

by Eric Lease Morgan



by Eric Lease Morgan

Availability

This guide is freely available. Download it. Copy it. Edit it. Do with it as you will. Give it away! In an effort to make this more possible, this text is available in a number of formats from a number of URLs:

1. PDF - This version is for printing, not reading on the screen
Location: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/pointers.pdf>>
2. ASCII Text - This is the simplest version in terms of formatting.
Location: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/pointers.txt>>
3. MacWrite - Here you will find a MacWrite Pro version.
Location: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/pointers.hqx>>
4. HTML - This is the online version.
Location: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/>>

Table of Contents

Introduction	i
The Internet	2
Data types	6
Search Engines	12
Search Engines	25
Fee Based Services	35
Other searching services	42
Browsable Lists	50
Knowledge Management and Information Evaluation	60
Ladder of Understanding	66
Appendix A: Simple webliography	71

Introduction

This, Pointers 4 Searching, Searching 4 Pointers, is an annotated bibliography (webliography). Its purpose is to provide you with starting points for methods and strategies for using the Internet to find academic information as well as become familiar with the advantages/disadvantages and strengths/weaknesses of Internet searching.

It does this through a series of topics:

1. **The Internet** - This is the briefest of introductions to the Internet.
2. **Data types** - A description of the various ways information formats available through the Internet include: HTML pages, mailing lists, newsgroup postings, books, serials, bibliographic citations, definitions, maps, government reports, etc.
3. **Search engines** - Searching techniques and output are similar and dissimilar to traditional online searching: Boolean logic, regular expressions, relevance ranking. As well as, some services like AltaVista, Hotbot, Excite, and DejaNews, attempt to index and provide searching mechanisms for subsets of Internet-based information.
4. **Other searching services** - There are many other search engines indexing and providing access to locally developed Internet information like WAIS, Harvest, mailing list archives, regular expressions, and commercial indexing applications.
5. **Fee-based services** - There are some information providers whose services are available through the Internet but require payment in order to be used: Dialog, JSTOR, EbscoHost, UMI ProQuest, SilverPlatter, Wilson Index, America Online, etc.
6. **Browsable lists** - Collections of Internet resources supplement and compliment Internet indexes: Yahoo, A2Z, WWW Virtual Library, etc.
7. **Information evaluation and knowledge management** - Methods describing how-to effectively keep your Internet information under control are outlined here.
8. **Ladder of Understanding** - The Internet effects the principles of librarianship, or does it?

Pointers 4 Searching, Searching 4 Pointers was originally designed as a half-day workshop to be augmented with group activities, presentations, and demonstrations. At the same time, this bibliography, the handout for the workshop, was written to be able to stand on its own as a reference for using the Internet to search for information in general.

Enjoy, and I hope you find what you seek.

Eric Lease Morgan, eric_morgan@ncsu.edu
<http://www.lib.ncsu.edu/staff/morgan/>
October 1997



THE INTERNET

Oddly enough, the human-computer interaction of traditional online searching is not very much unlike the client/server model of computing, the principle computing metaphor used for Internet communications.

The Internet

To truly understand how much of the Internet operates, and especially for information seeking on the Internet, it is important to understand the concept of client/server computing.

The client/server model is a form of distributed computing where one program (the client) communicates with another program (the server) for the purpose of exchanging data.

The client's responsibility is usually to:

1. Handle the user interface.
2. Translate the user's request into the desired protocol.
3. Send the request to the server.
4. Wait for the server's response.
5. Translate the response into "human-readable" results.
6. Present the results to the user.

The server's functions include:

1. Listen for a client's query.
2. Process that query.
3. Return the results back to the client.

A typical client/server interaction goes like this:

1. The user runs client software to create a query.
2. The client connects to the server.
3. The client sends the query to the server.
4. The server analyzes the query.
5. The server computes the results of the query.
6. The server sends the results to the client.
7. The client presents the results to the user.
8. Repeat as necessary.

This client/server interaction is a lot like going to a French restaurant. At the restaurant, you (the user) are presented with a menu of choices by the waiter (the client). After making your selections, the waiter takes note of your choices, translates them into French, and presents them to the French chef (the server) in the kitchen. After the chef prepares your meal, the waiter returns with your diner (the results). Hopefully, the waiter returns with the items you selected, but not always; sometimes things get "lost in the translation."

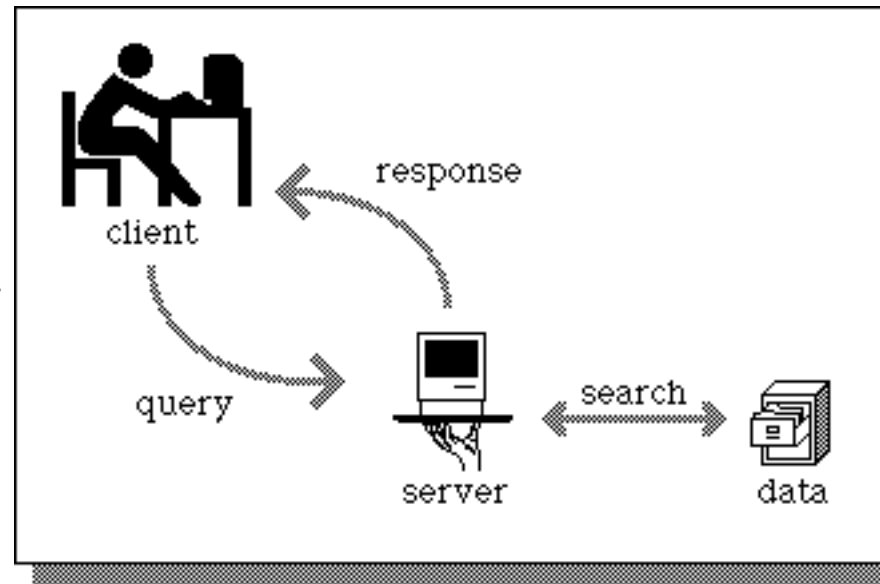
The process of providing traditional online searching services is also a lot like the client/server model of computing. For example, suppose you were to provide traditional online searching services.

This is how it would work:

1. A person would approach you (you're the client program) and express an "information need."
2. You would ask the person a number of questions allowing you to translate their need into a search strategy (the protocol).
3. Next, you, the client, would send the search strategy, the protocol, to the remote database (the server)
4. Wait for a response
5. Interpret it.
6. Present it to the person.

It is important to know about client/server computing in the Internet environment because it sets the stage for describing the sorts of data and information available through the Internet. It helps you diagnose where problems may be occurring when your Internet searching experiences prove fruitless. It helps you understand that the Internet can only provide you with data, and at most, information, but not knowledge and wisdom.

From a computing standpoint, flexible user interface development is the most obvious advantage of the client/server model. It is possible to create an interface independent of the server hosting data. Therefore, the user interface of a client/server application can be written on a Macintosh and the server can be written on a mainframe.



Client/server computing also allows information to be stored in a central server and disseminated to different types of remote computers. Since the user interface is the responsibility of the client, the server has more computing resources to spend on analyzing queries and disseminating information. This is another major advantage of client/server computing; it tends to use the strengths of divergent computing platforms

to create more powerful applications. Although its computing and storage capabilities are dwarfed by those of the mainframe, there is no reason why a Macintosh could not be used as a server for less demanding applications.



DATA TYPES

This is a description of the various information formats available through the Internet including: HTML pages, mailing lists, newsgroup postings, bibliographic citations, definitions, etc.

Data types

The Internet is made up of “hosts and hosts” of data. This data is only turned into information after it has been organized and given some sort of value. Consequently, if you want to glean any information from the Internet, then you have to locate data.

Since there are many different formats of data you have to learn what tools to use to locate the data you need. The definitions below describe the various types of data found via the Internet search engines in next sections.

Definitions

ASCII text files

As defined here, these are files encoded using the American Standard Code for Information Interchange (ASCII) format. They may or may not have any structure, but they can be read by the vast majority of computers anywhere. Frequently, this is the format of files that are created locally and indexed using some sort of tool like WAIS or Verity's Search '97.

bibliographic citations

Frequently, these are pointers to books, journal

articles, or other physical media containing information. The pointers usually include author(s), title(s), publisher name, a date, and pages. There are few standard ways this information is formatted, especially on the Internet and on fee-based services.

computer products

These are pieces of hardware and software, usually for purchase, whose primary purpose is to enhance the functionality of your computer.

current events

All types of news fit into the category of “current events.” This content may include USENET news, and or the sorts of news published from wire services and broadcasting companies.

directories

Directories are mostly used to locate people, places, or services. They are ordered lists of information usually consisting of names, addresses, telephone numbers and descriptions.

email addresses

Similar to telephone numbers, email addresses are used to facilitate communication with one or more

people or services via the Simple Mail Transfer Protocol (SMTP).

full-text articles

Full-text articles are journal or magazine publications existing in digital form. This form may include a rudimentary ASCII representation of the publication, an image of the original publication, or a combination of both ASCII and image representations. These forms can usually be delivered to your desktop via a simple download or through some sort of document delivery service via your 'fax machine.

HTML files

HTML (hypertext markup) files are structured ASCII text files implementing a particular Standard Generalized Markup Language (SGML) document type definition (DTD). HTML files represent the vast majority of Internet content index by the Internet search engines. HTML files, when correctly and thoroughly structured, can not only communicate a lot of information, but they can communicate a lot of information about the information. Consequently, well-constructed HTML pages are easier to locate using Internet search engines since many Internet search engines take advantage of HTML's internal structure to create their indexes.

image files

These are digital graphics or pictures usually in the Graphics Interchange Format (GIF) or JPEG (Joint Photographic Experts Group) format.

mailing lists

Mailing lists are one type of electronic bulletin boards for Internet. The other is USENET newsgroups. They are most often served via mailing list manager programs like ListProc, LISTSERV, ListSTAR, or Majordomo. The technology works by first submitting your email address to a database program. Another program then "listens" for new email and redirects it to every record in the database. This way people can share common problems, solutions, and interests via email. Some mailing list programs have the ability to archive the mail they receive. These archives are what may or may not be searchable through various Internet search engines.

news

News is much like current events except its content is generally less popular or trendy. News is more authoritative and contains less opinion than current events.

sounds

In this context, sounds are stimuli producing auditory sensation saved in a digital format. The most common format is AIFF (Audio Interchange File Format).

USENET newsgroup posting

USENET newsgroup postings are just like mailing list messages, except they are delivered to the end-user via the Network News Transfer Protocol (NNTP). Because they are the posting of humans, the type of content they contain mirrors human thought and interests. These are good places to find very “hot” topic or very timely information.



ABOUT SEARCHING

Searching techniques and output are similar and dissimilar to traditional online searching: Boolean logic, regular expressions, relevance ranking.

Search Engines

You've heard it a million times. "You can find anything on the Internet." Well, you also know this is overstatement. Furthermore, you know that finding things on the Internet is a lot like finding a needle in a haystack, or to turn the phrase, "drinking from a firehose."

On the other hand, if you know how to make Internet search engines work for you, then you might very well be able to find what you are looking for on the Internet. The trick is knowing how to use the tools properly. In general, you should apply traditional online search techniques to your information needs. This means you must:

1. articulate your information need as specifically as possible
2. choose a search tool whose features accommodate the articulated need
3. understand how to translate your need into terms the tool understands
4. apply your search query
5. evaluate the results
6. repeat as necessary or until resources are expired

The definitions below describe the most common features implemented by the search engines in the next section. The cited items following the definitions describe in more detail the Internet searching process. Of particular interest are the items by Duda, the Internet Scout Project, Campbell, and Zorn.

Definitions

Boolean logic

Articulated by George Boole, Boolean logic is a method of creating new sets of items by combining existing sets with logical union (or), intersection (and), and exclusion operations (not). The syntax for implementing these operations are different from service to service, but generally speaking the plus sign (+) or "and" will be used to specify intersection, a vertical bar (|) and "or" denotes union, and the minus sign (-) or "not" denotes exclusion. By default, most search engines treat multiple terms as phrase searches or with a logical "and" and not an "or."

"concept extraction"

Coined by Excite, Inc., "concept extraction" attempts to match search terms and phrases with other terms and phrases with similar meanings. Its a thesaurus.

date ranges

This feature allows you to specify two dates (days, months, or years) in your query and have the results contain only items whose "publication" is between those two dates.

domain

Domains are the names given to sets of computers with Transmission Control Protocol/Internet Protocol (TCP/IP) configurations. Every computer implementing TCP/IP will be associated with an IP number and/or name. These names are domains such as `www.whitehouse.gov`.

field searching

Databases, by definition, contain records, items with similar characteristics. These characteristics are generally implemented as fields. Typical fields in a database of bibliographic records include author, title, date, and pages. Field searching allows you to specify a query for items in one or more of these similar characteristics.

natural language

The Holy Grail of search features, natural language queries are queries allowing you to submit search strategies representing human speech. They usually

work by removing the stop words from the input and then performing some sort of Boolean operation on the remaining set.

nested queries

Nested queries are queries allowing you to perform multiple set manipulations with one command-line. Nested queries allow searchers to override the default precedence of query operations like Boolean operations. They are almost always implemented with the use of parentheses.

phrase searching

Phrase searching are queries where the input terms represent matches for consecutive strings of text. These strings of text can be case-sensitive or case-insensitive. Phrase searching is the default behavior of most Internet search engines.

proximity

Very similar to phrase searching, proximity searches allow you to specify the number of words are allowed to appear between two query terms. Proximity searches go one step beyond phrase searches since they allow you to specify a greater number of string combinations.

ranges

Ranges allow you to specify queries where two query terms can be represented in a linear fashion and result sets will contain items between the query terms.

regular expressions

Regular expressions are a symbolic means of describing strings of text. They are an abstract representation of words and terms based on positions with lines of text, casefulness, letters, digits, and symbols. Regular expressions have no usefulness when it comes to describing the meaning of words and phrases, only their "shape."

relevance ranking

Relevance ranking is a method of ordering lists of search results where items of the most statistical significance appear at the top and least significance appear at the bottom. A relevance ranking is a number calculated through a combination of number of term occurrences, position(s) in the located document, length of the located document, "weight" of the search terms, and number of term appearances in the entire database.

spelling options

My catalog is not necessarily your catalogue. Search

engines that account for spelling differences process queries in such a way that multiple spellings of search terms can be retrieved. This is done through the use of regular expressions or thesaurus terms.

truncation/stemming

Words in Western languages are built upon shorter string of letters with discrete meanings. Through the use of truncation/stemming techniques, search queries can be constructed to locate items containing variations of the shorter strings, thus increasing recall. Truncation/stemming implementations vary from tool to tool, but are usually performed by concatenating dollar (\$) or number (#) signs to the ends of search terms.

wild card

Wild card searches are a variation of truncation/stemming searches except queries are allowed to specify variations in spelling within terms and not necessarily at the end of terms.

About Internet searching

Title: Surfing and Searching

Remote: <URL:http://www.pathfinder.com/@@NYqtIQQAlBOFQHHZ/fortune/buyersguide_test/jobcsurf.html>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/jobcsurf.html>>

Description: This easy-to-read article from Fortune magazine outlines why it is important to not only to use Internet search engines, but to combine these tools with browsable lists. The article quotes Reva Basch who says, rightly so, that one good way to find information is to monitor USENET newsgroups for experts in a subject area and then query those experts directly. It just goes to show that people are the real sources of information, not computers.

Author: Bates, Mary Ellen

Title: Seven Deadly Sins of Online Searching

Remote: <URL:<http://www.onlineinc.com/oluser/NovOU/bates11.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/bates11.html>>

Cost: 0

Description: This article humorously defines seven things you should avoid when doing online searching: pride, haste, avarice, apathy, sloth, narrow-mindedness, and ignorance. These “sins” are

put into the perspective of online searching and are intended to be kept in mind when doing traditional online searches as well as Internet-based search strategies. Not only is this a quick and easy read. It makes a lot of sense too.

Author: Brandt, D. Scott

Title: Relevancy and searching the Internet.

Citation: Computers in Libraries ISSN:1041-7915 Sept 1996, v16, n8, p35(3)

Remote: <URL:<http://thorplus.lib.purdue.edu/~techman/CIL996.htm>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/CIL996.html>>

Description: Relevancy has long been important in the world of information gathering. It's vital in tasks such as locating the most relevant resources, using the most relevant retrieval methods, and making sure the information found is relevant to the need. Relevancy has many synonyms-applicability, correspondence, pertinence-and relates to many kinds of decisions we make in our lives. Because so many of these decisions require sifting through massive amounts of information, relevancy is an important factor in making the best decisions we can.

Author: Brooks, Monica

Title: Research on the Internet

Remote: <URL:<http://www.marshall.edu/library/facwww.htm>>

Description: In one (slightly lengthy) page, this Argus

Clearinghouse approved site outlines how to do research on the Internet. It lists objectives, Web basics, compares and contrasts search engines, describes Boolean logic, OPACS, FTP, and important aspects of government information. This is a fine research aid, and a great starter page for students in academe.

Author: Campbell, Karen

Title: Understanding and Comparing Search Engines

Remote: <URL:<http://www.hamline.edu/library/bush/handouts/comparisons.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/campbell.html>>

Description: This document lists reviews of Internet search engines. It is a good place to begin when you want to see what other people have said about the available tools.

Author: Clyman, John

Title: Finding your needle in the Web's haystack.

Citation: PC Magazine ISSN:0888-8507 July 1996, v15, n13, p39(3)

Remote: <URL:<http://www.zdnet.com/pcmag/issues/1513/pcmg0080.htm>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/pcmg0080.htm>>

Description: "Six popular World Wide Web search engines are reviewed. DEC's Alta Vista is the most powerful and comprehensive full-text search system, often generating five to

ten times as many matching documents as other engines. The default Simple Search is easy to use, but most users will want to learn the Advanced Search mode, which supports Boolean operators and case-sensitive and proximity searches. Excite is one of the oldest search engines and has fallen behind competitors. Its quirky interface is annoying, and users cannot modify the original query from the results page. InfoSeek generally finds the desired information and has a good query-by-example function. Lycos combines full-text searching with two searchable indexes, but neither index is comprehensive. ... WebCrawler is adequate but not especially powerful. Yahoo! is purely a searchable index built by real people instead of automated agents. Users can search only for words in the category name or site summary."

This article, while just slightly dated, provides a more than adequate overview of the most popular Internet search engines.

Author: Community Networking

Title: Research Works

Remote: <URL:<http://cn.net.au/>>

Cost: 0

Description: Like AskScott, this service models its format around a librarian reference interview. It frames the services it provides in the form of questions:

1. Are you looking for a particular research resource
2. Can we suggest the best resources to approach

3. Would you like some training in professional research
4. Or perhaps our past projects, business plan and background will interest you!

Number 2, "Can we suggest the best Resources to Approach", is the most interesting since it presents you with a tiny form to complete and based on your input suggests tools for searching.

Personally, I think this sort of service represents a future opportunity for librarians to share their expertise with information seekers. There is no reason why the sort of service represented here could not be expanded and improved to become a more comprehensive service.

Author: Duda, Andrea L. editor

Title: Untangling the Web

Remote: <URL:<http://www.library.ucsb.edu/untangle/>>

Description: This is a "Proceedings of the Conference Sponsored by the Librarians Association of the University of California, Santa Barbara and Friends of the UCSB Library" from April 26, 1996 University Center, University of California, Santa Barbara. The linked abstracts and papers describe all aspects of World Wide Web development in a library setting. Of particular interest are "Yahoo! Cataloging the Web" by Anne Callery, "Spinning a Web Search" by Mark Lager, and "Spiders and Worms and Crawlers, Oh My: Searching on the World Wide Web" by Ann Eagan and Laura Bender.

Author: Haskin, David

Title: Right Search Engine

Remote: <URL:<http://www.internetworld.com/1997/09/report.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/report.html>>

Description: "We tested six of the leading search engines: AltaVista, Excite HotBot, InfoSeek, Lycos, and WebCrawler. We found that each can find an enormous amount of information, but a few are clearly superior in the way they home in on the most relevant information and in the interface they offer."

Author: Hearst, Marti A.

Title: Interfaces for searching the Web.

Citation: Scientific American ISSN:0036-8733 March 1997, v276, n3, p68(5)

Remote: <URL:<http://www.sciam.com/0397issue/0397hearst.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/0397hearst.html>>

PDF: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/hearst.pdf>>

Description: "New user interfaces are being developed to help users find information on the Internet using an intuitive and explorative approach. This system places a topic within an information tree which users can follow toward the specific information they need."

This article outlines possibilities for spatially organizing information and search results for the purposes of better retrieval and analysis.

Author: Internet Scout Project

Title: Searching the Internet

Remote: <URL:<http://scout.cs.wisc.edu/scout/toolkit/searching/>>

Cost: 0

Description: This is an excellent set of pages! It not only lists search engines and general as well as specific browsable lists, but it describes these things in greater detail than most of the other "About Internet Searching" pages listed in the present guide. This is one of the better places to start if you want a no-nonsense introduction to searching the Internet.

Author: Lynch, Clifford

Title: Searching the Internet.

Citation: Scientific American ISSN:0036-8733 March 1997, v276, n3, p52(5)

Remote: <URL:<http://www.sciam.com/0397issue/0397lynch.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/0397lynch.html>>

Description: The increasing number of Web sites on the Internet will require changes in present day search engines to enable them to find the information that the user specifically

requires. This may also involve changes in the way data or information is formatted for entry into the Internet.

Author: Mauldin, Michael L.

Title: Searching the World Wide Web; Lycos: design choices in an Internet search service.

Citation: IEEE Expert ISSN:0885-9000 Jan-Feb 1997, v12, n1, p8(7)

Remote: <URL:[http://dlib.computer.org/dynaweb/ex/ex1997/@ebt-link;hf=0?target=if\(eq\(query\(%27%3CFNO%3E+cont+x1008%27\),0\),1,ancestor\(ARTICLE,query\(%27%3CFNO%3E+cont+x1008%27\)\)\)>](http://dlib.computer.org/dynaweb/ex/ex1997/@ebt-link;hf=0?target=if(eq(query(%27%3CFNO%3E+cont+x1008%27),0),1,ancestor(ARTICLE,query(%27%3CFNO%3E+cont+x1008%27)))>)>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/Mauldin.html>>

Description: Lycos is a search engine that can be used for collecting, storing and retrieving information about pages on the World Wide Web. Lycos is based on the LongLegs program, and incorporates Pursuit retrieval engine and Lycos Catalog of the Internet. It uses a proprietary spider program written in C for foraging. Its search is based on popularity heuristic and is biased towards more popular and useful Web pages. The Pursuit retrieval program uses an inverted file containing document identifiers. Lycos simplifies the search for relevant information in the Web.

Author: Munson, Kurt I.

Title: World Wide Web indexes and hierarchical lists: finding tools for the Internet.

Citation: Computers in Libraries ISSN:1041-7915 June 1996, v16, n6, p54(4)

Description: Indexes and hierarchical lists are two types of search tools for locating information on the World Wide Web and other Internet resources. Indexes, such as Lycos and Open Text, provide access to records by matching search terms against descriptive cataloging. Hierarchical lists, such as Yahoo!, use descriptive and subject cataloging to group common resources by location.

Author: Notess, Greg

Title: Search Engines Showdown

Remote: <URL:<http://imt.net/~notess/search/>>

Description: "This site summarizes, reviews, and compares the search features and database scope of the Internet search engines and finding aids." It does this by dividing its content to the following nine parts: search engines, directories, multi-search, USENET & others, strategies, statistics, reviews, definitions, and bibliography. The bibliography alone is worth the price of admission.

Author: Notess, Greg R.

Title: Searching the hidden Internet

Citation: Database ISSN:0162-4105 June-July 1997, v20, n3, p37(4)

Remote: <URL:<http://www.onlineinc.com/database/JunDB97/nets6.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/nets6.html>>

Description: There are several sites on the Internet that cannot be found through the automated indexing of various sites. Some of this information can only be found in the PDF file. A new level of Internet databases and smart searching techniques will make these sites more accessible. Some sites on the World Wide Web require registration or a log-in procedure.

Author: Notess, Greg R.

Title: Searching the Web with Alta Vista

Remote: <URL:<http://www.onlineinc.com/database/JuneDB/nets6.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/notess.html>>

Description: The following sentence from the article itself pretty well sums up its content:

I compared a single, non-truncated keyword search on Alta Vista with the same search on the best known of the other search engines: Inktomi, InfoSeek, Open Text Index, Lycos, Excite, and

WebCrawler. Searching on a fairly distinctive single word eliminates the disparity among the search engines in how they handle multiple word searches. In each of the five searches, the Alta Vista search resulted in a much higher number of hits. In fact, Alta Vista searches came up with two to six times the number of hits found by the second ranking search engine.

The article goes on to describe the strengths of AltaVista when compared to other search engines.

Author: Stix, Gary

Title: Finding pictures on the Web.

Citation: Scientific American ISSN:0036-8733 March 1997, v276, n3, p54(2)

Remote: <URL:<http://www.sciam.com/0397issue/0397lynchbox1.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/0397lynchbox1.html>>

Description: "Search engines are under developed to increase their ability to find graphic information on the Internet. Current search engines rely on text captions to access graphic information. Future developments would incorporate the ability to compare various visual features, such as contrast, coarseness, directionality, shapes and color."

The article describes some future possibilities for locating graphics on the Internet.

Author: Tweney, Dylan

Title: Searching is my business: a gumshoe's guide to the Web.

Citation: PC World ISSN:0737-8939 Dec 1996, v14, n12, p182(8)

Remote: <URL:http://www.pcworld.com/software/internet_www/articles/dec96/1412p182.html>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/1412p182.html>>

Description: The Web can be a powerful research tool, but users must know what they are looking for and focus carefully to avoid wasting time. Techniques for maximizing Web productivity are presented. Web directories such as Magellan and Yahoo are fast, no-nonsense tools that point directly to useful sites but cover only a small fraction of all Web content. Search engines use automated 'spider' programs to locate information but tend to generate too many irrelevant matches if the user is not careful. Techniques for narrowing a search include being specific and adding Boolean operators. There are also 'meta' search tools on the market that organize and consolidate search results by sending queries to multiple search engines simultaneously. Numerous search assistants are available, but few are useful; three of the better ones are Knowledge Discovery's More Like This, Symantec's Internet FastFind and Quarterdeck's WebCompass 2.0. Offline browsers such as FreeLoader and First Floor's Smart Bookmarks save time and money.

Author: Yahoo!

Title: Searching the Web

Remote: <URL:http://www.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Searching_the_Web/>

Cost: 0

Description: This site represents the most comprehensive collection in this guide. It includes pointers to hundreds of search engines, tutorials, and Internet directories. The collection is divided into many subject areas, including: Indices, All-in-One Search Pages, Comparing Search Engines, How to Search the Web, Indices to Web Documents, Regional Robots, Spiders, etc. Documentation, Search Engines, Web Directories.

Author: Zorn, Peggy; Emanoil, Mary; Marshall, Lucy; Panek, Mary

Title: Advanced web searching: tricks of the trade

Citation: ONLINE (WILTON, CONN), vol. 20, no. 3, 12ppp, 1996

Remote: <URL:<http://www.onlineinc.com/onlinemag/MayOL/zorn5.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/zorn5.html>>

Description: "The purpose of this report is to look closely at several Web search systems that provide advanced search features and search a comprehensive and authoritative database of Internet sites. Based on these two key requirements, the Alta Vista, InfoSeek, Lycos, and Open Text are considered for evaluation. The search features looked for include complex Boolean, duplicate detection, keyword(s) in context, limiting retrieval by

field, proximity and/or phrase searching, relevancy ranking of results, retrieval display options, search set manipulation, and truncation."



SEARCH ENGINES

Some services like AltaVista, Hotbot, Excite, DejaNews, and OPACs attempt to index and provide searching mechanisms for subsets of Internet-based information.

Search Engines

For the most part, all the Internet search engines provide the same sorts of features: Boolean logic, relevance ranking, and phrase searching. In fact, phrase searching and relevance ranking are these tool's strong points. They are designed so the end-user can put in a few words and get something (useful) out. The problem comes in when you have to know how to specify a phrase search. Do you surround your query with single quotes, double quotes, or does the engine default to phrase searching without any delimiters?

Relevance ranking made its first popular appearance with WAIS in 1991. Nobody ever understood exactly how it worked then, and few people understand how it works now. Most Internet search engines will not tell you how they determine relevance. Its "proprietary." In a nutshell, the services evaluate the number of times a particular term appears in a document and then compare that with the length of the document. This ratio determines relevance. Therefore, if a document was 1 word long and the document contained your single search term, then the located document is 100% relevant. Obviously, this sort of searching does little help when the theme of the document is never explicitly stated. Do we chalk this up to poor writing skills?

In general these services are weak on field searching and range qualifications. Remember, these tools are designed for popular culture, not necessarily power searches. Yet, the power search features of many of these services can be quite illuminating. They are usually implemented as more extensive forms for the end-user to fill out thus eliminating the need to know greater detail of the underlying database's structure.

Since each of the Internet search engines listed below are so similar, it is suggested you pick the service you like the most, read its instructions as thoroughly as possible, and use that service first moving to other services as needed. Personally, I like AltaVista the best since it has the most search features and since it indexes entire documents, not just parts of them.

Lists of search engines

Title: Collection of All the Search Engines You Will Ever Need!

Remote: <URL:<http://www.ts.umu.se/~hezze/>>

Cost: 0

Ease of use: Mindless

Description: This is a simple list of 24 search forms allowing you to search various Internet indexes. Rudimentary.

Title: Feeality Searches & Links

Remote: <URL:<http://www.freeality.com/>>

Cost: 0

Description: This is a simple list of search engine forms. Rudimentary.

Title: Librarians' Index to the Internet

Remote: <URL:<http://sunsite.berkeley.edu/InternetIndex/>>

Cost: 0

Description: "The Librarians' Index to the Internet is a searchable, annotated, subject directory of close to 3,000 Internet resources chosen for their usefulness to the public library user's information needs." Originally developed by a librarian, Carole Leita, this site has grown from a gopher bookmark file into what it is today. If you're nice to her, then maybe she will give you her Perl scripts and you can start your own collection.

Author: A2Z

Title: Internet indices, Directories, & How-To Guides

Remote: <URL:http://a2z.lycos.com/Internet/Indices__Directories_and_How-To_Guides/>

Cost: 0

Description: This break-your-browser page contains Lycos's list of search engines. Unlike Excite's list, this list does offer explanations, but many of the resources seem dated.

Author: Drudge, Bob

Title: My Search Engines

Remote: <URL:<http://www.refdesk.com/newsrch.html>>

Cost: 0

Description: This is a list of at least 260 search engines from all over the 'Net. The engines are divided into subject areas and available via an alphabetical list as well. Like many of the other lists of search engines, My Search Engines contains forms directly searching each entry. It also provides pointers to few Internet searching tutorials and briefly describes how to use each tool.

Author: Excite

Title: Searching

Remote: <URL:http://www.excite.com/xdr/Computing_and_Internet/Internet/Searching/>

Cost: 0

Description: Here is Excite's collection of Internet search engines. No commentary. No explanations.

Author: Haskin, David

Title: Right Search Engine

Remote: <URL:<http://www.internetworld.com/1997/09/report.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/report.html>>

Description: "We tested six of the leading search engines: AltaVista, Excite HotBot, InfoSeek, Lycos, and WebCrawler. We

found that each can find an enormous amount of information, but a few are clearly superior in the way they home in on the most relevant information and in the interface they offer.”

Author: HotWired, Inc.

Title: Wired Cybrarian

Remote: <URL:<http://www.wired.com/cybrarian/frame/tech/web.html>>

Cost: 0

Description: The sidebar of this web page contains links to search engines and browsable lists of interest to academic librarians. The page sports the usual “wired” look with lots of colors, graphics, and extreme HTML. At the same time, this page concisely lists many of the more popular items for searching and browsing. The commentary is, at the very least, entertaining.

Author: Macmillan Publishing USA

Title: Search the Internet

Remote: <URL:<http://www.mcp.com/general/search/srcheng.html>>

Description: This page lists and describes a few lesser known search engines, as well as a few golden oldies.

Author: McKinley Group, Inc.

Title: Search Engines

Remote: <URL:http://www.mckinley.com/magellan/Reviews/Computing/World_Wide_Web/Searching_The_Web/Search_Engines/index.magellan.html>

Cost: 0

Description: Here is another list of search engines, but this one is ordered by Magellan’s ranking system and reviews. Not a bad place to get an overview of searching mechanisms.

Author: McKinley Group, Inc.

Title: Searching the Web

Remote: <URL:http://www.mckinley.com/magellan/Reviews/Computing/World_Wide_Web/Searching_The_Web/index.magellan.html>

Cost: 0

Description: Here you will find reviews and links to resources evaluating resources. “Meta-data about meta-data?” Seriously, this site points the way to sites similar in purpose to the present guide.

Author: Mentor Marketing Services

Title: Eureka!

Remote: <URL:<http://www.best.com/~mentorms/eureka.htm>>

Cost: 0

Ease of use: Easy

Description: This is a list of 46 search engines and standard

forms for using them. This site also, briefly, describes each of the search engines.

Author: Netscape Communications, Inc.

Title: Netscape Net Search

Remote: <URL:<http://home.netscape.com/escapes/search/ntsrchrnd-4.html>>

Cost: 0

Ease of use: Easy

Description: Using very sophisticated HTML and JavaScript, this page provide direct access to the major Internet indexes. Parts of each index are displayed in a window allowing direct data input. The page provides context sensitive help based on the currently selected index. The page is also (slightly) customizable. It also points to the most popular browsable lists.

Author: Nicholson, Scott

Title: AskScott

Remote: <URL:<http://www.askscott.com/>>

Cost: 0

Ease of use: Easy

Description: “Just as the reference librarian in a library helps you find the best reference work for your search, AskScott helps you find the most appropriate Internet reference tool for your search.”

This set of pages frames sets of browsable lists and Internet search engines with in the reference interview model. By asking

themselves simple questions, a user of AskScott is lead to various Internet tools that may help provide the answers to their questions. The entire set of pages is also available for download and use in a local library. The whole idea is a good one and represents a fresh approach to locating information on the Internet.

Author: Zamboni

Title: Zamboni's Search Engines

Remote: <URL:<http://www.iglou.com/zamboni/search.html>>

Cost: 0

Description: This is a very simple list of remote search engines with no explanation nor forms to complete.

Internet search engine

Title: Excite

Remote: <URL:<http://www.excite.com/>>

Cost: 0

Ease of use: Easy

Data types: HTML files; news; USENET newsgroup posting;

Search features: phrase searching; Boolean logic; relevance ranking; “concept extraction”;

Description: Excite's search engine is much like everybody else's. Its power search form simply spells out the search features on a more descriptive manner. The features include phrase

searching, Boolean logic, and relevance ranking. Excite uses the same “concept extraction” technique of locating documents a Magellan because Magellan is owned by Excite. The site also hosts pointers to other search engines as well as collections of Internet resources on popular subjects.

Title: Hotbot

Remote: <URL:http://www.hotbot.com/>

Cost: 0

Ease of use: Easy

Data types: HTML files; image files; sounds; USENET newsgroup posting;

Search features: date ranges; field searching; phrase searching; Boolean logic; domain;

Description: HotBot uses an almost purely forms-driven interface to its search engine. A unique feature of HotBot is its ability to select a wide range to Internet media types to seek. For example, it includes: Image, Java, Acrobat, audio, JavaScript, VB Script, Video, ActiveX, Shockwave, VRML. Some these media types can be located in other services through the command line, but not all. HotBot also provides access to a number of specialized databases on the subjects of USENET, top news sites, classified ads, domain names, stocks, discussion groups, shareware, businesses, people, email addresses. HotBot’s browsable lists of Internet resources merely points to Wired’s Cybrarian <URL:http://www.wired.com/cybrarian/frame/tech/web.html>. While HotBot has a lot of features, feel a bit out of control when I

am limited to only a forms interface.

Title: InfoSeek

Remote: <URL:http://www.infoseek.com/>

Cost: 0

Ease of use: Easy

Data types: HTML files; news; USENET newsgroup posting;

Search features: phrase searching; Boolean logic; relevance ranking; domain; field searching;

Description: This site does not have “power” search feature, nor does it have complicated forms interface for completing queries. All queries are from a simple command line. The command line supports the usual suspects: Boolean operations, phrase searching, and a few field searches. Its collection of Internet resources looks just like everybody else’s too. The only difference is they publish a collection policy; “InfoSeek Select sites are chosen by our editors from the index based on their editorial value, traffic and the number of links to the sites.” Like Yahoo, once you are browsing a particular subject area you can use the searching mechanisms to find resources from that area.

Title: Lycos

Remote: <URL:http://www.lycos.com/>

Cost: 0

Ease of use: Easy

Data types: HTML files; sounds; image files; bibliographic citations;

Search features: phrase searching; Boolean logic; natural language;

Description: Starting out in academia (Carnegie Mellon University) and then going commercial, Lycos is the grand daddy of the Internet search engines. It is also the service that really built a name for itself by rating sites as the “Top 5%” and awarding sites the distinction of displaying the Top %5 badge. The service’s search features rely greatly on variations of phrase searching, but the service also has a reputation of locating more hits than the other services. As evidenced by a visit to its home page, Lycos is becoming more and more like commercial service offering more than simple Internet searching and advertisements. Lycos also offers search links to Barnes and Noble as well as UPS. Obviously, this service is thinking the bigger picture and is quickly leaving academia behind.

Title: MetaCrawler

Remote: <URL:http://www.metacrawler.com/>

Cost: 0

Data types: HTML files; USENET newsgroup posting; computer products;

Search features: domain;

Description: This is one of the true all-in-one search engines. By supplying MetaCrawler with query, MetaCrawler searches the most popular Internet indexes, collates the results, removes the duplicates, and displays them on your screen. Since the remote index’s search features are diverse, the search features of

MetaCrawler are a bit generic. You can limit your search by domains, type of data, and phrases (or not). This may very well be a service to keep an eye on.

Title: Northern Light Search

Remote: <URL:http://www.nlsearch.com/>

Cost: \$1-10

Ease of use: Easy

Data types: bibliographic citations; HTML files; full-text articles;

Search features: phrase searching; truncation/stemming;

Description: This is a newcomer into the fray of Internet search engines. Like the resources described in the article “Searching the Hidden Internet” by Notess, this search engine not only provides access to a database of broad WWW documents, but articles from selected magazines and journals (Special Collections) as well. Presently the search features are weak. On the other hand, this engine analyzes the results of queries and creates groups of documents, clusters, much in the same way AltaVista searches are “refined.” When you locate a document in the Special Collection you are interested in, you can pay as you go for the article or set up an account. Presently, articles range in price from \$0-10 per copy.

Title: Open Text

Remote: <URL:http://index.opentext.com/>

Cost: 0

Ease of use: Easy

Data types: HTML files; USENET newsgroup posting; email addresses; current events;

Search features: field searching; Boolean logic; proximity;

Description: It's just a guess, but I believe OpenText's search engine is really an advertisement for its commercial document storage and retrieval software. Like a number of the other services, OpenText's interface is entirely menu-driven. It consists of a number of blank fields surrounded by qualifiers. The qualifiers are of two types: fields and operators. The fields include: summary, title, first heading, URL, anywhere. The operators include: and, or, but not, near, followed by. It is obvious from this description that OpenText is strong on phrase searching and the proximity of words. Field searching is a bit weak.

Author: Digital Equipment Corporation

Title: AltaVista

Remote: <URL:http://altavista.digital.com/>

Cost: 0

Ease of use: Challenging

Data types: HTML files; USENET newsgroup posting; image files;

Search features: Boolean logic; field searching; phrase searching; relevance ranking; truncation/stemming; nested queries;

Description: Of the free Internet search engines, AltaVista is the one with the most features and seemingly widest coverage. While the searching syntax is a bit obtuse, it is the most powerful and when used correctly can provide result sets that are easy to manage. Like a growing number of search engines, AltaVista employs statistical methods for refining the results of queries. When result sets are "refined" they are groups into categories based on the most frequently used words in the results. This method is similar to Verity's services as well as North Light Search. If you like AltaVista, then buy the book AltaVista Search Revolution and you will like it even more.

Author: McKinley Group, Inc.

Title: Magellan

Remote: <URL:http://www.mckinley.com/>

Cost: 0

Ease of use: Easy

Data types: HTML files;

Search features: Boolean logic; nested queries; phrase searching; relevance ranking; "concept extraction";

Description: The strength of this service lies in its Internet reviews. Using a 4 point scale, Magellan reviews Internet sites and ranks them in subject lists accordingly. Unfortunately, we don't know what their "collection management" policy is, so we can not determine how or why they have picked particular items for the collection nor what they look for in particular resources. The search engine itself is divided into three parts: reviews, entire

web, “green light”. When you select reviews you only search the reviews. When you select the entire web, you search that. The green lights section are items that have been deemed appropriate for children. An uncommon feature of this service is its “concept extraction.” This is billed as a sort of controlled vocabulary/thesaurus that locates items on “senior citizens” when you search for “elderly people.” Again, Magellan does not describe how this works, so you have to take their word for it. Literally.



FEE BASED SERVICES

There are some information providers whose services are available through the Internet but require payment in order to be used: Dialog, Lexis/Nexis, JSTOR, EbscoHost, UMI ProQuest, SilverPlatter, Wilson Index, America Online, etc.

Fee Based Services

In the “old” days, there was the MEDLARS family of databases, DIALOG, and BRS. Since then the MEDLARS databases have gotten bigger, DIALOG has gotten more expensive, and BRS has gone out of business. There are also a growing number of newcomers on the block who provide similar information services. Generally speaking, all these services are more authoritative, offer superior searching mechanisms, and index more specific subject areas than their Internet counterparts. Of course this extra level of service does not come for free.

This section reviews a numbers of these services in terms of their strengths/weaknesses as well as their advantages/disadvantages.

Fee-based services

Title: America Online

Remote: <URL:<http://www.aol.com/>>

Cost: \$1-10

Ease of use: Mindless

Data types: bibliographic citations; full-text articles; ASCII text files; image files;

Search features: field searching; Boolean logic; phrase searching; truncation/stemming;

Description: Don't overlook America Online (AOL) as a potential source of your online information. This commercial service is geared for the masses and therefore fits best into a public library sort of collection. At the same time, the sorts of information available through AOL's services are quite impressive, especially considering the **Cost:** reference materials in all subject areas, (maga)zines of all types, microcomputer software that's easy to find, and business information galore. The services on AOL are well documented online as well as in books. The interface represents the simplest around. The search features vary from service to service within AOL, and while they are adequate, they do not offer very much for the “power user.”

Title: JSTOR

Remote: <URL:<http://www.jstor.org/>>

Cost: \$ Thousands

Ease of use: Easy

Data types: full-text articles;

Search features: Boolean logic; truncation/stemming; phrase searching; nested queries; proximity; field searching; date ranges; relevance ranking;

Description: This service is unique among all the other services reviewed here in that it provides full-text access to older materials instead of newer materials. Furthermore, since this service represents a non-profit organization, emphasis seems to be on cost-recovery as opposed to excessive profit. The search engine provides the usual features plus a bit of relevance ranking. The

titles in the database are, for the most part, for the liberal arts researcher. The images produced from the service's scans are of high quality and more than readable online. JSTOR is filling a much needed niche in the bibliographic database world.

Title: Northern Light Search

Remote: <URL:<http://www.nlsearch.com/>>

Cost: \$1-10

Ease of use: Easy

Data types: bibliographic citations; HTML files; full-text articles;

Search features: phrase searching; truncation/stemming;

Description: This a newcomer into the fray of Internet search engines. Like the resources described in the article "Searching the Hidden Internet" by Notess, this search engine not only provides access to a database of broad WWW documents, but articles from selected magazines and journals (Special Collections) as well. Presently the search features are weak. On the other hand, this engine analyzes the results of queries and creates groups of documents, clusters, much in the same way AltaVista searches are "refined." When you locate a document in the Special Collection you are interested in, you can pay as you go for the article or set up an account. Presently, articles range in price from \$0-10 per copy.

Title: Ovid

Remote: <URL:<http://preview.ovid.com/libpreview/start.cgi>>

Ease of use: Challenging

Data types: bibliographic citations; full-text articles;

Search features: Boolean logic; field searching; truncation/stemming; wild card; proximity;

Description: Using the BRS search software underneath, Ovid is providing bibliographic and full-text access to a number of databases. The search engine, like just about everybody else, is divided into basic and advanced modes. In either case you can specify BRS queries. Like other services, you can browse and search the authority indexes. Unlike a few of the services, free-text searches can be mapped to controlled vocabulary terms for higher precision/recall ratios. An extra nice part of this service is its ability to output its results into formats that are easily imported by personal bibliographic utilities. Ovid also offers the ability to search other database using the search strategies of previous queries.

Title: SilverPlatter

Remote: <URL:<http://www.silverplatter.com/>>

Data types: full-text articles; bibliographic citations;

Search features: Boolean logic; date ranges; field searching; nested queries; phrase searching; proximity; truncation/stemming; wild card;

Description: SilverPlatter offers sets of bibliographic databases for leasing. These databases can be accessed locally on your computer or remotely on theirs. These databases can also be accessed using any one or a number of client applications for Windows, Macintosh, Unix, or Web browsers. It supports the usual

compliment of search features and only lacks the relevance ranking features of the more popular Internet search engines. On the other hand, the ability to browse authority indexes makes for higher precision/recall ratios.

Author: Ebsco

Title: EbscoHost

Remote: <URL:<http://www.epnet.com/host.html>>

Data types: bibliographic citations; full-text articles;

Search features: Boolean logic; field searching; nested queries; date ranges; truncation/stemming; wild card;

Description: Ebsco seems to branching out from it traditional jobber services into this bibliographic and full-text index. After all, they probably had a lot of this data to begin with. The service offers the usual suspects when it comes to searching. It is nice to be able to browse the authority indexes. Ebsco has put together a collection of library-related titles for browsing and searching in order to demonstrate its product. Try also The Library Reference Center at <URL:<http://www.epnet.com/lrc.shtml>>.

Author: H. W. Wilson

Title: WilsonWeb

Remote: <URL:<http://www.hwwilson.com/tecfaq7.html>>

Ease of use: Easy

Data types: bibliographic citations; full-text articles;

Search features: Boolean logic; date ranges; field searching; nested queries; phrase searching; proximity; truncation/stemming; wild card;

Description: WilsonWeb is the collection of H.W. Wilson indexes and abstracts made available via SilverPlatter ERL database search engine. Consequently, it has all the same search features as the SilverPlatter line of products.

Author: IAC

Title: InfoTrac SearchBank

Remote: <URL:<http://www.informationaccess.com/>>

Ease of use: Difficult

Data types: bibliographic citations; full-text articles;

Search features: Boolean logic; wild card; field searching;

Description: Despite the fact that this is the publisher of some very popular bibliographic indexes, this service is a bit difficult to use. The query language is obtuse. The advanced search is not that advanced. Queries returned unexplained results, and there aren't very many options for creating queries. Weak. Especially compared to other choices.

Author: Knight-Ridder

Title: DIALOG@Carl

Remote: <URL:<http://uncweb.carl.org/>>

Ease of use: Easy

Data types: bibliographic citations; full-text articles; news;

Search features: Boolean logic; date ranges; field searching; proximity; truncation/stemming; wild card;

Description: DIALOG@Carl provides “kinder, gentler” interface to a subset of the DIALOG family of databases. Using the simple search feature you enter a controlled vocabulary term, keyword, or phrase and the engine searches for the query terms in its default indexes. That’s about it for the simple search. The advanced search allows you to more exactly specify what fields to search as well as supplying date ranges, output numbers, and whether or not you want to limit your search to full-text items. A very nice feature of this service is its ability to search multiple databases simultaneously. To make this even easier, DIALOG@Carl has divided its databases into subject area for searching. Another nice feature “automagically” allows you to search other databases using the query from the previous search.

Author: Knight-Ridder

Title: DialogWeb

Remote: <URL:<http://dialog.krinfo.com/>>

Cost: \$ Hundreds

Ease of use: Difficult

Data types: bibliographic citations; full-text articles; news;

Search features: Boolean logic; date ranges; field searching; nested queries; proximity; truncation/stemming; wild card;

Description: This URL points to a Web-based interface to the DIALOG family of databases. The service includes all the features of command-line searching of any DIALOG database plus a few other features. Furthermore, the online help is some of the best around. Somebody did a really good job of porting the command-line interface to a HTML environment. This access method may even be lesser expensive than traditional DIALOG searching since every command is immediately followed by a “search hold” command. With this interface you have access to more than 450 databases of all types. But this power and flexibility comes at a cost, a rather steep cost. Sign-up fees are around \$250 per account and there is an almost \$150 annual charge just to keep the password. Database rates are around \$45/hour + a couple dollars for each full-text record downloaded. At the same time, the wealth of information found in DialogWeb may be what a professional information seeker really needs. Let’s hope they know the command syntax!

Author: OCLC

Title: FirstSearch

Remote: <URL:http://gilligan.prod.oclc.org:3050/html/fs_homepage.htm>

Ease of use: Easy

Data types: bibliographic citations; full-text articles; directories;

Search features: Boolean logic; field searching; phrase searching; proximity; truncation/stemming;

Description: This service provides access to an unusual array of bibliographic, full-text, and directory indexes. Like everybody else, it supports a simple search as well as a power search. The power search, like everybody else, explodes out various fields making it possible to create more elaborate queries. Like the other services, FirstSearch has divided its collection of databases into subject areas hopefully making it easier for patrons to locate the best database for their particular need. Unlike some of the other services, you can not search more than one database at a time, nor can you save searches between database selections.

Author: UMI

Title: ProQuest Direct

Remote: <URL:<http://www.umi.com/proquest/>>

Ease of use: Easy

Data types: bibliographic citations; full-text articles;

Search features: Boolean logic; proximity; phrase searching; wild card; field searching; ranges; nested queries; “concept extraction”; spelling options;

Description: This service provides the “full meal deal” with one search field interface; there is not power search feature with multiple input fields. To specify field searching you enter the field name you want to search and then the term inside parentheses. It is mnemonic. The Search Assistant is a nice feature where it queries you for term and qualifiers and then builds a search strategy for you.



COMMON SEARCHING PROTOCOLS

There are many other search engines indexing and providing access to locally developed Internet information like WAIS, Glimpse/Harvest, LISTSERV and ListProc, Z39.50, Archie, Finger, etc.

Other searching services

There is more to Internet searching than HTML files. For example, “hosts” of information exist in the archives of USENET newsgroup postings and mailing list archives. There exist more than a few Internet services offering the ability to search newsgroup posting. There are fewer services allowing you to search mailing list archives. Real “news” like the sorts created by Associated Press and the major broadcasting networks is also available from quite a number of services.

Many search services found on the Internet employ “regular expressions.” Regular expressions are a way to describing the shape of words as opposed to their meaning. Harvest (Glimpse) services as well as ListProc mailing lists rely on regular expressions for searching syntax. Regular expressions are also used the infamous grep program of Unix computers. Used effectively, grep can locate information lost on just about any file on your file system. The time spent learning regular expression syntax is not time lost.

You might also consider creating your own search engine. There exist quite a number of these tools. Some of them are as free as a free kitten. Other cost thousands and thousands of dollars. WAIS was one of the first to popularize the idea. It is still a strong tool. Harvest allows you to index the content of remote sites.

OpenText and Verity’s solutions are expensive, but offer features the other tools don’t like technical support. ROADS is a nice piece of software designed for the collecting and indexing of Internet resources. If you are tired of maintaining your collections by hand, consider ROADS.

Other services

Title: Deja News

Remote: <URL:<http://www.dejanews.com/>>

Cost: 0

Ease of use: Easy

Data types: USENET newsgroup posting;

Description: Billed for business persons, this service collects and archives USENET posting and makes them available through a simple searching and browsing interface. Qualifying results is done through subsequent forms making the process longer, but possibly more efficient for the uninitiated. One especially nice feature is it ability to limit searches to specific parts of the USENET hierarchy. This is done by first browsing the subject classifications and then searching. The service could be improved if the searching mechanisms were less form driven.

Title: Excite

Remote: <URL:http://www.excite.com/>

Cost: 0

Ease of use: Easy

Data types: HTML files; news; USENET newsgroup posting;

Search features: phrase searching; Boolean logic; relevance ranking; “concept extraction”;

Description: Excite’s search engine is much like everybody else’s. Its power search form simply spells out the search features on a more descriptive manner. The features include phrase searching, Boolean logic, and relevance ranking. Excite uses the same “concept extraction” technique of locating documents a Magellan because Magellan is owned by Excite. The site also hosts pointers to other search engines as well as collections of Internet resources on popular subjects.

Title: Open Text

Remote: <URL:http://index.opentext.com/>

Cost: 0

Ease of use: Easy

Data types: HTML files; USENET newsgroup posting; email addresses; current events;

Search features: field searching; Boolean logic; proximity;

Description: It’s just a guess, but I believe OpenText’s search engine is really an advertisement for its commercial document storage and retrieval software. Like a number of the other services, OpenText’s interface is entirely menu-driven. It consists

of a number of blank fields surrounded by qualifiers. The qualifiers are of two types: fields and operators. The fields include: summary, title, first heading, URL, anywhere. The operators include: and, or, but not, near, followed by. It is obvious from this description that OpenText is strong on phrase searching and the proximity of words. Field searching is a bit weak.

Title: Regular expressions

Local: <URL:http://www.lib.ncsu.edu/staff/morgan/pointers/texts/regular-expressions.txt>

Data types: ASCII text files;

Search features: regular expressions;

Description: Regular expressions represent the Unix way of locating text within documents and they frequently raise their ugly head in many search interfaces. In a nutshell, regular expressions locate “patterns” of text within documents, not concepts; regular expressions help you locate data and information based on the shapes of words, not their meaning or relationships. The document cited above describes regular expressions in terms of the Unix grep command. An understanding of regular expressions will come in handy when using many alternative search tools like Harvest and ListProc mailing list archives.

Title: ROADS

Remote: <URL:<http://www.ukoln.ac.uk/roads/>>

Cost: \$0

Ease of use: Challenging

Search features: Boolean logic; field searching; nested queries; phrase searching; relevance ranking; truncation/stemming;

Description: ROADS is a suite of software whose purpose is to provide a means for collecting, organizing, and searching Internet resources. Requiring a Unix computer, the software is written completely in Perl and therefore completely open. Once installed, it provide the means for data entry into a database via HTML forms or simple ASCII text files. These forms/text files are flexible, structured records. Each record can have a unlimited number of fields and field lengths. Consequently it is easy to assign multiple subject heading to Internet resources. ROADS also provide the means for automatically creating HTML files for browsing as well as an interface for searching your collection. It doesn't stop there. The software also includes a built-in link checker so you can keep your data fresh. Lastly, it is possible to not only search your local data via the ROADS software, but you can search other ROADS collections via the same interface. Thus, it would be possible to create a single-user interface to global collections of Internet resources. This is the best free database software I've seen for collecting and maintaining Internet resources.

Author: Internet Research Task Force Research Group on Resource Discovery (IRTF-RD)

Title: Harvest

Remote: <URL:<http://www.tardis.ed.ac.uk/harvest/>>

Cost: 0

Ease of use: Easy

Data types: ASCII text files; full-text articles; HTML files; USENET newsgroup posting; image files;

Search features: Boolean logic; field searching; nested queries; phrase searching; regular expressions; truncation/stemming; spelling options;

Description: Harvest combines an Internet spider application with a database search engine (WAIS or Glimpse). Using this tool you can "gather" data from remote sites and have it indexed in a well-documented format. This gathered data can then be "brokered" to end users for searching. Its greatest strength is its ability is to combine multiple gatherers from local and remote sites and broker them from one location reducing network traffic. Its greatest limitation is its output features; the Harvest output is, generally speaking, difficult to display in a pleasing way interpret.

Author: Kotsikonas, Anastasios

Title: ListProc

Remote: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/listproc-searching.txt>>

Cost: \$0

Ease of use: Easy

Data types: mailing lists;

Search features: phrase searching; regular expressions;

Description: ListProc was originally designed to be a mailing list server completely duplicating the functionality of LISTSERV except ListProc was to be run on Unix computers. Like LISTSERV mailing lists, ListProc mailing lists can be archived and searched if the mailing list administrator has turned on these features. Once turned on you can search the archives using regular expression syntax and “get” documents from the resulting search strategies. ListProc comes with an client that can be used to search the archives interactively instead of via email, but few administrators turn this feature on. Mailing lists are a good source of timely information and are good ways of identifying self-pronounced subject experts.

Author: L-Soft International, Inc.

Title: LISTSERV

Remote: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/listserv-searching.txt>>

Cost: \$0

Ease of use: Challenging

Data types: mailing lists;

Search features: Boolean logic; date ranges; field searching; nested queries; phrase searching; truncation/stemming;

Description: People are the real sources of information, and as you may or may not know, the archives of many mailing lists are searchable. LISTSERV was the first real mailing list software to be

put into wide spread use. Administrators of LISTSERV mailing lists may or may not turn on the archiving of mailing list distributions. Additionally, they may or may not turn on searching mechanisms for these archives. If these features are turned on, then you can search the archives of the mailing lists using either a WWW front-end supplied by the administrator, or you can search the archives via email. The reference URL above points to a simple text file describing how to search LISTSERV mailing lists. For more information, see <URL:<http://www.l-soft.com/>>, the home page of the now-commerical version of LISTSERV.

Author: Pfeifer, Ulrich

Title: FreeWAIS-sf

Remote: <URL:http://ls6-www.informatik.uni-dortmund.de/ir/projects/freeWAIS-sf/fwsf_toc.html>

Cost: 0

Ease of use: Challenging

Data types: ASCII text files; HTML files; image files; image files;

Search features: Boolean logic; field searching; nested queries; relevance ranking; truncation/stemming;

Description: This application indexes text files and the file names of non-text files. It is a powerful application brought to maturity after spending much of its life in various software shops. If you wanted to index some of you data locally, and you had a Unix computer at your disposal, then this application would be one to consider doing the job.

Author: Verity

Title: Search '97

Remote: <URL:http://www.verity.com/>

Cost: \$ Thousands

Ease of use: Challenging

Data types: ASCII text files; HTML files;

Search features: Boolean logic; field searching; nested queries; phrase searching; relevance ranking; truncation/stemming;

Description: If you want to provide your own searching interface to some of your local data, then you might consider purchasing Search '97. This software indexes local ASCII files (and therefore HTML source) as well as PDF and WYSIWYG data. If your source data is well structured, then Search '97 provides the means for powerful field searching. The software also includes an Internet spider/robot. This application allows you to "feed" the spider a set of URLs which it will then go out and index. Consequently, using Search '97 you could build your own Internet search service. More realistically, you could use this software to index the top N levels of the Internet resources you point to or you could index things like electronic serials. Search '97 is very much an "information science" sort of tool since much of its output is based on the statistical relevance of query terms. The software also uses these statistical methods to allow you to "find more like this one." Search '97 is a powerful tool, but not necessarily for the faint of heart.



BROWSABLE LISTS

*Collections of Internet resources
supplement and compliment Internet
indexes: Yahoo, A2Z, WWW Virtual
Library, etc.*

Browsable Lists

Subject-based, hierarchical collections of Internet resources have a number of advantages over searchable indexes.

First of all, the end-user does not have to mentally articulate any sort of information need; they do not have to know what they are looking for in order to find something. Second, the end-user does not have to know any sort of searching syntax. All they have to do is select items from the screen. Third, browsable collections offer the ability to see the entire collection at a glance and it is very easy to find “other items like this one.” Finally, browsable lists are easy to construct since they require little technology to create. Anybody can do it, and almost everybody does.

But there are a number of disadvantages as well. Browsable lists assume an underlying classification scheme. These classification schemes represent conceptual models of information/knowledge that may or may not fit the end-user’s conceptual model. Consequently, browsable lists may force end-user’s into a framework they don’t understand. Furthermore, these conceptual models must be flexible enough to accommodate the addition of new ideas but at the same time the models must be formalized enough to be consistent. The models must be dynamic and thus any model time limited.

The items below describe and list collections of browsable Internet resources. Of particular interest are the items by Duda, Notess, INFOMINE, the Argus Clearinghouse, and the Librarians’ Index to the Internet since these items speak more to the academic subject listings.

About browsable lists

Author: Duda, Andrea L. editor

Title: Untangling the Web

Remote: <URL:<http://www.library.ucsb.edu/untangle/>>

Description: This is a “Proceedings of the Conference Sponsored by the Librarians Association of the University of California, Santa Barbara and Friends of the UCSB Library” from April 26, 1996 University Center, University of California, Santa Barbara. The linked abstracts and papers describe all aspects of World Wide Web development in a library setting. Of particular interest are “Yahoo! Cataloging the Web” by Anne Callery, “Spinning a Web Search” by Mark Lager, and “Spiders and Worms and Crawlers, Oh My: Searching on the World Wide Web” by Ann Eagan and Laura Bender.

Author: Munson, Kurt I.

Title: World Wide Web indexes and hierarchical lists: finding tools for the Internet.

Citation: Computers in Libraries ISSN:1041-7915 June 1996, v16, n6, p54(4)

Description: Indexes and hierarchical lists are two types of search tools for locating information on the World Wide Web and other Internet resources. Indexes, such as Lycos and Open Text, provide access to records by matching search terms against descriptive cataloging. Hierarchical lists, such as Yahoo!, use descriptive and subject cataloging to group common resources by location.

Author: Notess, Greg

Title: Search Engines Showdown

Remote: <URL:http://imt.net/~notess/search/>

Description: "This site summarizes, reviews, and compares the search features and database scope of the Internet search engines and finding aids." It does this by dividing its content to the following nine parts: search engines, directories, multi-search, USENET & others, strategies, statistics, reviews, definitions, and bibliography. The bibliography alone is worth the price of admission.

Browsable list

Title: C&RL NewsNet Internet Resources

Remote: <URL:http://www.ala.org/acrl/resrces.html>

Description: This is a list of Internet subject guides regularly published in College & Research Library News since June of '96. The Internet resources from these columns, chosen by subject experts, could easily form the basis of any college or university's collection of Internet links.

Title: Excite

Remote: <URL:http://www.excite.com/>

Cost: 0

Ease of use: Easy

Data types: HTML files; news; USENET newsgroup posting;

Search features: phrase searching; Boolean logic; relevance ranking; "concept extraction";

Description: Excite's search engine is much like everybody else's. Its power search form simply spells out the search features on a more descriptive manner. The features include phrase searching, Boolean logic, and relevance ranking. Excite uses the same "concept extraction" technique of locating documents a Magellan because Magellan is owned by Excite. The site also hosts pointers to other search engines as well as collections of Internet resources on popular subjects.

Title: Hotbot

Remote: <URL:<http://www.hotbot.com/>>

Cost: 0

Ease of use: Easy

Data types: HTML files; image files; sounds; USENET newsgroup posting;

Search features: date ranges; field searching; phrase searching; Boolean logic; domain;

Description: HotBot uses an almost purely forms-driven interface to its search engine. A unique feature of HotBot is its ability to select a wide range to Internet media types to seek. For example, it includes: Image, Java, Acrobat, audio, JavaScript, VB Script, Video, ActiveX, Shockwave, VRML. Some these media types can be located in other services through the command line, but not all. HotBot also provides access to a number of specialized databases on the subjects of USENET, top news sites, classified ads, domain names, stocks, discussion groups, shareware, businesses, people, email addresses. HotBot's browsable lists of Internet resources merely points to Wired's Cybrarian <URL:<http://www.wired.com/cybrarian/frame/tech/web.html>>. While HotBot has a lot of features, feel a bit out of control when I am limited to only a forms interface.

Title: INFOMINE

Remote: <URL:<http://lib-www.ucr.edu/>>

Cost: 0

Description: "INFOMINE is intended for the introduction and

use of Internet/Web resources of relevance to faculty, students, and research staff at the university level. It is being offered as a comprehensive showcase, virtual library and reference tool containing highly useful Internet/Web resources including databases, electronic journals, electronic books, bulletin boards, LISTSERVs, online library card catalogs, articles and directories of researchers, among many other types of information."

Title: InfoSeek

Remote: <URL:<http://www.infoseek.com/>>

Cost: 0

Ease of use: Easy

Data types: HTML files; news; USENET newsgroup posting;

Search features: phrase searching; Boolean logic; relevance ranking; domain; field searching;

Description: This site does not have "power" search feature, nor does it have complicated forms interface for completing queries. All queries are from a simple command line. The command line supports the usual suspects: Boolean operations, phrase searching, and a few field searches. Its collection of Internet resources looks just like everybody else's too. The only difference is they publish a collection policy; "InfoSeek Select sites are chosen by our editors from the index based on their editorial value, traffic and the number of links to the sites." Like Yahoo, once you are browsing a particular subject area you can use the searching mechanisms to find resources from that area.

Title: Librarians' Index to the Internet

Remote: <URL:<http://sunsite.berkeley.edu/InternetIndex/>>

Cost: 0

Description: "The Librarians' Index to the Internet is a searchable, annotated, subject directory of close to 3,000 Internet resources chosen for their usefulness to the public library user's information needs." Originally developed by a librarian, Carole Leita, this site has grown from a gopher bookmark file into what it is today. If you're nice to her, then maybe she will give you her Perl scripts and you can start your own collection.

Title: Lycos

Remote: <URL:<http://www.lycos.com/>>

Cost: 0

Ease of use: Easy

Data types: HTML files; sounds; image files; bibliographic citations;

Search features: phrase searching; Boolean logic; natural language;

Description: Starting out in academia (Carnegie Mellon University) and then going commercial, Lycos is the grand daddy of the Internet search engines. It is also the service that really built a name for itself by rating sites as the "Top 5%" and awarding sites the distinction of displaying the Top %5 badge. The service's search features rely greatly on variations of phrase searching, but the service also has a reputation of locating more hits than the other services. As evidenced by a visit to its home page, Lycos is

becoming more and more like commercial service offering more than simple Internet searching and advertisements. Lycos also offers search links to Barnes and Noble as well as UPS. Obviously, this service is thinking the bigger picture and is quickly leaving academia behind.

Title: Yahoo

Remote: <URL:<http://www.yahoo.com>>

Cost: 0

Description: This is the most recognized of the browsable collections of resources. Surely its collection seems to be one of the largest catering to popular culture as well as many academic needs. What more needs to be said?

Author: A2Z

Title: Sites from A2Z

Remote: <URL:<http://a2z.lycos.com/>>

Cost: 0

Description: This site, in my opinion, is well-designed if you have a particularly fast network connection. It is one of the older collections based on one of the grand daddies of Internet search engines, Lycos. Like is other commercial brethren, this service divides it resources among popular culture as well as more academic needs. What every happened to Dewey?

Author: Aldea Communications

Title: InterNIC Academic Guide to the Internet

Remote: <URL:<http://www.internic.net/aldea/attframes2.html>>

Cost: 0

Ease of use: Easy

Data types: HTML files;

Description: “The InterNIC Academic Guide to the Internet is the Internet guide created especially for the higher education community. Our mission is to develop the primary Internet resource of and for the research and academic community.”

This is an organized collection of Internet resources chosen for the academic community. The collection surrounds the sciences rather than the humanities. Each resource is accompanied with text and icons describing the content. The service is indexed with Harvest and provides opportunities for user commentary.

Author: Argus Associates

Title: Argus Clearinghouse

Remote: <URL:<http://www.clearinghouse.net/>>

Cost: 0

Ease of use: Easy

Data types: HTML files;

Description: The Argus Clearinghouse started out in a classroom of the Library and Information Studies school of the University of Michigan. It was here that Lou Rosenfeld assigned students the task of organizing Internet resources into study aids. The

collections he and his students created became quite popular for their thoroughness and focus.

Since then the Argus Clearinghouse has become a commercial enterprise specializing in authoritative resource guides. The guides are intended to be academic in nature and (for the most part) up-to-date. If you have a specific topic of interest and that topic is listed at the Clearinghouse, then you will have a head start on locating the information you seek.

Author: Duda, Andrea L. editor

Title: Untangling the Web

Remote: <URL:<http://www.library.ucsb.edu/untangle/>>

Description: This is a “Proceedings of the Conference Sponsored by the Librarians Association of the University of California, Santa Barbara and Friends of the UCSB Library” from April 26, 1996 University Center, University of California, Santa Barbara. The linked abstracts and papers describe all aspects of World Wide Web development in a library setting. Of particular interest are “Yahoo! Cataloging the Web” by Anne Callery, “Spinning a Web Search” by Mark Lager, and “Spiders and Worms and Crawlers, Oh My: Searching on the World Wide Web” by Ann Eagan and Laura Bender.

Author: Louisiana State University Libraries

Title: Webliography: A Guide to Internet Resources

Remote: <URL:<http://www.lib.lsu.edu/weblio.html>>

Cost: 0

Description: This is a list of Internet resources organized by academic subjects: business, engineering, government, humanities, science, social science, reference. The business, government, and humanities listings have the Argus Clearinghouse “seal of approval.” What is exciting about this site is that it demonstrates that you do not have to have the largest of physical collections or budgets to create significant collections of Internet resources.

Author: McKinley Group, Inc.

Title: Magellan

Remote: <URL:<http://www.mckinley.com/>>

Cost: 0

Ease of use: Easy

Data types: HTML files;

Search features: Boolean logic; nested queries; phrase searching; relevance ranking; “concept extraction”;

Description: The strength of this service lies in its Internet reviews. Using a 4 point scale, Magellan reviews Internet sites and ranks them in subject lists accordingly. Unfortunately, we don’t know what their “collection management” policy is, so we can not determine how or why they have picked particular items for the collection nor what they look for in particular resources. The

search engine itself is divided into three parts: reviews, entire web, “green light”. When you select reviews you only search the reviews. When you select the entire web, you search that. The green lights section are items that have been deemed appropriate for children. An uncommon feature of this service is its “concept extraction.” This is billed as a sort of controlled vocabulary/thesaurus that locates items on “senior citizens” when you search for “elderly people.” Again, Magellan does not describe how this works, so you have to take their word for it. Literally.

Author: Munson, Kurt I.

Title: World Wide Web indexes and hierarchical lists: finding tools for the Internet.

Citation: Computers in Libraries ISSN:1041-7915 June 1996, v16, n6, p54(4)

Description: Indexes and hierarchical lists are two types of search tools for locating information on the World Wide Web and other Internet resources. Indexes, such as Lycos and Open Text, provide access to records by matching search terms against descriptive cataloging. Hierarchical lists, such as Yahoo!, use descriptive and subject cataloging to group common resources by location.

Author: Notess, Greg

Title: Search Engines Showdown

Remote: <URL:<http://imt.net/~notess/search/>>

Description: "This site summarizes, reviews, and compares the search features and database scope of the Internet search engines and finding aids." It does this by dividing its content to the following nine parts: search engines, directories, multi-search, USENET & others, strategies, statistics, reviews, definitions, and bibliography. The bibliography alone is worth the price of admission.

Author: Secret, Arthur

Title: WWW Virtual Library

Remote: <URL:<http://www.fisk.edu/vl/>>

Cost: 0

Description: Here you will find a simple, flat, alphabetical list of 10's, if not 100's of subjects. For the most part, the subjects are academic in nature and represent collections of Internet resources. Each collection is loosely maintained by volunteers who are subject experts but not necessarily information specialists. This is not a bad place to consider looking for subject guides when beginning an Internet search.



INFORMATION EVALUATION AND KNOWLEDGE MANAGEMENT

Methods describing how-to effectively keep your Internet information under control are outlined here.

Knowledge Management and Information Evaluation

The continual development of the United States' (and the world's) service-based economies has increased the total volume of available information. Instead of manufacturing and selling widgets, we are developing and selling skills. These skills inherently require the ability to observe situations and perform actions. The process of observing situations and performing actions requires more than simple data and information. It requires knowledge, but knowledge is recorded as data and information. Consequently, business is increasingly concerned with its accumulation of knowledge, or as it is frequently called, its "intellectual capital."

This realization has given rise to the popular idea of "knowledge management." There are a small, but growing number of knowledge managers, knowledge workers, knowledge systems, and Chief Knowledge Officers (CKO) within business today. The purpose of these groups of people and things is to make the most of the knowledge any one company has and leverage it better than its competitors. It is believed that the business who better leverages its accumulated knowledge will be the business who prospers.

Obviously business is feeling the need to garner better control over its intellectual capital.

Concurrently, the Internet and networked information in general as also increased the total volume of available information. Now anybody can be a "publisher" and many people do. Yes, much of the information is not necessarily what you may want or need, but one person's sour milk is another person's cheese. These factors have contributed to greater feelings of information overload. Information overload is nothing new, but in the increasingly specialized environment in which we live information overload makes itself more readily apparent.

Effective use of your computer can help you cope with information overload. Email is a good thing to work on first. One principle that can be diligently applied is the ABC Rule. Classify your email into three categories. The A category you will act upon and delete it when you are done. The B category you will file and possibly index (using something like WAIS). The C category you will delete right away. Your mailbox is not a to-do list. Don't keep things there as reminders of projects. Add those reminders to your calendar (be it electronic or not) and delete the original.

When subscribing to mailing lists, subscribe in digest mode. This will reduce your volume of email considerably. When you get the digested mailing, scan the summaries for interesting content. If you see a topic that is discussed again and again, then consider

reading it even though it may not particularly interest you because it apparently interests many other people. When you are done with it, delete it. Remember, you can almost always retrieve again.

Finally, concerning email, try using your email program's filtering features. This will separate your personal/work-related mail from the more general email noise.

When reading journal articles in printed or electronic form, read the first two paragraphs, the last two paragraphs, and the footnotes (or URLs) first. If the content of these sections interest you, then read the balance of the text. This will work most of the time since these sections are good places to find summaries of the document's content.

When browsing online documents or sets of documents (like browsable lists), use judiciously your browser's back button and not the navigation links found within the document. This way you will leave a "trail of bread crumbs" and make it easier for you to go back from whence you came. Additionally, it makes it easier for you to traverse all possible paths. Furthermore, using the back button will reinforce the content you browsed since you will be seeing it at least twice.

Consider "surfing the 'Net" without loading images. This not only downloads pages faster but it also eliminates much of the

superfluous data bombarding your senses. Good HTML design does not require lots of images.

Use your browser's bookmark features. Take the time to organize your bookmarks, but also regularly take the time to weed its collection.

Some businesses are trying to deal with these same sort of problems with knowledge management. The sites listed below are collections of pointers to knowledge management definitions, texts, and technologies. Read them and compare what they say to the traditional purposes of librarianship.

Knowledge management

Author: Ash, Jerry

Title: Knowledge Management Links

Remote: <URL:<http://aviary.share.net/delphi/ash/ckolinks.html>>

Description: Short essays surrounding the topic of knowledge management and how it effects business are collected here.

Author: Bellinger, Gene

Title: Knowledge Management

Remote: <URL:<http://www.radix.net/~crbnblu/kmgmt.htm>>

Description: After an introductory essay, this page includes pointers to references, introductory and advanced articles, books, organizations, and products all dealing with knowledge management.

Author: Inference Corporation

Title: Knowledge Management Resource Guide

Remote: <URL:<http://www.inference.com/welcome/knowledge.html>>

Description: Created by a company providing consulting services, this is a list of knowledge management resources created for the perspective of business managers.

Author: Public Service Commission of Canada, Strategic Planning, Analysis and Research Branch

Title: Corporate Intelligence and Knowledge

Remote: <URL:<http://www.psc-cfp.gc.ca/library/knowledge.htm>>

Description: Here you will find a collection of books, articles, and links to information on corporate intelligence and knowledge. The items are many times in French since this list is generated in Canada and they are officially bilingual.

Author: Special Library Association

Title: Selected References on Knowledge Management

Remote: <URL:<http://www.sla.org/membership/irc/knowledg.html>>

Description: Lists of articles, websites, and software dealing with the issue of knowledge management, this site seems to be maintained. "If its created by librarians, then it's got to be good."



LADDER OF UNDERSTANDING

The Internet effects the principles of librarianship, or does it?

Ladder of Understanding

An alternative title for this section could be “The Internet effects the principles of librarianship, or does it?”

Often times librarianship is described as the process of collecting, organizing, archiving, and disseminating data and information. The advent of the Internet with its ability to share vast amounts of data and information have made much of traditional librarianship seem mundane and outmoded. Yet there remains a core of the profession that has not changed. A core element that is rarely articulated. That element is evaluation and facilitation of knowledge.

Librarians do not collect, organize, archive, and disseminate information just for the sake of it. These processes are a means to an end, not the end themselves. Why collect things if no one is going to use them. Why organize them unless some one was suppose to browse them?

Why archive anything if there is not the desire to retrieve it later? Data and information are the manifestations of the profession. The creation of knowledge, the spread of wisdom, and the

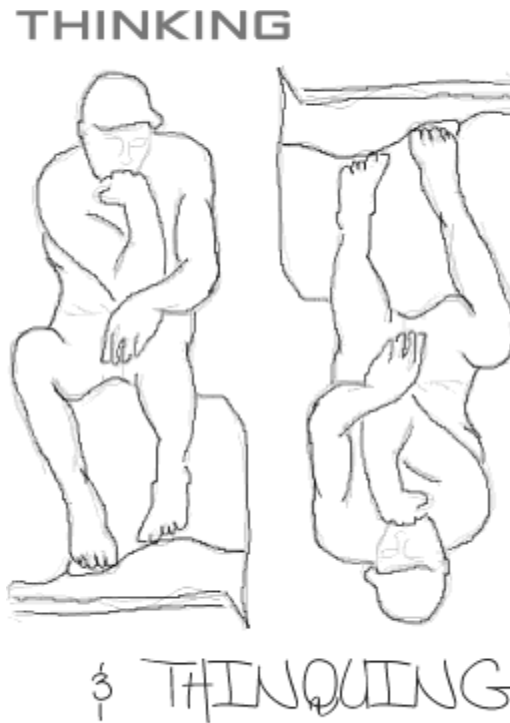
advancement of understanding are closer to its ultimate goals.

The Internet with all its advantages and disadvantages has had profound effects on librarianship. Yet if the profession reflects

upon itself, then the profession will understand the Internet and its accompanying technology is simply another tool to fulfill those goals.

The process of amalgamating traditional library skills (and ethics) with Internet technology requires a certain type of thinking as well as something else I have coined as “thinqing.” In this setting, “thinking” is an intellectual process characterized by methodical, systematic, left-brain activities. In many ways (but not all) this sort of activity is characterized through things like mathematics and computer programing. The other half of the process, “thinqing,” is intuitive, creative, and unsystematic. Many people characterize artistic endeavors in this manner.

Both of these intellectual processes, thinking and thinqing, are necessary for the libraries of today (and even yesterday) to manage technology effectively. Thinking must be used to analyze



the needs of our clientele. It must be applied when drawing up a budget. Thinking is a necessary activity when learning how to use the newest piece of software. Similarly, thinquing must be a part of the process when evaluating how to use computer technologies for library services. Thinquing must be taken into account when asked a new reference question and the answer is not readily apparent. Thinquing is the process you use when you encounter a new problem and must come up with some sort of solution. The problem with the profession today is it tends to ignore obvious problems and consequently it rarely employs the practices of thinquing.

Put another way, it does not only behoove libraries to continually be aware of new computer technologies (thinking), but they must also be able to discover possibilities for improving services with these technologies (thinquing). Then, and only then, will librarians be effectively using the Internet. The entire process requires an fundamental understanding of library principles and, at the same, it requires individual librarians to thinque “outside the box” for the purposes of enhancing methods of applying our fundamental principles.

In today's world of networked information, more and more information seeking activities can be accomplished without the

need of a librarian. Frequently our clientele can do real, significant information seeking without ever stepping into a library. Many of our profession (as well as lay people) see this sort of environment as a prelude to the demise of libraries. While the future of libraries will not be constant with their past, I do not see libraries fading away. Rather, I see the current environment

fostering a means for evolution and an enhancement of library services. Like a caterpillar, libraries can use the current environment to foster growth, turn upon itself for the purposes of reorganization, and emerge as a beauty unto itself and for others.

In conclusion, as more and more people gain access to more and more information, these same people will have to come to terms with methods for evaluating and using this information. This process, the process of evaluating and using information is, in my opinion, the future of librarianship. This process moves the library from one of dispensing information to fostering knowledge and understanding. It has been said

that understanding is like a four-rung ladder. The first rung on the ladder represents data and facts. As the data and facts are collected and organized they become information, the second rung on the ladder. The third rung is knowledge where knowledge is information internalized and put to use. The last



rung is wisdom, knowledge of a timeless nature. Technology has enabled more people to climb between the first and second rungs of the ladder with greater ease. Similarly, technology may enable libraries and librarians to climb higher on the ladder as well and provide knowledge services instead of simply information services.

Appendix A: Simple webliography

This is a simple list of the URLs and citations found in the text of this document. The references are listed first by subject and then alphabetically by author and title.

About browsable lists

Duda, Andrea L. editor. Untangling the Web.

Remote: <URL:<http://www.library.ucsb.edu/untangle/>>

Munson, Kurt I. World Wide Web indexes and hierarchical lists: finding tools for the Internet. Computers in Libraries ISSN:1041-7915 June 1996, v16, n6, p54(4).

Notess, Greg. Search Engines Showdown.

Remote: <URL:<http://imt.net/~notess/search/>>

About Internet searching

Surfing and Searching.

Remote: <URL:http://www.pathfinder.com/@@NYqtIQQAlBOFQHHZ/fortune/buyersguide_test/jobcsurf.html>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/jobcsurf.html>>

Bates, Mary Ellen. Seven Deadly Sins of Online Searching.

Remote: <URL:<http://www.onlineinc.com/oluser/NovOU/bates11.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/bates11.html>>

Brandt, D. Scott. Relevancy and searching the Internet. Computers in Libraries ISSN:1041-7915 Sept 1996, v16, n8, p35(3).

Remote: <URL:http://thorplus.lib.purdue.edu/~techman/CIL996.htm>

Local: <URL:http://www.lib.ncsu.edu/staff/morgan/pointers/texts/CIL996.html>

Brooks, Monica. Research on the Internet.

Remote: <URL:http://www.marshall.edu/library/facwww.htm>

Campbell, Karen. Understanding and Comparing Search Engines.

Remote: <URL:http://www.hamline.edu/library/bush/handouts/comparisons.html>

Local: <URL:http://www.lib.ncsu.edu/staff/morgan/pointers/texts/campbell.html>

Clyman, John. Finding your needle in the Web's haystack. PC Magazine ISSN:0888-8507 July 1996, v15, n13, p39(3).

Remote: <URL:http://www.zdnet.com/pcmag/issues/1513/pcmg0080.htm>

Local: <URL:http://www.lib.ncsu.edu/staff/morgan/pointers/texts/pcmg0080.htm>

Community Networking. Research Works.

Remote: <URL:http://cn.net.au/>

Duda, Andrea L. editor. Untangling the Web.

Remote: <URL:http://www.library.ucsb.edu/untangle/>

Haskin, David . Right Search Engine.

Remote: <URL:http://www.internetworld.com/1997/09/report.html>

Local: <URL:http://www.lib.ncsu.edu/staff/morgan/pointers/texts/report.html>

Hearst, Marti A. Interfaces for searching the Web. Scientific American ISSN:0036-8733 March 1997, v276, n3, p68(5).

Remote: <URL:http://www.sciam.com/0397issue/0397hearst.html>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/0397hearst.html>>

PDF: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/hearst.pdf>>

Internet Scout Project. Searching the Internet.

Remote: <URL:<http://scout.cs.wisc.edu/scout/toolkit/searching/>>

Lynch, Clifford . Searching the Internet. Scientific American ISSN:0036-8733 March 1997, v276, n3, p52(5).

Remote: <URL:<http://www.sciam.com/0397issue/0397lynch.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/0397lynch.html>>

Mauldin, Michael L. Searching the World Wide Web; Lycos: design choices in an Internet search service. IEEE Expert ISSN:0885-9000 Jan-Feb 1997, v12, n1, p8(7).

Remote: <URL:[http://dlib.computer.org/dynaweb/ex/ex1997/@ebt-link;hf=0?target=if\(eq\(query\(%27%3CFNO%3E+cont+x1008%27\),0\),1,ancestor\(ARTICLE,query\(%27%3CFNO%3E+cont+x1008%27\)\)\)](http://dlib.computer.org/dynaweb/ex/ex1997/@ebt-link;hf=0?target=if(eq(query(%27%3CFNO%3E+cont+x1008%27),0),1,ancestor(ARTICLE,query(%27%3CFNO%3E+cont+x1008%27))))>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/Mauldin.html>>

Munson, Kurt I. World Wide Web indexes and hierarchical lists: finding tools for the Internet. Computers in Libraries ISSN:1041-7915 June 1996, v16, n6, p54(4).

Notess, Greg. Search Engines Showdown.

Remote: <URL:<http://imt.net/~notess/search/>>

Notess, Greg R. Searching the hidden Internet. Database ISSN:0162-4105 June-July 1997, v20, n3, p37(4).

Remote: <URL:<http://www.onlineinc.com/database/JunDB97/nets6.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/nets6.html>>

Notess, Greg R. . Searching the Web with Alta Vista.

Remote: <URL:<http://www.onlineinc.com/database/JuneDB/nets6.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/notess.html>>

Stix, Gary . Finding pictures on the Web. Scientific American ISSN:0036-8733 March 1997, v276, n3, p54(2).

Remote: <URL:<http://www.sciam.com/0397issue/0397lynchbox1.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/0397lynchbox1.html>>

Tweney, Dylan. Searching is my business: a gumshoe's guide to the Web. PC World ISSN:0737-8939 Dec 1996, v14, n12, p182(8).

Remote: <URL:http://www.pcworld.com/software/internet_www/articles/dec96/1412p182.html>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/1412p182.html>>

Yahoo!. Searching the Web.

Remote: <URL:http://www.yahoo.com/Computers_and_Internet/Internet/World_Wide_Web/Searching_the_Web/>

Zorn, Peggy; Emanoil, Mary; Marshall, Lucy; Panek, Mary. Advanced web searching: tricks of the trade. ONLINE (WILTON, CONN), vol. 20, no. 3, 12ppp, 1996.

Remote: <URL:<http://www.onlineinc.com/onlinemag/MayOL/zorn5.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/zorn5.html>>

Browsable list

C&RL NewsNet Internet Resources.

Remote: <URL:<http://www.ala.org/acrl/resrces.html>>

Excite.

Remote: <URL:http://www.excite.com/>

Hotbot.

Remote: <URL:http://www.hotbot.com/>

INFOMINE.

Remote: <URL:http://lib-www.ucr.edu/>

InfoSeek.

Remote: <URL:http://www.infoseek.com/>

Librarians' Index to the Internet.

Remote: <URL:http://sunsite.berkeley.edu/InternetIndex/>

Lycos.

Remote: <URL:http://www.lycos.com/>

Yahoo.

Remote: <URL:http://www.yahoo.com>

A2Z. Sites from A2Z.

Remote: <URL:http://a2z.lycos.com/>

Aldea Communications. InterNIC Academic Guide to the Internet.

Remote: <URL:http://www.internic.net/aldea/attframes2.html>

Argus Associates. Argus Clearinghouse.

Remote: <URL:http://www.clearinghouse.net/>

Duda, Andrea L. editor. Untangling the Web.

Remote: <URL:http://www.library.ucsb.edu/untangle/>

Louisiana State University Libraries. Webliography: A Guide to Internet Resources.

Remote: <URL:http://www.lib.lsu.edu/weblio.html>

McKinley Group, Inc. Magellan.

Remote: <URL:http://www.mckinley.com/>

Munson, Kurt I. World Wide Web indexes and hierarchical lists: finding tools for the Internet. Computers in Libraries ISSN:1041-7915 June 1996, v16, n6, p54(4).

Notess, Greg. Search Engines Showdown.

Remote: <URL:http://imt.net/~notess/search/>

Secret, Arthur . WWW Virtual Library.

Remote: <URL:http://www.fisk.edu/vl/>

Fee-based services

America Online.

Remote: <URL:http://www.aol.com/>

JSTOR.

Remote: <URL:http://www.jstor.org/>

Northern Light Search.

Remote: <URL:http://www.nlsearch.com/>

Ovid.

Remote: <URL:http://preview.ovid.com/libpreview/start.cgi>

SilverPlatter.

Remote: <URL:http://www.silverplatter.com/>

Ebsco. EbscoHost.

Remote: <URL:http://www.epnet.com/host.html>

H. W. Wilson. WilsonWeb.

Remote: <URL:http://www.hwwilson.com/tecafaq7.html>

IAC. InfoTrac SearchBank.

Remote: <URL:http://www.informationaccess.com/>

Knight-Ridder. DIALOG@Carl.

Remote: <URL:http://uncweb.carl.org/>

Knight-Ridder. DialogWeb.

Remote: <URL:http://dialog.krinfo.com/>

OCLC. FirstSearch.

Remote: <URL:http://gilligan.prod.oclc.org:3050/html/fs_homepage.htm>

UMI. ProQuest Direct.

Remote: <URL:http://www.umi.com/proquest/>

Internet search engine

Excite.

Remote: <URL:http://www.excite.com/>

Hotbot.

Remote: <URL:http://www.hotbot.com/>

InfoSeek.

Remote: <URL:http://www.infoseek.com/>

Lycos.

Remote: <URL:http://www.lycos.com/>

MetaCrawler.

Remote: <URL:http://www.metacrawler.com/>

Northern Light Search.

Remote: <URL:http://www.nlsearch.com/>

Open Text.

Remote: <URL:<http://index.opentext.com/>>

Digital Equipment Corporation. AltaVista.

Remote: <URL:<http://altavista.digital.com/>>

McKinley Group, Inc. Magellan.

Remote: <URL:<http://www.mckinley.com/>>

Knowledge management

Ash, Jerry. Knowledge Management Links.

Remote: <URL:<http://aviary.share.net/delphi/ash/ckolinks.html>>

Bellinger, Gene. Knowledge Management.

Remote: <URL:<http://www.radix.net/~crbnblu/kmgmt.htm>>

Inference Corporation. Knowledge Management Resource Guide.

Remote: <URL:<http://www.inference.com/welcome/knowledge.html>>

Public Service Commission of Canada, Strategic Planning, Analysis and Research Branch. Corporate Intelligence and Knowledge.

Remote: <URL:<http://www.psc-cfp.gc.ca/library/knowledge.htm>>

Special Library Association. Selected References on Knowledge Management.

Remote: <URL:<http://www.sla.org/membership/irc/knowledg.html>>

Lists of search engines

Collection of All the Search Engines You Will Ever Need!.

Remote: <URL:<http://www.ts.umu.se/~hezze/>>

Feeality Searches & Links.

Remote: <URL:<http://www.freeality.com/>>

Librarians' Index to the Internet.

Remote: <URL:<http://sunsite.berkeley.edu/InternetIndex/>>

A2Z . Internet indices, Directories, & How-To Guides.

Remote: <URL:http://a2z.lycos.com/Internet/Indices__Directories_and_How-To_Guides/>

Drudge, Bob . My Search Engines.

Remote: <URL:<http://www.refdesk.com/newsrch.html>>

Excite. Searching.

Remote: <URL:http://www.excite.com/xdr/Computing_and_Internet/Internet/Searching/>

Haskin, David . Right Search Engine.

Remote: <URL:<http://www.internetworld.com/1997/09/report.html>>

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/report.html>>

HotWired, Inc. Wired Cybrarian.

Remote: <URL:<http://www.wired.com/cybrarian/frame/tech/web.html>>

Macmillan Publishing USA. Search the Internet.

Remote: <URL:http://www.mcp.com/general/search/srcheng.html>

McKinley Group, Inc. Search Engines.

Remote: <URL:http://www.mckinley.com/magellan/Reviews/Computing/World_Wide_Web/Searching_The_Web/Search_Engines/index.magellan.html>

McKinley Group, Inc. Searching the Web.

Remote: <URL:http://www.mckinley.com/magellan/Reviews/Computing/World_Wide_Web/Searching_The_Web/index.magellan.html>

Mentor Marketing Services. Eureka!.

Remote: <URL:http://www.best.com/~mentorms/eureka.htm>

Netscape Communications, Inc. Netscape Net Search.

Remote: <URL:http://home.netscape.com/escapes/search/ntsrchrnd-4.html>

Nicholson, Scott . AskScott.

Remote: <URL:http://www.askscott.com/>

Zamboni. Zamboni's Search Engines.

Remote: <URL:http://www.iglou.com/zamboni/search.html>

Other services

Deja News.

Remote: <URL:<http://www.dejanews.com/>>

Excite.

Remote: <URL:<http://www.excite.com/>>

Open Text.

Remote: <URL:<http://index.opentext.com/>>

Regular expressions.

Local: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/regular-expressions.txt>>

ROADS.

Remote: <URL:<http://www.ukoln.ac.uk/roads/>>

Internet Research Task Force Research Group on Resource Discovery (IRTF-RD). Harvest.

Remote: <URL:<http://www.tardis.ed.ac.uk/harvest/>>

Kotsikonas, Anastasios. ListProc.

Remote: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/listproc-searching.txt>>

L-Soft International, Inc. LISTSERV.

Remote: <URL:<http://www.lib.ncsu.edu/staff/morgan/pointers/texts/listserv-searching.txt>>

Pfeifer, Ulrich . FreeWAIS-sf.

Remote: <URL:http://ls6-www.informatik.uni-dortmund.de/ir/projects/freeWAIS-sf/fwsf_toc.html>

Verity. Search '97.

Remote: <URL:<http://www.verity.com/>>