm is proposed for Semantic indexing of web pages. Semantic web search is the search of meaning for user's query. The meaning of query is hidden in the query itself. Words like What, Why, When affects the query meaning badly. In the traditional search these interrogational words removed or ignored from query and page ranking performed on the basis of main keywords. The search results in traditional search are not according to user satisfaction. This paper proposed a system that provides a result in order to relevance of keywords. The system is divided into three modules as: User Interface Module: User enter the query through user interface whatever wants to search and gets the corresponding results. Query Analysis: Query entered by the user will be interrupted by preprocessing [4].

Implementation

Web generation 3.0 extends the traditional search to Machine learning concepts where pre assigned knowledge used to refine query meaning. Semantic Search Knowledgebase provides knowledge keywords with specific pre assigned weight. This weight further helps to calculate the web page weight for quality measurements.

Table 1: Knowledge Base of Literature Related Keywords

Terms	Weight	Terms	Weight
Novel	8	Themes	1
Essay	9	News	3
Punjabi literature	3	Poems	4
English Literature	2	Poets	7
Hindi Literature	2	Literature books	6

When query execute for particular Semantic query, all knowledgebase keywords also searched with query. Semantic Search can perform on all web pages documents instead of particular highlighted contents. As table 1 shows the literature related knowledgebase where each term has pre assigned weight. If user search for 'poem' then all knowledgebase keywords will includes to query and search will continue with all keywords.

Algorithm for the Semantic Search:

The algorithm describing the sequence of computation steps:
Input Query
Output Ranked Results

Step 1: For each Keyword y in query
- Find equivalent keywords of key y in knowledgebase.
- Store all equivalent in to an ekeys table.

Step 2: For each equivalent x in an ekeys table
- Match x against web page keywords in database.
- Calculate the frequency of matched x keywords with in a webpage.
- Store keyword frequency in keyword2 table.

Step 3: Sort the web pages stored in keyword2 table from high to low frequency.

Step 4: Show ranked results of sorted web pages from high to low frequency.

Semantic Search algorithm works same as traditional search algorithms, but with the difference of knowledgebase, meaning of query and searching criteria extends. After searched all the web page related to query the weight of each page calculated as:

*Web page weight = (1term frequency * term weight) + (2^{nd} term frequency * term weight) + (3^{rd} term frequency * weight) + (N^{th} term frequency * weight)*

Page weight defines a quality of web page and all terms inside page. Each term's weight provides importance of terms in webpageand item frequency describes a term's frequency in a web page. In Semantic Search ranking of web page depends on the quality of web page with query related quality contents. A web page with high weight gets high rank in SERP. Spamming can be avoided with more knowledge of query keywords.

IV. Discussion

Semantic Search Engines provides more relevant results as compare to traditional Search Engines. Semantic Search provides enough knowledge to extend the query for more accurate search on web.Traditional search just compare the query keywords with the web page keywords for search the web pages related to query. Ranking on traditional search can be affected by spammers with some Black Hat techniques. Semantic Search overcomes this with extends the search area. Spammer can't change all the web page keywords to mislead the Search Engine. Traditional search doesn't know the meaning of query, if user put a wrong query with no meaning search will continue with same set of keywords. Semantic Search extracted the meaning of query and then search with meanings.

Reference

[1] Junaidah Mohamed Kassim, Mahathir Rahmany, Introduction to Semantic Search Engine, International Conference on Electrical Engineering and Informatics IEEE, 2009.

[2] Yi Jin, Zhuying Lin, HongweiLin,"The Research of Search Engine Based on Semantic Web", International Symposium on Intelligent Information Technology Application Workshops IEEE, 2008.

[3] DuyguTümer, Mohammad Ahmed Shah, YıltanBitirim, An Empirical Evaluation on Semantic Search Performance of Keyword-Based and Semantic Search Engines: Google, Yahoo, Msn and Hakia, Fourth International Conference on Internet Monitoring and Protection IEEE, 2009.

[4] Robin Sharma, AnkitaKandpa and Priyanka Bhakuni, Web Page Indexing through Page Ranking for Effective Semantic Search, Proceedings of7'h International Conference on Intelligent Systems and Control IEEE, 2012.

[5] R. Guha, Rob McCool and Eric Miller, Semantic Search, www2003, 2003.

[6] Bettina Fazzinga, Thomas Lukasiewicz,"Semantic search on the Web", IOS Press, 2010.