International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 5; Issue: 9; September -2016

# An Intensive Study on Conventional Portal against linguistic Portal

Dr. Ram Kumar Talasila
Assistant Professor,
Department of Computer Science
Debre Tabor University,
Debre Tabor City, Ethiopia

Venkata Krishna Bandi
Lecturer,
Department of Computer Science
Debre Tabor University,
Debre Tabor City,Ethiopia

*Abstract*-The World Wide Web ("WWW", or simply "Web") is an information space that allows us to share information from global data repositories. To find out user specific data the web uses specialized tools known as web search engines. These search engines are a remotely accessible program that does keyword searches for information on the Internet. As there is tremendous growth in the volume of data or the information it is difficult to get syntactically relevant documents with in less time using conventional search engines.It can be possible with semantic web by providing sufficient context about resources on the web and building the semantic search engines that use the context so that machines can find out the meaningful documents. In this paper we present study on the general search engines and semantic search engines and have done a survey on how the keyword based search engine work for a user query practically and how semantic search engines provides results differently depending upon their specific performance

## I. INTRODUCTION

In Current world wide web, Thousands of billions of documents availablewhich are awaiting to present information on an amazing variety of topics. To retrieve documents related to the user query we make use of search engines.There are various search engines available today but working of each search engine is different from the other.Everything is machine-readable, it is not machine-understandable in conventional search engines.[1]They use keyword based searching which retrieves all the relevant documents published all over the Web but it lags inprecision and recall[2]

We need the Semantic Web to express information in a precise, machine- interpretable form, ready for software agents to process, share, and reuse it, as well as to understand what the terms describing the data mean. That would enable web-based applications to interoperate both on the syntactic and semantic level. The Semantic Web will bring structure to the meaningful content of Web pages, creating an environment where software agents roaming from page to page can readily carry out sophisticated tasks for users

Semantic web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework, which integrates a variety of applications using XML for syntax and URIs for naming.'[3].To make this possible we make use of a concept called semantic searching with help of semantic search engines.The goal of semantic searching is to deliver information in a meaningful context rather than having to sort through lists of documents bound by loosely-related keyword.

Semantic Web is about how to implement reliable, large-scale interoperation of Web services, to make such services computer interpretable, i.e., to create a Web of machine-understandable and interoperable services that intelligent agents can discover, execute, and compose automatically [4].

This paper is organized as follows: Section 2 describes the keyword based search engines, section 3 describes the Semantic search enginesand Section4 and 5 reports and discusses the experimental findings related to key word based search engines and semantic search engines respectively and the last section concludes the paper.

## II. SEARCH ENGINES:

Search engine is a program Which takes Input as user query does the process called searching and retrieves the related documents as output generally the web results are based on relevancy of the search query given by the user.

Even Though there are differences in the ways various search engines work, they all perform the basic functions like web searching, building indexes and building search queries. In web searching, before a search engine tells userswhere a file or document is, itmust befound first.Using the concept called webcrawling, the spiders also called special software robots build the lists of words found on web. In indexing the search engine must store the information gained by spiders in such a way that the gathered data must be accessible to users. To build the index we make use of following things:i)information stored with the dataii)method of indexing.[5]

The purpose of indexing is to retrieve the data as quickly as possibleand the most effective way of indexing is

hashing.In building search queries the user builds a query of his own interest and submit it to the search engine.The query may either simple or complex.currently in search engines we make use of Boolean operators such as AND,OR,NOT,QUOTES,NEAR,FOLLOWED BY etc to build the complex queries.The Booleanoperators allow us to refine and extend the terms of search.

The conventional search engines such as Google, Yahoo, and Bing (MSN) uses keyword based search but still dominate the present markets of search engines.

Google determines relevancy by using the famous Page Rank algorithm.According to this,the site which contains more number of inbound links will be the better site and should be denoted as higher rank.Coming to the working of Google,once the pages are crawled and indexed they are returned to Google for ranking.With the help of thousands of servers Google assigns ranking for the documents by considering hundreds of factors and using hundreds of algorithms.For popular queries Google response is very fast and It is also famous for decentralization and redundancy.

Google now provides individual and focused search interfaces over images, videos, locations, news articles, books, research papers, blogs, and real-time social media. Google does not support complex queries which require aggregation from multiple resources. This is because lack of proper structure in HTML documents for connecting information. Eventhough it uses the best of the limited structure available and produces better search results. [7]

Yahoo search Technology is similar to Google and analyze documents using many factors to determine relevance of a user query or search query.In Yahoo ranking of inbound links is different from Google. Yahoo initially uses a search service Inktomi or sometimes from Google later on it launches it's own version of searching algorithm for document retrieval.[6]

## III. SEMANTIC SEARCH ENGINES

Semantic searches can overcome the limitations of keyword searches because they use an ontology to retrieve information about objects.Unlike traditional search engines, which crawl the Web for gathering Web pages, Semantic Web search engines index context relevantdata stored on the Web and provide an interface to search through the crawled data.The ideal search engine would be able to match the search queries to the exact context and return results within that context

In general, Functionalities of semantic search engine are interpretation of user query, extracting the relevant concepts from the sentence, building a user query using the predefined concepts that is launched against the ontology and finally the results are presented to the user. [8]Ontology [9] is one of the most important concepts used in the semantic web infrastructure, and RDF(S) (Resource Description Framework/Schema) and OWL (Web Ontology Languages) are two W3C recommended data representation models which are used to represent ontologies.

The Semantic Web will support more efficient discovery, automation, integration and reuse of data and provide support for interoperability problem which cannot be resolved with current web technologies. Currently research on semantic web search engines are in the beginning stage.[10].The advantages of semantic search are i) easy to locate relevant information to the user's subject of interest which saves the user time ii)it can handle large queries and bring search results with the accurate context.[11,12]. Compared to search engine,semantic search engine provides the following[17]:

i)Provides proper structure for information connection using stach of technologies.

ii)provision for automatic information transfer

iii)Ability to handle huge number of users

iv)Defining universal format

## IV. RESULTS FROM TRADITIONAL SEARCH ENGINES

We have analyzed various search engines that how the results are varied with the same user query and found that all the search engines produces related documents pertaining to the user query present in the web. We make use of Google, Yahoo,Microsoft Bing etc Search Engines for this experiment. We observed that the results from different search engines are different and the results from same search engine are varying from data to date and time to time. We also observed that results from Yahoo and Bing are almost identical because there is transfer of organic search betweenYahoo and Microsoft as Yahoo Search is now powered by Microsoft Bing.We also found that the 10% variation is present between Yahoo and Bing search engine result pages and less than 2% variation for first page rankings.[25,26]

The Following table 1 shows the variations in searching using same user query:

User Query for searching : 1. Narendra modi may be the next prime minister of India

| S. No | Name of the search engine | No Of Web Links/web results Retrieved | Web links |
|---|---|---|---|
| 1 | Google | 19,100,000 | Rank1. http://www.theguardian.com/world/2014/feb/13/us-restores-ties-narendra-modi-tipped-indian-pm<br><br>Rank2. http://online.wsj.com/news/articles/SB10001424052702304703804579380760810245856?mg=reno64-wsj&url=http%3A%2F%2Fonline.wsj.com%2Farticle%2FSB10001424052702304703804579380760810245856.html |
| 2 | Yahoo | 1,390,000 | Rank1 http://www.indiatvnews.com/politics/national/narendra-modi-the-man-the-next-prime-minister-of-india--9942.html<br><br>Rank 2<br><br>http://www.sunday-guardian.com/analysis/the-rise-and-rise-of-tomorrows-prime-minister-narendra-modi |
| 3 | Bing | 25,80,000 | Rank 1<br><br>http://www.indiatvnews.com/politics/national/narendra-modi-the-man-the-next-prime-minister-of-india--9942.html<br><br>Rank 2<br><br>http://www.sunday-guardian.com/analysis/the-rise-and-rise-of-tomorrows-prime-minister-narendra-modi |

TABLE 1

## V. Results from Semantic Search Engines

Unlike conventional search engines, semantic search engines provide lesser number of results for a user query by considering the context of the query.For example if we use Hakia semantic search engine for the above user query we got only 50 results asHakia considers semantic contextof the given query.The result pages are also different from tradionalSearch Engines and semantic search engine provides their own result pages according to specific features owned by each search engine. Depending on the context of the user query some SSE's directly retrieves documents related to the query,some produces summary of the documents related to the query, some uses semantic keys to retrieve the documents and some displays bites from the documents along with document references and some does not give any result if the context of the user query is not related to predefined ontologies.

| S. NO | Semantic Search Engine | WebReference | Author, Year | Search Methodology | Result Summary |
|---|---|---|---|---|---|
| 1 | Hakia | http:// pubmed.hakia.com/ | Rıza Can Berkan,2004 | pure analysis of content | Gives only relevant documents |
| 2 | Exalead | http://www.exalead.com/search/web/ | François Bourdoncle and Patrice Bertin,2000 | semantic processing and faceted navigation to Web repositories | Gives relevant Result Documents based on image,web,video,Wikipedia along with advanced search option |
| 3 | Sencebot | http:// www.sensebot.net/ | Dmitri Soubbotin,2007 | It uses text mining and multidocument summarization to extract sense from Web pages | Generates a text summary of multiple Web pages |
| 4 | DuckDuckGo | www.duckduckgo.com | Gabriel Weinberg,2006 | It is a mata search engine gathers information from multiple search engines. | Produces documents related to relevant contexts of the user query |
| 5 | Cognition Search | www.cognition.com | Dr. Kathleen Dahlgren, Scott Jarus(With the support from cognition technologioes),2006 | Natural language processing | Retrieves formula for meaning in search |
| 6 | FactBites | http: // www.factbites.com/ | Luke Metcalfe, 2005 | Searches for authoritative and informative content | provides meaningful summaries for all the resultant documents |
| 7 | Lexee | http://www.lexxe.com/ | Dr. Hong Liang Qiao,2005 | Uses semantic key for the search process | Provides answers for the user query |
| 8 | Kosmix | www.kosmix.com | Anand Rajaraman and Venky Harinarayan,2008 | Content categorization | Meaning in search query |
| 9 | Swoogle | http:// swoogle.umbc.edu/ | PhD thesis work of Li Ding advised by Professor Tim 2004 | Indexes documents using RDF(resource description format) | Semantic web results |

## V. CONCLUSION

From this survey, we learn that there are a huge number of promising methods to semantic document retrieval.The conventional search engines also moving towards semantic retrieval by upgrading their search schemes. Many semantic search engines are available as mention in section 5 with various features and various retrieval mechanisms. The future face of search is semantic as well as Graph based search.To bring semantic document retrieval to its full potentialthe research community move a step forward to know and analyze the existing semantic search engines and situate a path for the development of promising concepts relatedto their research areas using semantic search engines

## REFERENCES

[1] Lassila, O. Web Metadata: A Matter of Semantics. IEEE Internet Computing 2(4), 30- 37,1998

[2] Celino, Irene, (----).Squiggle: a Semantic Search Engine for indexing and retrieval of multimediacontent. Politecnico of Milano, Italy.

[3] Matthews, Brian. "Semantic web technologies." E-learning 6.6 (2005): 8,JISC Technology and standards Watch

[4] McIlraith, S.A., Son, T.C., & Zeng, H. Semantic Web Services. IEEE Intelligent Systems 16(2), 46-53, 2001.

[5] http://computer.howstuffworks.com/internet/basics/search-engine.htm

[6] http://www.searchenginejournal.com/google-and-yahoo-search-engine-technology-comparison/2267/

[7] Hogan, Aidan, et al. "searching and browsing linked data with swse: the semantic web search engine." (2010).

[8] Rodrigo, L., et al (2005). A Semantic Search Engine for the International Relation Sector.Intelligent Software Components, S.A.

[9] 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

[11] Semantic Search, Available: http://www.netlingo.com/lookup.cfm?term=semantic% 20search

[12] Berkan,C.,Riza,(2007).SemanticSearch:An Antidote for Poor Relevancy. http://www.readwriteweb.com/archives/semantic_search_antidote_for_poor_relevancy.php

[13] http:// swoogle.umbc.edu/

[14] http:// www.sensebot.net/

[15] http:// www.factbites.com/

[16] http:// pubmed.hakia.com/

[17] Madhu, G., A. Govardhan, and T. K. V. Rajinikanth. "Intelligent Semantic Web Search Engines: A Brief Survey." *International Journal of Web & Semantic Technology* 2.1 (2011).

[18] $Gómez-Pérez, A., & Corcho, O. Ontology Languages for the Semantic Web. IEEE Intelligent Systems 17(1), 54-60, 2002.

[19] www.kosmix.com

[20] www.duckduckgo.com

[21] www.cognition.com

[22] Dong, Hai, Farookh Khadeer Hussain, and Elizabeth Chang. "A survey in semantic search technologies." (2008).

[23] Mangold, Christoph. "A survey and classification of semantic search approaches." *International Journal of Metadata, Semantics and Ontologies* 2.1 (2007): 23-34.

[24] Lausen, Holger, et al. "Semantic web portals: state-of-the-art survey." *Journal of knowledge Management* 9.5 (2005): 40-49.

[25] http://www.vtdesignworks.com/blog/yahoo-search-powered-by-bing-but-not-identical

[26] http://getstat.com/are-yahoo-bing-now-the-same/

[27] http://newsbreaks.infotoday.com/nbReader.asp?ArticleId=35805

[28] Véronis, Jean. "A comparative study of six search engines." (2006).

[29] Kassim, Junaidah Mohamed, and Mahathir Rahmany. "Introduction to semantic search engine." *Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on*. Vol. 2. IEEE, 2009.

[30] Tumer, Duygu, Mohammad Ahmed Shah, and Yiltan Bitirim. "An empirical evaluation on semantic search performance of keyword-based and semantic search engines: Google, yahoo, msn and hakia." *Internet Monitoring and Protection, 2009. ICIMP'09. Fourth International Conference on*. IEEE, 2009.

[31] Guha, Ramanathan, Rob McCool, and Eric Miller. "Semantic search." *Proceedings of the 12th international conference on World Wide Web*. ACM, 2003.

[32] Jansen, Bernard J., and Amanda Spink. "How are we searching the World Wide Web? A comparison of nine search engine transaction logs." *Information Processing & Management* 42.1 (2006): 248-263.

[33] Seymour, Tom, Dean Frantsvog, and Satheesh Kumar. "History Of Search Engines." *International Journal of Management & Information Systems* 15.4 (2011).

[34] Ding, Li, et al. "Swoogle: a search and metadata engine for the semantic web." *Proceedings of the thirteenth ACM international conference on Information and knowledge management*. ACM, 2004.

[35] Lazonder, Ard W., Harm JA Biemans, and Iwan GJH Wopereis. "Differences between novice and experienced users in searching information on the World Wide Web." *Journal of the American Society for Information Science* 51.6 (2000): 576-581.

Dr. Ram Kumar Talasila, presently working as Assistant Professor in Debre Tabor University, Debre Tabor City, Ethiopia. He has a total of 11 years of rich experience comprising teaching, research and industry. He received his Ph.D degree in Computer Science and Engineering and M.Tech degree in Computer Science and Engineering from Bharath University. He has guided 35 post Graduated and 61 graduate projects. He has published 19 papers in International Journals and presented 2 papers in National /International Conference Proceedings. He has Academic participation in 11. He is a member of Computer Society of India (CSI) and Indian Society for Technical Education (ISTE), New Delhi. His current research interests are in the areas of Software Engineering, Operating System, and Information Storage and Retrieval systems.

Venkata Krishna Bandi is working as Lecturer at Debre Tabor University, Debre Tabor City, Ethiopia. He has received B.Tech Degree in Computer science and M.Tech Degree in Software Engineering. He is having 8 years of teaching experience. His main research interest includes Data Mining, Software Engineering, cloud computing, and Network Security.