

UNAK – 2008

Information: Diversity and Awareness

On

9th – 11th October 2008

At

Yasar University,

Izmir - Turkey

Sub Theme: Search Engine

By

Dr. (Mrs.) SHAISTA PARVIN

Asstt. Librarian

Dr. Babasaheb Ambedkar Marathwada University,
Aurangabad

Maharashtra – India.

Phone: +91-0240-2393999

Mobile: 9890201232

E-mail: muqista@yahoo.co.in

SEARCH ENGINES

Dr. Shaista Parvin
Asstt. Librarian
Dr. Babasaheb Ambedkar
Marathwada University,
Aurangabad.

ABSTRACT

The present Article places emphasis on Web Technology. It deals with sub types of Search Engines it elaborates the details of search engines.

Keywords: *Web Technology, Search Engines.*

□ **Introduction:**

The World Wide Web, popularly known as www or web. It is the most versatile and fast growing communication medium on the Internet. The web was created in the early in 1990s by the European Laboratory for particle physics and had, as its primary goal, that of allowing researchers to work together on projects and to make project information easily accessible. The World Wide Web has revolutionized the way we access information. It is a navigational tool, which enables the users to access hundred of sites without actually knowing the addresses. These sites provide thousands of documents which are hyper linked to each other in the form of a network, so that the user from the local system can access to any document anywhere in the world immediately and can surf from one location to another without any difficulty. The web is a vast collection of text graphics sound, and video files linked together that make it possible for the user to jump or travel from one document to another.

□ **Web Technology**

‘Web Technology is a very fast growing field of knowledge. It encompasses all the activities related to World Wide Web. It has a body of

software, a set of protocols and conventions. It uses hypertext and multimedia techniques to make the web easy for anyone to use. The web is based on client server mechanism where the end user uses client software to access information or the data kept on a server. The computer and its web server software, which can host a web page, act as 'server' where as the end users system and its browser act as the client. The web pages or documents are developed by web servers according to a particular format i.e. HTML (Hyper Text Mark-up Language) which provides tags to connect the different elements of web pages which are available on the same system or on the remote computers into a web. HTML tags facilitate accessing of web pages by the clients on the servers and transferring of web pages to client's system. Transferring of web documents from the web server to web client is performed by an Internet Protocol (IP) called HTTP (Hyper Text Transfer Protocol). The location of documents on the web pages is specified by standard format called URL (Uniform Resource Location).' (Mahapatra & Ramesh, 2004)

An on going research in web technology enables facilitates us a further accessibility and better navigation. 'The enormous amount of information available on the thousands of web servers is not easy to access by the users. The information spread out in large number of pages on the web will cause great difficulty to a user in the search of for a particular piece of information in the area of his interest. Rapid growth of the web makes it difficult from individual to locate information and navigate to access these. There are basically two ways to search the web; by conducting a search using a search engine, or by following the links in a specially designed called a directory. The difficulty in searching the web can be overcome with the help of search engines. Search engines are employed which help pinpointedly located the information we need on the Web. Search engines are key means to finding resources on the web. Any program that locate

required information on the internet is known as a search engines locate web sites on a particular subject, reduce the surfing time and provide a list to find the sites of ones interest. Engine this is right to say in this text that, the web without search engines is equal to a library without a catalogue. Each Search Engine has its own format and charlatanistic features for searching.’ (Vishnu & Anilkumar, 2000.) Further advancement in the field will enable more specific and faster results.

□ **Search Engines**

A Search engine is useful for locating information in the vast space of the web. Basically, it is a searchable database of Internet files collected by a computer. As a librarian we may say that search engine in nothing but catalogue of information on the Internet. Search engines are on of the most effective means for the user to reach page. Search engines allow the user to enter search terms-keywords and/or phrases-that are run against a database containing information on the web pages collected automatically by programs called spiders. The search engine retrieves web pages from its database that match the search terms entered by the searcher. It is important to note that when a user conducts a search using a search engine, the latter does not search for the information across the entire web at the given instance. Instead, it searchers a fixed database, located at the search engine’s website containing information on selected web pages. It is updated at regular intervals according to the specific criteria employed by the search engine.

Architecture of Search Engine

‘A search Engine Consists of three major components. A spider, an index and a search engine mechanism.

The spider

To find information on the millions of web pages that exist, a search engine employs a special program, called a spider, to build lists of the terms found on websites. A spider automatically fetches web pages for search engines; it is called a spider because it crawls over the web. The Spider (also called a crawler, a robot, a worm or robot) is a program that traverses the web from link to link, identifying and reading pages and then storing them in the index. Every thing the spider finds goes to the index.

The index

All the terms that a spider finds go into the index. If the content of a web page changes, then the spider notices it, brings back the new information and thus updates the index with the new information. The index is a database containing a copy of each web page gathered by the spider.

The search engine mechanism

The search engine mechanism is the software that enables to query the index and that usually returns results ranked in order of relevancy. It is the information retrieval program that performs two major tasks, first it search through eh millions of terms recorded in the index to find matches to a search, and secondly it ranks the retrieved records in order of what it believes is the most relevant. The program receives search request, compare it to entries in the index and returns the results in a schematic order.

□ **Types of Search Engines**

Search engines are of two types

- 1) Individual Search Engines
- 2) Meta Search Engines

An individual engine uses spider/crawler to collect its own searchable index.

Meta (also termed as parallel or multithreaded mega search engines or combined search engines). Engine, searches query simultaneously various

engines and display results collected by the spider of other search engines. In a Meta search engine, the keyword that we are entering to its search box is simultaneous by transmitted to most of the popular search engines and the databases of web pages. Within seconds the searcher gets the compilation of the results containing. Matching sites from all of the search engines queries.

Meta search engines are of two types

- 1) Separate retrieval for comprehensive result.
- 2) Collated retrieval for select results.

Former one displays results as separate list of results from each engine that was searched. It does not collate the results in one screen. With this type of Meta search engine, we can retrieve comprehensive, even overwhelming results. Latter one displays results in single screen, often with the duplicate link removed. This type of engines search queries in various sites up to a certain amount of time or up to a certain number of documents. Meta search engines are useful when you want to retrieve a relatively small number of relevant results or for obscure topics. To know over all picture of availability on web for a topic these engines are useful.

□ **Important Search Engines**

- Individual Search Engine

❖ **AltaVista** (www.altavista.com)

Altavista, opened in 1995, is one of the largest comprehensive & favourite search engines among researchers. AltaVista offers simple as well as Advanced Search. The simple interface includes a single search box and a pull-down menu that allow limiting search to one of 25 languages. Advanced query mode offers Boolean and near operator. For Full Boolean Searching with nesting and parentheses, Alta Vista Advanced Search offers the most options.' (Mukherjee, 2005.)

❖ **AOL Search** (www.aol.com/netfind/)

‘It was launched in March 1997. AOL Search provides users with editorial listings that come from Goggle’s crawler-based index. Indeed, the same search on Goggle and AOL Search will come up with very similar matches. The ‘internal’ version of AOL Search provides links to content only available within the AOL online service.

❖ **Ask Jeeves** (www.askjeeves.com.)

Ask Jeeves initially gained fame in 1998 and 1999 as being the ‘natural language’ search engine that let you search by asking questions and responded with what seemed to be the right answer to everything. The company at one point had about 100 editors who monitored search logs. They then went out onto the Web and located what seemed to be the best sites to match the most popular queries.

Humans are still used at Ask Jeeves, though the number of editors is now only around ten. Nevertheless, the human-provides answers may still be the selling point for why some people, especially those new to the Web, may want to use Ask Jeeves. For popular queries, the human-selected matches in the ‘Click Ask below for your answers’ sections of the results may feel very relevant. If shown, these appear at the very top and bottom of the search results page.

❖ **Google** (www.Google.com)

Google has a well-deserved reputation as the top choice for those searching the web. The crawl-based service provides both comprehensive coverage of the web along with great relevancy. Google was originally a project by Stanford university students Larry Page and Sergey Brin Called Back Rub by 1998; the name had been changed to Google.

Google provides the option to find more than web pages, however. Using tabs on the top of the search box on the Google homepage, you can easily seek out images from across the web. Google is also known for the wide range of features it offers, such as cached links that let you resurrect dead pages or see older versions of recently changed ones. It offers excellent spell checking, easy access to dictionary definitions integrated stock quotes, street maps, telephone numbers and more. The toolbar has also won Popular acclaim for the easy access it provides to Google and its features directly through the Internet Explorer browser.

❖ **Hot Bot** (www.hotbot.com.)

It was launched in 1996. Offers simple as well as expert search. Simple search supports the Boolean AND and OR operators, phrase searching and choices “the person” or “the URL”. Expert search also supports date limits, media type and locations. Enclosed phrases in quotation. Results are displayed with the document size, and date of unknown origin.

❖ **Info seek** (www.infoseek.com.)

It was launched early 1995. Offers a simple query option, but search words may be limited to particular fields (such as within document title). Operator (-) used for elimination and (+) for addition of two keywords. Displays results by title, file size, URL, a brief abstract and rank relevancy.

❖ **Inktomi**

Among the major search engines, Inktomi is the second-oldest crawler. It briefly operated as an experimental search engine at UC Berkeley. However, the creators then formed their own company in 1996 with the same name and gained their first customer, HotBot, in the middle of that year.

❖ **Lycos** (www.lycos.com.)

Lycos is one of the oldest search engines on the Web, launched in 1994. It ceased crawling the Web for its own listings in April 1999 and instead uses crawler-based results provided by FAST. Relevant categories of human-compiled information from the Open Directory appear at the bottom of the search results page. At the top of the page, Lycos will suggest other searches related to your original topic right under the search box.

❖ **MSN search** (www.msn.com.)

Microsoft MSN is known for constantly reworking its software products until they get them right, and MSN Search is a shining example of the company putting that same effort into an online product.

Overall, MSN Search provides a blend of human-powered directory information and crawler coverage different from any of the other top choices listed above. It's a high-quality resource that provides its own unique view of the Web and is one worth checking. Microsoft's MSN service went live in Oct 1998 with its Inktomi results, although it had existed in various formats and under different names previously.

❖ **Northern Light** (www.nlsearch.com.)

It was launched in August 1997. Boolean operator OR and NOT are supported as are quotes around phrase and "+" to precede a word that must be present in the results, and "-" to precede a word that must not appear in the results. Initial results are presented in order by computer relevance, along with folders and subfolders of results by concepts or site of origin; information provided for each link includes documents address, title, relevance rank in percent, and a short abstract.

❖ **Teoma** (www.teoma.com)

Teoma is a crawler-based engine owned by ASK Jeeves. It has an extremely small index of the web, only about one tenth the size of crawler-

competitors Goggles, All The web com, Inktomi and Altavista. All common techniques such as use of sign in front of each word, use of quotation for searching whole word, use of-sign for elimination of word from a query string, etc. are some of the techniques with this search engine follows.

❖ **Yahoo** (www.yahoo.com)

Launched in 1994, yahoo is the web's oldest 'directory' a place where human editors organize websites into categories. Search techniques based on simple search box. Boolean operation and string searching are also supported. Results returns along with their descriptive Text and the subject hierarchy under which it can be found in yahoo is also displayed.

❖ **Wise Nut**

Like Teoma, WiseNut is a crawler-based search engine that attracted attention when it appeared on the scene in 2001. Like Teoma, WiseNut features good relevancy. Unlike Teoma, WiseNut has a large database, making it nearly as comprehensive as Goggle, AllTheWeb.com and Inktomi.

▪ **Meta Search Engines**

Meta search engines may turn out to be useful in the following situations:

- When researching an obscure topic;
- When you are not having luck finding anything when you search;
- When you want to retrieve a relatively small number of relevant results.

A Meta search engine does not search the Web itself, but it uses the data and the services of other search engines and display them in a new form. Meta Search engines are a combination of a number of search engines. A search typed into a Meta search engine is sent on to the collaborating search engines have a special back door opening to the other search engines. There are then two ways a Meta search engine displays the search results.

By one method, the user gets the same results as if he had searched all the search engines separately, only faster. Some engines require visiting each site to view results, while others will fetch the results back to their own sites. When results are brought back to the site, a certain limitation is placed on what is allowed to be retrieved. With this type of Meta search engine, one can retrieve comprehensive, and sometimes overwhelming, results.

By the second method, the Meta crawlers take the results from the separate search engines and process them using their own logic system. For example, the results may be assessed, put in a new order or have added notes. The user gets therefore more than the sum of the collected results. This type is more common and returns a single list of results, often with the duplicate hits removed. This type of Meta engine always brings the results back to its own site for viewing. In these cases, the engine retrieves a certain maximum number of documents from the individual engines it has searched, cut off after a certain point as the search is processed. The cut-off may be determined by the number of documents retrieved or by the amount of time the Meta engine spends at the other sites. Some of these services give the user a certain degree of control over these factors. All of this has two implications:

- These Meta search engines return only a portion of the documents available to be retrieved from the individual engines they have searched.
- Results retrieved by these engines can be highly relevant, since they are usually grabbing the first items from the list of hits ranked by relevancy returned by the individual search engines.

□ **Metacrawler**

At the time of writing, Metacrawler is usually cited as the most efficient and most popular Meta search engine. Metacrawler also offers a web directory. The main headings of their directory are:

- . Arts & Entertainment
- . Autos
- . Business & Money
- . Computers & Internet
- . Games
- . Health
- . News & Media
- . Recreation
- . Reference
- . Regional
- . Science & Technology
- . Society
- . Sports
- . Travel

Metacrawler advertise themselves by saying: ‘Get the most relevant results you are looking for fast by searching multiple search engines at once’. The ‘advanced search’ allows the setting of ‘keyword default’ (search for any of the supplied words, all of the words or the exact phrase provided). The user can indicate which search engines should be searched by the crawler. It is also possible to choose the number of results to be displayed on each page and the maximum number of results from each engine. The results can either be sorted by relevance or source (where source means search engine). The maximum amount of time to wait for results can also be chosen.’ (Muller, 2003.)

□ **CONCLUSION**

It is observed that during search on websites the users need to be equipped with search engine. Search engines run from special sites on the web and are designed to help people find information stored on other sites.

They are the most fascinating information retrieval tools. They are constantly improving their web search facilities. ‘There are differences in the ways the various search engines work, but they all perform the following three basic tasks:

- They search the internet-or select parts of the internet-based on a set of criteria.
- They keep an index of the words/phrases they find, with specific information such as where they found them, how many times they found them, etc.
- They allow users to search for words/phrases or combinations of words/phrases found in that index.’ (Chowdhury & Chowdhury, 2001.)

Every search engine claims its position as world’s number one search engine, and the person handling the search engine had to decide what suits him the best.

BIBLIOGRAPHY

1. Chowdhury, G. G. and Chowdhury Sudatta: (2001) Information Sources and Searching on the World Wide Web. London, Library Association Publishing, P.15-16.
2. Mahapatra, M., and Ramesh, D. B.: (2004) Information Technology Applications in Libraries A Textbook for Beginners. Bhubaneswar, Reproprint, PP. 464-465.
3. Mukherjee, Bhaskar: (2005), User's Query Through Search Engines: Librarian's view SRELS journal & Information Management, 42(4) PP.487-488.
4. Muller Jeanne Froidevaux: (2003) A Librarian's guide to the Internet, searching and evaluating information. Oxford, Chandos Publishing, PP. 35-39, 68-72
5. Vishnu, V. V., and Anilkumar, C.: (2000) India- specific Internet search Engines. In T. Ashok Babu and Others (eds) vision of Future Library and Information systems. New Delhi, Viva Books, Pvt. Ltd., PP.149.