

## APPLICATION OF SEARCH ENGINE IN DIGITAL LIBRARIES

Rakesh Kumar Mishra  
Central Library, D.D.U.Gorakhpur University,  
Gorakhpur

### Abstract

*In this paper we have to discuss about the meaning of the search engine, how search engine works, different categories of search engine, various search techniques, and important search engines. Actually Search engines are programs that search documents for specified keywords and returns a list of the documents where the keywords were found. A search engine is really a general class of programs; however, the term is often used to specifically describe systems like Google, Bing and Yahoo! Search that enable users to search for documents on the World Wide Web.*

**Keywords:** *FTP, WebCrawler, Database, Gateways, Boolean Operator*

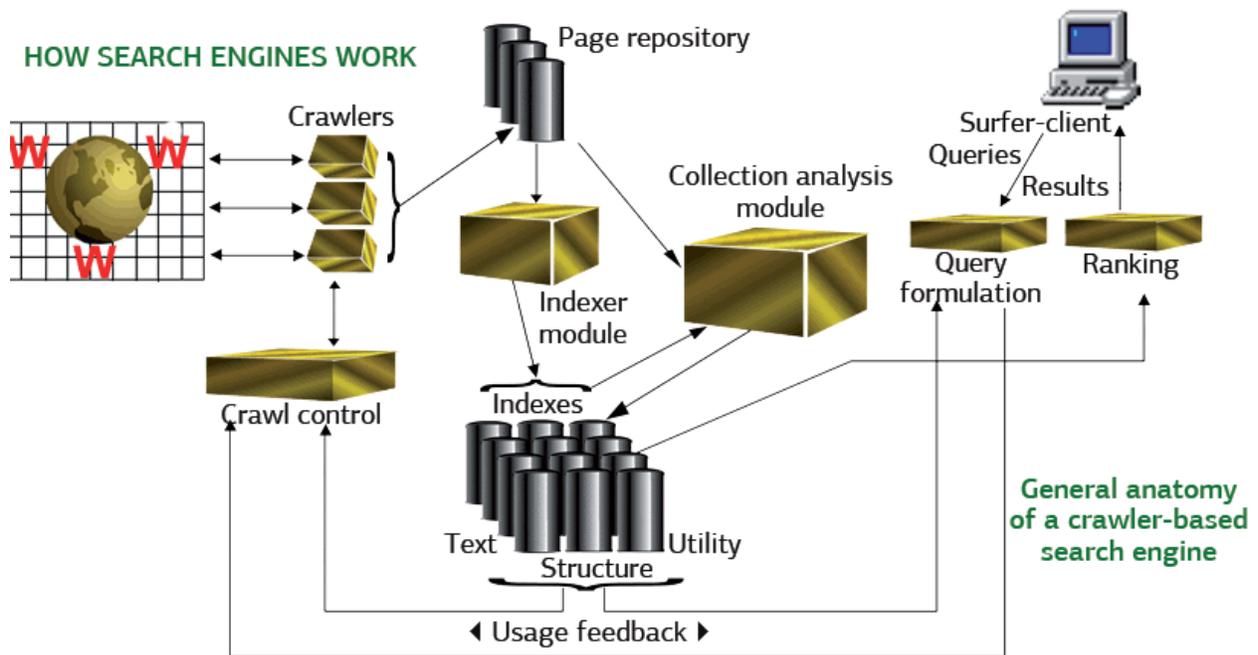
### INTRODUCTION

The Archie, developed in 1990 by Alan Emtage, a student of McGill University in Montreal, can be considered the first search engine that was used for indexing and searching files on file transfer protocol (FTP) server, which is the service of the internet. VERONICA was developed at the university of Nevada to search all menu items on Gopher server. Excite was developed in 1993 which is used for

statistical analysis of word relationship to make searching more efficient. WebCrawler developed in 1994 that indexed entire page. Lycos, Infoseek and Open text is one of the important search engines, which was appeared soon after Web Crawler was launched. Google was launched in 1998, which is most powerful search engine to till date. Search engine is a generic term used for the software that searches the web for pages relating to specific query. Actually search engine is a computer program that searches documents on the internet containing term being searched by a user and it can be defined as a tool for finding, classifying and storing information on various websites on the Internet. Different search engines work in different ways. Search engines are also defined as online utilities that quickly search thousands of web documents for a word or phrase being searched. Some search engines are case sensitive i.e. it discriminate upper case from lower case. Search engines are usually accessed using web clients (web browsers).

### WORKING OF SEARCH ENGINES

Search engines is not really the search of WWW, instead they search their databases consisting of keywords or full text of web pages. Typical search engine has the following three components.



**a. Robot or Spider or crawler**

This is a computer programs that traverse the World Wide Web (WWW) information space. Crawler uses the HTTP to connect to remote web pages. Main function of Crawler/Spider is indexing of web pages, link validation, mirror websites & identifying new information.

**b. Databases**

This is nothing but the collections of information like URLs, abstract, even full text etc. Crawler harvesting the information from databases by using certain phrases, keys, etc.

**c. Agent or User Interface**

It is a software program that accepts queries from users and searches them through database (repositories of pages). The user interfaces matches the query with the database, finds hits and rank them in order of relevance.

**CATEGORIES OF SEARCH ENGINES**

Different types of search engines are exist based on their types of information user want to retrieve. Mainly four types of search engines are categories, which are as follows-

**1. Primary Search Engines**

It deploy computer programs called web crawlers or spider to traverse the web and scan websites for words, phrases or whole site so as to generate a database of web pages. This is one the most commonly used search engines. Primary search engines vary in terms of database size & content, syntax used (word search, Boolean search or phrase search). ‘Any word’, ‘all words’ and ‘Exact phrase’ are mainly three basic options available in primary search engines.

**2. Meta Search Engines**

This search engines are called as mega indexes and not used for exhaustive search. Meta search engines is the search engine that searches the databases of several other search engines at the same time to locate web pages that match search term given by user. Google, AltaVista, Excite, Lycos are most common Meta search engines used to retrieve the information.

### 3. Web or Subject Directories

It is the yellow pages of the Internet and Yahoo is the most common subject directory. Directories can be searched using keyword and it differs from indexes. Subject directories retrieve relevant results and increase the possibility of finding high quality and reliable websites.

### 4. Subject Portals or Gateways

It is mainly Meta resources, subject-based gateways, virtual libraries guided to Internet resources which allow easier access to network-based resources in particular subject area. Subject portals are often independent web sites or library's websites. LibrarySpot, Librarian's Index to Internet, etc are the common examples of subject portals.

Ease of use, Comprehensiveness, Quality of Content, Control over the search, Flexibility in searching, Information presentation of results and Assessment of Relevance are one of the criteria which may be used to assess the usefulness of a search engine in all above categories of search engines.

## SEARCH TECHNIQUES

Mainly two types of interfaces are used to search the databases are first **basic** and second one is **advanced** searching. In Basic Search user enters only "Keywords of search term" and search engine looks through its indexes in the database for matches. To refine the search users uses different types of operators like Boolean operators, Phrase Searching, Proximity Searching, natural language searching etc.

**Boolean Operators** allow uses of AND (+), OR (\*) and NOT (-) so that user can retrieve relevant search results. For example *Library AND Automation* give the result that contains both words. Similarly *Library OR Automation* will retrieve pages that include any of the term in the search query. *Library AND Automation NOT Gorakhpur*, will retrieves pages that exclude a word specified after 'NOT'.

In **Phrase Searching**, search engine only retrieve the information that are included in the Phrase. It is very important search engine, which can be used to increase the chance of retrieving relevant results. For example "**RFID Based Circulation Services in Libraries**", it will search exactly only term which is included in double quotation marks.

In **Proximity Searching**, proximity operators are used to specify the relative location of words in a document. This operator helps us to search for words within a certain distance of one another in databases. Symbols 'w' (with/Within) & 'n' (near) are mainly used in this types of searching.

Library 'near' Automation would retrieve documents containing "Library Automation"

and ‘Automation of Library’. (Library Computerization or Library ‘near’ Automation) and Gorakhpur is particularly useful for clustering synonyms or for searching specific terms before other terms are searched.

In *Truncation & Wildcards*, search engine search for multiple endings of a word. Symbol called as wild cards \*, ? or # are the end of the word root not to be indicated a truncated search. For Example catalog\* retrieves catalog, cataloging, cataloguer, catalogued. Dir example lib\*y retrieves library or liberty.

The *Natural Language Searching* allows searching in the same language as it is spoken. For example suppose user want to know what is library digitalization, then his/her query should be written as “*What is library Digitalization*”. According to this query search engine will search the relevant web pages that would answer this question.

## IMPORTANT SEARCH ENGINES USED TO SEARCH DOCUMENTS IN LIBRARIES

There are various search engines as discussed above, but the practical details of all search engines described are as follows-

### 1. Primary Search Engines

- a. **Google** (<http://www.google.com>) is a next generation search engine and its biggest strength is the size of its database and link-based relevance ranking. It supports Boolean searching, proximity searching, field searching, character searching, numbers and number range searching.

- b. **MSN Search** (<http://search.msn.co>) is the search engine for the MSN portal site. It supports Boolean searching, proximity searching, truncation, field searching, and restriction to media/file types.
- c. **Fast Search and transfer** (<http://www.fastsearch.com>) created the search engine and database for AlltheWeb and Lycos. It consists merely of a navigation bar and a search entry box and a number of links at the page bottom.
- d. **Alta Vista** (<http://www.altavista.com>) is a leading provider of search services and technology. It was the first search engine to launch image, audio and video i.e. multimedia search capabilities, Web pages are evaluated for relevance i.e. its ranking system in not as effective as that of other search engines.
- e. **Lycos** (<http://www.lycos.com/>), in this search engine searches can be restricted to language and desire locations in web pages i.e. title, description or URL.

### 2. Web Directories

- a. **Yahoo!** (<http://www.yahoo.com>) is most popular internet portals. It support Boolean searching, proximity searching, field searching, limit to language, date, filetype. Truncation is supported by this search engine. Retrieve results are stored by relevance algorithm.

- b. **Galaxy** (<http://www.galaxy.com>) is a hierarchical, topically organized search engine. It provides for restricting the search to title, description, body or other locations in a page. Boolean operators (AND, OR, NOT) are used in galaxy search engine to retrieve the information.

### 3. Meta Search Engines

- a. **WebCrawler** (<http://www.webcrawler.com>) is Meta search technology which highlights and uses the strengths of major search engines to deliver more relevant and comprehensive results. It also make easy to refine the search so that a user can find the most meaningful results right away.
  - b. **MetaCrawler** (<http://www.metacrawler.com>) is a search engine that has no internal databases. It sends the query from a user to the search engines, and then put them into a uniform format for display.
  - c. **Ask Jeeves** (<http://www.askjeeves.com>) instead depends on crawler-based technology to provide results to its users.
4. **Subject Portals** is the most important directories of Internet resources in particular subject areas, compiled and organized by specialists and LIS professional.
- a. **LibrarySpot.com** (<http://www.libraryspot.com>) is free library resource centre for librarians and their users, students, educators, etc for

information. It is the first vertical information portals designed to make finding the best topical information on the internet quickly & easy.

- b. **Librarian's Index to the Internet** (<http://lii.org>) is one of the important portals for the benefits to users of public libraries. This site provides both browsing and searching interfaces.
- c. **Virtual Library** (<http://www.vlib.org>) is the oldest catalogue of the web. The virtual Library pages are widely recognized as being amongst the highest-quality guides to particular sections of the web.

Other subject portals like BUBL, BIOME, Edinburgh Engineering Virtual Library, Social Science Information Gateway, Academic Info etc are the most popular portals for searching the information for particular related subjects.

### CONCLUSION

In this Internet age search engines play a vital role for harvesting the information for their research and development. One can search any types of information in any form by using the appropriate techniques of search engine. Search engines are websites that facilitate users to search resources on the Internet.

### REFERENCES

1. Bates, M.E. (1997), The Internet: Part of a Professional Searcher's Toolkit, Online, 21(1).

2. Tyner, Ross: Internet Search Tools and Techniques, 2001.
3. How to use Web Search Engines (<http://www.monash.com/spidap4.html>).
4. Dong, L.T. and Su, L.T. (1997), Search Engine on the World Wide Web and Information Retrieval from the Internet.
5. Brandt, D.S. (1996). Relevancy and Searching the Internet

### **Author Profile**



Dr Rakesh Kumar Mishra working as Programmer-Cum-System Incharge (Information Scientist), Central Library, D.D.U.Gorakhpur University,Gorakhpur(UP)-273009 since 10/04/1997 to

till date.I have done M.C.A., [M.L.I.Sc](#), & Ph.D.(Topic: Some contributions of data mining techniques in management decision in libraries). Dr Mishra is a author of 6 books, published several articles in edited books, attended/presented a papers in several national/international seminars/conferences and published several papers in national/international journals. Dr Mishra participate two times in INFLIBNET, Ahmedabad for training purpose.Dr Mishra has vast experience of teaching of UG/PG (Compter Science and Library Science Displines) in Gorakhpur University as well as paper setters and evaluators. Dr Mishra worked as a ACS in Evaluation Centre in Gorakhpur University in many times. Dr Mishra has life membership of Uttar Pradesh Library Association(UPLA) as well as act as chairman of the selection committee at NIELIT, Gorakhpur.