A Survey on Different Search Engines

J. Murugachandravel
Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India

***Abstract*: Today web has become an essential part of everyone's life. Whenever any information is needed, the web provides a helping hand with plenty of information. People can submit their needs in the form of queries, which are usually in the form of text to a search engine via browser. The search engine processes the queries and provide much relevant information, in any form, either text or images. There are lots of search engines available for satisfying users' need for information. This paper explores the features of different search engines and makes a survey.**

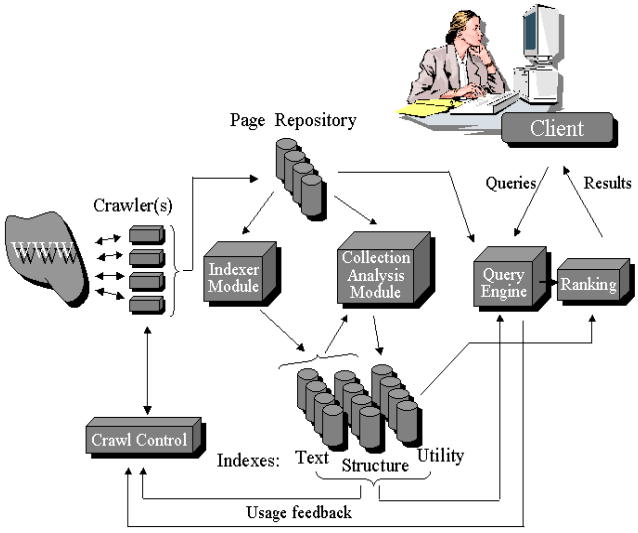
***Key words*: Search Engines, keyword based search engine, semantic based search engine.**

I. INTRODUCTION

Nowadays, internet has become a major source of information for people. It provides information in a faster manner. It significantly provides many services, like facilitating e-mails, means of communication among groups, online banking, online shopping, travelers guide, social networking, weather forecasting, video conferencing, etc. For all, the information is available on the web in the form of text, images, sounds, videos, etc. Sitting at one place, we can do lots of work. By entering into the web, we can read the newspapers, know about the current affairs, latest information in science and technology, and so on. There are various search engines which facilitate the above features. Some of the most popular search engines are google, yahoo, Bing, Ask, etc. Each of them performs the common function of retrieving information. But the way they retrieve information may differ. This leads to differences in their performances. First of all it is necessary to know the common functioning of every search engines.

II. WORKING OF SEARCH ENGINES

Search engines are involved in retrieval of information on the web. Information retrieval deals with the representation, storage and access of information on the web. The search engine gathers information on the web using a software called ***crawler*** and stores them in a repository. Then an ***indexer*** is used to create an index on the crawled information for faster information retrieval. Then a ***query processor*** is used to accept queries from users and process them. The query processor generates a set of keywords which are looked up in the index to select the documents which contain most of the keywords of search query. The documents are ranked according to their relevance to search query and provided to user. Given below is the architecture of the generic search engine.



III. TYPES OF SEARCH ENGINES

Based on the way the searching operation is carried out, search engines may be classified into ***Keyword*** based Search Engine and ***Semantic*** Search Engine. Keyword based Search engines accept queries from users and process them by using the keyword rules of stemming. Keyword search engines generally consider every word in the search query except the articles, a, an and the, or / and , etc. They do not look into the meaning of the text. Hence, the users may feel inconvenience in using these types of search engines. Some of the search engines that use keyword based searching strategy are google, yahoo, etc. Semantic based search engines consider all the words of the query, understand the meaning and then starts the searching process. This may display a better result and users may feel convenient. Some of the search engines that fall under this category are Swoogle, Hakia, Bing, DuckDuckGo, etc.

IV. FEATURES OF SEARCH ENGINES

Let us know about the features of some of the search engines.

GOOGLE: It is the most commonly used search engine. It handles more than 9.002 billion search queries every day all over the world. It also handles searching in many different languages. It was founded by the company Alphabet

Incorporation in the year 1998. It indexes around 40 billion pages. It tracks the IP Address and facilitates information sharing.

YAHOO! SEARCH: It was founded by Yahoo!. It was launched in 1995. It indexes around 10 billion pages. It tracks IP Addresses of the users and facilitates information sharing.

BING: It was founded by Microsoft. Bing provides various services like web, video, image, map searching, Bing was founded by Steve Ballmer in 2009. MSN consists of a search engine, index and a crawler. It has made nearly 13.5 billion indexes. It tracks user's IP Address. It uses dedicated servers.

DUCKDUCKGO: This search engine was founded by Gabriel Weinberg in 2008. Duckduckgo is built with search APIs from different vendors. It got its name from the children's' game duck, duck, goose. It has a good user interface and it is more beneficial than other search engines. It gives importance to producing best results, rather than producing many unwanted results. It emphasizes on protecting searchers' privacy and avoids filtering personalized search results. It overcomes the challenges of polysemy words in the search queries by analyzing the context. Polysemy refers to words having more than one meaning. For example, Fast, Kite, etc., have more than one meaning. Hence, it is considered as a semantic based search engine. This search engine does not track the user's IP Address. It does not use any dedicated servers.

ASK: Ask was known as Ask Jeeves. Jeeves refers to "Gentleman's Personal Gentleman", which means answering to any query. Ask Jeeves is a Keyword based search engine. Ask Jeeves is currently in the hands of *InterActiveCorp* (IAC). The performance of Ask.com was better than other search engines like Microsoft's MSN. Ask uses a question/answer format where other users can also answer the user's queries. Hence, it may provide results that are of lesser quality when compared to other search engines, google, yahoo, etc.

Alta Vista: Alta Vista was founded in 1995 and it was one of the famous search engines. AltaVista was developed in the Digital Equipment Corporation. Its architecture consists of a web crawler and an indexer. It used a multithreaded web crawler that was able to retrieve more results than other search engines. In 1998, it was first in handling the user queries.

Wolfram Alpha: This search engine is a computational search engine. This search engine provides a computational platform that comprises of numerical computations. Wolfram alpha uses a text box, where the users type their queries. After receiving the query, the search engine performs computations using a knowledge base and a collection of datasets. If needed, the search engine connects to other materials where data is available and uses them to perform computations. Wolfram codes contain 15 million lines and they run on more than 10000 processors. It locates the user but does not identify the user.

V. CHALLENGES

While learning about search engines, we have to consider certain challenges. Usually, most of the users view only the first page of the results page. This leads to huge traffic on the first ten websites on the result page. Hence, some of the web authors try to change their placement in the ranking order of the search engines. This process is called ***search engine spam***. To face this problem, we can use a spam classifier to identify which are spam pages and which are not spam pages.

The web may contain low quality and unreliable information. Evaluating the quality of the information and different ranking algorithms is very difficult. Search engines must avoid crawling and indexing duplicate pages and information. The structure of the data influences the techniques to be used for retrieval.

VI. CONCLUSION

Hence, we have explored the architecture of the search engines. We have seen about the functioning of crawlers, indexers and query processors. Next, we have learnt about the features of some of the search engines. We have also identified the challenges that prevail in the search engines.

REFERENCES

[1]. J. Griesbaum, Evaluation of three German Search Engines: Altavista.de, Google.de and Lycos.de. Information Research, Vol.9, No.4, 2004. http://www.informationr.net/ir/9-4/paper189.html
[2]. C.Silverstein, M.R. Henzinger, J. Marais, and M. Moricz, Analysis of a very large AltaVista query log.ACM SIGIR Forum,33,6-12,1999
[3]. Suruchi Chawla, Search Engines: Information Retrieval on the Web, Everyman's Science, Vol I.I No.4, October '16 – November '16.
[4]. S. Lam, the Overview of Web Search Engines, University of Waterloo, 2001
[5]. https://en.wikipedia.org/wiki/Bing_(search_engine)
[6]. https://en.wikipedia.org/wiki/Comparison_of_web_search_engines