SEARCHING THE WORLD WIDE WEB

Wilfred W. Fong
Assistant Dean
School of Library and Information Science
University of Wisconsin-Milwaukee
Milwaukee, Wisconsin
U.S.A.

ABSTRACT

This article reviews some of the search engines which are currently available on the World Wide Web. The amount of information sources on the Web are growing exponentially daily. It is extremely difficult to keep track of what is available. The search engines provide an attempt to organize various Web information sources. The article discusses the differences and similarities among all these search engines and presents the pros and cons on each search engine described. Some of the options or features described in this article may not be valid since the search engines can be changed without notice. This also makes it more difficult to search for information on the Web.

In the recent years, the Internet has become an important component in many libraries and information centers. Not only because of its electronic communication features (email) but also because of the vast amount of information which is available on it. Many libraries even have positions such as the "Internet Librarian."

As the computer technology improves, graphical interface has gradually become popular among computer users. In the same token, the Internet has gone "graphical" from textual interface by implementing the concept of "point-and-click" based on the principles of hypertext, or hypermedia. This "point-and-click" portion of the Internet is known as the World Wide Web, (WWW or Web). It was started by Tim Berners-Lee and others at CERN, the European Particle Physics Laboratory in Geneva,
Switzerland, as a way to organize information for researchers. According to Berners-Lee, "World Wide Web is a wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents. The project merges the techniques of information retrieval and hypertext to make any easy but powerful global information system." 

Initially, many Internet users treated the Web as an information browser. In less than a few years, the Web has become an information provider. One can read newspapers, do shopping or banking, listen to a radio broadcast, watch a video clip, or simply conduct a search at a library on the Web.

It is estimated that there are currently more than 20,000 sites on the Web. No directory is able to cover the entire Web. New sites may be added and old sites may be removed on a daily basis. However, there are directories and search engines available on the Web to help users browse through the Web to locate information. This article examines such features on the Web.

**Alta Vista (http://www.altavista.digital.com)**

"Alta Vista is the result of a research project started in the summer of 1995 at Digital's Research Laboratories in Palo Alto, California, U.S.A. By combining a fast Web crawler with scaleable indexing software, the team was able to build a large index of the Web in the Fall of 1995." By the end of the testing period, a large index consisting of the full text of over 16,000,000 pages was produced. The site was then made public on December 15, 1995. By the end of February, Alta Vista had grown to more than 21,000,000 pages, ten billions words, and four million daily HTTP requests.

Alta Vista includes a Web Indexer which is approximately 30GB. In addition, this site has a News Indexer which keeps an up-to-date index of the Newsnet.

To find documents using the "Simple Query," a user will enter as many words and phrases as possible and select whether to search the Web or Newsnet. A phrase is any string of contiguous words. One way to link words into a phrase is to use quotes. For example, the term, "American Library Association" (with quotation marks), is considered as a phrase which will search the Web for all documents that contain this phrase. Figure 2 shows the results of this search. However, the term, American
Library Association (without quotation marks) is considered as three words which will yield Web documents that contains any of these three words. Alta Vista will then rank the documents with the most matches first. Figure 3 shows the results of this search.

Users can also eliminate similar words or phrases by using the plus sign, "+", and minus sign, "-". Any words or phrases preceded with "+" are considered as required words while the ones with "-" are to be excluded from the search. An asterisk (*) is for truncation. Any of these search options can be combined in one search statement. Alta Vista is not case-sensitive.

![Figure 1: Home Page of Alta Vista](image)

Alta Vista also provides users with an Advanced Queries option which allows the use of binary operators, AND, OR, NEAR and NOT, to combine words and phrases. Parentheses can also be used to prioritize the search sequence in a series of words and phrases with the binary operators. However, in order to use this option, the words and phrases plus the binary operators must be in quotes. In addition, the users can select the ranking method by selecting the "Results Ranking Criteria" which will first list the documents containing the words or phrases used in the search statement. More detailed information on both options, Simple Query and Advanced Queries, are available by clicking the Help button on the respective page.
Journal of Library and Information Science (April 1996)

Search and Display the Results
Tip: To find links to my page: link: http://prince.com/cinderella.html
Word count: Association:1397327, Library:2070040, American:265687
Documents 1-10 of about 600000 matching some of the query terms, best matches first.

Caddecott Medal
Caddecott Medal. American Library Association, Association for Library Service to Children. Established 1938. A medal presented annually to the illustrator of the most... http://www.ucalgary.ca/757Edelbrown/caddecott.html - site SK - 22 Jan 96

Newbery Medal
Newbery Medal. American Library Association, Association for Library Service to Children. Established 1922. A medal presented annually to the author of the most... http://www.ucalgary.ca/757Edelbrown/newbery.html - site SK - 28 Feb 96

TEACHER RESOURCES
http://www.research.org/teacher/teacher.html - site SK - 3 Dec 93

ATLANTIS: American Theological Library Association Discussion
http://www.hbc-arw.de/hbs/tools/scholar/sections/5_21.html - site SK - 10 Mar 93

ALAWON - American Library Association Washington Office Newsletter
ALAWON - American Library Association Washington Office Newsletter. Resource Type: Electronic newsletter. Keywords: Newsletter, library, library-related U.S. federal...
http://dx.internic.net/cgi-bin/enthtml/newsletter/alawon.html - site SK - 28 Feb 96

Thomas B. Wall
American Library Association Home Page. General Library Sources. Chicago Public Library. Library JobSearch -- University of Illinois, GSLIS. WEB OPAC. Information and...
http://earth.library.pitt.edu/~dbw/library.htm - site SK - 7 Feb 96

American Library Association
Education Association. ALA. American Library Association. 50 East Huron. Chicago, IL 60611. Peggy Sullivan, Executive Director. (312) 280-2153. 1-800 545-2433 ext. 2153....
http://www.ala.org/markets/resources/inherited93.html - site SK - 24 Dec 93

p. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 [Next]

Figure 2: Alta Vista Search Results of the Phrase, "American Library Association."
Search and Display the Results

Tip: Do not use AND or OR to combine words, simply type a few words or phrases.

Word count: American Library Association: about 8000

Documents 1-10 of about 6800 matching some of the query terms, best matches first.

Caldcott Medal
Caldcott Medal. American Library Association, Association for Library Service to Children. Established 1938. A medal presented annually to the illustrator of the most...

http://www.ucalgary.ca/75/Edkbrown/caldcottlt.html - size 5K - 22 Jan 96

No Title
BOOK [american:library:association, Author = "{American Library Association--Young Adult Services Division--Intellectual Freedom Committee}", Title = "Hit list...

http://www.illnet.illinois.edu/C47/books/american_library_association - size 774 bytes - 30 Apr 95

Newbery Medal
Newbery Medal. American Library Association, Association for Library Service to Children. Established 1922. A medal presented annually to the author of the most...

http://www.ucalgary.ca/75/Edkbrown/newbery.html - size 5K - 28 Feb 96

TEACHER RESOURCES

http://gromercy.ios.com/~fooldber/teacher.html - size 4K - 3 Dec 95

John Newbery Medal

http://www.csc.gov/nj/ librarians/Lit_Prizes/Newbery.html - size 21K - 8 Jan 96

Access to Electronic Information
Item 11: American Library Association Access to Electronic Information. Access to Electronic Information, Services, and Networks: an Interpretation of the LIBRARY BILL...

http://cedar.evansville.edu/~library/pmosscss.html - size 8K - 14 Feb 96

ALAWON - American Library Association Washington Office Newsletter
ALAWON - American Library Association Washington Office Newsletter. Resource Type: Electronic newsletter. Keywords: Newsletter, library, library-related U.S. federal...


No Title
We added the American Library Association gopher. I can be found under Resources for Librarians. (7/26/94)


Thomas B. Wall
American Library Association Home Page. General Library Sources. Chicago Public Library. Library JobSearch – University of Illinois, GSLIS. WEB OPAC. Information and...

http://fourth.library.pitt.edu/~dsw/library.htm - size 699 bytes - 7 Feb 96

Figure 3: Alta Vista Search Results of the words, American Library Association.
CUI W3catalog (http://cuiwww.unige.ch/w3catalog)

This search engine is a collaborative web service of Centre Universitaire d'Informatique (CUI) of the University of Geneva and the Software Composition Group at the Universität Bern Institut für Informatik und angewandte Mathematik (SCG). The CUI provides a dedicated server for the W3catalog while the SCG builds the catalog and maintains the convert scripts in addition to providing a mirror version of the catalog. The CUI W3catalog consists of a huge collection of Web resources (Figure 4). It is created daily by concatenating a list of resource databases collected by other net providers:

- NCSA What's new
  (http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-new.html)
  This site consists of a daily updated listings of new or changed sites on the Web. It also contains the Top 5 Sites of the Day, followed by the regular listings.

Figure 4: CUI W3 Catalog Home Page

- NCSA's NCSA Starting Points
  (http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/StartingPoints/NetworkStartingPoints.html)
This site contains hyperlinks to many common Internet-based information resources. It can be used by itself to conduct simple searching of information available on the Web.

- CERN's W3 Virtual Library Subject Catalog and selected sublists
  (http://www.w3.org/hypertext/DataSources/bySubject/Overview.html)
  This is where the Web was born. This site is considered as one of most comprehensive resources databases. It links to more than 100,000 Web sites using hyperlinks—point-and-click—instead of entering subject term(s) or phrase(s). An experimental site using Library of Congress Subject Headings to classify Internet resources can also be found here.

- Martin Koster's Aliweb Archie-like Indexing for the Web
  (http://www.nexor.co.uk/public/aliweb/aliweb.html)
  ALIWEB uses an automatic programs which can simply pick up descriptions of new Web sites, and combine them into a searchable database. Web site owners may also submit a description of their sites in a standard format used by ALIWEB.

- Scott Yanoff's Internet Services List
  Scott Yanoff's list is one of the oldest indexes to the Internet information resources. Before the era of the World Wide Web, Yanoff's list was a popular list on the Gopher. This list is very unique in that it maintains the old connections to Gophers, telnetable Internet sites, and ftp sites in addition to the Web sites.

- Simon Gibbs' list of Multimedia Information Sources
  (http://viswiz.gmd.de/MultimediaInfo/)
  This list contains extensive information sources on multimedia, including multimedia computer products, multimedia access to the Internet such as RealAudio, newsgroups, and listservs that are on the subject of multimedia.

- John December's list of Computer-Mediated Communication Information Sources (CMC) and Internet Tools Summary (ITS).
W3 Catalog

Please enter a search word/pattern or provide a Perl regular expression:

NB: Searches are case-insensitive.

Result of search for "librar* and map*":

: United Nations Scholarly Workstation
  Yale Univ. Library and Social Science Statistical Lab, New Haven, CT, US
  A collection of digitized texts, finding aids, data sets, maps, and pointers to print and electronic
  information pertinent to the United Nations.
  http://www.library.yale.edu/un/unhome.htm (now)

: UC San Diego Maps
  Research Services Dept., Univ. Library, La Jolla, CA, US
  Links to major WWW sites featuring land, ocean, atmospheric and planetary data. In addition, we mount
  local GIS projects which depict geographic and demographic aspects of the San Diego/Tijuana region.
  Currently featured are a series of graphics produced by San Diego Association of Governments, and
  digital cartography produced at El Colegio de la Frontera Norte in Tijuana. (noM)

: Hampshire College
  Academic & Administrative Computing, Amherst, MA, US
  Hampshire College's WWW server information is for everyone in the Hampshire community, including
  students, faculty, staff, and alums. Come look at our current events, our campus map, what's new in the
  library art gallery, our 5-college searchable on-line course catalog, and contributions from students. (now)

October 25, 1994
  The PCL Map Collection, General Libraries, The University of Texas at Austin, has added
detailed maps of Haiti including a street map of Port-au-Prince and topographic maps. (now)

October 17, 1994: VIBE, the magazine of urban music (Jazz, Blues, Reggae, Hip Hop, Rap) and youth
culture is proud to announce VibeOnline! -- a free service on the World Wide Web. Featuring isMapped
covers as navigation interfaces, full text and photos excerpted from the paper version and a library of the
latest music releases, VibeOnline! breaks new grounds in Internet publishing. VIBE is a joint venture
between affiliates of Quincy Jones/David Salzman Entertainment and Time publishing Ventures, Inc. (now)

August 25, 1994: VIBE, the magazine of urban music (Jazz, Blues, Reggae, Hip Hop, Rap) and youth
culture is proud to announce VibeOnline! a free service on the World Wide Web. Featuring isMapped
covers as navigation interfaces, full text and photos excerpted from the paper version and a library of the
latest music releases, VibeOnline! breaks new grounds in Internet publishing. VIBE is a joint venture
between affiliates of Quincy Jones/David Salzman Entertainment and Time Publishing Ventures, Inc. (now)

June 29, 1994: Library Information Technologies at Case Western Reserve University (CWRU)

Figure 5: Sample Page of W3catalog Search Results
(CMC: http://www.december.com/cmc/info/index.html)
This two lists are similar to the Simon Gibb's list of Multimedia Information Sources.

Marcus Speh's User Documents for DESY and HEP
(http://info.desy.de/general/users.html)
This list is made possible by the Deutsches Elektronen-Synchrotron (DESY) in Hamburg for the High Energy Physics community. The DESY laboratory performs basic research in high-energy and particle physics as well as in the production and application of synchrotron radiation. All documents indexed in this list are submitted by the users for the benefit of others. The documents are basically in the following areas: General Physics; C++ Virtual Library; GNU's Not Unix; Literate Programming Library; Lattice Field Theory; and Multigrid Algorithms.

One of the unique features of this search engine is that the users are encouraged to mirror the W3catalog on their site.\(^5\)

The W3catalog also supports Boolean operators, AND and OR, and truncation (*). Figure 5 shows a page of a search result using the phrase "librar* and map*". W3catalog includes the initials of each resource databases identifying where the document(s) are originally indexed.

Unlike most search engine sites, the W3catalog does not provide detailed documentation describing the databases and search features/options. In addition, users cannot submit URLs directly to the W3catalog as this is a concatenation of the nine resource databases as listed above.

**InfoSeek Guide (http://guide.infoseek.com/)**

Infoseek Guide gives the users access to many popular Internet resources, including World Wide Web pages, Gophers, FTP sites, Usenet newsgroups and the Frequently Asked Questions (FAQs) lists posted to the news.answers newsgroup.

There are two options to use the Infoseek Guide to search for information resources on the Internet. The first one is by browsing while the second one is searching via search terms inputted by users. The topics
for browsing are located on the left hand side of the Infoseek Guide home page (Figure 6).

There are six options available for searching:

1. World Wide Web - this option will search the entire Web, or those sites that are known to the Infoseek Guide;
2. Infoseek Select Sites - this is to limit the search to those sites selected and categorized by the Infoseek Guide;
3. Categories of Sites - this is to search for a category within Infoseek Select by choosing Categories of Sites;
4. Usenet Newsgroups - this option is to search through the vast information available on the Internet newsgroups;
5. E-mail Addresses - this option will allow the users to find someone's email address;
6. Web FAQs - this can be a handy option for ready-reference on a specific topic.

Although Infoseek does not offer the support of Boolean operators, it has several search options. The double quotation marks (" ") implies to search the entire term as a phrase but not as individual words. The hyphen, "-", between two words signifies that the two words must be adjacent to each other. Other options and features of the searching capabilities are available from the "Search Tips" option on the Infoseek homepage.

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

Lycos is the first five letters of the Latin term for the Wolf Spider which began as a play on words with the Web/spider. Lycos basically consists of three services. Similar to Infoseek Guide, Lycos provides a browsing and searching mode to its database for the Internet resources. Lycos has sixteen categories to explore the new directory services known as A2Z (Figure 7). Lycos also has a "Point NOW" option connecting to current headlines from newspapers on the Web. Users can submit their sites to be included in the Lycos database.

Lycos allows the use of Boolean operators, AND and OR, in searching the database. The "enhance your search" option will prompt the user for a search form. By using a search form, a user can narrow or widen the search very easily. The user first has to enter a series of search terms and then select how the terms should be combined, i.e. using AND, OR,
Figure 6: Infoseek Guide Home Page
or matching up to seven terms. The user can further adjust the selectivity of the Lycos search engine by choosing one of the five levels from loose match to strong match. The result display option controls the numbers of hits matching a query. The default is 10 hits on each results page. The last option is for the results detail display which limit the amount of information Lycos will display.

There are three levels of detail display: standard, detailed, and summary. The standard is the default setting which displays the most relevant and frequently-used information about a site including its title, abstract, and URL. The detailed option will display the following for a site: title, ranking (with 1 being the highest rank among all hits), links to outside resources, words matched in page, abstract, and URL. The summary results display provides only the title and ranking. All titles listed are hyperlinked to the document. Figure 8 shows the search results of documents on "Architecture" and "Greek" using the "enhance your search" option with the setting of "strong match" and "standard results" display.

Lycos has detailed documentation on how to use its searching features and different search options. There is one advertisement on every page in Lycos. It is a current trend of the Web that many search engines have made their page(s) available for advertisement from companies which are attracting businesses from the Web users.

Figure 7: Home Page of Lycos
Lycos search: greek architecture
47,834,280 unique URLs
Found 83,704 documents with the words greek (26,563), architecture (18,168)
Megallen (http://www.mckinley.com)

The name of this site, Megallen, is taken from Ferdinand Magellan, a Portuguese explorer who navigated the Strait of Magellan in 1520. Megallen, produced by The Mckinley Group, is a Web guide to the Internet resources. It is different from the other search engines introduced in this article as Megallen consists of sites that are rated and reviewed by a team of editors and writers from the Mckinley Group. The users can find editorial content, a directory of rated and reviewed Internet sites, a vast database of yet-to-be-reviewed sites, and a powerful search engine. Megallen includes Web sites, FTP sites and Gophers, newsgroups, and Telenet sites.

The rating system is used in evaluating each Web site, awarding from one to ten points in three areas:
1. Depth: Is it comprehensive and up-to-date?
2. Ease of Exploration: Is it well organized and easy to navigate?

Figure 9: Home Page of Magellan
A points system is built based on the above criteria resulting in an overall rating from one to four stars. This rating shows next to each of the “hits” for a query along with a review.

Although Magellan does not support the use of Boolean operators in its searching option, it does allow the use of “-” to exclude a search term, i.e. the NOT Boolean operator. This search engine, as many other search engines, ranks the sites in the results according to the “relevancy” which is the frequency of the word or words used in the query.

Some of the documents may contain a green dot as assigned by the Magellan. It indicates that, at the time of review, it contained content apparently only intended for mature audiences according to the subjective reviews.

![Magellan Internet Guide](image)

**Search for:**
Search rated and reviewed sites only entire database

Related Topics
- Humanities & Social Sciences / Library & Information Studies
- Interdisciplinary Studies / Area & Cultural Studies
- Languages
- Education / Public Libraries
- Travel & Tourism / Europe
- Entertainment & Pop Culture / Books

31009 results returned, ranked by relevancy.
(Displaying results 1 through 10)
Narrow to the 3 matches from the rated and reviewed database
p. 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 [Next]

****
**Chinese American Librarians Association (CALA)**

Review: A superbly designed interface and colorful collection of introductions introduces the mission and history of its producer to the world via the Web. Here is as much information about the organization as anyone is likely to desire. Find an outline of contents:
http://www.lib.sdu.edu/awww/cala

***
**Chinese-Language-Related Information Page**

Review: The Chinese-Language-Related Information Page is a comprehensive navigational tool pointing you to Chinese-language resources on the Internet. Visitors will find links to online language courses and teaching materials, sources of Chinese text files, linguistics.
http://www.webcom.com/~bamboo/chinese/chinese.html

**Chinese-Language-Related Information Page**

Chinese-Language-Related Information Page The Chinese-Language-Related Information Page will be continually improved to make it a comprehensive navigational tool pointing you to Chinese-language-related resources. The intended audience is anyone who...

http://www.webcom.com/~bamboo/chinese/chinese.html

---

**Figure 10: Sample Search Results of Magellan using two terms: Chinese and Librarians.**
Yahoo! (http://www.yahoo.com)

Yahoo! is one of the first search engines available for the Web. It is a searchable hierarchical subject-oriented guide for the World Wide Web and Internet. It was started by two Stanford University computer science doctoral students, David Filo and Jerry Yang, in 1995. Yahoo! maintains a database which contains titles, URLs, and comments. All of these three items are searchable. The original index database was constructed by its authors. It also contains an option to allow users to submit URLs for inclusion in the database.

Yahoo! lists sites and categorizes them into 14 appropriate subject categories: Arts, Business and Economy, Computers and Internet, Education, Entertainment, Government, Health, News, Recreation, Reference, Regional, Science, Social Science, Society and Culture (Figure 11). Sub categories are used under each main subject category (Figure 12).

There are basically three ways to use Yahoo! to search the Web and the Internet. The first one is to browse through its hierarchical structural by using the subject and sub categories. Figure 12 shows the second level of the sub-categories of Business and Economy. A new item is indicated by a New! symbol. The number in the parenthesis represents the total numbers of URLs which are available under this item. By selecting one of the sub categories, Yahoo! will show the listing of the URLs under that sub categories. Figure 13 shows the sample page of a listing under the sub category of Consumer Economy. The hierarchical structure is shown on the top of the page.

The second way to use Yahoo! is to use its search capabilities to lookup information available on the Web and the Internet. The search features allow a user to enter a keyword or set of keywords. Yahoo! is not case-sensitive. By using the keyword(s), it will search its entire database or the user may limit the search within a subject category. Yahoo! will then compile a list of not more than 100 matches. The list is similar to the one shown in Figure 13.

Yahoo! also provides the users with another way to browse its database. There are several options available on the top of the Yahoo! Home Page (see Figure 11): What's New, What's Cool, What's Popular, and Random Link. The Headlines option connects the users to the latest news section which is updated hourly. Users can submit URLs by clicking on the “Add URL” option. The “Info” button contains useful information on Yahoo!
Figure 11: Categories of Yahoo

Figure 12: Sub-Categories in Yahoo (Business and Economy)
Business and Economy: Consumer Economy

Options
Search all of Yahoo Search only in Consumer Economy

- Automobile@ 
- Corporate Protests (7)
- Consumer Financing (2) 
- Public Interest Groups@ 
- Consumer Law@ 
- Research (7)

- Consumer Information Center - You've seen the TV commercials, now view & download the publications online, free! The Consumer Information Catalog lists over 200 free and low cost publications covering health, business, home and money matters, and more!
- Nolo Press' Self-Help Law Center - online legal assistance for the little guy.

- Ambler's Central
- Consumer Brochures - IIE
- Consumer Brochures Show 95@
- Consumer Information - Info available includes how to protect yourself against sweepstakes scams, work-at-home schemes, phony investment deals, confidence swindles, etc.
- Consumer News (UWGC)
- Consumer World - 600+ "consumer" links.
- Consumers for World Trade - world trade issues as they affect consumers. The page is official information of Consumers for World Trade, an advocacy group.
- Flea Market @ IUW
- FTC Consumer Brochures
- Home Recording Rights Coalition - In October, 1981, within weeks of a court declaring sales of VCRs illegal, the Home Recording Rights Coalition began fighting to protect consumers' rights to use home recording products.
- Internet/Advocacy Center - Internet repository for consumer protection information.
- PriceWeb - a list of nationally advertised prices on many popular computer peripherals. This makes it easy to find the lowest price.
- The EnviroProducts Directory - This directory should help eco-conscious consumers find companies that they will feel comfortable doing business with.
- Ticketmaster Numbers
- U.S. Consumer Product Safety Commission

Yahoo! Have you seen Yahoo! Internet Life yet?
It's Free when you click here!

Copyright © 1994-96 Yahoo! All Rights Reserved.

Figure 13: Sample Page of a Yahoo! Search Result
(Consumer Economy)
Conclusion

There are more than 50 Web search engines available on the Internet. Many other search engines such as Open Text (http://www.opentext.com), Web Crawler (http://webcrawler.com), and Architext's Excite (http://www.excite.com) are quite heavily used by many Web users.

Recently, some Web sites provide an option known as an "All-in-one" search page, which is to perform a search using several search engines at the same time. For instance, Netscape's NetSearch (http://home.netscape.com), Microsoft's Internet Explorer's search option (http://www.msn.com/access/allinone.htm), or William Cross' Search page (http://www.albany.net/allinone/) have such multiple engine capabilities. The users can select one or more search engines to perform a search query.

How does one rank the best search engines? This author believes that it is extremely difficult to rank Web search engines. It has been noted that different search engines index different Web documents. For example, Digital's Alta Vista includes USENET messages in their database; Megellan and InfoSeek make FTP sites and Gophers searchable, etc. This explains why some search engines yield more hits than others. For years online information retrieval researchers has been arguing the question, "Is more better?" Unlike some online periodical indexes, a "hit" from a Web search engine may never be found because Web documents are "dynamic." The documents may be available from the Web today but not on the next day. The URLs may be changed or deleted from the Web sites. In addition, many of these search engines accept submissions of URLs sent by individual Web users. Therefore, a document may appear in one search engine but not the other. An extensive and comprehensive search of a subject can be tedious and time consuming on the Web.

Some search engines allow users to use Boolean operators, i.e., AND, OR, NOT, while others use words only, or a combination of both. For example, Lycos uses a minus "-" as NOT. Search results are also displayed differently among search engines. Some engines have developed a ranking system to display their search results in the order of the best match of the search terms entered.
The WWW Virtual Library

This is a distributed subject catalogue. See Category Sitemaps, Library of Congress Classification (Experimental), Top Ten most popular fields (Experimental), Statistics (Experimental), and Index. See also arrangement by resource type, and other subject catalogues of network information.

Mail to maintainer of the specified subject or www-request@nsl.org to add pointers to the list, or if you would like to contribute to administration of a subject area.

See also how to put your data on the web, and what's new in Asian Studies and Social Sciences. All items starting with % are NEW and rarely maintained. New this month: Astronomy & Astrobiology North America Resources, Astronomy, Solar-Terrestrial Physics and Chemistry, Australian Indigenous Languages, Buddhist Organisation & Evaluation of Information Resources, Electronic Reference & Scholarly Citation of Internet Sources, Held & Korean Collection, Tibetan Reference & Cultural Centre, Tibet Buddhism.

Aboriginal Studies
This document keeps track of leading information facilities in the field of Australian Aboriginal studies as well as the Indigenous Peoples studies.

Astronomy and Astrobiology

Astronomy, Solar-Terrestrial Physics and Chemistry

African Studies

Agriculture

AIDS

Animal health, well-being, and care

Anthropology

Applied Linguistics

Archaeology


Figure 14: The WWW Virtual Library Homepage

The WWW Virtual Library

This is a distributed subject catalogue, arranged by the Library of Congress standards. See home page, Category Sitemaps, top ten most popular fields (Experimental), Statistics (Experimental), and Index. See also arrangement by resource type, and other subject catalogues of network information.

Mail to maintainer of the specified subject or add pointers to the list, or if you would like to contribute to administration of a subject area.

Agriculture

Agricultural Sciences, of History

Agriculture, Library Science

Agricultural Economics

Agricultural Science, Food, and Related Arts

Agricultural Science, General Works

Here are classified works too general or comprehensive to be classified with any particular subject, however broad.

Geography

Anthropology

Archaeology

History

World History

General and Local History

Including geography and description of individual regions and countries

Languages and Literature

Law

Medicine

Library and Information Science

Music and Books on Music

Philosophy

Psychology

Religion

Figure 15: The WWW Virtual Library

Experimental Subject Tree based on the Library of Congress
Another problem in searching the Web via these search engines is that they are not using a "controlled- vocabulary" system, or standardized subject headings. Many of the search engines index their systems using words as "concepts," e.g., InfoSeek. Different search engines may use different search terms. This makes Web searchers unable to make a uniform search among the search engines impossible. Some engines use subject trees and hierarchies (e.g., Yahoo), which are best for browsing materials on the Web. Many subject trees also have searching options available. A comprehensive subject tree is the WWW Virtual Library (http://www.w3.org/pub/DataSources/bySubject/Overview.html), which can be accessed via many search engines including the W3Catalog. In the Virtual Library site, there is an experimental subject tree based on the Library of Congress Subject Headings (http://www.w3.org/pub/DataSources/bySubject/LibraryOfCongress.html).

As of this date, many of these search engines are still available on a free basis as they are currently supported by advertisements from companies which have their products marketed on the Internet.

![Internet Searches](image)

**Figure 16: Search Option in the Microsoft Internet Explorer**

*(Sample All-in-One Search Page)*
Searching for information on the Web is not an easy task. Although many of these search engines have good help instructions to assist novice searchers, search strategy is still an important component. Browsing may seem to be easy, but the users may lose their direction after several hyperlinks.

There are, however, several Web sites developed by librarians to help users evaluate and understand the techniques searching the Web. These help pages can be found at the Berkeley Digital Library SunSite (http://sunsite.berkeley.edu/Help/searchdetails.html); "Understanding WWW Search Tools" by Jian Liu (http://www.indiana.edu/~librcsd/search/); "Beyond Surfing: Tools and Techniques for Searching the Web" by Kathleen Webster & Kathryn Paul (http://magi.com/~mmelick/it96jan.htm); and "World Wide Web Searching Tools-An Evaluation" by Ian R. Winship (http://www.bubl.bath.ac.uk/BUBL/IWinship.html).

The implications of the rapid Internet development to the library profession have greatly increased. Librarians should be trained in improving their skills in formulating effective search strategies in searching the Web. Although it is difficult to keep up with the development of the Internet, continuing education for librarians should be emphasized more than ever. Library school's curriculum should be reconsidered and reconstructed to prepare their graduates for these new kinds of reference sources and services.

NOTES

3. URL: http://www.altavista.digital.com/cgi-bin/query?pg=q&what=web
4. URL: http://iamwww.unibe.ch/~scg/W3catalog/doc/
5. In order to mirror the W3catalog site, a user has to install the W3catalog Mirroring Suite once. Instructions are available at the URL listed in note #4.
7. URL: http://www.mckinley.com