Chapter-III

World Wide Web Search Engines: Strategies for Effective Information Retrieval

1. Introduction
2. What is a Search engine?
3. Types of Search Engines
4. Major Search Engines: An Overview
5. Comparative Features of Major Search Engines
6. Meta Search Engines
7. General Guidelines for Successful Search
8. Sample Queries and the Test Environment
9. Conclusion
1. Introduction

The growth of World Wide Web has been one of the most remarkable developments, which has rapidly gained popularity and become most widely used application of the Internet. Today, the World Wide Web has emerged as the most powerful medium for information publishing and access. A plethora of information sources for education and research are available on the Web including scholarly journals, technical reports, theses, courseware, content page, data set, patents, reference sources, discussion forums etc. The publicity of World Wide Web has gained so much importance that many people all over the world naively equate World Wide Web with the Internet. The friendly user interface and hypertext features of Web have been attracting a significant number of users as well as information providers. As a result, the Web has become an ocean of all kinds of information or data, making any query into the huge information reservoir extremely difficult. In order to overcome this difficulty in retrieving information from WWW, several companies and institutions quickly developed various search aids called as search engines.

However, the search engines require different search strategies to retrieve information on the Internet. Searcher ability to find the information on the Internet is a function of how precise his queries are and how frequently he uses search strategies for different search engines. These issues play a very important role in retrieving relevant information on the Web. There are several effective ways to use special operators and search strategies to target the result.
In this regard this chapter discusses the different major search engines and their respective search strategy used for successful information retrieval on the Internet.

2. What is a Search engine?

Search engine is a tool, which helps in the retrieving information from the Internet. It is programmed in such a way that it indexes the Web and accordingly builds their databases. When the query has entered in the search engine, it checks its index with the query. Then relevant matches are retrieved and returned as ‘hits’ or ‘search results’. In other words search engine acts as a searchable index of Web pages of the world.

2.1. Primary Components of Search Engines

When using a search engine, the user is searching a database of indexed Web sites. All search engines have three primary components:

a) Search engine spider

b) Search engine index/database

c) Interrogation/retrieval software.

a) Search Engine Spider

The spider visits a Web page, reads it and then follows links to other pages within the site. The spider returns to the site on a regular basis depending upon the individual search engine policy, to look for changes.
Figure-3.1: Search Engine Spider

b) Search Engine Index or Database

This is the main element of any search engine—it is what the user interrogates. Once it could be said that these indexes were built along similar guidelines, with the location and frequency of words the primary determining factors in results relevance ranking. However, during 1998, a number of new search engine providers appeared. These companies built their indexes according to differing criteria. For example the Direct Hit index is based on the popularity of a Web site. Google index is based on the number of links between pages and sites, whilst the Real Names index is a pay-for service that enables companies to register keywords to protect their brands and company identity.
c) Interrogation or Retrieval Software

All search engines have their own customised software to interrogate their databases. Essentially, though, they operate according to similar principles: any Web site which contains words or terms that match the user’s search query will be presented in the list of results presented on screen to the user. Ranking each of these matching Web sites by relevance is determined by algorithms that analyse the location and frequency of the user’s search terms against this list of matching Web sites. The nuances of how these algorithms works varies between search engines, which is one reason for the different results that users usually experience when running the same search across different search engines. However, much more important reason for these search results differences is that the [content] overlap between the engines remains relatively low.

3. Types of Search Engines

Search engines can be divided into four categories: robots, directories, metasearch engines and software tools. Inevitably, some search engines combine characteristics of more than one of these categories. Some authors have also classified the engines in a very similar way: as robots, directories, metasearch engines, geographic specific resources and subject specific resources.

3.1 Robot-driven Search Engines

These robots are programs that reside on a host computer and retrieve information from sites on the Web using standard protocols. In effect, they automatically travel the Internet following links from documents and collecting information according to the HTML (Hypertext Markup Language) structure of the document (such as Universal Resource Locator also called as URL, document title, keywords in the text) about the
resources they come across. No doubt the different search engines follow different algorithms to index the information on the Web and to output results to a users’ query. However, it is also noted that users are often unaware of the comprehensiveness or currency of the material that search engines have indexed.

Examples of such search engines include some of the best-known engines, such as AltaVista (www.altavista.com), Excite (www.excite.com), Lycos (www.lycos.com), HotBot (www.hotbot.com), Google (www.google.com) and Infoseek (www.infoseek.com)

3.2 Directory based Search Engines

These engines are also known as subject collections or as ‘subject gateways’ and are collections of links to relevant URLs created and maintained by subject experts, or by means of some automated indexing process. Some of these services also include a robot driven search engine facility, but this is not their primary purpose. They rely mainly on people to identify and group resources. These directories offer access to information that has been classified into certain categories. Hypertext links lead the user to the most appropriate URL. There is an implication that only high quality URLs are provided. Such engines typically search manually created structured metadata rather than full text of the target Web sites. Yahoo, SOSIG, EEVL, Biz/Ed and UK directory are examples of such directories.

3.3 Metasearch Engines

These search engines utilise databases maintained by other search engines. A metasearch engine accepts a single query from the user and sends it to multiple search engines in parallel. The array of Web databases is accessed (sometimes in parallel and
sometimes in sequence) to return the hits from each search engine. After the underlying search engines return retrieved items, they are further processed and presented to the user. The functionality of these search engines depends largely on the performance of those engines, mentioned above, that create and manage their own databases. They are attractive because users do not have to visit multiple search engines and re-enter their search terms. The most popular metasearch engines include:

MetaCrawler (www.metacrawler.com), Dogpile (www.dogpile.com), Metafind (www.metafind.com) and ProFusion (www.profusion.com)

3.4 Software Tools

These tools are sometimes classified as ‘browsing companions’ or ‘browser searchbots’. They are similar to metasearch engines but need to be installed on the user’s workstation. One of their features is that results can be saved to hard disk for subsequent retrieval. They may also be able to check for broken links and prevent duplication of results. Most of these products are shareware and require payment. Examples of such software tools include Websleuth, Coprenic 98, f.Search 1.3.1, Query-N Metasearch and Meta Hari.

A question that naturally arises is that when to use a robot driven search engine or directory based search engine or metasearch engine and software tools? If searcher is an experienced in finding information on the Web, it would be ideal to use a robot-driven based search engine to reduce the time and efforts incurred in finding the relevant information. However, if he is novice user, then he may start with a directory and, as he become experienced, he could switch to robot-driven search engine. Another situation in which searcher may use a directory is that when he is unfamiliar with the relevant
keywords of a particular topic and/or a situation that requires him to explore as much information as possible on the topic. There is some overlapping between search engines and directories. However, in both the types of search tools there is a provision for the keyword search as well as category-based search.

Some time when users are not getting the desired results or when the results are inadequate, they probably try with another search engine. Even when that does not yield results to users satisfaction, they would try another search engine, and so on. This process is laborious and time consuming. Fortunately there is an effective way of using multiple search engines and that too simultaneously! By using what is known as metasearch engine, they can submit their search query to more than one search engine. What the metasearch engine does is that it takes searcher search query as the input and submits it to several search engines.

Each search engine then gives its result or output, in a collated or uncollated format, back to the metasearch engine, which in turn displays the results. Searchers not only saves Internet connection time and also telephone cost, but also eliminate the effort of accessing individual search engines separately and typing the same search query again. However, the disadvantage of using metasearch engine is that searcher search query syntax may not be supported by all search engines to which the metasearch engine submits your query.
Chapter III  World Wide Web Search Engines: Strategies for Effective Information Retrieval

Figure 3.2: Relationships among Different Types of Search Engines

- **Tangibles:**
  - Information well organised
  - Different search methods available
  - A large amount of information available
  - Can narrow search topics

- **Reliability:**
  - Good syntax consistency for the key words in searching
  - Search results are relevant to the query

- **Responsiveness:**
  - Search results are provided quickly

- **Assurance:**
  - No repetition of pages/sites
  - No dead links
  - Information is up-to-date

- **Empathy:**
  - The layout upon first impression is easy to understand
  - Offers natural language searching
  - There are help screens, introductory pages or sample searches to guide users
  - Offers language selection for documents written in specific language

Figure 3.3: Fourteen Items of Customer Requirements on Search Engine
4. Major Search Engines: An Overview

The staggering volume of information on the Web is loosely held together by hyperlinks. Every day about a million electronic pages are added to the World Wide Web. This is impressive but chaotic growth of Internet has evolved into a global mess, making the job of high quality information retrieval a mind-numbering task. A super hub called search engine maintains a list for every word of all known Web pages containing that word. Such a collection of lists is known as ‘index’. When a search button is pressed only this index list is accessed, not the entire World Wide Web. The link returned by search engine that is actually a stale copy of the Web page. Only when the link clicked the current page is retrieved. So in this regard an attempt has been made to discuss the search features, strategies and other characteristics of different search engines.

4.1. Yahoo (www.yahoo.com)

Yahoo is probably the most popular and best known Internet resource. Yahoo search engine was launched in late 1994. It has become one of the most popularly used search engines on the Web. It is the Web’s largest human compiled directory, listing over more than one million sites. These directory listings are also supplemented by search results derived from Inktomi’s 110 million-page search index.

Yahoo is divided into fourteen broad subject categories, including arts, business and economy, education, government, health, reference, science etc. It also includes categories for new Web sites and includes a section ‘just launched on the Web’ to publicise new Web sites. The feature ‘My Yahoo’ allows users to create their own personal yahoo guide to favorite sites and topics of interest. Additionally, yahoo is developing regional guides for some commentaries (New York city, San Francisco etc) as
well as local guides for some other countries (France, German and Japan) in their respective languages.

4.2. Google (www.google.com)

Google search engine was launched in 1998 and it was developed by students at Stanford University. Google focuses on the link structure of the Web to determine relevant results for the user. It’s proprietary technology, PageRank (named after co-founder Larry page), crawls the Web analysing both the links between Web sites and the accompanying text around each hyperlink. The company estimates that its index is between 70-100 million pages, but through the links analysis enables users to reach an estimated 300 million Web pages which is currently a much greater reach than any search engine provider.
Like most other second-generation search providers, the company is focusing on co-branding its technology rather than building its own search portal. In August 1998 the company signed a deal with AOL subsidiary Netscape to be main search provider on the Netcentre portal.

4.3. **HotBot (www.hotbot.com)**

Launched in May 1996 by Wired and it was acquired by Lycos in October 1998, but continues to be run as a separate service from the Lycos search engine. Accesses the Inktomi search engine index, rather than compiling its own index. However, primary results are derived from Direct Hit, the popularity-based search provider. Directory listings are derived from the Open Directory.
4.4 Alta Vista (http://www.altavista.digital.com/)

Alta Vista began to be developed in the Summer of 1995 at Digital's Research Laboratories in Palo Alto, California, and was formally delivered to the Web on December 15, 1995. It indexes the full text of over 16,000,000 Web pages (by January 1996) with unspecified update frequencies. According to its documentation, Alta Vista can fetch 2.5 million pages a day following the Robots Exclusion Standard, and index 1 GB of text per hour. Alta Vista supports Boolean searching, term as well as phrase searching (i.e., proximity searching with the NEAR operator), field searching (eg, title:steelhead; url:home.html), right-hand truncation with some restriction, and case-sensitive searching if only the first letter of a word is capitalised.
Alta Vista provides three display options: compact, standard, and detailed. Although the latter two are the same, the display order or relevancy ranking of search results is determined by the location (e.g., in title or the body of text) of matching words, occurrence frequencies of matching words, and distance (i.e., how many words apart) between the matching words. However, only the first few words of a document found are displayed, which may limit users' ability to judge its relevancy without referring to the full version of the document. In addition, general search terms such as "computer" and analysis are automatically ignored in Alta Vista.

4.5 **Excite (http://www.excite.com/)**

Excite was developed by Architext Software and it was launched in late 1995. This search engine was immediately popular with users due to its large index and
integration of non-Web material such as company information. It claims 1.5 million fully indexed Web pages, and its index is updated approximately once a week\textsuperscript{8}.

Excite allows keyword searching as well as concept searching since it is able to determine related concepts from document collections, eliminating the need for external manually-defined representations such as thesauri. An example of concept searching given by Excite is that a search query about intellectual property rights will retrieve all documents about the topic even if terms such as software piracy or copyright law rather than the actual matching words appear in the document. In other words, the search engine itself handles synonyms and related terms, taking the burden of vocabulary control off users' shoulders. As for keyword search, query terms are both AND'ed and OR'ed in each
search, but a higher weight is given to results with terms AND'ed. However, Excite does not support at present other advanced search options than those being described already.

Equipped with automatic abstracting capability, Excite is able to generate an abstract for each of the Web pages it indexes, which is a very unique and fine feature that many of its counterparts do not have. But, there are no different formats for displaying search results. In addition, its online documentation appears somewhat unorganised.

4.6. Lycos (http://www.lycos.com/)

Lycos, representing the first 5 letters of the Latin name for wolf spider, was originally designed at Carnegie Mellon University and launched as a search engine in May 1994. It was later sold to America Online and became Lycos, Inc. at which Michael Mauldin, the person who has overseen Lycos’ development is still a full-time employee. Although commercialised, Lycos continues to provide free services to the Internet community.

By the end of January 1996, Lycos has indexed over 95% (ca. 19 million unique URLs including FTP and Gopher) of Web resources, making it the largest Web search engine in its family. Nevertheless, it does not index the full text of a Web page. Rather, it only extracts the title and a portion of a document (e.g., the smaller of the first 20 lines or 20% of the document). This practice has been singled out by Lycos’ competitors as its most salient weakness. Around 50,000 documents are added, deleted, or updated in the Lycos index everyday.
Lycos supports Boolean logic, and furthermore, it incorporates that feature in such a way that the users do not have to type the Boolean operators when conducting a search. For example, one only needs to select the search option "Match all terms (AND)" to use the AND operator. Another search feature Lycos provides is to match query terms against Web documents at 5 different levels, namely, Loose match, Fair match, Good match, Close match, Strong match. Nevertheless, no specific explanation is given as to how the different levels of match are determined. Truncation is automatically done in Lycos during a search, which may result in some unwanted search outcome. Phrase search is not supported by Lycos so any queries with phrases cannot be appropriately executed.

On the other hand, Lycos implements a wide variety of display options. Users are given the choices of viewing 10, 20, 30, or 40 research results at a time. In addition, each
search result can be displayed using the summary, standard, or detailed format. The detailed format corresponds with the long abstracts Lycos prepares, which include URL, title, outline, keys, abstract, description, date, and other related information. The summary format contains what Lycos' short abstracts have - URL and descriptions. In terms of coverage, the standard formats lies somewhere between the summary and detailed formats. The online documentation available at Lycos' Web site describes the composition of each output segment (e.g., outline and keys) in detail.

4.7. Ask Jeevs: (www.askjeevs.com)

It was launched in June 1998 as the first natural language search agent on the Internet. It operates by matching a user’s query against a database of 7 million template questions, presenting variant questions if there is no match. It will also conduct a metaserach across Altavista, Go (Infoseek), Lycos and Yahoo. It has been licensed by Altavista for its own search site www.askjeevs.com (See figure-3.10)

4.8. Direct Hit (www.directhit.com)

Direct Hit search engine was launched in April 1998. It offers co-branded search solutions to other search engine providers. It operates by providing a second level ranking of the user’s search results on the basis of popularity. The company currently licenses its technology to ten search sites including AOL, HotBot, Lycos, MSN and LookSmart. In August 1999 the company announced that it had raised almost $27 million from venture capital firms and private inventors (See figure-3.11).
Chapter-III  World Wide Web Search Engines: Strategies for Effective Information Retrieval

Figure 3.10: Ask Jeeves Home Page

Figure 3.11: Directhit Home Page
4.9. FAST (www.alltheweb.com)

Norwegian Company launched in May 1999 with the largest ever search engine index at the time-over 200 million pages. It ambitiously aims to index all of the Web—hence it’s URL (Universal Resource Locator) is www.alltheweb.com. Unlike other search engine companies, who use mainframe computers to power their services, FAST has linked together a few hundred Dell PCs (Dell has a 5 percent stake in the company) and uses parallel processing to deliver its service.

![Alltheweb Home Page](image)

Figure 3.12: Alltheweb Home Page

4.10. Go (Infoseek) (www.go.com)

The Infoseek search engine was launched in 1995. The Disney Corporation acquired a large stake in Infoseek in June 1998, and in January 1999, Infoseek, was re-launched and re-branded as a portal site known as Go. Like many other search portals,
Go offers users the option of searching the index or browsing through a human-compiled Web directory.

![Figure 3.13: Infoseek Home Page](image)

4.11. Inktomi (www.inkomi.com)

Founded in February 1996, Inktomi is probably the most famous search engine index. It powers the search results for several famous portals and search sites including HotBot, Yahoo, AOL, MSN Search and SNAP. However, not all of these companies access Inktomi’s full 110 million page index and there are variations in results between the different search sites due to the different filtering and relevance ranking algorithms Inktomi provides to each partner company. It is not possible to interrogate the inktomi index directory.
4.12. LookSmart (www.looksmart.com)

Launched in October 1996, like Yahoo, LookSmart is a human-compiled directory. In addition to providing its own search site, the company also licenses its directory to other companies including Alta Vista (who in turn provide search results to LookSmart whenever there is no match to a user’s query within the directory) and in August 1999 with Excite (replacing excites own directory). During the same month, the company raised US$ 92.4 million on its public listing of 7.7 million shares at US$ 12 each.
4.13. **Northernlight** ([www.northernlight.com](http://www.northernlight.com))

Northern was launched in August 1997 has continually been one of the largest indexes, gradually increasing in size until it became the biggest search engine. This leading position has since been superseded by the launch of FAST in May 1999. The company also offers a special collection of non-Web material such as newspaper and magazine articles. Whilst it is free to search within the special collection, users must pay a charge (up to $US 4) to view any articles from this collection. Search results are clustered in folders by topic. Like Altavista, this search engine is popular with researchers due to its scope and functionality.

This search engine was launched in June 1998. This directory uses volunteer editors to catalogue the Web. This initiative quickly gained prominence and was acquired by Netscape later in November of that year. Netscape pledged to allow anyone to use the directory service. In April 1999 Lycos re-launched itself as a directory service, deriving its primary results from the Open Directory.
4.15. **RealNames** ([www.realnames.com](http://www.realnames.com))

Launched in 1998 and formerly known as Central crop. RealNames charges companies an annual US $100 to register individual key words, such as company name, or a brand name. Obviously many companies want to, and do, register many keywords to protect their brands etc. This has proved a very successful economic model for the company and in August 1999 it successfully raised over US$ 70 million from venture capitalists in a third round of financing. Although the index is directly available as a download from the company’s Web site and is incorporated within Microsoft’s Internet Explorer 5 browser, its most notable success has been access from search engines that license its index, such as Altavista and Go ([infoseek](http://www.infoseek.com)).
4.16. Magellan

Magellan is an online Internet guide and search engine. The Internet guide contains a directory of rated and reviewed Internet sites as well as a database of sites awaiting review. Sites are rated on depth, ease of use, and net appeal. Magellan includes Web sites, FTP, and Gopher servers, newsgroups and Telnet sites in its database and searches. Magellan offers a browse capability within its database through a directory and subdirectory hierarchy which is divided into 26 broad categories including arts, communications, daily living, entertainment, food etc. Searchers may be limited to only the rated and reviewed database or can include the entire database. Magellan has been translated into French and German and will later be available in other languages. Additional information may be on the World Wide Web at http://www.mckinley.com/

4.17. Open Text Index

Open Text Index is a search engine with two options: simple search and power search. The simple search engine is a single output field, which allows users to enter a group of words or single phrase without further qualification. This is similar to many other search engines. The power search options allows users to choose where in Web pages they want to search for words and phrases. For example, a user can specify that he wants to search everywhere in Web pages or to limit the search by summary, title, first heading or URL. In addition, the power search option allows the use of proximity operators (and, or, but not near) to combine words and phrases in multiple data entry fields, each of which can be limited as described above. Open text index is currently available in several languages, including Japanese, Portuguese and Spanish. As do many of the others services described here, Open Text Index offers additional specialised
search services including searching current events from a variety of newspapers, searching newsgroup postings, and e-mail address locators. It also offers several lists of 10 to 12 selections within a variety of categories such as cartoons, columnists, and cool sites. Information on open text index may be located on the World Wide Web at http://index.opentext.net/

4.18. WebCrawler

WebCrawler is a search engine and directory operated by Excite Inc. It builds a selective but comprehensive index of the World Wide Web, eliminating combinations of letters and numbers and certain common words, like ‘www’ or ‘Web’, from the index to help keep it reasonably small; such words are considered not informative enough in a query. The basic search features search this index using 'natural language searching' of words or phrases in plain English. The advanced feature also offers a range of Boolean operators. Any or all search terms are matched and documents are returned with a ranking indicating how relevant that document is to a user’s search. Documents with higher relevance scores appear higher in the list. A browse feature, WebCrawler selects, provides access to directories and subdirectories of selected ‘best’ resources, each with a review produced by the WebCrawler editorial team. The directory is divided into fifteen broad categories, including arts and entertainment, business, computers, daily news etc. Information on WebCrawler may be located from http://www.webcrawler.com

4.19. Deja News

Although Deja News is not a search engine or catalogue but a World Wide Web interface to Usenet newsgroups which allows users to search Usenet postings as well as read and post Usenet newsgroups. Users can search newsgroup postings and use author
profiling to see to which newsgroups a particular person has posted. Deja News provides a Usenet browser for users to search newsgroup. Deja News archives Usenet postings and users can search them by keyword, filter for certain criteria or use query profiling to locate newsgroups where specified keywords appear most often. Information on Deja News may be located at www.dejanews.com/

4.20. Galaxy

Galaxy is search engine and directory. As other search engines the Galaxy also divided into broad categories including business and commerce, community, government, humanities, law etc. Professional information specialists were hired to organise Galaxy and oversee the classification process. Only pages that have been submitted to Galaxy are listed within the directory. With the search engine, users can search all Web pages refereed by Galaxy on all text, title, or link text matching on any or all search terms. In addition to Web resources, Galaxy may also be used to search the Galaxy pages index, Gopher titles, and telnet resources. It also provides a useful listing of only its directory pages, which contain only new items. Information on Galaxy may be located from its home page at http://galaxy.tradewave.com/galaxy.html.

5. Comparative Features of Major Search Engines

The comparative features of search engines are presented in table-3.1. It clearly shows about the selected search engines features such as size of the content, fulltext, Boolean operators used, truncation, phrase searching and use of upper or lower case letters. It is observed that different search engines have different search strategies for successful information retrieval on the Internet.
### Table 3.1: Comparative Features of Major Search Engines

<table>
<thead>
<tr>
<th>Scope</th>
<th>Yahoo</th>
<th>Google</th>
<th>Altavista</th>
<th>Excite</th>
<th>Lycos</th>
<th>Hotbot</th>
<th>Northern Light</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content Size</strong></td>
<td>-</td>
<td>1.25 billion sites</td>
<td>250M pages</td>
<td>250M pages &amp; media objects</td>
<td>50M pages</td>
<td>110M sites</td>
<td>200M sites</td>
</tr>
<tr>
<td><strong>Full Text</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Default Word</strong></td>
<td>AND</td>
<td>AND</td>
<td>OR</td>
<td>OR</td>
<td>and</td>
<td>AND</td>
<td>and</td>
</tr>
<tr>
<td><strong>Boolean Connectors</strong></td>
<td>AND, OR</td>
<td>Limit including and excluding words</td>
<td>AND, NOT, NEAR</td>
<td>AND, NOT</td>
<td>or, not, adj., near, before, far</td>
<td>OR, NOT</td>
<td>Or, not</td>
</tr>
<tr>
<td><strong>Phrase Search</strong></td>
<td>Use quotation mark i.e., &quot; &quot;</td>
<td>Use quotation mark i.e., &quot; &quot;</td>
<td>Use quotation mark i.e., &quot; &quot;</td>
<td>Use quotation mark i.e., &quot; &quot;</td>
<td>Use quotation mark i.e., &quot; &quot;</td>
<td>Use quotation mark i.e., &quot; &quot;</td>
<td></td>
</tr>
<tr>
<td><strong>Truncation</strong></td>
<td>Automatic</td>
<td>Automatic</td>
<td>No, use *</td>
<td>No</td>
<td>No</td>
<td>No, ? for one right hand letter; * for left hand</td>
<td>Automatic</td>
</tr>
<tr>
<td><strong>Case Sensitive</strong></td>
<td>Yes (Upper case retrieve exact match)</td>
<td>No</td>
<td>Yes (Upper case retrieve exact match)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td><strong>Word Included</strong></td>
<td>Use +</td>
<td>Use +</td>
<td>Use +</td>
<td>Use +</td>
<td>Use +</td>
<td>Use +</td>
<td>Use +</td>
</tr>
<tr>
<td><strong>Word Elimination</strong></td>
<td>Use -</td>
<td>Use -</td>
<td>Use -</td>
<td>Use -</td>
<td>Use -</td>
<td>Use -</td>
<td>Use -</td>
</tr>
<tr>
<td><strong>Duplicate Detection</strong></td>
<td>Yes</td>
<td>Grouped under one title</td>
<td>Grouped under one title</td>
<td>Yes</td>
<td>Yes</td>
<td>Grouped under one title</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>Special Features</strong></td>
<td>-</td>
<td>Search any language</td>
<td>Limit by date, language, or format field followed by a colon</td>
<td>Concept searching suggest terms</td>
<td>Search for image and sound files</td>
<td>Limit by date, language, location, and page depth.</td>
<td>Custom folders</td>
</tr>
</tbody>
</table>

Source: searchenginewatch.com
6. Meta Search Engines

Unlike the individual search engines and directories, metasearch engines do not have their own databases and they do not collect web pages. Moreover they do not accept URL additions and they do not classify or review web sites. Instead, they send queries simultaneously to multiple Web search engines and/or Web directories. Many of the metasearch engines integrate search results, duplicate findings are merged into one entry, some rank the results according to various criteria and some allow selection of search engines to be searched.

Before conducting a metasearch engine search, it is important to find out which search engines are included in the metasearch engine search. Most metasearch engines default to the major search engines, such as AltaVista, Excite, Lycos, and Infoseek. Others will also include Usenet searches and other specialized databases. Negotiations between the metasearch engine companies and the individual search engine companies may also result in a major search engine being excluded from a metasearch engine. For example, Northernlight would not allow any of the metasearch engines to robotically search its index since this process drains its resources. Development of metasearch engines lags behind development of search engines. Some metasearch engines still include defunct search engines. Some metasearch engines allow one to make his choice of which to use.

Successful use of a metasearch engine depends on the status of each of the individual search engines used. Some may be heavily loaded at the time and some may be unreachable. The added features mentioned above require further resources from the metasearch engines, resulting in slower response time, a serious problem with many of
the metasearch engines. Many of them, therefore, have a timeout period, so that attempt
to work with a particular search engine can be abandoned if no response comes from it
within a set period of time.

Remember too that a query submitted to a metasearch engine, with its uniform
search interface and syntax, is to be applied against the diversity of individual search
ingines. It is therefore impossible for metasearch engines to take advantage of all the
features of the individual search engines. Boolean searches, for example, may produce
varied results. Phrase searches may not be supported. Other features, such as query
refinement, are sacrificed.

Moreover, metasearch engines generally do not conduct exhaustive searches and
they do not bring back all the pages from each of the individual search engines. They
only make use of the top 10 to 100 hits from each of them. While this is sufficient for
most searches, individual search engines must be consulted if one needs to go beyond the
top hits as determined by the metasearch engines. Some metasearch engines facilitate this
need by providing query links back to the individual search engines.

Should one use a metasearch engine instead of an individual search engine? There
is no definitive answer to this question. Much depends on what one is seeking for. For a
specific, obscure search term, it is recommend to start with a metasearch engine, as it will
search many sites at the same time, thus saving users a lot of time, and making their
search less tedious. On the other hand, if users are reasonably confident that any major
search engine will return the page they are looking for, starting with an individual search
engine would be recommended.
The following metasearch engines, listed alphabetically, are among the major ones currently available. Brief discussion of search tips and major features of each of them are provided.

6.1. Ask Jeeves (http://www.askjeeves.com/)

It has simple syntax and results are presented in pull-down menus. It reports number of matches from each search engine and there is neither integration nor ranking. It has interesting design, fairly good response time and limited number of search engines used.

6.2. Debriefing (http://www.debriefing.com/)

This metasearch engine searches AltaVista, Yahoo, Infoseek, Excite, Webcrawler, Lycos and Hotbot in the English version. Its French version searches Yahoo France, PagesWeb, Ecila, Infoseek France, Excite France and Lokace. It supports Boolean (+ -) and phrase searches (" "), collates the results, ranks them and removes duplicates. It provides the most significant domain name for a search; in the advanced search mode, it allows for searches within a particular site (no need to provide a complete URL).
6.3. Dogpile (http://www.dogpile.com/)

It is relatively new which searches Web sites, Usenet, FTP sites and new files. For novices start with "Custom Search" where one can set the order and the number of the 25 search engines, so that results from one's favorite sites return first, and/or exclude certain sites (skip) from the search engine list, a very handy feature. Timeout can be set from ten to 60 seconds and it searches three sites at a time and if there are enough results (ten hits), the search will stop, otherwise it will continue on to the next three sites. Ten records from each of the three sites will be displayed. Further hits from the three sites can be retrieved with a click, and the next three sites can be searched with a click as well. Search results are displayed with summaries. Number of hits from each site is reported, Boolean searches are supported, response time is very good and no integration of results.
6.4. Highway 61 (http://www.highway61.com/)

Searches only Yahoo, Lycos, Webcrawler, Infoseek, Excite and AltaVista. It also supports AND and OR searches and number of hits from each site is reported. Results displayed with summaries, sites coming from most search engines are ranked higher and interesting way of presenting options.

6.5. Internet Sleuth (http://www.isleuth.com/)

One of the largest collections of searchable sites, divided into several major categories: Web search engines and directories, reviewed sites, news, business and finance, software and Usenet- and very flexible selection of search engines to be included. Maximum search time can be set between ten seconds and two minutes. No integration of results and display of search results can be customised to show titles only or titles with summaries. The number of results from each site can range from 10 to 100.
Here there is a convenient arrangement for retrieving more records from individual search engines and response time is moderate.

![Internet Sleuth Home Page](image)

**Figure 3.20: Internet Sleuth Home Page**


Sculps the Web, Usenet, news, stock symbols, company names, MP3 files, pictures and sound. It also supports optional phrase searches and searches limited to titles only, optionally shows summaries, Boolean operators can be used (+ and -) and it claims to present results in a uniform format by relevance and source. A limited number of search engines is supported: AltaVista, Excite, Infoseek, Lycos, WebCrawler, and Yahoo. No arrangement for further searches in the individual search engines and response time is moderate.
6.7. **MetaCrawler** ([http://www.go2net.com/search.html](http://www.go2net.com/search.html))

It is one of the earliest metasearch engines, purchased by go2net from University of Washington. It is best to customize it before using set default interface (regular, power, or low bandwidth). It also selects the default Boolean operators to be used (OR, AND, or as a phrase). It may limit results from Web pages from North America, Europe, Asia, Australia, South America, Africa, Antarctica, or U.S. educational, commercial or government sites. Results are displayed with summaries, integrated and ranked and response time is fairly good. Web search includes only the major search engines: Lycos, Infoseek, WebCrawler, Excite, AltaVista, and Yahoo. Many other types of databases have been added recently - computer products, usenet, files and stock quotes.
This is another highly recommended site for its flexible configuration, clean display of search results, and good response time. For a quick search of the major search engines, it should be considered the number one metasearch engine.

6.8. MetaFind (http://www.metafind.com/)

From the same company that produces Dogpile, Metafind searches six search engines, returning links and organizing the results. It retrieves ten links from AltaVista twice, ten from Excite twice, 50 from HotBot, 25 from Infoseek, 30 from Planetsearch, and 50 from Webcrawler. User can use AND, OR, NEAR, NOT, ( ), and "" in his search. AND is the default connector. Timeout period can be set from 10 to 60 seconds. Search results can be sorted by keyword, by domain, alphabetically or not sorted. This is another highly recommended site for its overall performance and for its sorting capabilities.

6.9. ProFusion (http://www.profusion.com/)

This metasearch engine is an excellent option in search engine selection. One can choose the best three, the fastest three, all or any of the available search engines. Boolean and phrase searches are supported and it searches the Web or Usenet. Search results can be displayed with summaries or without, one can have up to 50 links of search results checked to make sure they are live. Results are integrated and number of hits from each search engine is reported and search terms can be saved for future reruns. Unfortunately, ProFusion tends to be very slow in response time, but with recent address change, speed has dramatically improved.
6.10. SavvySearch (http://www.savvysearch.com/)

It is another pioneer of metasearch engines, rich with features and including a great number of search engines. It covers Usenet groups, and many other specialized databases. Boolean operators (AND and OR) and phrase searches are supported. It allows for setting the number of retrievals from each search engine (from 10 to 50) and can display search results in brief, normal and verbose formats and optional integration of results. The most inclusive search is possible by selecting all the available types of resources: Web indexes, directories, Usenet, software, people, reference, entertainment, commercial, academic, images, and technical reports. Response time is poor.
6.11. **Verio Metasearch** ([http://search.verio.net/](http://search.verio.net/))

The advanced query interface has a very powerful scoring feature, allowing one to decide which individual search engine's results carry more weight than others. Maximum delay time can be arbitrarily set, number of search results can range from ten to all. It returns the most meta-information about a site, including relevance rank and score, and number of search engines ranking a site in its top ten hits. Slow response time.

<table>
<thead>
<tr>
<th>Metasearch Engine</th>
<th>Number of Search Engines Covered</th>
<th>Response Time (in seconds)</th>
<th>Boolean Operator</th>
<th>Phrase Search</th>
<th>Title/Summary</th>
<th>Number of Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>AskJeeves: <a href="http://www.askjeeves.com/">http://www.askjeeves.com/</a></td>
<td>4</td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
<td>Both</td>
<td>Yes</td>
</tr>
<tr>
<td>Debriefing: <a href="http://www.debriefing.com/">http://www.debriefing.com/</a></td>
<td>7</td>
<td>-</td>
<td>Yes (+,-)</td>
<td>Yes</td>
<td>Both</td>
<td>Yes</td>
</tr>
<tr>
<td>Dogpile: <a href="http://www.dogpile.com/">http://www.dogpile.com/</a></td>
<td>25</td>
<td>60</td>
<td>Yes</td>
<td>Yes</td>
<td>Both</td>
<td>-</td>
</tr>
<tr>
<td>Highway61: <a href="http://www.highway61.com/">http://www.highway61.com/</a></td>
<td>5</td>
<td>-</td>
<td>Yes (AND,OR)</td>
<td>Yes</td>
<td>Both</td>
<td>-</td>
</tr>
<tr>
<td>InternetSleuth: <a href="http://www.isleuth.com/">http://www.isleuth.com/</a></td>
<td>6</td>
<td>10-120</td>
<td>Yes</td>
<td>-</td>
<td>Both</td>
<td>-</td>
</tr>
<tr>
<td>Mamma: <a href="http://www.mamma.com/">http://www.mamma.com/</a></td>
<td>6</td>
<td>-</td>
<td>Yes (+,-)</td>
<td>-</td>
<td>Only title (summary optional)</td>
<td>-</td>
</tr>
<tr>
<td>MetaCrawler: <a href="http://www.go2net.com/search.html">http://www.go2net.com/search.html</a></td>
<td>6</td>
<td>-</td>
<td>Yes (AND, OR)</td>
<td>-</td>
<td>With Summary</td>
<td>-</td>
</tr>
<tr>
<td>MetaFind: <a href="http://www.metafind.com/">http://www.metafind.com/</a></td>
<td>6</td>
<td>60</td>
<td>YES (AND, NOT,NEAR, NOT)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ProFusion: <a href="http://www.profusion.com/">http://www.profusion.com/</a></td>
<td>-</td>
<td>-</td>
<td>YES</td>
<td>YES</td>
<td>Optional</td>
<td>-</td>
</tr>
<tr>
<td>SavvySearch: <a href="http://www.savvysearch.com/">http://www.savvysearch.com/</a></td>
<td>-</td>
<td>-</td>
<td>Yes (AND,OR)</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Source: [http://www.indiana.edu/~librcsd/search/meta.html](http://www.indiana.edu/~librcsd/search/meta.html)
7. **General Guidelines for Successful Search**

In the pursuit of finding information on the Web, a most important requirement is that a searcher should know search strategy prior to starting the search. Such a strategy facilitates him to get qualitative results in a considerably less amount of time and efforts.

Here are a few guidelines that would assist a searcher in successfully completing the search task:

- Determine search needs in mind or on paper and refine this search need to make it more specific.
- Select an appropriate search engine and think whether using a specialised search engine would be better or not.
- For broad or general topics one must start with the directory-based search engines like Yahoo so that he can see the range of what is available and take advantage of subject heading structure to move towards more relevant document.
- For highly specific or obscure topics, use the search tools. Often it can take a long time to figure out where in a subject hierarchy a particular term belongs, a large keyword searchable database can home in on the term quickly.
- Look for help page because search tools handle combining terms and searching for phrases in different ways and the help page will describe the correct way to state search for that search tool.
- Formulate search query and it should be as precise as possible. And, the search query should preferably contain as many keywords (or search terms) as possible. Try plurals, synonyms, and British & American spelling and, more importantly, change
the search query by using different related set of keywords or search terms with the help of subject thesaurus and indexes.

- Select the search tool with a large database, to start with. Use specific keyword and rare or unusual words are better than common ones.

- Search for phrases and Boolean operators when ever possible.

- If the first strategy does not give good results, repeat the search by varying the terms and combinations. Try this in different ways in different search tools or metasearch engines. Be aware of the differences in the search syntax.

- Some search tools are sensitive to word order. So the most important words should be entered first.

- Use the NOT operator to exclude unwanted Web pages.

- Learn how to use search syntax for different search engines, whether it supports Boolean operators, advanced search options or not.
Algorithm for Effective Informational Retrieval

For successful search here an algorithm is given which is most helpful to the beginner in searching information on the Web.

*Step: 1.* Go to favorite search engine

*Step: 2.* Type in keywords

*Step: 3.* Click the search button

*Step: 4.* Wait for the results

*Step: 5.* Retrieve too many results

*Step: 6.* Click on the first result

*Step: 7.* Wait for the document to load

*Step: 8.* Don’t find answer

*Step: 9.* Back up

*Step: 10.* Click on the another result

*Step: 11.* Wait for the document to load

*Step: 12.* Don’t find the answer

*Step: 13.* Repeat until frustrated

*Step: 14.* Try new key words

*Step: 15.* Wait for some more time

*Step: 16.* Pick some more results

*Step: 17.* Still don’t find answer

*Step: 18.* Try another search engine

*Step: 19.* Repeat until get desired results
Figure 3.23: Flowchart for Effective Information Retrieval on the Internet
8. Sample Queries and the Test Environment

As a sample search queries, the following five Physics research topics are selected from University News - a weekly journal, for testing various features of each search engine as well as to represent different levels of searching complexity.

1. X-ray crystallographic studies on compounds of biological and medical interest.
2. Dynamic properties of some diluted magnetic semiconductors and rare earth materials at high pressure.
3. A study of gravitational instability of gaseous plasma.
4. Photoneutron production and irradiation studies using microtron facility
5. Optical and electrical switching properties of praseodymium thin films

8.1. Search Results

Search was carried out using different search engines viz., Yahoo, Google, Altavista, Lycos, Excite and Hotbot. In order to retrieve relevant data from each search engine, the search strategy of respective search engines has been used to get good result and the collected data has been given in the following tables. From the study it can be observed that, the result obtained from each search engine significantly varies among the search engines. In this study an attempt has been made to collect the data on total number of sites retrieved, selected sites, relevant sites out of selected sites, irrelevant sites, relevant links, irrelevant links and sites can not be accessed. Search results obtained from different search engines has been discussed in the following tables.
Yahoo is probably the most popular and best known Internet resource. Hence here an attempt has been made to retrieve information on above mentioned five queries using the search strategy. The data regarding the information retrieved from the Yahoo search engine is given in table-3.3. It can be seen from the table that 20,696 sites are retrieved from the search engine and out of which only 50 sites are selected for the study. Since more number of sites are retrieved from the search engine for each query, it is decided to select only first 50 sites. It is evident from the table that 54% of sites are irrelevant while 26% of sites are relevant. It is also observed that 13.6% and 3.6% of sites are irrelevant links and relevant links respectively.

### Table-3.3: Yahoo Search Engine

<table>
<thead>
<tr>
<th>Search Queries</th>
<th>Total Number of Sites</th>
<th>Selected Sites</th>
<th>Relevant Sites</th>
<th>Irrelevant Sites</th>
<th>Relevant Links</th>
<th>Irrelevant Links</th>
<th>Sites Can't be Accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search#1</td>
<td>983</td>
<td>50</td>
<td>23 (46)</td>
<td>14 (28)</td>
<td>2 (4)</td>
<td>8 (16)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Search#2</td>
<td>9260</td>
<td>50</td>
<td>14 (28)</td>
<td>27 (54)</td>
<td>-</td>
<td>7 (14)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Search#3</td>
<td>6150</td>
<td>50</td>
<td>9 (18)</td>
<td>23 (46)</td>
<td>3 (6)</td>
<td>13 (26)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Search#4</td>
<td>263</td>
<td>50</td>
<td>7 (14)</td>
<td>34 (68)</td>
<td>4 (8)</td>
<td>5 (10)</td>
<td>-</td>
</tr>
<tr>
<td>Search#5</td>
<td>4040</td>
<td>50</td>
<td>12 (24)</td>
<td>37 (74)</td>
<td>-</td>
<td>1 (2)</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>20,696</td>
<td>250</td>
<td>65 (26)</td>
<td>135 (54)</td>
<td>9 (3.6)</td>
<td>34 (13.6)</td>
<td>7 (2.8)</td>
</tr>
</tbody>
</table>

Note: Number given in parenthesis represents the percentage.

From the review of literature, it is observed that Meghaghab and his co-workers (1998)\(^\text{11}\) examined the effectiveness of five World Wide Web search engines: Yahoo, WebCrawler, Infoseek, Excite and Lycos. The study involved five queries that were checked against each of the search engines in original and refined formats, a total of fifty searches. The queries varied in terms of broadness, specificity and level of difficulty in
finding Internet resources. The results of the study revealed that Yahoo had the highest precision ratio for both original and refined queries. Infoseek maintained second place with respect to refined queries and Lycos with respect to original queries.

Thus it is observed that the findings of the present study are in line with the findings of the Meghaghab and others.

**b) Google**

Google in another most popular search engine on the Internet and also most popularly used by the Internet users, because Google focuses on the link structure of the Web to determine relevant results for the users. Thus for the present study, Google search engines also selected to know the relevancy of information which are retrieved for each query. As in case of Yahoo, here also only 50 sites are selected for each query for the analysis.

<table>
<thead>
<tr>
<th>Search Queries</th>
<th>Total Number of Sites</th>
<th>Selected Sites</th>
<th>Relevant Sites</th>
<th>Irrelevant Sites</th>
<th>Relevant Links</th>
<th>Irrelevant Links</th>
<th>Sites Can't be Accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search#1</td>
<td>1110</td>
<td>50</td>
<td>19 (38)</td>
<td>23 (46)</td>
<td>3 (6)</td>
<td>5 (10)</td>
<td>-</td>
</tr>
<tr>
<td>Search#2</td>
<td>1700</td>
<td>50</td>
<td>17 (34)</td>
<td>24 (48)</td>
<td>-</td>
<td>6 (12)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Search#3</td>
<td>6890</td>
<td>50</td>
<td>14 (28)</td>
<td>12 (24)</td>
<td>1 (2)</td>
<td>23 (46)</td>
<td>-</td>
</tr>
<tr>
<td>Search#4</td>
<td>441</td>
<td>50</td>
<td>12 (24)</td>
<td>29 (58)</td>
<td>2 (4)</td>
<td>6 (12)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Search#5</td>
<td>4530</td>
<td>50</td>
<td>8 (16)</td>
<td>31 (62)</td>
<td>-</td>
<td>10 (20)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Total</td>
<td>14671</td>
<td>250</td>
<td>70 (28)</td>
<td>119 (47.6)</td>
<td>6 (2.4)</td>
<td>50 (20)</td>
<td>5 (2)</td>
</tr>
</tbody>
</table>

Note: Number given in parenthesis represents the percentage

The table-3.4 illustrates the level of relevancy of search engine and it is clear from the table that 51.6% of sites are irrelevant and only 28% of sites are relevant. It is also
observed that 20% of sites are irrelevant links and only meager percent of sites (2.4%) are relevant links.

c) **Altavista**

In case of Altavista search engine the total number of sites are 45,448 out of which only 250 sites are selected for the study (50 sites for each query). Among the selected sites the irrelevant sites are amounted to 54.8% and the about 20% of sites are relevant. About 19.6% of sites are irrelevant links and 3.2% of sites are relevant links (Table-3.5).

In case of Leonard study (1996) some selected search engines were tested on their accuracy of results, ease of use and provision of advances options using 15 queries specifically composed for evaluation. Study found that Altavista seems to be the best search for information search on the Internet compared to other search engines, which they selected for the study. From the present study it also found that relatively good number of relevant sites are found in case of Altavista.

<table>
<thead>
<tr>
<th>Search Queries</th>
<th>Total Number of Sites</th>
<th>Selected Sites</th>
<th>Relevant Sites</th>
<th>Irrelevant Sites</th>
<th>Relevant Links</th>
<th>Irrelevant Links</th>
<th>Sites Can't be Accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search#1</td>
<td>20,520</td>
<td>50</td>
<td>14 (28)</td>
<td>33 (66)</td>
<td>-</td>
<td>2 (4)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Search#2</td>
<td>4076</td>
<td>50</td>
<td>17 (34)</td>
<td>24 (48)</td>
<td>2 (4)</td>
<td>6 (12)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Search#3</td>
<td>575</td>
<td>50</td>
<td>6 (12)</td>
<td>14 (28)</td>
<td>3 (6)</td>
<td>23 (46)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Search#4</td>
<td>3458</td>
<td>50</td>
<td>4 (8)</td>
<td>39 (78)</td>
<td>1 (2)</td>
<td>6 (12)</td>
<td>-</td>
</tr>
<tr>
<td>Search#5</td>
<td>16815</td>
<td>50</td>
<td>9 (18)</td>
<td>27 (58)</td>
<td>2 (4)</td>
<td>12 (24)</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>45,444</td>
<td>250</td>
<td>50 (20)</td>
<td>137 (54.8)</td>
<td>8 (3.2)</td>
<td>49 (19.6)</td>
<td>6 (2.4)</td>
</tr>
</tbody>
</table>

Note: Number given in parenthesis represents the percentage.
d) Lycos

It is very interesting to note that from the Lycos search engine only 1,514 sites are retrieved. It is also observed that for 5th and 2nd search queries only 7 and 2 search results are found respectively. Even for the 4th search query there is no result. Overall majority of sites are irrelevant (59.63%) and only 17.43% of sites are relevant. It is also observed from the table-3.6 that 19.26% and 3.6% sites are irrelevant sites and sites can not be accessed respectively.

<table>
<thead>
<tr>
<th>Search Queries</th>
<th>Total Number of Sites</th>
<th>Selected Sites</th>
<th>Relevant Sites</th>
<th>Irrelevant Sites</th>
<th>Relevant Links</th>
<th>Irrelevant Links</th>
<th>Sites Can’t be Accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search#1</td>
<td>1411</td>
<td>50</td>
<td>9 (18)</td>
<td>33 (66)</td>
<td>-</td>
<td>6 (12)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Search#2</td>
<td>7</td>
<td>7</td>
<td>3 (42.85)</td>
<td>4 (57.14)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Search#3</td>
<td>94</td>
<td>50</td>
<td>6 (12)</td>
<td>27 (54)</td>
<td>-</td>
<td>15 (30)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Search#4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Search#5</td>
<td>2</td>
<td>2</td>
<td>1 (50)</td>
<td>1 (50)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>1514</td>
<td>109</td>
<td>19 (17.43)</td>
<td>65 (59.63)</td>
<td>-</td>
<td>21 (19.26)</td>
<td>4 (3.6)</td>
</tr>
</tbody>
</table>

Note: Number given in parenthesis represents the percentage

In case of Chu and Rosenthal' (1996)^13, study used sample search queries that were based upon real reference questions. These queries were structured to evaluate the search engines' abilities to deal with variety of query syntax, for example different Boolean logic searches. The authors tested the first ten results from three search engines for precision and concluded that Altavista out performed Excite and Lycos in both search facilities and recall, although Lycos had the largest claimed coverage of Web resources.
Another factor that was taken into consideration was the response time of the search engines, which perhaps surprisingly did not vary between search engines. But in the present study it is found that relatively less number of relevant sites found in case of Lycos when compared to Google, Yahoo and Altavista. Since the study used different queries to retrieve information generally the result varies from one study to another.

e) **Excite**

Very few number of sites found in case of Excite search engine compared to other search engines. As the data presented in the table-3.7 which illustrates the total number of relevant sites, irrelevant sites, relevant links and irrelevant links. It is clear from the table that the relevancy is amounted to 18.3% and 40.8% of sites are irrelevant. Meager percent of sites (5%) are relevant links.

<table>
<thead>
<tr>
<th>Search Queries</th>
<th>Total Number of Sites</th>
<th>Selected Sites</th>
<th>Relevant Sites</th>
<th>Irrelevant Sites</th>
<th>Relevant Links</th>
<th>Irrelevant Links</th>
<th>Sites Can't be Accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search#1</td>
<td>59</td>
<td>50</td>
<td>13 (26)</td>
<td>17 (34)</td>
<td>3 (6)</td>
<td>14 (28)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Search#2</td>
<td>58</td>
<td>50</td>
<td>7 (14)</td>
<td>33 (66)</td>
<td>2 (4)</td>
<td>7 (14)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Search#3</td>
<td>58</td>
<td>50</td>
<td>9 (18)</td>
<td>16 (32)</td>
<td>1 (2)</td>
<td>20 (40)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Search#4</td>
<td>40</td>
<td>40</td>
<td>4 (10)</td>
<td>21 (52.5)</td>
<td>2 (5)</td>
<td>13 (32.5)</td>
<td>-</td>
</tr>
<tr>
<td>Search#5</td>
<td>68</td>
<td>50</td>
<td>11 (22)</td>
<td>11 (22)</td>
<td>4 (8)</td>
<td>23 (46)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Total</td>
<td>283</td>
<td>240</td>
<td>44 (18.3)</td>
<td>98 (40.8)</td>
<td>12 (5)</td>
<td>77 (32.08)</td>
<td>9 (3.75)</td>
</tr>
</tbody>
</table>

Note: Number given in parenthesis represents the percentage

In an important paper Clarke and Willett (1997)\(^4\) compared the effectiveness of Altavista, Excite and Lycos using thirty different searches. The paper is important both because of the critical evaluation of earlier research that it provides and because it offers a realistic and achievable methodology for evaluating search engines. The authors
developed a method for comparing the recall of the three sets of searches, despite the fact that they are carried out upon non-identical sets of Web pages, because each search engine has indexed a different set of documents. They developed an algorithm for calculating relative recall by checking how many of the so called relevant found by one search engine were present at all in the universe of documents covered by the other search engines. The study clearly shows that Alta Vista is significantly better at retrieval performance than either Lycos or Excite. Thus the present study results also shows that Altavista is significantly better compared to Lycos and Excite in retrieving relevant information. Because relatively a few relevant sites (18.3%) found in case of excite.

f) Hotbot

Efforts are also made to retrieve information using Hotbot search engine with using the search query, which are used for previous search engines. The number of sites retrieved from the Hotbot search engines for each query has been discussed in the table-3.8.

<table>
<thead>
<tr>
<th>Search Queries</th>
<th>Total Number of Sites</th>
<th>Selected Sites</th>
<th>Relevant Sites</th>
<th>Irrelevant Sites</th>
<th>Relevant Links</th>
<th>Irrelevant Links</th>
<th>Sites Can't be Accessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search#1</td>
<td>336</td>
<td>50</td>
<td>7 (14)</td>
<td>41 (82)</td>
<td>-</td>
<td>2 (4)</td>
<td>-</td>
</tr>
<tr>
<td>Search#2</td>
<td>3,171</td>
<td>50</td>
<td>11 (22)</td>
<td>34 (68)</td>
<td>-</td>
<td>5 (10)</td>
<td>-</td>
</tr>
<tr>
<td>Search#3</td>
<td>2496</td>
<td>50</td>
<td>9 (18)</td>
<td>27 (54)</td>
<td>3 (6)</td>
<td>8 (16)</td>
<td>3 (6)</td>
</tr>
<tr>
<td>Search#4</td>
<td>147</td>
<td>50</td>
<td>6 (12)</td>
<td>21 (42)</td>
<td>4 (8)</td>
<td>17 (34)</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Search#5</td>
<td>1996</td>
<td>50</td>
<td>8 (16)</td>
<td>31 (62)</td>
<td>5 (10)</td>
<td>6 (12)</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>8146</td>
<td>250</td>
<td>41 (16.4)</td>
<td>154 (61.6)</td>
<td>12 (4.8)</td>
<td>38 (15.2)</td>
<td>5 (2)</td>
</tr>
</tbody>
</table>

Note: Number given in parenthesis represents the percentage
The table-3.8 illustrates the search results of Hotbot search engine and it is evident from the table that 61.6% and 16.4% of sites retrieved are irrelevant and relevant respectively. 15.2% of sites are irrelevant links and only few percent of sites are relevant links (4.8%).

8.2. Results and Discussion

As it is observed from study that the search results of each search engines varies significantly and overall Altavista search engine searched more number of sites (45,448) while Excite searched least number of sites (283). In case of relevancy of search engines it is found that majority of relevant sites found in case of Google (28%) search engine followed by Yahoo (26%) and Altavista (20%). Thus it is observed that the findings of the present study are in accordance with the findings of the some authors as discussed above.

It is also observed that more number of irrelevant sites found in case of Hotbot (61.6%), Lycos (59.6%) and Altavista (54.8%). Thus Google, Yahoo and Altavista are rated as best search engines for retrieval information on the net. After considering the relevancy, accuracy and user-friendliness of each search engine, it is decided to consider the Yahoo search engines for retrieving Web based sources for evaluation purpose. The analysis and interpretation of data regarding the evaluation of Web based sources in Physics and its sub divisions in different forms has been discussed in the chapter-V (Part-A).
9. Conclusion

With the development of information technology, search engines are widely used as tools to find information on the Internet. Variety of information sources are available on the Internet and searching this huge information on the Web is very difficult. With the advent of World Wide Web search engines, information access is made easy. These are highly popular tools to search the Web information effectively. They are special sites on the Web that are designed to retrieve relevant information to a searcher. Thus before searching information on the Internet, searcher should select best search engine and it is also very essential to know the search techniques for different search engines, which facilitate to retrieve the exhaustive and pinpointed information and save the time of the information searcher by avoiding the irrelevant documents.
References


5. Oppenheim (C) and others. The evaluation of WWW search engines. *Journal of Documentation*. Vol. 56 (2), March 2000, pp190-211.


   Available at: http://www.cnet.com/Content/Reviews/Search/.

   Also available at: http://www.asis.org/annual-96/ElectronicProceedings/chu.html