CHAPTER -5
SEARCH ENGINES IN INTERNET
The Internet is a global network of computer networks. Thousands of networks are linked together forming a worldwide network. Internet has made possible to provide information, any bit of knowledge, records, at any time, kept at any part of the world if it is available in digital media just a click away, and lakhs and lakhs of libraries are now on the side of information superhighway and millions of people are joining the rush in the superhighway seeking information from the net. Internet is the network formed by the cooperative interconnection of computer network to achieve a goal and to communicate with each other.

1. Development of Internet

The first true Internet was Arpanet a military network established in the United States in the late sixties. The Advanced Research Project Agency (ARPA) of the United States Department of Defense established ARPA net in 1969. It was an experimental four-computer network for the research scientists to communicate among themselves. Since then the internet has grown and is used for many more purposes. The first test network built on these principles was installed in National Research laboratory in Great Britain in 1968. Soon after its pentagon’s advanced research project agency (ARPA) of U.S. department of defense decided to launch a network based on the principle of R&D Corporation. Thus in 1969 ARPANET was created. Initially there were four super-computer stations, called nodes, on this high-speed network. This network was created in order to share information and research results and to collaborate more easily.
2. **Minimum requirement of hardware and software for Internet**

For harnessing most of the possibilities of today's Internet, the minimum resource able computer should have the following requirement that I think advisable based on my experience.

- **Processor**

  A Pentium processor is recommended for its capability to support the required operating system. Pentium IV processor with 1.5GHz or above. But Pentium IV processor with 800 MHz makes a workable.

- **Operating system**

  The latest version windows is the best choice for Internet use because that will feature or will accommodate latest built in web browser like Internet Explorer and it will be the easiest to use programmed for selling up our Internet connection.

- **Display**

  *The minimum display advisable for web browsing is configured to run at 800 X 600 resolution and 256 colors. Enhanced color modes, such as 16 bit color are fine but few require those modes. Resolutions higher than 800 X 600 are not recommended because they will make some items in web pages appear too small.*
• **Memory**

The advisable minimum RAM for supporting latest windows version and either of the leading web browsers is 128 MB. Latest Internet explorer version will run in lesser configurations but not in their full installation and net well. The reasonable minimum memory for comfortable internet browsing on windows XP version or above is 256 RAM.

• **Hard Disk**

*The size of the hard disk we have to select, depend on load. Requirement of the software we want to be at least 25 percent empty. Windows web browsers need lots of free disk space for temporary data storage. If it is not available the system will give poor performance and cannot be relied in time of need. If the system is going to be used in the reference section or the services section 40 GB hard disk apt capacities are available.*

• **CD Rom Drive**

For Internet access CD- Rom drive is not an essential component. But we will need one to install the Internet software required for starting things if we acquire that software on a CD. The present speed is 52 X for CD Rom drives.

• **Modems**

Modem is modulator and demodulator. If our computer is to communicate through telephone lines with another computer electrical impulses that can be sent over the connection. The device is called modem. The higher the number of bits pare second modem will transmit; the faster we will receive the information.
• **Other Peripherals**

Now a lot of sound and music files are available one line. To hear them we will need a sound card speaker or headphones. If you want to make a long distance phone call through the Internet, you can connect a microphone to our sound card. If you went video conferencing, you have to add a PC video camera. If you have to create our own WebPages. So you add a digital camera.

• **Internet software**

Getting Internet software is like borrowing money. It is difficult when we really need it. If you already have money or internet software, getting more of it is easy. Os the trick is getting started. We need two types of software first the communication software to establish the TCP/IP connection between our computer and our internet service provider and taken the client software for each activity you want to perform through the net, using the web browser, email client, newsreader and so on. *The popular operating systems will usually have some software in windows 98/ME/2000/NT/XP etc.*

The best place to get our startup Internet software is forms our Internet service provider. Then after you can go online, you can easily select and acquire any software you want. All you need from are startup software is a way to begin. Most prepackage Internet software in inexpensive and often comes with step program to conveniently sing up for one or more Internet service providers. Two major Internet client suites Internet explore and Netscape Navigator are available on CD ROM at any software store.
3. **Internet connectivity**

There are many ways to obtain access to the Internet. The method varies according to the type of computer system being used and the types of connection offered. Some connections, such as dialup, must be initiated every time you desire Internet access. *Other connection types remain available 24/7. These “always on” or “full time” connection make Internet access as simple as opening your browser or e-mail program.*\(^7\) One can connect to the Internet services mainly following three types.

- Dial – up Connection
- ISDN (Integrated services digital networks)
- DSL (Digital Subscrver Line)

### 3.1 Dial – up connections

In many homes and small businesses individual users connect to the Internet by using a telephone line and a 56kbps modem. The easiest way to create this kind of connection is by setting up an inexpensive account with an Internet service provider. *The ISP maintains banks of modems at its facility to process the incoming dial up requests form customers.*\(^8\) The ISP is servers rout traffic between customer’s computers and the Internet.

### 3.2 ISDN Integrated Services Digital Network

In a Digital telephone service that simultaneously transmits voice, data and control signaling over a single telephone line. ISDN service operates on standards telephone lines but requires and special modem and phone service, which adds to
the cost. ISDN data connection speed: 128 Kbps. The benefits of ISDN being able to connect a pc telephone and fax machine to a single ISDN line and use them simultaneously.

3.3 DSL – Digital Subscriber Line

It also uses regular phone lines to transmit data in megabits per second. Incoming data is significantly faster than outgoing data. That is your computer can receive data at the rate of 1.5 – 8.4 Mbps but it can send data at only 16–640 kbps. This arrangement may be fine however if yours principally interested in obtaining very large amounts of data rather that in bending them to others with DSL you can download that 6-minute music video in only 11 minutes (compared to an hour with ISDN). A big advantage of DSL is that is always and unlike cable (discussed shortly), its transmission rate is consistent.

4. Internet protocol

A formal description of message formats and the rubles tow or more computers must follow the computer-to-computer interfaces or high – level exchanges between application programs. TCP/IP (transmission control protocol/Internet protocol) is the basic communication language of protocol of the internet. In order to allow the two computers to communicate over the both should have TCP/IP copy loaded on their systems.

4.1 TCP/IP Protocol

The TCP/IP protocols suite was developed prior to the OSI model. Therefore the layer is the TCP/IP protocol suite
does not match exact with those in the OSI model. The TCP/IP protocol suite is made of five layers. Physical data link, network, transport and application. The first four layers provide tended, network interface, internetworking and transport functions that correspond to the first four layers of the OSI models. The three top most layers in the OSI model, however, are represented in TCP/IP by a single layer called the application layer.

The protocols used are World Wide Web

- **SMTP** — *Simple mail transfer protocol*
- **FTP** — *file Transfer protocol*
- **DNS** — *Domain Name system*
- **SNMP** — *Simple network management protocol*
- **NFS** — *Network file system*
- **TFTP** — *Trivial file transfer protocol.*

4.1.1 SMTP

The TCP/IP protocol defining electronic mail services on the Internet.

4.1.2 FTP

File transfer protocols (FTP) is the Internet tool used to copy files from one computer to another.

4.1.3 DNS

A TCP/IP application service that converts user-friendly names to IP address.

4.1.4 SNMP

The TCP/IP protocol that specifies the process of management in the Internet.
4.1.5 NFS

A TCP/IP application protocol that allows a user to access and manipulate remote file system as if they were focal. It uses the services of remote procedure cell protocol.

5.1.6 TFTP

Trivial file transfer protocol unreliable TCP/IP protocol file transfer that does not require complex interaction between client and server.

5. Internet Services Provider (ISP)

A company that connects you through your communications line to its server or central (Host) computer, which connect you to that internet via another called online services, are America online, (AOL) Earth link, Microsoft network (MSN), AT&T world net and prodigy. There are also many local ISPs. The ISP will assign you a user name and a password, as well as e-mailbox. The ISPs local access number for your area is called its point of presence. Few selected national ISPs of India.

BSES Telecom www.powersurfer.com
Dish net DSL www.eth.net
Free diallin www.freedialin
Global online www.gol.net.in
Mantra online www.mantraonline.com
MTNL www.boy.net.in
Nanda net com www.livewire.net
Only smart www.onlysmart.com
Pionet online www.pionetindia.com
Sampark online www.samparkonline.com
6. Internet services

The most common used Internet services.

- **WWW**
- **E-mail**
- **News**
- **FTP**
- **IM Instant Messaging**
- **Online serves**
- **Chat**

6.1 World Wide Web

The World Wide Web is a repository of information spread all over the world and linked to other. The www has a unique combination of flexibility portability and user-friendly features that distinguish it from other services provided by the Internet. www today is client server service in which a client using a browser can access a service using a server. The service provided is distributed over many locations called websites.

6.2 Electronic mail (E-mail)

The only Internet services that is more electronic mail, or E-mail. It a is system for exchanging messages through a computer network. People most commonly use email to send and are receiving text messages, but depending on the software you use; you may be able to exchange audio or video massage
with someone else. When you send an e-mail message, the message is stored a server until the recipient can retrieve it. This type of server is called a mail server. Many mail servers use the post office protocol (pop) and are called pop server. The most common way to create send and receive e-mail is by using an e-mail program and an Internet connection through ISP or LAN.

6.3 News

In addition to the message distribution to mailing list by e-mail, the Internet also supports a form of public bulletin board called news. There are tens of thousands of active Internet news groups. Each devoted to discussion of a particular topic. Many of the most widely distributed news groups are part of system called Usenet, but others are targeted to a particular region or the users connected to a specific network or institution, such as a university or a large corporation.

6.4 File Transfer Protocol (FTP)

It is the Internet tools used to copy files from computer to another. An FTP site is a collection of files, including data files and for programs that are housed on FTP servers. FTP site that are offer called archives may contain thousands of individual programs and files. Public FTP archives permit anyone to meter copies of their files by using special FTP client software. It is not always necessary to use an FTP client to download files from an FTP site kick browsers also.
6.5 **Instant Messaging - IM**

Type of chat software that restricts participation to specific user. There are several popular IM services possible such as windows manager, A O I - Instant manager and others. In buddy list - a list of other uses with whom you would like to chat. Wherever your IM program is running and you are online, the software lets you know where your “buddies” are also online you can chat with than.

6.6 **Online services**

An On Line service is company that offer access, I have rely on subscription basis to small discussion group, data bases on various subjects (such as weather information, stock quotes, new payer articles and so on), an other services ranging form electronic banking and investing of online games. Online service also offers access to the Internet functioning as an ISP to their subscribers. The most popular online services are American online compu-serve and prodigy. In addition to Internet access online services offer other features that typical ISPS do not.

6.7 **Chat**

It is a popular way for Internet users to communicate in real time with other user. Its is often referred to as the “C B radio” of the Internet because it enables a few or many people to join a discussion. Its is a multi-user system where people join channels. Channels are discussion groups.
7. Search Engines

The web is only as useful as the information being found. While it can be a great deal of the just browsing from one website to another, it is not very efficient for locating specified information. In fact, we never locate useful pages; the web will be of little value. The trick, then, is to find websites and pages with the information we want. To do this there are a variety of search tools. The Internet provides a link to many valuable information sources with an centralized database for organization and searching. Essentially, search engines have two parts: the seeking part it does and the searching part you do. Communication Technology is catching up at a very rapid and continuous pace because of the demand for move information. INTRANET has a very vital role to play in today's information communication system. It has become very easy to store information and provide it through world wide Technology to the user who can access informs all over the world. However, owing to this availability of excessive information it has become very difficult for the common user to fitting out the information required by them. Search Engines are most important among the most popular tools for resource discovery on the world wide web. Using the various search engines on the web is chanced by knowing how they are actually designed and especially by knowing the special rules all too open quite different from each other. I have tried to address both these needs ad provided a comparative study of the five most popular search engines in order of usefulness. This articles objective is to hold the new user get the most useful hits when using the various search engines. At the end I have placed a handy little label which summaries center common characteristics among the search.
So search engines use software robots to survey the web and built there database web document are retrieved and indexed. There are two primary methods of text searching, keyword and concept.

- **Keyword** - This is the most common form of text search on the web. Most search engine does their text search on & retrieval using keywords. Most search engines index every word on every page. Others index only part of the document such as the title, headings, subheadings, hyperlinks to other sites & the first 20 lines of the text. Keyword searches have a tough time distinguishing between words that are spelled the same way but mean something different i.e. hard cider, hard stone, hard exam, & the hard drive on our computer.

- **Concept-based searching** - Unlike keyword search systems, concept based search systems try to determine what we mean not just what we say. In the best circumstances a concept based search returns hits on documents that are 'about' the subject/ items we are exploring, even if the words do not precisely match the words we enter in to the query. *Excite is currently the best known general purpose search engine site on the web that relies on concept based searching.*

### 7.1 Types of search engines

There are two major types of search engine.

- **Individual** - An individual engine uses a spider to collect its own searchable index.
• **Meta:** A meta engine searches multiple individual engines simultaneously. It does not have its own index but uses the indexes collected by the section on meta search engines.

### 7.2 Function of the search engine

Search engines have two major elements:

- **Spider:** Program that traverses the web from link to link, identifying and reading pages.
- **Index:** Database containing a copy of each web page gathered by the spider. It is common for search engines to offer indexes on their contents to users. When we go to a search engine site, we can press the index rather than entering search terms. This index may not have human overseeing; the robot might automatically assign a new site to place in the index.

- **Keywords:** A search engine will give us a form to fill up. We enter in words as we would like it to look for. These words might be called keywords or search terms. However, in a broader sense, keywords are any search term we put in.

- **Boolean Logic:** We would use "and" to narrow a search, "or" to broaden it, and "not" to refine it. For example, "Cowboys and Football" would find us a site that deals with the old west herds, and not football teams which are also named cowboys.
• **Plus & minus signs** - Plus & minus signs are another way in which we can refine our research.

• **Search results** - It's then a matter of matching numbers to your keywords.

• **Result list** - Generally we receive a list of 15 sites. At the bottom of the list will be an option to click for more hits.

### 7.3 Major Search Engines

#### 7.3.1 AltaVista

http://www.altavista.digital.com

*The AltaVista search engine developed by digital equipment corporation in palo alto California is one of the most powerful and flexible of the major global www search engines on the net today*. It combines a fast web crawler “Scooter” with cable indexing software. AltaVista claims the largest web index with over 31 million entries to www pages on over 620000 servers world wide it also indexes approximately 4 million articles posted to 14000 UseNet news group daily and maintains them for several weeks. The site is accessed for searches over 30 million limes every week. **Special features** -

Very large database. Default to phrase searching i.e. if you type more than one world in window the worlds will be searched as a phrase if they are found in AltaVista’s phrase dictionary, otherwise the default Boolean logic or. Offers result from the real names service if your search term matches a company product
or service registered at real names. Offers a freshness guarantee that all pages in its index are less than 28 days old. Offers spell checking with some searches. Advanced search offers term proximity searching. Related search feature offers hyper linked topic for executing a subsequent more focused search relating to the initial query. Offers searches within numerous fields such as anchor applet host image link test little URL. Allows results to be retrieved in numerous languages. Offers retrieval of pages by their last modified date this is useful depending on the frequency and comprehensiveness of the AltaVista index. Translates any text or pages retrieved as search results this service is known as Babel fish and may used as a standalone service located at http://babeefish.altavist.com

7.3.2 Hot Bot

http://www.hotbot.com

Hot boot is a relative newcomer to the array of global www.search engine it was developed by inkotomi corporation formerly part of the network of workstation project at the University of California Berkeley Hot boot claims to have searched and indexed the entire web Hotbot has 110 million web sites and 4 million newsgroup articles. It has a lot of advertisements.

Special Features

Offers helpful users fill in template that easily handles complex searches. Offers unusual search options with often accurate results including media type URL, geographic location and personal name. Offers
the option to conduct related searches on predefined topics. Supports automatic truncation. Cluster results by presenting on hit per site. Offers the options to search within these results to further relive result from an existing list of hits. Supports complex boolean searches with parentheses ic behavior and offers limited of pages by their lost modified date this is useful depending on the frequency and comprehensiveness of the hotbot index update.

7.3.3 Excite

http://www.excite.com

The excite search engine developed by architect software company. Uses a wed crawler to build its indexes to WWW sites. Is also an offers wed review of more than 60000 of the highest rented sites on the net in a variety of more categories. Excite claims to be the most comprehensive search tools on the net with over 55 million wed sites indexed. The Excile uses artificial intelligence technology styled “ICE” (intelligent concept Extraction) to establish relationships among the terms its wed Gawler finds on indexed pages.

Special Features

Offers to search for more documents like this one to retrieve pages related to your search results. This is helping full when you find relevant hits and want to see more on the topic. Offers the options to search within existing results to further narrow your topic. Searches for concepts related to your search terms. Database is claimed to be rather current: top pages are visited weakly by the channel content with results for searches on broader popular term. Offers an index of useful
information related to certain topics such as public companies, sports teams, TV shows, popular bands cities, colleges, autos and more. With list of hits offers the option to add displayed related terms to the original search. Offers multiple syntax options including Boolean operations implied Boolean and from-based Boolean searching.

7.4 More Few Search Engines

- **Yahoo! (NASDAQ: Yahoo)**  [www.yahoo.com](http://www.yahoo.com)
- **Google**  [http://www.google.com](http://www.google.com)
- **WEBINDIA123**  [www.webindia123.com](http://www.webindia123.com)
- **Education info India**  [www.educationinfoindia.com](http://www.educationinfoindia.com)
- **123word**  [www.123word.com](http://www.123word.com)
- **All academy**  [www.allacademi.com](http://www.allacademi.com)
- **Box mind**  [www.boxmind.com](http://www.boxmind.com)

8. Use of Internet & search engines in library environment

As no library can be self-sufficient due to shrinking budget & rising prices of library materials, equipments, manpower, etc. Knowledge explosion has related a need for network-based resource of material. In order to satisfy the clients need % also from the economic point of view, a number of work to build up a library network through organized collection of resources for supplying planned coordinated services to a large extent of users come up.

There are 5 major areas of use of Internet by the library professionals.
8.1 Bibliographic verification

Bibliographic services given information about locating information sources and information itself few sites which are useful for bibliographic services in library.

www.myspace.com/index.cfm
www.geekcutture.com/ultimatebb
www.ets.org/toelf.edoois12.html
www.dbmoz.com/m/language
www.everynule.com/index
www.carl.org/uncover

8.2 Acquisition

Acquisition work has become greatly simplified with use of Internet few sites, which is useful for acquisition, work in library.

www.library.vanderbitt.edu
www.cs.cum.edu/web/booktitles.html
www.icweb.loc.gov/globat/text
www.topt.uced.edu/aquistion/index.html
www.uni.com/press/hyt.html
8.3 Reference resources

The reference library can extend the scope of their reference services with the information available in FTP achieves, Gopher servers and world web servers. Few sites which are useful for reference resource in library.

www.alphadictionary.com

www.yourdictionary.com/languagre/romance.html

www.iai.com/glassaries.html

www.history_us.com/the_international encyclopedi

www.encyclopedia.com/glossaries.html

www.uni_paderborm.de/html/dictionary

www.eb.com

www.mhreference.com/EST.html

8.4 Library without walls

The library can be a depository of local information / resources to the worldwide community and at the same time provide processes to the information available all over world to the local user depending on their need and priority.

8.5 Online public access catalogue

OPACs are one of the most important services that are available on the Internet.

www.opac97.bl.uk

www.niss.ac.uk/reference/opacs.html
REFERENCE

1. Raman Nair, R. Internet for Information services. New Delhi: Ess Ess Publication
2. Obenaus, G. *The Internet: An Electronic trave Aslib proceeding*. 1995, 46(4) p 95
4. Raman Nair, R. Accessing Information through INTERNET. New Delhi: Ess Ess Publication, p. 28-32
5. Raman Nair, R. Accessing Information through INTERNET. New Delhi: Ess Ess Publication, p. 28-32
6. Raman Nair, R. Accessing Information through INTERNET. New Delhi: Ess Ess Publication, p. 28-32
7. Norton, Peter. Introduction to Internet
8. Norton, Peter. Introduction to Internet
15. Arents, Hansc, Search Engines 1996
17. http://www.Hotbot.com dtd 05.092005