

SEARCH ENGINES AND WEB DIRECTORIES: AN OPINION

Vivekanand Jain*

Abstract *For basic information needs seekers of information are using popular search engines on internet, but for subject specific queries, it is advised to use the web directory or subject specific search engines. The present paper deals with features of search engines, web directories and search techniques for effective information retrieval. It also correlates online searches with traditional library search techniques. The information seeking behavior is also varied from person to person or new generation to old generation. Search engines make the internet popular among users. It is a tool to fulfill the information need of the users and fill up the digital divide in society and specially in academic community.*

Keyword: *Web Directory, Search Engines, Subject Directories, Information Retrieval*

A web search engine is an interactive tool to help people locate information available on World Wide Web. Web search engines are actually databases that contain the reference to thousands of resources. Users interact with the database, submitting questions that ask the database if it contains resources that match specific criteria. There are many search engines available on the web. A web search engine provides an interface between the user and the database. The interface presents the user with an opportunity to type in a search string, which may be a word, a phrase, a date, or some other criterion, and a way to submit the request.

Search engine and Web directories are the tools to retrieve the information from Internet. The search engine searches sites available on the web, but web directory searches only selected, reviewed sites. Search engine generally will find more sites than directories, because search engines automatically add more and more sites to their databases as they find them with the help of robots/ crawler. But directories generally have higher quality listing because they index websites after proper review by human being.

Some search engines provide ranking to web pages on the basis of relevancy. Librarians can use their experience to satisfy the information needs of the users, but for the search engines all users are same. Search engines can't distinguish between the users as layman, subject expert, child, adult etc., so the output is based on the search string only. Search engines don't have the ability to ask a few questions to focus your search as a librarian can. They also can't rely on judgment and past experience to rank web pages, in the way human can.

Library professionals have been providing information to their patrons for a long time from printed to online resources.

Librarians and Internet search engines are having the same mission of organizing the information and fulfilling the information needs of the users in a very efficient manner.

From Librarian's point of view the Internet search engines are information retrieval systems. So, the same method of evaluation of Information retrieval system can be applied to Internet Search Engines. Library catalogue, Indexing systems (Pre coordinate, post coordinate, Chain indexing, subject headings, etc) are tools for information retrieval. In the same manner Internet Search Engines are search tool to retrieve online information from Web servers. Lancaster and other information scientists have evaluated the precision and recall value of Information Retrieval Systems which can be applied to Internet Search Engines.

Some important role played by Internet search engines are as follows:

1. To search and collect the information available on web servers in distributed environment.
2. To index and organize the information in a systematic manner.
3. To provide retrieval facility for users.
4. It helps users to know the location of resources.
5. To know ranking of available information/ web pages.
6. To know the research trend on particular topic.
7. It helps in finding information, products and business on the web.
8. The effective use of Internet as ready reference source.

Internet Search Engines are specific in their features because it is more interactive. Users can define its search string as per their query/needs. They can put a query in any format: direct

* Deputy Librarian, Central Library, Banaras Hindu University Varanasi, Uttar Pradesh, India.
Email: vivek_digora@rediffmail.com

question, simple keyword, advanced search, etc. Search engines are meant to access the global information. But to fulfill the local information needs of the users in depth, most of the search engines provide area based/country based local search facility. For example: few search engines make special provision to search in India, i.e.

Google has option as	google.co.in
Yahoo has option as	Yahoo.co.in

There are number of search engines which are dedicated to particular region/ country. Global search engines are also providing localized search facility as mentioned above.

LIBRARY SCIENCE AND INTERNET SEARCH

There are a number of library search techniques, classification and cataloguing features adopted by Web organizers. Few of them are as follows:

1. Use of classification schemes to organize Internet resources.
2. Subject approach on Internet.
3. Cataloguing of Internet resources.
4. Use of library retrieval techniques
5. Boolean search
5. Indexing pattern of storage and retrieval
6. Pre coordinate and post coordinate indexing.
7. Creation of metadata.
8. Use of standard terminology: thesauri, subject headings etc.

I found that there is a similarity in library catalogue and Internet search engines. Like one can not get a book from the huge collection of library without catalogue. In the same manner, one can't get information from Internet without help of search engines.

Catalogue of Library	= Search Engine for Internet
Subject Catalogue	= Web/Subject Directories
Union Catalogue	= Meta Search Engines
Books/Document	= Online Information Resources

In the earlier database searches, usually access was limit to few fields only like title, author etc. But now they provide access to all fields including running text known as free text search. The free text search provides large number of output. It creates problems to users rather than satisfying them. The right amount of information at the right time can be possible after interaction and mediation of the library professionals in organisation of Internet resources and their retrieval process.

MECHANISM OF SEARCH ENGINES

Search engines allow the user to enter keywords that run against a database. Based on a combination of criteria (established by the user and/or by the search engine), the search engine retrieves web resources from its database by matching the keywords entered by the searcher. The search engines do not search the Internet directly, but search a fixed database that have been compiled some times before the search.

All search engines are intended to perform the same task in a different way, which leads to sometimes amazingly different results. Factors that influence results include size of the database, frequency of updating it, retrieval techniques, and search capabilities. Search engines also differ in their search speed, design of the search interface and the way in which they display results.

TYPES OF SEARCH ENGINES

There are a number of search engines. Most of them are available free of cost, but few are on payment basis (pay per click e.g. Karnak, Tracerlock, Informant, etc.) available to its subscribers through user id or IP address of the institution. Search engines can be classified into the following categories on the basis of their indexing, retrieval systems and other characteristics:

Simple Search Engines (Crawler based):

Crawler based search engines create their listing automatically with the help of "spider" or "robot". If you modify your webpage, the crawler based search engines eventually find these modifications/ changes, and that can affect on listing. Title of the Web page, headings, body content, font size, hyperlinks and other elements all play important role in ranking of the webpage. There are three major elements to crawler based search engines: spider, index and the search engine software.

Example: AltaVista, Google, MSN.

Subject Search Engines

These are search engines dedicated to specific subject areas like Medicine, Science, Law etc. For example:

Science: Biocrawler; Sciseek; Search4science.

Medical Search Engine: Medexplorer, Mednets,

PubMed, Medlineplus, WRAPIN,

MedHunt, WebMD.

Legal Search Engine : WestLaw, Quicklaw, Lexis.

Meta Search Engines

Meta-search engines don't have its own database, but they provides search results from other public search engines, collate those results and present the combined result before the user. They search a number of search engine databases together.

Example: Metacrawler (9 SEs) Webcrawler, Mamma, Dogpile; Savvysearch (28 SEs); Excite.

Web Directories (Human Powered Directories):

Information is available in selected subject directories from broader categories to specific one. It is created by human editors rather than robots or crawlers. They index selected websites after proper review.

Web directory is a service that offers a collection of links to internet resources submitted by the creators or evaluators and organized into subject categories. For example Internet Public Library, INFOMINE, Virtual Library, dmoz, etc. are very useful directories for academic purpose.

The editors write the review for web site or web site itself submits a short description to the directory for entire site. Changes in the web pages will have no immediate effect on listing. Things that are useful for improving a listing with a search engine have nothing to do with improving a listing in a directory. The only exception is that a good site, with good content, might be more likely to get reviewed for free than a poor site. Example: Yahoo Directory.

Hybrid Search Engine (Web Directories + Search Engines):

In hybrid search engines both facilities are available of search engines as well as directories. They present both type of results together: crawler based results and human-

powered listing. A hybrid search engine will provides basic search results with information in web directory.

Example: Google, Yahoo.

The search results in Yahoo provide normal output with its location in directory.

For example: If user search "Jain Religion" in Yahoo, then results with thousands of hits with category Religion>Jainism.

BASIC SEARCH TECHNIQUES

To retrieve information from the web, different basic search techniques are used by users. The exact search techniques and the corresponding search operators will depend on the search system chosen. The following are the basic search techniques applied in search engines for information retrieval:

A. Simple Search

A search can be conducted by entering a single search term or a phrase comprising more than one term. The keyword search is the simplest form of search facility offered by most of the search engines. The query can be quite simple, a single word at minimum.

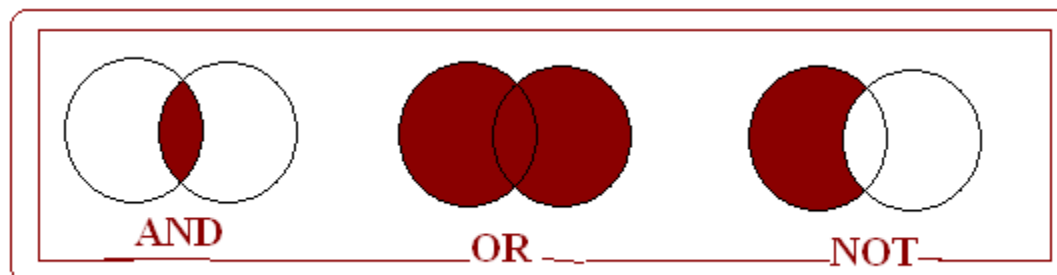
Search

B. Boolean Search

Building a more complex query requires the use of Boolean operators that allow you to refine and extend the terms of the search. This is a very common search technique that combines search terms according to Boolean logic using AND, OR and NOT as operator with two or more search terms.

- **AND** - All the terms joined by "AND" must appear in the pages or documents. Some search engines substitute the operator "+" for the word AND.

Diagram 1: Presentation of Boolean Operations



- **OR** - At least one of the terms joined by “OR” must appear in the pages or documents.
- **NOT** - The term or terms following “NOT” must not appear in the pages or documents. Some search engines substitute the operator “-” for the word NOT.

The following diagrams illustrate the basic Boolean operations.

Using Quotation Marks “ ” or Phrase search

Most of the search engines allow searcher to enter a phrase within double quotes. In the advanced search interfaces of search engines, users are required to select the options for phrase searches. With the help of “phrase search” users can get much better results. The result shows only those items which contains exact phrase only.

Digital Library of India

Truncation

Truncation is a search facility that enables a search to be conducted for all the different forms of a word having the same common root. As an example, the truncated word ‘Librar*’ will retrieve items containing the terms ‘Library’, ‘Libraries’, ‘Librarian’, etc.

Concept-Based Searching

Using the idea of a thesaurus, a search engine can expand upon the keyword that a user may input. In this manner, users do not have to know the exact words to use to retrieve relevant documents.

Information search and retrieval is of major importance in locating relevant materials. The ability to aid and assist a user in finding relevant information is the goal of librarians and information scientists. On the Web, search engines have made the process easier by incorporating a number of newer techniques which include artificial intelligence, Bayesian statistics and probability theory, weighting, and query by example. With the goal of finding relevant materials, these new techniques locate information and also refine the search query. Since search engines have different criteria in creating the indexes, it is most useful to use more than one engine in searching the Web to gain relevant information. With advances in the tools for information retrieval, the future holds exciting possibilities for searching on the World Wide Web.

CONCLUSION AND SUGGESTIONS

Internet is essentially an Information Storage and Retrieval System characterized by its enormous size, hypermedia structure, and distributed architecture. In this age of competition all search engines are trying to improve their features day by day but not disclosing their indexing method as well as ranking algorithms. Popularity of search engines depends on usefulness of retrieved information to the users as well as simplicity in searching of information.

As per study by BBC (UK), they found Google generation (borne after 1993) users are very fast in searching of information from web using search engines but they do not know the basic search techniques. So, the Librarian can teach the users specially the young users of Google generation about the search techniques for effective information retrieval as well as evaluation of online information resources.

REFERENCES

- What is a web search engine? <http://www.columbia.edu/acis/fifteen/search/search/sld02>
- Jain, V., & Prasad, H. N. (2003). The role of search engines in information retrieval from Internet. *Proceedings of Society for Information Science (SIS-2003) Roorkee*, 372-379.
- Indian Search Engines. Retrieved from http://www.stylusinc.com/Indian_search_engines
- Search Engine Showdown. Retrieved from www.searchengineshowdown.com
- Sullivan, D. (2002). How search engines work. Retrieved from <http://www.searchenginewatch.com/webmasters/article.php/2168031>
- Lucas, W., & Topi, H. (2001). Form and function: The impact of query term and operator usage on web search results. *Journal of American Society for Information Science and Technology*, 53(2), 95-108.
- Chowdhury, G. G., & Chowdhury, S. (2001). *Information sources and searching on the World Wide Web*. London, Library Association.
- Griffiths, P. (2000). *Managing your Internet and Intranet services: the information and library professional's guide to strategy*. London, Library Association. pp. 111-123.
- Jain, V. (2011). *Retrieval Efficiency of Internet Search Engines*. Saarbrucken, LAP Lambert Academic Pub. Pp.15-36.
- Nicholas, D., Rowlands, I., Clark, D., & Williams, P. (2011). Google Generation II: web behaviour experiments with the BBC, *Aslib Proceedings*, 63 (1), 28-45.
- Subject Directories : <http://libguides.utpb.edu/subjdir>