The World Wide Web is estimated to contain more than two billion pages of publicly accessible information, and web continues to grow at an exponential rate, triple in size over past two years. There are estimated 800 million web page on the web and it may double every year. *(Faulkner's Cyberspace Digest - 08/06/99).*

Web lacks bibliographic control, no analogous system are available as developed by Library of Congress, of cataloguing or classification. Instead Web offers the choice of different search tools, they have a variety of ways for you to refine and control your searches. Internet search engines are special sites on the web that are designed to find information on other sites, they all perform three basic tasks:

1. They search the Internet- or selected piece of Internet – based on important words.
2. They keep an index of the words they find and where they find them.
3. They allow users to look for words or combination of words found in that index.

Search Engines and Subject Directories:

The two basic approaches to searching the web are:

**Search Engines**

**Subject directories**

To find information on the hundred of million Web pages that exists, a search engine employ special software robots, called spiders to build lists of the words found on the web site. List building process is called Web crawling.

All search engines are intended to perform same task, but each does it in a different way, which leads to different results produced by search engines.

The factors that influence results

Size of database

The frequency of update

Search capabilities

Examples of search engines:

AltaVista (http://www.altavista.com)

Excite (http://www.excite.com/search)
There is variety of search engines has led to create a “meta” search tools. Often referred to as multi-threaded search engines. These search engines allows the user to search multiple databases simultaneously, via a single interface.

Popular multi-threaded search engines include:

- Metacrawler
- Ixquick
- Surfwax
- Dogpile
- Profusion

The idea of meta search engine is much better than the reality in most cases. They cannot be better than the database query. For complex search meta search engine will not work, need to know rules of the search engine. Most do not search Google.

**Subject specific search engines** do not attempt to index the entire Web. Instead they focus for web site or pages within defined subject area, geographical area, and type of resources. These search engines aims for depth of coverage within a single area, rather than breadth of coverage’s across subject, they are often able to documents not included in largest search engines.

- Beaucoup!
- Search Engine Colossus
- Searchengines.com

**Jobs**: [http://www.monster.com](http://www.monster.com)

**Games**: [http://gamesdomain.com](http://gamesdomain.com)


**Education**: [http://www.education-world.com](http://www.education-world.com)

**Subject Directories** are hierarchically organized indexes of subject categories that allow the web searchers through list of web sites by search of relevant information. They are manually
maintained and compiled and many include a search engine for searching their own databases. They are smaller than search engines but they produce more relevant results and links to home pages only whereas search engine typically indexes every pages of a given web site. Directories are useful for the structure they give to the Internet.

Examples of subject directories include:

Looksmart (http://www.looksmart.com)
Open Directory (http://dmoz.com) (acts as the Google directory and is the largest most comprehensive human edited directory of the web. It is maintained by a vast global community of volunteers)
Yahoo (http://www.yahoo.com)

Specialized subject directories

There are search engines complied by subject specialists to important resources in his or her area of expertise to produce more relevant information and is more comprehensive. Such guide exists for virtually for every topic. Like Voice of Shuttle (http://www.vos.ucsb.edu) provides excellent starting point for humanities search.
Some web sites act as collections or clearinghouses of specialized subject directories.

Examples of Clearinghouses include:

Argus Clearinghouse (http://www.clearinghouse.net)
About.com (http://about.com)

Choosing Search Engine

Google and Altavista are the popular search engines, having largest index, and returns few good hits fast. Altavista displays the most common words from your results which you can add to your query for more focused results. Google's popularity ranking often makes pages worth looking at rise near the top of search results. You learn Google and Alltheweb and may be Altavista and teoma

www.alltheweb.com AlltheWeb indexes billions of web pages, hundreds of millions of multimedia, Audio and FTP files, as well as tens of millions of PDF and MS Word® files.

For broad academic subjects you can use Librarians’ Index to the Internet (www.lii.org) and Infomine (www.Infomine.ucr.edu/Main.html). Index to the Internet is to provide a well-
organized point of access for reliable, trustworthy, librarian-selected Internet resources, serving the world.

For popular or commercial topic **www.yahoo.com** can be a better choice.

For visual view of related subjects **www.webbrain.com**.

**www.vivisimo.com** Vivísimo is recognized internationally as the leader in document clustering technology automatically categorizes textual information into crisp, meaningful, hierarchically sorted category folders

**www.surfwax.com** identify promising results using preview summary "add "Focus Words" to narrow next search.

**www.virtuallrc.com** Virtual LRC indexes approximately 5,000 quality information web pages, all selected by librarians and teachers. The VLRC includes selected sites in a growing list of subject/information areas including: full-text magazines, electronic text archives, biography, biology, career information, , Careers, Education, Mathematics, Science, Technology, Tutorials on the Web, and Writing Style Guides etc.

**www.teoma.com** Teoma means "expert" in Gaelic. And it is our expert ability to analyze and understand the Web as it occurs naturally. It provides ranks by subject-expert popularity (link from authority pages) , suggests expert hub sites. Teoma adds to search approach, known as **Subject-Specific Popularity** ranks a site based on the number of same-subject pages that reference it, not just general popularity.

**www.infoplease.com** search 30,000 people (encyclopedia , almanac, dictionary).

**www.cnet.com** meta search engine for free and inexpensive software.

**Search Strategy**
Regardless of the search tool being used, the development of an effective search strategy is essential if you hope to obtain satisfactory results. Most search engines index every word of a document, this method increase the number of search results retrieved while decreasing the relevance of theses results. Most engines allow you to type in a few words, and then search for occurrences of these words in their database. Each one has their own way of deciding what to do about approximate spellings, plural variations, and truncation.

**Search Logic**
Search logic refers to the way in which you are using combine your search term. For example Banaras Hindu University could be interpreted as a search for any of the three search terms. Depending on the logic applied, the results of the three searches would differ greatly. All search engines have default method.

Boolean logic, the basic Boolean operator are AND, OR, NOT. Variations on this operator are called proximity operators that are supported by some search engines include ADJACENT, NEAR, FOLLOWED BY.

**Boolean AND NOT**

“biomedical engineering” AND cancer AND NOT “Department of” AND NOT “School of”

“biomedical engineering” cancer –“department of” –“school of”

A type of operator used by some search engines to improve search constraints by instructing the search to look for words that are within a short distance of each other in a document. If looking for information on the inventor Thomas Alva Edison,

“Thomas Alva Edison” OR "Thomas A. Edison" OR "Thomas Edison"

Thomas NEAR Edison

How near is NEAR? That depends. In AltaVista the words used to be less than 10 words apart.

**dogs** **near/3** **cats**

finds documents in which dog and cat occur within three words of each other, in either order.

### Search Tips

<table>
<thead>
<tr>
<th>Ctrl-F</th>
<th>Ctrl-F search for the text in the current document.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bookmark your results</strong></td>
<td>For repeated search on later date, add a bookmark (or Favorite) to your current results.</td>
</tr>
</tbody>
</table>
| **Wildcards** | Some searching allows “wildcards” character in search statements. The common ones are #, * and ?. Some search engine allows right truncation (e.g. psycholog*) while other support middle truncation. At AOL Search, the? Symbol serves as a wildcard and will replace any single character, such as:

s?ng matches sing, sang, song |

### Search Engine Maths

The important thing to note is search for exactly what the problem is
For example if to know about Windows 98 bugs, search for “Window 98 bugs” not “Windows”

**Using the + Symbol to Add**

Put a plus sign in front of words that *must* be present on the webpage.

+banaras +hindu +university

Only pages contain all the words would appear in your results.

**Using the – Symbol to Subtract**

A minus sign in front of a word will tell the search engine to *subtract* pages that contain that particular word.

rice –university

**Using quotation Marks to Multiply**

Multiplying a terms through a “phrase search” can be much better way to get the answer you are looking for.

+ world +health +organization

may not guarantee the nearness of the two words instead

“world health organization”

In capital initial letter will cause the terms to be searched as a phrase “World Health Organization”

Now only pages that have all the words and in exact order will be listed.

"plant disease" + india -west bengal

This means that the pages the search engine shows you must include the phrase plant disease they must not include the word west, and they should preferably include the word bengal. If there is no sign in front of a word, most search engines will read a + sign. In other words: it will default to AND if it finds no "mathematical signs".exception is AltaVista, which will interpret the lack of a sign as an OR operator.

Search engines may get confused. What does the following search imply, really?

(Women OR females) AND networking

(Sarajevo OR Sarayevo) AND Peace

(Literature OR Litterature) AND (French OR Fracaise)
Field Search

The main fields that can be accessed in field searching are:

**Title:**

The title normally contains important keywords referring to the content of the page. If you restrict your search to the page titles, you will get fewer -- but more focused -- hits. Several AllTheWeb, AltaVista and Inktomi (as accessed via HotBot) use the *title:* command, where you follow the command with the word you want to find in the title of documents. For example, if you wanted to find all the pages listed that had the word “graphics” in their titles, you would do this:

*Computer AND title:graphics*

*Title: “affirmative action”*

**URL:**

You may restrict you search to pages with addresses that contain a certain word. example,

*information technology –url:nic.in*

Would work to list pages about “information technology” but would remove any that came from nic.in.

Google’s advanced search page uses the allinurl command for finding URLs that contain certain words. For example you want to PDF find files about notification. You can use the inurl command to specify that URL’s that must have word “pdf” in the document, the “allinurl” command takes its name because when using it, you are requiring that ALL the words appear IN the URL. In contrast, the inurl command means that ANY of the words you specify should appear.

**Domain:**

Some search engines allow you to restrict your search to a specific domain. By doing a domain-search you may for instance restrict your search to pages in a specific country. British pages normally end in the letters *.uk. A search for Jaguar AND car AND domain:.uk should give you British pages containing information on the Jaguar car.*
There are also some top level domains (com, org, net etc.) that are not restricted to specific countries, although they are predominantly American. You can use these endings to restrict your search to commercial (.com), US educational (.edu), US governmental (.gov) or US military (.mil) site.

If you wanted to see all the pages from the BHU web site. At AltaVista, you could use this command:

**host: bhu.ac.in**

In response, AltaVista would display all the pages it has indexed from bhu.ac.in domain. You can even combine other commands

**host:nic.in bhu Electronics**

In response, AltaVista to list pages with words BHU and electronics within nic.in site.

**host:nic.in – “information technology”**

AltaVista will list all pages within nic.in which do not contain exact phrase “information technology.

You can get rid of nic.in pages by doing this:

**“information technology” – host:nic.in**

We could decide to see all pages containing physics in in (Indian) site.

**“physics” + host:in**

Similarly

**education + host:au**

The above command that works at AltaVista. The same examples work at Google.

**Search Engine Specific issues**

Several search engines offer the ability to search within the text of a URL. For Google also has a command that lets you narrow your search to find documents in particular formats, such works better than forcing the URL command into this role. The command is `filetype:`, and you follow it with the extension you want to search for. For instance:

**Computer graphics filetype:pdf**

Would bring all pdf files containing “computer graphics”.

By searching only in titles, one can eliminate pages with only brief mentions of a concept, and only retrieve pages that really focus on your concept. Search engines offer the ability to search for all the pages linking to a particular page or domain.

To title search at Google and Teoma, you need to use the `intitle:` command. This means to find a single word like “graphics” within the title of documents, you would enter:

**intitle:graphics**
With Google, use the *allintitle:* command, which means that Google will find documents that have ALL the words you specify in their page title. The command would be used like this:

`allintitle:graphic design`

*AltaVista* retrieves upper and lower case. For example, *dodge* retrieves *dodge* and *Dodge*; *Dodge* retries *Dodge* only.

For most searches, Google is the best place to start, Google is Biggest, Fastestm, and Most comprehensive search

**Important Search Guideline**

If you are clear what you want to search use search directory.

If you want to include stop-words in your search either use + sign just before the word or put the word within quotes.

**“to be not to be”**

Search for common terms should be avoided if it is not part of the search term, use directory instead.

More keyword will allow you to narrow down your search.

Use wildcards (*, ?) for plurals.

Put the main subject first.

Use synonyms like (dogs OR canines).

**Invisible Web**

The Invisible Web, on the other hand, consists of files, images and Web sites that, for a variety of reasons, cannot be indexed by popular search engines like Google, FAST and Alta Vista. Some of the most sought after information on the Internet is contained within databases and indexes that are considered part of the Invisible Web. In late 2000, search company *Bright Planet* released a study that suggested that the portion of the Web that could not be indexed by search engines was 500 times larger than the portion that could be indexed by search engines. The first is the Web pages that search engines are unable to access. The second is the Web pages that search engines are able to access, but choose not to include.

**How to Find the Invisible Web**

Use *Google* and other search engines to locate searchable databases by searching a subject term and the word "database". If the database uses the word database in its own pages, you are likely to find it in Google. The word "database" is also useful in searching a topic in *Yahoo!*, because Yahoo! used the term to describe searchable databases in its listings.
EXAMPLES for Google & Yahoo!:
plane crash database
languages database
toxic chemicals database

Citation: Communication Support for Sustainable Development. Eds. Dipak De and Basavaprabhu Jirli, GangaKaveri Publishing House, Jangamawadi Math, Varanasi - 221001