ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (422-426), Month: January - March 2015, Available at: www.researchpublish.com

Enhanced Search Engine

Akash Kosambia[1], Prof. Tarik El Taeib[2]

Abstract: **In this era of the World Wide Web clients of information assistances faces not only an extremely distributed but also varied data space with unrelated data foundations and also with "value added" sections in information systems and with the more important question of whether they should still, search literature in the focused databases of information centers like the Information-zentrum Sozialwissenschaften, which is also known as IZ in Germany, or whether they should use www search engines like AltaVista or Fast (Krause, 2001). The main aim is to provide elevated proper search results over a quickly growing World Wide Web and also to develop sensible system which can make use of the added information which is there in the hypertext.**

Keywords: **Enhanced Search Engine, World Wide Web.**

1. INTRODUCTION

To make search engine work for millions of demands is a very challenging a task. Search engine ripostes millions of queries daily there are differences in the way numerous search engine works but there are elementary steps that every search engine achieves that are they search internet created on particular words, they keep catalog of the word from where they discover them and they always favor ontology. An information is found by the search engine form its database by considering listings which is sent in by creators who want associate, or by considering the related information from their web crawlers or spiders etc, programs that wander the web storing links to and information about every sheet they visit. For a web search engine a web crawler moves and keeps Web pages quite often. Approximately, crawler begins by enlisting an first set of URLs So, in a line, where each URLs to be improved are stored and arranged. As of From this particular queue, the crawler gets a URL , it downloads the page, it abstracts any URLs in the transferred page, and thus puts the new URLs in the queue. This course is reiterated until the crawler chooses to end. Pages are collected for advanced used for other applications like Web search engine and Web cache. The majority of the necessary steps for a search engine is the search functioning itself, main quality of its results and ability to crawl, and its guide the web resourcefully. The principal focus is to assume high quality search outcomes over a quickly enhancing World Wide Web. A number of much-admired Search engines are Google and Yahoo, which allocate some common descriptions and are synchronized to some extent.

2. SEARCH ENGINE USE AND FEATURES

Search engines are proposed to curtail the time you need to spend time looking for the information on the World Wide Web. Considering the path in which searches are performed and the ways different information can be used is a valuable skill for students of geography to develop. Although search engines can look very diverse, a number of features are common to a large amount of them, these include:

■ a display screening shows how many 'hits' or web pages the search engine has found that is equivalent your search (the more web pages found the more hits you have)

■ A list of the hits (a number of hits include a short explanation of each one with your search words highlighted)

■ A help file (this will direct you through the search process)

■ an better search facility (a search engine is a bit like a book's contents page, it gives you an idea of what page a topic begins and ends, whereas an superior search facility is the index in that it helps you recognize specific information more precisely).

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (422-426), Month: January - March 2015, Available at: www.researchpublish.com

Search engines can be split into two individual types: those with topic categories and those without. Using both types to search the Internet will result in a huge array of sites or individual web pages that you or your students must trawl through to select applicable material. By considering the two main types of search engines it is possible to decrease the amount of time you spend searching (Broad, 2003).

3. IMPROVED SEARCH OPTION

A few of the search engines propose the chance to observe a real client search behavior. A small amount of time spent presenting entries by the client using Meta Spy certifies idea that the simple search-engine user allows only not more than two searches sometimes typing wrong spelling and making use of very little advanced options. In order to motivate better searches, it provides a list of terms by itself which are then selected by the client to process and reconsider their original search. This list of newly selected terms is presented, it is drawn from the evaluation of related files, at the starting of the page of the search results. The client only needs to choose checkboxes to add one or more terms and then click the option of "Search again" to move on. The search test which is "Clinton impeachment" makes a list of the terms which includes variant. A same refined option to help client in order to search the results and currently unveiled an extra feature called "Related Searches" is possessed by Alta vista. For example When a user writes a simple query, like "pets" which links to search the narrow part of the expression like house pets and pet health etc are given at the beginning of the result page of AltaVista (O'Hanlon, 1999).

4. BUSINESS OF SEARCH ENGINE

Expression toward ubiquity is continued by the Google, most of the libraries still fights with context, comparison with, and clear explanation to their customers about the free web. As international market has been achieved by the Google, with the use and broad adoption, previous famous search engines like Yahoo and AltaVista has gone under various modifications and business changes. The search engine business are not in the trade of making very much appropriate and precise search results. Their trade is the nothing but the version of search traffic promotion revenue. This variation is only improved when search engine is trendier but the good search results in the search engine are the bait. Hit lists of all the search engine convey users to promote sites. Few of the search engine does it more ingeniously than other search engine, but Google also contains many sponsored links (Pace, 2004).

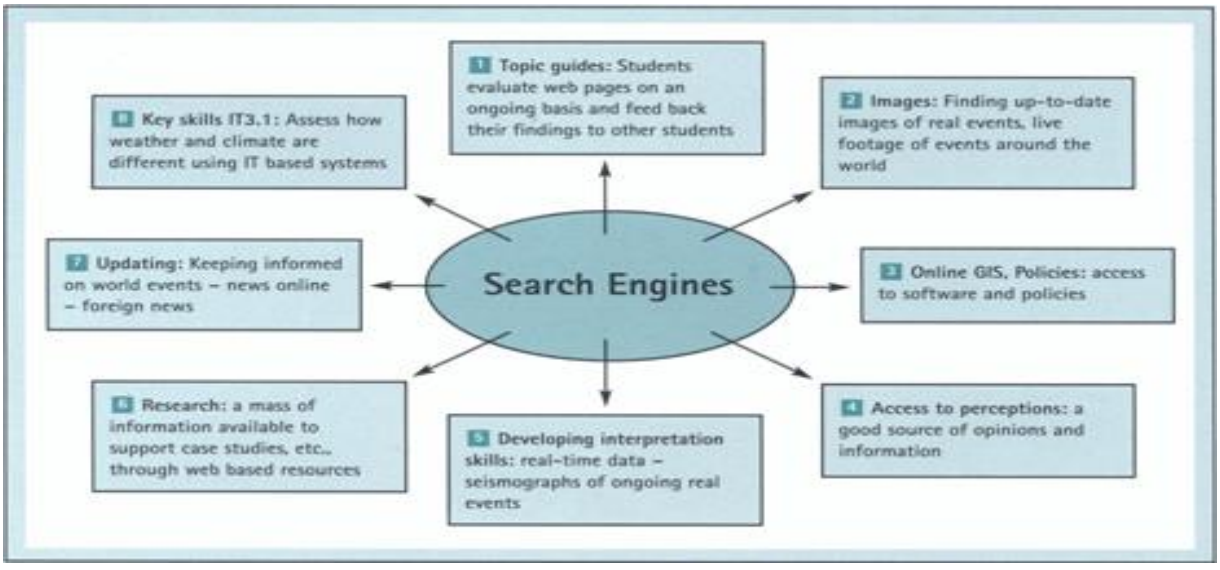

Figure 1: (Broad, 2003)

5. SYSTEM ARCHITECTURE OF SEARCH ENGINE

The general idea of how the entire system of a search engine works is in this section. One of the most important functions of the search engine are crawling, indexing and searching are explained in detail. It is a software design which consists of software components, the interfaces which are given by those components describes a specific level of abstraction of a system and relationships among them. The Performance of search engine only depends on class of result and its efficiency which is the response time and throughput of the respective search engine.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (422-426), Month: January - March 2015, Available at: www.researchpublish.com

To look for the information from a particular search engine relevant web pages and files need to be searched, a software is engaged by the search engine known as "*spiders"*, in order to build the list of words on the web sites. The course of building the list is called the *Web crawling*. This program which usually negotiate the web with the help of downloading identification and also follows relations from web page to web page. They are usually used by search engines to group information for indexing. page validation, visualization, update announcement, mirroring etc. is contained by supplementary potential apps however, other name for Web crawlers is spiders or worm or robot etc. They are nothing but the automated programs which catches the links that are contained on the web page.

The lists of URLs are sending by the URL Server so that it is obtained by the crawlers. This particualr web pages are then stored in the store server. This fetched web pages are then packed in and stored into the storage area by the store server. Every web page contains a linked Identification (ID) number known as doc ID, now this Id is given to a new URL whenever it is determined out of a web page. With the help of indexer and sorter indexing function is carried out. several functions are performed by the indexer from the repository fetched web pages are read, then the fetched document is uncompressed and parsing is done. In the series of word occurrences which is known as 'hits' documents are altered. In document this hits witness the search location and also it performs the assessment of font size. Now the set of "barrels" gets this hits by the indexer, so that a partly sorted forward index is created. An added important function is executed by the indexer, which is nothing but it parse's out every single links which are present in the web page and all the significant information regarding the file is collected. This particular file encloses necessary information which decide whether every link points to and from or it points to the text of the link.

The anchors file is interpreted by the URL resolver. The anchor text is kept into the forward index which in turn linked by the document ID which the anchor points to. It then creates a document full of links. The database file for every single document is work out for Page Rank.

The barrels are taken by the sorter which is then sorted by the doc ID and they are restored again by word identification "ID" so that it creates the inverted index. After this small brief space is needed for this particular operation. A catalog of word IDs and offsets which are interested in the inverted index by the sorter. This directory is taken together with the help of lexicon which creates by the indexer and it also creates a new lexicon which can then be used by the searcher. All the terms which are taking place in the index along with some term-level statistics are cataloged by the lexicon, that are used by the ranking algorithms. The web server performs the search and it uses the lexicon with the inverted index and the Page Ranks to react queries.

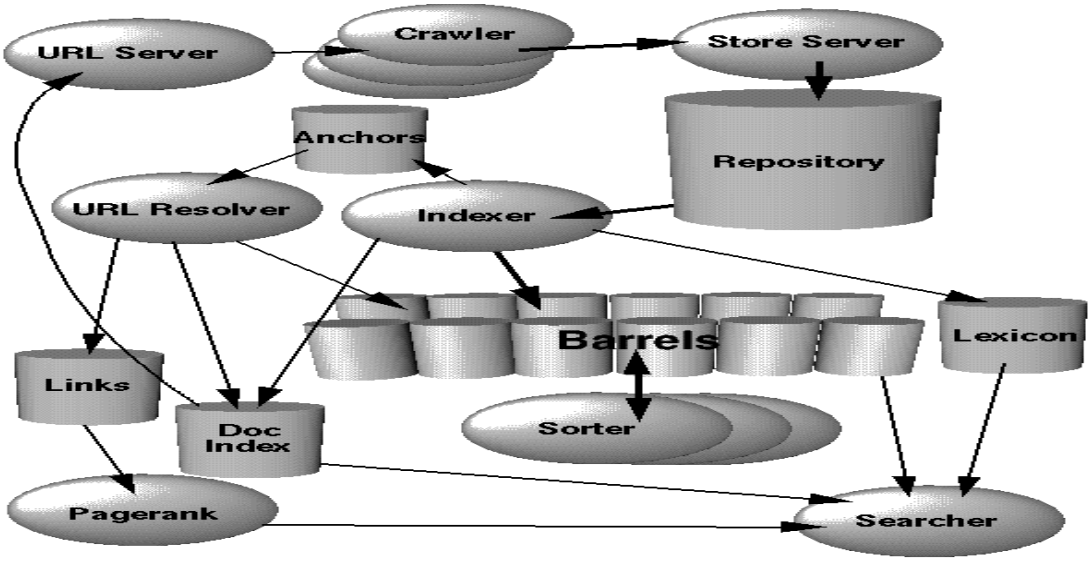

Figure 1: Architecture of Search Engine

6. WORKING OF WEB CRAWLER

A vital part to the search engine is web crawler, to begin a web crawler is a tricky job. There are many risky presentation issues involved and some real community issues are also related to web crawlers. Crawling involves interacting with vast

number of web servers and frequent name servers which sometimes gets ahead of the control of the system, which makes crawling one of most frail application. Crawling speed is looked not only by the Internet connection but also by site speed. There are various applications of web crawlers, which essentially works on the same root. Steps concerned in functioning of a web crawler are:

* The web pages are downloaded.
* Go all the way through the downloaded web page and each links are retrieved.
* This process is repeated again and again for every retrieved link.

Crawling through a web crawler can be done on the complete web site on both Internet and intranet. A URL which starts is particular and the web crawler keeps track of the links established in the HTML Web page. This will lead to keep track of some more links that is followed by the web crawler. A site creates a tree structure where start-URL is the root and root-HTML-pages are the nodes straight attached to the root. Consequent links are the nodes attached to the preceding nodes.

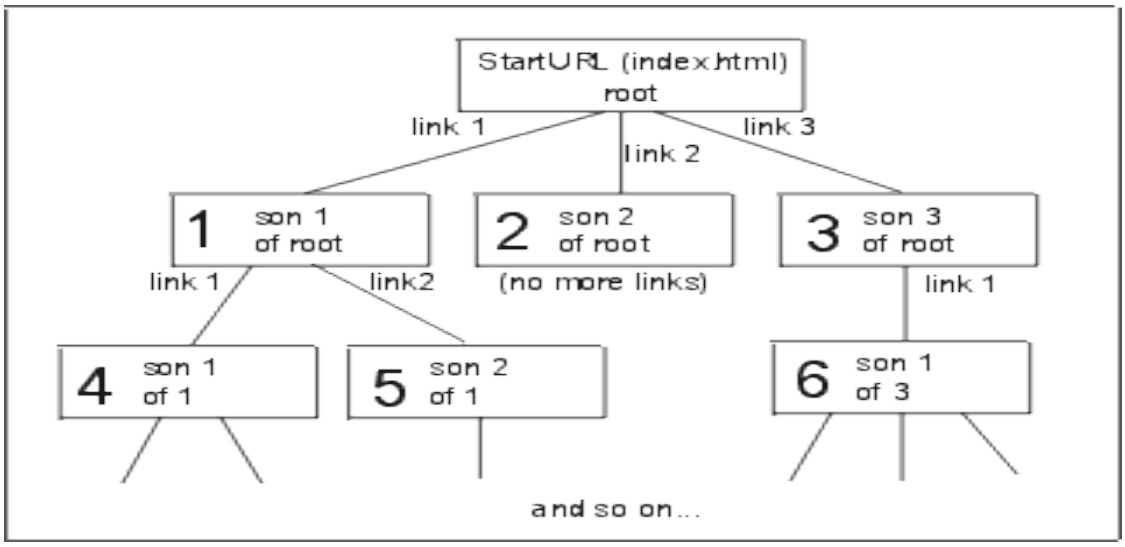

Figure 2: Working of a web crawler

Web crawler software does not roam around to various systems on the Internet as viruses. A crawler usually keeps approximately 300 connections open at a time. Web crawler begins its functioning by initiating a particular web page, making a note of hyperlinks that points to other web pages continuing the same process for the new links that points to some other new web pages. A Web crawler stays on a single particular machine. To retrieve a document residing on some other machine on the internet, a web crawler simply sends HTTP request just like a user clicks on link while browsing. A web crawler can be considered as processing items in a file.

7. META SEARCH ENGINE

A Meta Search usually does not have its own database of web pages; it uses other queried search engine's information to create their own sets of results. These search engines are officially responsible in getting the results from free search engines and a series of directories which can be frequently small or highly marketable. The algorithm working by meta search engine is somewhat differs from other search engines. The result of this is displayed like every other search engine's query which is posed in a currently opened web browser. Some enhanced meta search engines works the outcome of one screen in many frames or in one frame but in orders. Some hard meta search engines give permission to the users to prefer shortest search engines in the query input process where at the same time using filters and extra algorithms in order to process the arrival of the query outcome, showing them to the users. Meta search engines are very useful if the user is trying to search a special keyword or a particular phrase. Few of the meta-search engines only go by search language to the primary direct search engine and if a search consist of more than one or two words or a difficult logic, most of them will be lost. Some of the search engines can support this logic. Few of the existing meta search engines with a particular direct search engines like Alta Vista and yahoo which includes Dogpile and Meta Crawler.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 1, pp: (422-426), Month: January - March 2015, Available at: www.researchpublish.com

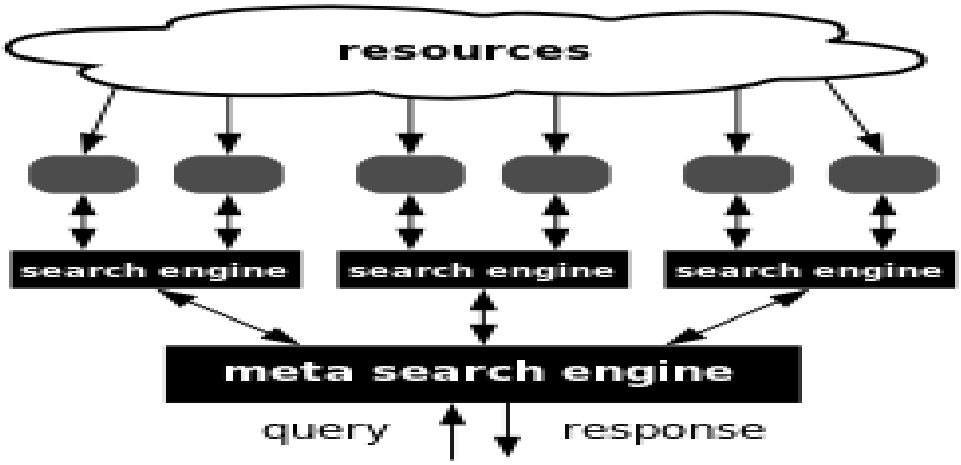

Figure 3: Architecture of Meta-Search Engine

8. WAYS TO IMPROVE WEB SEARCH ENGINES

There are a number of ways through which the performance of web search engines can be improved. For this firs we need to improve query input on user. Secondly, we got to see the query results towards filtering. and Thirdly, in order to give the solution to the algorithms of web page indexing, output and collecting. This third method is the main method for any direct search engines to control the problems of asymmetrical accessing and obsolete information. It is mainly significant to look at the method one and two in order to deal with user interface issues like How to control user queries and demonstrate the particular results successfully.

9. CONCLUSION

Most on-line users uses On-line searching as the main part of the web search and search engines are the only and famous way to do this. Therefore search listings have become very peculiar. Sellers are very much interested in advertising on search engines. The result of these ideas has gone down from about 2% to less than 0.5% in recent time. Due to this the sellers realized that banner ads do not bring the traffic volume. (Sen, 2005).

This idea explains the structure of search engine typically and gives the proper working of a web crawler which is developed in Java. It also explains the features of all the components related for searching some kind of information on the Web. The design of the search engines are useful to grade websites which are based on few grouping of their popularity and practical studies which designates many political and economic and social biases for the information they give. This information can be the direct result of economic and profitable process.

REFERENCES

[1] Broad, J. (2003). Using search engines. Teaching Geography, 28(3), 118-121. doi: 10.2307/23754336

[2] Krause, J. (2001). WWW -- SEARCH ENGINES, AN ALTERNATIVE TO THE SOCIAL SCIENCE DATABASES? BMS: Bulletin of Sociological Methodology / Bulletin de Methodologies Sociologique(72), 37-46. doi: 10.2307/23891704

[3] O'Hanlon, N. (1999). Web Search Engines Evolve to Meet Challenges. Reference & User Services Quarterly, 38(3), 247-251. doi: 10.2307/20863557

[4] Pace, A. K. (2004). Technically Speaking: The Business of Search Engines. American Libraries, 35(5), 60-61. doi: 10.2307/25649179

[5] Sen, R. (2005). Optimal Search Engine Marketing Strategy. International Journal of Electronic Commerce, 10(1), 9-25. doi: 10.2307/27751172.