

A Study on Web Search Engines and User Interfaces

By

Asoke P. Chottopadhyay

Reader

Dept. of Chemistry

University of Kalyani

Kalyani-741235

Swapan K Dasgupta

In-charge Internet Centre,

Dept., Registrar

University of Kalyani,

Kalyani-741235, WB, India.

P.K Panigrahi

Reader

Deptt. of Library & Information Sc.

Calcutta University

Calcutta, India.

E-mail: asoke@klyuniv.ernet.in

E-mail: dasgupta_swapan@yahoo.com

E-mail: panigrahipk@yahoo.com

ABSTRACT

A search begins with a search tool's Web site, reached by means of its address or URL. Each tool's Web site comprises a store of information called a database. Then the database has links to other databases at other Web sites, and the other Web sites have links to still other Web sites, and so on and so on. Thus, each search tool has extended search capabilities by means of a worldwide system of links. A *search tool* is a computer program that performs searches and retrieves information. There are different types of search tools

In the search engine the different forms based user interface provides access to sophisticated searching capabilities that allow users to find references in their fields. Menu GUI (Graphical Users Interface) pops up asking user to select either the "X" Database or the "Y" Database. Accordingly, the appropriate database contents are loaded into the search engine. e.g. When one of the database buttons in the above step is clicked, the basic user interface for the search engine appears as shown.

This pops up another user interface that contains the list of images that satisfied users criteria. If the search criteria are incorrect, then an appropriate error message is posted on the results GUI.

KEYWORDS: World Wide Web, Search Engines, User Interface, Evaluation of Search Engine

0. INTRODUCTION

Development of the subject of 'Web Search Engine', or simply search engines, has just crossed the state of a newborn baby; there have been significant developments during the past decade. This subject is growing up exponentially in the Universe of Knowledge. Actually it falls under the broad heading of Computer Science in a particular group viz. 'design, creation and use of web pages and related information'.

Imagine users searching for information in the world's largest library, where the books and journals (stripped of their covers and title pages) are shelved in no particular order, and without reference to a central catalogue. Then it will be a nightmare for the researchers to get any information from that library. The Internet or the World Wide Web (these two words will be used synonymously, although they have slightly different meanings) is exactly the same thing. Instead of a central catalogue, the web offers the choice of dozens of different search tools, each with its own database, command language, search capabilities, and method of displaying results.

Search engines do just that: they provide users a guided tour of the net, providing access to valuable information, links to databases. In short provide the most valuable item in the world today – information. Naturally, they have become as popular as the Internet. A large number of users go on surfing web sites via search engine every second.

1. ORGANISATION OF THE INTERNET

In order to understand how search engines work, we need to understand a little bit about the Internet or the web first. It is the network of hypertext links that connect its pages. The term hypertext derives from the Greek words for text-over-text. It is essentially non-linear, although it can allow user to create a multiplicity of linear trails. The information may be accessed by alternative paths.

The most flexibility in hypertext structure is facility for an object-oriented design, which also invokes facilities for audio and video applications including Java applets. Just by clicking on the hypertext button, the user can play sound or video clips or jump to a text window. The entire knowledge domain is open to him and necessary links between ideas can be made easily.

This is accomplished by creating links, called hyperlinks, between information. These links are provided so that readers may "jump" from one bit of information to further information about a specific topic (which may have more links, leading each reader off into a different direction). For instance, a user reading an article about bio-acoustics of marine mammals, may be interested in seeing the picture of a dolphin. On the other hand, the user may want to hear the sound of it. He/she may even be interested in seeing what the sound of a marine mammal "looks like" in a spectrogram. The user might even want to find out more about sounds made by other animals in the sea, thus leading him on a completely different, more detailed path. Following these links allows a user to travel effortlessly from one document to another. Since such links have to be built explicitly into each web-page, they reflect the opinions of the designer(s) of a particular page about which other pages are pertinent (and note that a designer does not have to ask permission of the owner of another page to build a link to it). But the link is not reciprocal: building a link from "A" to "B" does not create one from "B" to "A". So any document that has a lot of these hyperlinks pointing to it is obviously well regarded by many page designers and thus likely to be important.

As users can see by these examples, that the medium is not limited simply to text. It can incorporate pictures, sound, even video. So it requires a multimedia approach to gaining information on the net, giving rise to the term hypermedia.⁸ What is also important, especially for a person wishing to sell or display information on the net, is to register his web-page. This is

known as **domain registration**, whereby the particular page or site is given a unique identity on the net. It helps if the **domain name** reflects the character of the site. As search engines mainly list sites or web-pages, they would definitely **not** be able to display a particular page or site if it lacks domain registration.

2. SEARCH ENGINES

A software package for searching particular information or specific topic, in an enormous body of information available on the Internet, is called a search engine. It can also be used to refer to a tool for searching in a locally fixed body of information (intranet, or a database). A search engine thus behaves like a reference librarian who helps user to find information sources.² The most common search engines are Google, Lycos, Yahoo, Webcrawler, Altavista, Aliweb, Infoseek etc.³

Software, such as Netscape, Internet Explorer, etc., which are used to browse items or bits of data or news on the internet, are programs that either call on specific search engines or use similar technology to search for the presence of specified terms on the net.⁴ This is commonly referred to as 'navigating' the net.. Some sites on the net do nothing but help user to find additional information. These sites contain software such as search engines.

3. HISTORY OF SEARCH ENGINES

3.1 Primitive search engine or browser

In the early days of web exploration, the medium was bulky and difficult to use. It was not that the hyperlink concept was deficient or that http did not do this job. A user, in order to find a resource, had to keep moving through hyperlinks, without any central organisation to lead him to his destination.

Lynx was one of the first web browsers. It was originally developed by Lou Montulli, Michael Grobe and Charles Rexae of Distributed Computing Support Group in Academic Computing Services Department at the University of Kansas.

To start up lynx one just had to type "lynx" after logging on to the system. To access a particular URL, for example to get a particular homepage, one had to type "lynx b <http://www.w3.org/>". To move up and down the page, one had to use the up and down arrow keys.⁵

The software developed by a team at the University of Minnesota, a revolutionary development on Internet, first appeared in late 1991. It was the first widely available, easy- to-use, client application for finding applications on internet server, called gopher. With gopher, users could search for information from all over the world with just a few keystrokes. At that time, for file transfer the ftp command had to be used as it is used today, and there was no provision for cut and paste of information as we have today in the graphical users interface or GUI made famous by Windows.⁶

Each of the major search engine differs from the others in its approach to provide services to its users. Still indexing is performed, with retrieval based on particular categories.

3.2 Query-based system for browsing full- text databases

A full-text retrieval engine developed at the University of Waterloo and the Open Text Software Corporation called pat was a very fast text search tool designed for use with the (electronic) New Oxford English Dictionary.^{7, 8}

4. ELEMENTS OF SEARCH ENGINES

A search engine consists of the following four elements:

A program that roams the net or area to be searched, collecting data records (typically, web pages) and links to more such data. These are variously known as spiders, worms, crawlers, or by other colourful epithets.

A database for storing records or collection of information retrieved / recovered by the spiders or other types of collectors. Such databases may contain thousands, or millions or various information, mainly references, links or phrases. Commercial databases, on the other hand, may collect data records in other ways, such as systematically entering full texts of newspapers or journals.

Semi indexing of the database, to enable fast access to terms that users can search for along with their supporting records. Sometimes irrelevant records are displayed due to improper indexing system. It may also happen due to the user's being unaccustomed with the searching methods available. Indexes may be enhanced by controlled vocabularies.

A search interface - the form in which the user must enter his search terms and the software behind it that queries the index, retrieves matches, ranks them for relevance and organizes the data for follow-up searches.

5. SERVICES OF SEARCH ENGINE

A search engine provides service on the Internet that enables the user to search for items of interest. Some such services are free and attempt to capture information from the whole range of material available on the net. Others are subscription-based but in return provide access to specialist publications, full-text retrieval capabilities, or other value-added services.⁹

There exists a variety of sites on the web; each site has its own address, and may allow users to jump to additional sites. Some sites contain only text, while some also contain pictures, sounds, movies, or any combination of these.¹⁰

Search engines allow users to type one or more keywords so that they can search for the information they are looking for. For instance, if a user is trying to find information on the Golden Gate Bridge of San Francisco, he can enter "Bridge", "Golden Gate", or "San Francisco". Once a search is complete, a list of web sites, the user may be interested in, is displayed. The user can then click on any one of the sites or decide to narrow down his search to more specific item(s).¹

6. DEVELOPMENT

In the information age we live in, being able to find information is almost as important as creating or analysing it. People who have information or know how to locate them provide valuable services to a corporation, a farmer, a researcher, a professional such as a doctor or a lawyer or even to an ordinary housewife. Entrepreneurs who are computer literate and have a good feel for search engines, online database and imaginative querying can do well in industries such as

- Ø Yellow Pages which provide service-oriented, technical or commercial information
- Ø News agencies, newspapers or information bureaus
- Ø Event organizers who organize media events, exhibitions, fairs etc.
- Ø Travel or tourism industry, etc.

Nowadays there are some bindings and some rules and regulations have to be followed for storing information on the web. Obviously, it is a good sign for maintaining purity and verifiability of information.¹²

7. TYPES OF SEARCH ENGINES

Among the various search engines available for use on the Internet, some are commercial, while others are freely available as open source software. There are thus two ways of searching. If the user is interested, he can configure and maintain his own search engine, or he can be a partner to a full-service search engine provider.

Nowadays, almost all Internet sites include many kinds of content viz. plain texts, html pages, word-processor documents, object files (e.g. .pdf files), and so on. Therefore, the user's system also needs to support entities such as sgml and extended character sets, such as Unicode. This is because there may also be items in non-English languages which need to be indexed according to local file systems or via the net through search engines.

Most search engines operate on the principle that pre-indexed data is easier and faster to search than raw text. The form and quality of the index created from users original html pages is of paramount importance to how the searches are performed, i.e. how fast, how accurately, and with which advanced features. For most search engines, the index takes the form of a highly optimised look-up database.

8. MULTI-THREADED SEARCH ENGINES

The growth in the number of search engines has led to the creation of "meta" search tools, often referred to as *multi-threaded search engines*. These search engines allow the user to search multiple databases simultaneously, via a single interface. While they do not offer the same level of control over the search interface and search logic, as do individual search engines, most of the multi-threaded engines are very fast. Recently, the capabilities of meta-tools have been extended to include such useful features as the ability to sort results by site, by type of resource or by domain, the ability to select which search engines to include, and the ability to modify results. Such additional features have greatly increased the effectiveness and utility of the meta-tools.

9. HOW SEARCH ENGINES WORK

Search engines allow the user to enter keywords that are run against a database (most often created automatically, by "spiders" or "robots"). Based on a combination of criteria (established by the user and/or the search engine), the search engine retrieves net documents from its database that match the keywords entered by the searcher. It is important to note that when using a search engine it is not searching the Internet, as it exists at this very moment. Rather, it is searching a fixed database that has been compiled some times before the search.

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

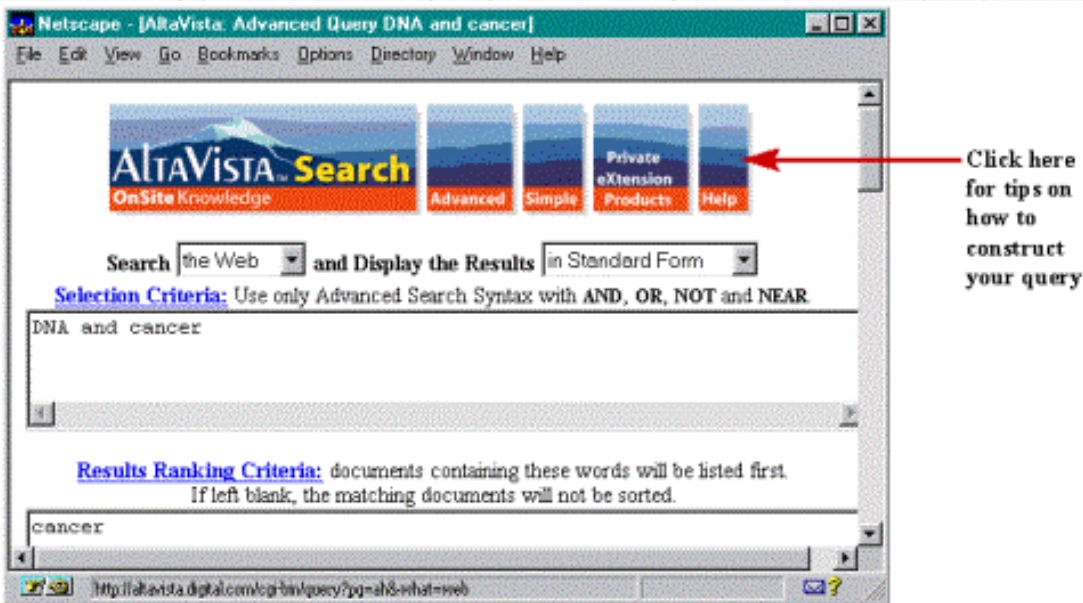
A web search engine runs the search string against the database, returns a list of resources that match the criteria, and display the results for the user. Many search engines use "fill-out" forms as an interface, and support complex queries. Many also include instructions and tips to search the database more effectively. Because search engines can use hypertext, users are able to link directly to resources listed in the result display.

In most cases, search engines are best used to locate a specific piece of information.⁹

10. SEARCH ENGINE INTERFACE

Users interact with the search engine through an interface, submitting questions that "ask" the database if it contains resources that match a specific criteria.¹⁴ The interface presents the user with a place to type in a search string, which may be a word, a phrase, a date, or some other criterion, and a way to submit the request. The search engine runs the search string against the database, returns a list of resources that match the criteria, and displays the results for the user.

Many web search engines use "fill-out" forms as an interface, and support complex queries. Many also include instructions and help in searching the database. Here is an example from the search engine AltaVista



With hypertext and hyperlink, users are able to link directly to resources listed in the results display. For example, a query "DNA and cancer" yields on the same search engine (Altavista)



Most web search engines use automated tools and programs to gather resources. These tools, often referred to as worms, spiders, crawlers, and robots, search thousands of websites worldwide, collect information, and store the information in the database of the search engine. Each web search engine maintains its own database, interface, and special features. In addition, each search engine collects its resources differently; the same query typed into several search engines is likely to produce different results.

11. EXAMPLES OF SOME SEARCH ENGINES

[Google](#)

Web, Images, Groups, Directory, News-New! • Advanced Search • Preferences • Language Tools. Advertise with Us - Search Solutions ...

Description: Enables users to search the Web, Usenet, and images. Features include PageRank, caching and translation...

Category: [Computers](#) > [Internet](#) > [Searching](#) > [Search Engines](#) > [Google](#)

www.google.com/ - 3k - 24 Nov 2002 -

[Yahoo!](#)

Listen to Radio - My Station, Today's Hits, R&B, Rap, Country, Rock, more. • advanced search • most popular. New! Holiday Gift ...

Description: The first large scale directory of the Internet.

Category: [Computers](#) > [Internet](#) > [Searching](#) > [Directories](#) > [Yahoo](#)

www.yahoo.com/ - 28k - 24 Nov 2002 -

[Lycos Home Page](#)

Skip to Search. SERVICES. Chat Clubs Downloads Email Horoscopes, Maps Mobile Movies MP3s Multimedia, Personals Stocks White Pgs Yellow Pgs MORE ». ...

Description: Portal with search powered by Fast, channels, and a directory.

Category: [Computers](#) > [Internet](#) > [On the Web](#) > [Web Portals](#)

www.lycos.com/ - 35k - 24 Nov 2002 -

[AltaVista](#)

AltaVista, Web Image MP3/Audio Video Directory News. AltaVista USA. More Precision,

SEARCH: Worldwide US RESULTS IN: All languages English, Spanish. ...

Description: The default search results consist of Overture and results from the Altavista spider. Displays related...

Category: [Computers](#) > [Internet](#) > [Searching](#) > [Search Engines](#)

www.altavista.com/ - 11k - 24 Nov 2002 -

[AlltheWeb.com](#)

Any language ...

Description: Search with a simple interface and huge database. Also offers news, picture, video, MP3 and FTP search....

Category: [Computers](#) > [Internet](#) > [Searching](#) > [Search Engines](#)

www.alltheweb.com/ - 11k - 24 Nov 2002 -

[GO.com](#)

Provided by Overture Search for: ...

www.infoseek.com/ - 51k - 24 Nov 2002 -

[Metacrawler](#)

Description: Searches the major search engines, throws out the duplicates and summarizes the results. Finds web...

Category: [Computers](#) > [Internet](#) > ... > [Search Engines](#) > [Metasearch Tools](#)

www.metacrawler.com/ - 24 Nov 2002 -

[My Excite](#)

Excite is the leading personalization Web portal, featuring world-class search, content and functionality. From financial portfolios to sports scores, local ...

Description: Portal offering a search service including search of a directory from the ODP, news, and links.

Category: [Computers](#) > [Internet](#) > ... > [Sites Using ODP Data](#) > [E](#)

www.excite.com/ - 48k - 24 Nov 2002 -

[WebCrawler Index - WebCrawler](#)

WebCrawler.com. Web News Photos. Yellow Pages White Pages, New! Advanced Search. No Buttons, No Banners, No Pop-Ups! Learn More, Autos ...

Description: Web directory from ODP and meta search, displaying matching categories.

Category: [Computers](#) > [Internet](#) > ... > [Sites Using ODP Data](#) > [W](#)

www.webcrawler.com/ - 14k - 24 Nov 2002 -

[Search.com](#)

CNET | Price comparisons | Product reviews | Tech news | Downloads, Quicklinks:
Shopping · Maps · Yellow Pages · Jobs · News · People · more... ...

Description: Customizable metasearch. Provides directory channels and news.

Category: [Computers](#) > [Internet](#) > ... > [Search Engines](#) > [Metasearch Tools](#)

www.search.com/ - 20k - 24 Nov 2002 - [Similar pages](#)

12. SEARCH ENGINE POSITIONING

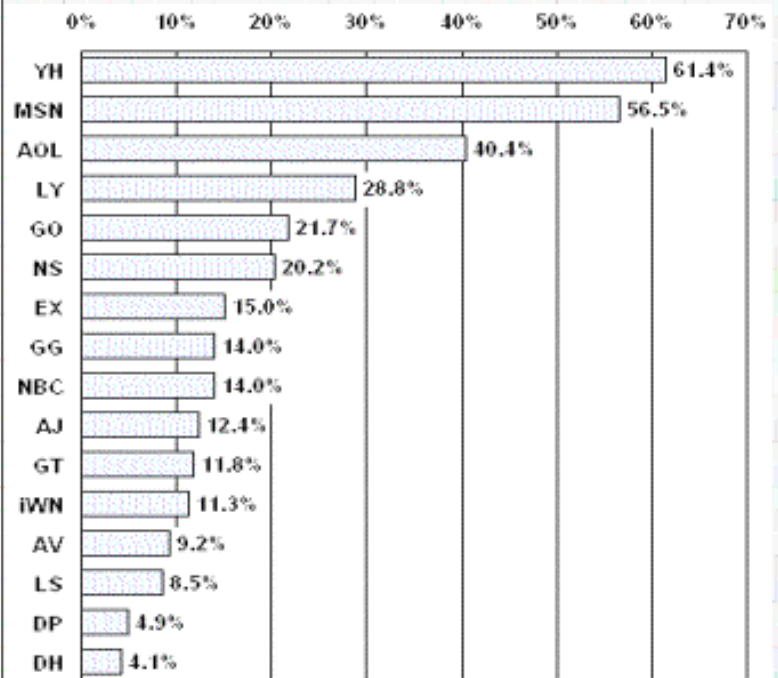
Search Engine positioning is very important to web sites. Since most new visitors come to a site by way of search engines, a position in the top ten rankings on the first page of major search engines (such as Google or Lycos) may quickly increase the number of visitors to a particular web site. It will then be able to attract more advertisements and thus generate more revenues. Often, Internet service providers charge a customer based on actual volume of data transfer to and from his site. This includes keeping a tab on the number of clicks that site receives. Such data, taken by search engines, also help in raising the rating for that particular site.

This may also apply to search engines themselves. A team which ranks search engine performance knows which search engines provide data to how many users, and the best order when registering websites. Top search engine rankings depend on many elements, which may be quite complex.

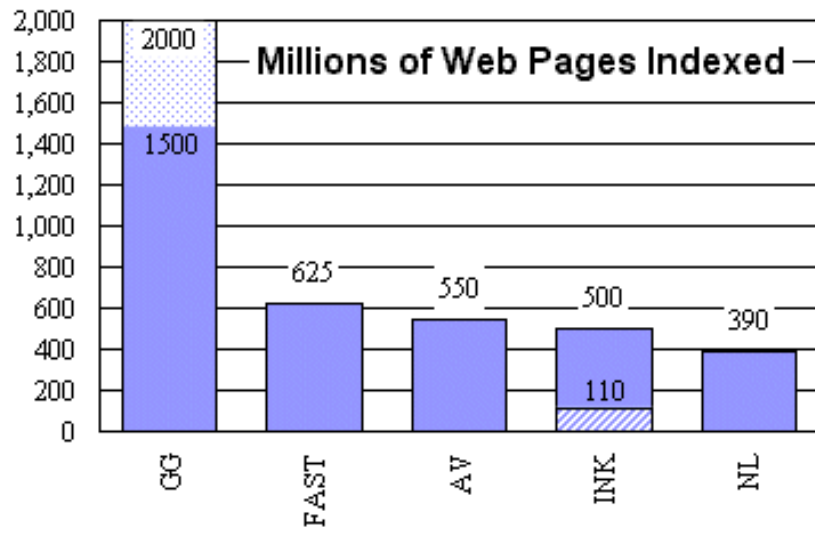
13. SEARCH ENGINE TREND

The figures below show the reach or clientele of several search engines, which is roughly the percentage of web surfers estimated to have visited each search engine or portal during the month. Because a web surfer may visit more than one service, the combined totals exceed 100 percent.

YH=Yahoo,
MSN=MSN,
AOL=AOL,
LY=Lycos,
GO=Go.com,
NS=Netscape,
EX=Excite,
GG=Google,
NBC=NBCi,
AJ=Ask Jeeves,
GT=GoTo,
iWN=iWon,
AV=AltaVista,
LS=LookSmart,
DP=Dogpile,
DH=Direct Hit



Search Engines as of December 11, 2001.



KEY: GG=Google, FAST=FAST, AV=AltaVista, INK=Inktomi, NL=Northern Light.

A user looking for *unusual* or *hard-to-find* information may wish to try any search engines with a large index, as this means that more of the web is covered; consequently, the user has a greater chance of finding what he is seeking. However, for general searches or for seeking information on popular topics, a large index does not necessarily imply better results.

Among the latest trends these days, crawler-based search engines flaunt both how many pages they have in their index, as also the much larger number of pages they visited to create that index. AltaVista says its collection of 250 million pages came from an original set of 400 million. FAST says its 400 million page index was developed from a set of 700 million. Excite's 250 million pages were retained after reviewing 920 million pages. Inktomi claims its core index of 110 million pages was created after analyzing over 1 billion across the web.

Google, the current leader, has indexed 1.5 billion pages, but because of the way it uses link data, it can actually return listings for additional pages that it has never actually visited. This gives it coverage of 2 billion pages. The 1.5 billion "pages" of Google also includes some PDF files, Microsoft Office documents and other text-oriented material. The 1.5 billion figure does not include the Google Groups discussion posts (about 700 million in number) or about 330 million image files.

Determining which search engine has the biggest index is a difficult task. Each of them provides figures it claims as true, but that does not automatically guarantee the truth of these self-reported numbers. This is especially an issue when a search engine claims to be much larger than the others. Users may want a third party to prove these claims. The following gives a relative assessment of their performance, with the best at the top of the list.

Search Engine	Reported Size	Expected Score	Actual Score	Rank
Google	560	1.0	1.0	1
FAST	340	2.0	1.8	2
Northern Light	265	3.0	2.3	3
HotBot	110	4.0	2.3	3
iWon	110	4.0	2.3	3
AltaVista	350	2.0	2.5	4

Yahoo-Google	560	1.0	3.0	5
Excite	250	3.0	3.0	5
Yahoo-Inktomi	110	4.0	4.3	6

14. SCOPE OF FURTHER DEVELOPMENT

Due to advances in information and communication technologies, fundamental changes are heralded in the way learning takes place. "Life-long learning" demanded by "knowledge society" may now be facilitated through anywhere any time e-learning solutions. Simply put, e-learning refers to learning and other supportive resources that are available through a computer connected to the internet / intranet with the help of search engines. The web environment facilitates interaction between the user and

the subject

his colleague through email / chat / mailing list / discussion forum.

video / audio / voice clipping

electronic or digital cash (e-cash) - a new concept to execute cash payments using search engine

an expert / facilitator through email/chat

any other expert in the world on a related or kernel subject.

problem solution, course registration, reporting etc. ^{77, 98}

Therefore, the present work on search engine may cite different opportunities in several subjects or fields of knowledge. The detailed scope of search engines is not possible to discuss in a short paper. For the purpose of compilation of this information data have been collected from various media e.g. print media such as journals, books, newspapers, dailies, weeklies, encyclopedia etc., and also from other references.⁷

15. CONCLUSION

If users today wish to obtain information from any knowledge repository in the global village, they have to first familiarize themselves with a variety of search tools and develop effective search techniques. If they wish to take advantage of the resources from the Internet without spending hours, it is next to impossible. This is where search engines come into the picture, to sift, sort and present before the users what they desire from a sea of irrelevant, uncharted and often, unverified information.

Modern search engines have boosted up research, e-commerce and other academic activities and enable all sections of users' communities to get more resources for their purpose. In order to cope with different search engines, their tools and techniques, the users must become familiar with them. Over and above all, students of Library and Information Science, who are the future managers of our knowledge-driven society, should be adequately prepared for such tasks, which implies the necessity for immediate revision and upgrading of course curriculum for Library and Information Science throughout our country.

REFERENCES

1. RANDALL (E). *Fortune*, January 1, 1998.
2. PANIGRAHI (P K). A language to create web page. Vidyasagar University Journal of Library and Information Science; 1997; 2; p61-70. *In*. Markup languages in web authoring.
3. DASGUPTA (SK). Search Engines: tools for exploring the Internet. Cliber-98
4. The New Penguin Dictionary of Science, © M. J. Clugston 1998
5. ALLEN (DW). Learning guide to internet. 1997. BPB; New Delhi. p 240.
6. GILSTER (P). The internet navigator. 1995. John Willey & Sons In; New York. P-483
7. Fawcett, (H J). New Oxford English Dictionary. 1988
8. *PAT*: The user's guide. Centre for the New Oxford English Dictionary, Waterloo, Ontario, Canada. 1988
9. CEDENO (N). The internet tool kit. 1996. BPB; New Delhi. p-140
10. Billion Docs; Google Upgrades Appliance, *SearchDay*, June 17, 2002.
<http://searchenginewatch.com/searchday/02/sd0617-update.html>
11. BRIN, (S) and PAGE (L). The anatomy of a large-scale hypertextual Web search engine. *In* 30 (1/7) Apr 1998, p.107-17
12. BURKE (MA). Organization of multimedia resources: principles and practice of information retrieval. 1999. Gower; England. p 123-6.
13. CHANG (K). "NCSA Mosaic Exploits Internet." NCSA. Access. Fall 1993.
14. DASGUPTA, (SK). Seminar on: Resource Sharing and Application of IT. Kalyani University, Dept. of Lib. & Inf. Sc
15. Dasgupta, SK. Dissertation paper on; Search Engine: a state-of-the-art report. Kalyani University, Dept. of Lib. & Inf. Sc
16. Dictionary of Business, Oxford University Press, © Market House Books Ltd 1996
17. DYSON (P). Pocket dictionary. 1996. 2nd ed. BPB Pub; New Delhi
18. Fawcett, (H J). New Oxford English Dictionary. 1988
19. Gene Golovchinsky. Master's Thesis. 1993.
20. PANIGRAHI (P K). Web authoring and information services in university libraries. *In*: Academic libraries in internet era, *ed.* P S G Kumar and C P Vashishth. Proceedings of 6th National Convention for Automation of Libraries in Educational and Research (CALIBER-99), 18-20 Feb, 1999, Nagpur. Ahmedabad, Information and Library Network Centre, 1999. p.23-37
21. WARD (P) and KRISTOPHER (D). "Empowering Students in the Information Age.". 10/ 1994.
<http://archive.ncsa.uiuc.edu/SDG/IT94/Proceedings/Educ/ward/ward.html> / wardp@gvsu.edu
22. CEDENO (Nancy). The internet tool kit. 1996. BPB; New Delhi. p-140.

BRIEF BIOGRAPHY OF AUTHORS



***Swapan K Dasgupta* is In-charge of Internet Centre & Spdt., Library Service at University of Kalyani. He has more than 30 years of experience in the Library profession.**



***Asoke P. Chattopadhyay* is the Reader at Department of Chemistry, Kalyani University and has more than 12 years of experience as a University Teacher**



***Pijush Kanti Panigrahi* is presently Reader at Kolkata Department of Library and Information Science at University of Calcutta**