Resource Discovery on the Internet

H.-M. Adorf

Space Telescope – European Coordinating Facility, European Southern Observatory, Garching b. München, Germany

Abstract

The prime sources for astronomical information resources on the Internet are the AstroWeb and the Star*s databases. For topics not covered by these databases, the Internet hosts a bewildering variety of resource discovery services including AliWeb, Harvest, InfoSeek, Lycos, WebCrawler, and the WWW Worm. These and other discovery tools are reviewed. They can be used to locate e.g. on-line library services, books and CD-ROMs, software, and people's e-mail addresses.

"Information isn't truly free if you can't find it."
— *Otis Systems 1995*

1 Introduction

We are witnessing an explosive growth of the Internet [1] and particularly its World Wide Web (WWW) service, which is now (Spring 1995) being offered by more than 12,000 hosts (Fig. 1) – a hundred-fold increase within less than 2 years from a mere 130 servers back in June 1993 [2]. While initially one could keep track of interesting sites by using personal 'hotlists', these times are long gone. Instead numerous directories and search engines have been created facilitating "resource discovery", which is a topic of active research [4].

In the astronomy domain, the task of collecting relevant pointers to Internet resources has been undertaken in various places. The most comprehensive collections of astronomical Internet resources seem to be found in the *AstroWeb* database [3] and in the *Star*s Family* products [52], [53], [54]. Fully WWW-oriented, AstroWeb holds more than 1,500 Uniform Resource Locators (URLs) of astronomical institutes, projects, individuals, etc., while StarWorlds and StarHeads offer about 4,500 URLs among other data and entries. For locating non-astronomical resources, however, one has to resort to other means.

Fig. 1. The growth of the World Wide Web from 130 servers in June 1993 to almost 12,000 servers in January 1995.

Fortunately several Internet services have recently appeared helping to locate WWW information resources of interest. Firstly, there are a number of good WWW *home pages* and *small collections* specifically designed to act as WWW 'starting points' for Internet exploration. Secondly, there are browsable and often searchable *catalogues*, which group WWW resources into categories. Thirdly, and in the long run most importantly, there are various *search tools*, which can be queried and which deliver 'clickable' lists of relevant URLs.

In the following I will in turn review each of these means of increasing power. For an overview of information retrieval tools and techniques see the accompanying article by Accomazzi, Murtagh & Rasmussen elsewhere in this volume.

2 Home Pages and Small Collections

It has become a good custom for providers of WWW clients to offer well-designed WWW home pages which can serve as starting points for general Internet exploration. Examples are the NCSA Mosaic, the EINet MacWeb, and the Netscape Communication Corporation home pages. Similarly, small topical collections, though usually incomplete, can serve as a good start for Internet novices.

2.1 Astronomy-Oriented Resources

Someone specifically interested in astronomy may advantageously turn to topical pages maintained at various sites. I can recommend e.g. Alan Richmond's *Astronomy & Astrophysics on the Web* [5] on WebStars; or home pages such as *Astronomy Information* [6] at the University of Manchester; the *Astronomy Home Pages* [7] at Cambridge University; or the *Astronomy* [8] entry in *Kevin's Internet Encyclopedia* [9].

2.2 Library-Oriented Resources

Librarians will find the *Libraries* [10] entry in the WWW Virtual Library a good starting point for web explorations. The *Library Catalogs with Web Interfaces* [11] comprises catalogues of Internet resources, new library catalogues with built-in Web interfaces, catalogues reachable through Z39.50 gateways, old library catalogs with Web interfaces added post-facto, and catalogues with Gopher interfaces.

The *libweb – Library Information Servers via WWW* [12] aims at providing a world-wide list of libraries (the catalogue of) which are on-line, but some librarians deliberately chose to have their libraries not appear there for fear of misuse. The St. Joseph County Public Library (SJCPL) *List of Public Libraries with Internet Services* [13] attempts to list all public libraries world-wide that either have their own Gopher, WWW, or Telnet server, or are serving information on someone else's Gopher, WWW, or Telnet server. The Graduate School of Library & Information Science, Univ. of Texas at Austin, provides the *UT Austin – GSLIS Internet Information Resources* [14]. Other noteworthy collections of interest to librarians are the *Library Resources* [15] at the University of Minnesota, and the *Links Mostly For Librarians* [16] at Washington & Lee University.

There are also regional collections such as the *Nordic Libraries Information Servers* [17] covering Denmark, Estonia, Finland, Iceland, Lithuania, Norway, and Sweden. The *Lund University Electronic Library* [18] project strives for an "intelligent" gateway to information resources on the Internet. Finally, a rich resource is the *iBS (Internet Book Shop)* [19] – it claims to be the largest on-line book shop in the world – which offers a searchable index to the valuable comprehensive *Books in Print* catalogue [20].

2.3 General Resources

A rapidly increasing number of Internet guides are currently being compiled facilitating explorations not restricted to the astronomy and library fields. Whether a collection of resources should be called a "guide" or a "directory" (see below) depends on the amount of explanatory text and the presentation style, but the distinction is certainly blurred.

An excellent resource is the *General Guides to the Internet* [21] provided by the US Library of Congress. Alternatively one may consult e.g. *Cruising the Web* [22], the *Internet Bookshelf* [23], or the collection of *Internet guide books* [24] which comprises Ed Krol's commendable *The Whole Internet User's Guide* [25], the Electronic Frontier Foundation's (extended) *Guide to the Internet* [26] (formerly "The Big Dummy's Guide") by Adam Gaffin, John December's searchable *Internet and Computer-Mediated Communication* [27] guide, Odd de Presno's *Online World Resources Handbook* [28], Brendan Kehoe's *Zen and the Art of the Internet* [29], or Kevin Hughes' *Entering the World-Wide Web — A Guide to Cyberspace* [30].

The *NNCS Internet Resource Guide* [31] is a directory of network resources, which is also available as (compressed and tar'd) file [32]. CERN's *W3 servers* [33] page offers a useful world-wide list (> 1 *MByte*) of registered WWW servers, sorted alphabetically by continent, state and city; statistics on the most popular Virtual Library pages are also provided. The *Comprehensive List of Sites* [34] contains more than 15,000 hosts; it is generated by the WWW Wanderer and the net.Index prototype (an autonomous content analysis-based search engine), and is searchable via *WANDEX (WWW Wanderer Index)* [35].

The *Clearinghouse for Subject-Oriented Internet Resource Guides* [36] catalogue is a result of the semi-official Internet Resource Discovery Project at the University of Michigan; the catalogue is searchable [37]. Other noteworthy guides are the *WebGuide Internet Resources* [38] by Mecklermedia; *Who's Who on the Internet* [39] by City University London (searchable by registered sites only) [40]; the *Internet Hypertext List* [41] originally based on Yanoff's Special Internet Connections (see below) and maintained at Indiana University; and the *Internet Resources* [42] kept at the University at Buffalo.

Some people prefer to sort their collections not by subject, but by geographic location. The *Virtual Tourist* [43] is a geographic directory of WWW servers in the world; it operates in close association with CERN's Master Web Server Directory. The *Country maps from W3 servers in Europe* [44], the *European Home Page* [45], and the *European Community Home Page* [46] permit to find European WWW servers by physical location.

Finally there is the specialised *Goldsite Europe* [47] server providing information on nearly 3,000 Internet connected companies and organisations worldwide. Goldsite claims to be "probably Europe's busiest and largest commercial Web Server" with over 15,000 indexed files and over a million accesses per month.

3 Resource Directories/Virtual Libraries

There are several catalogues and directories of URLs which have been collected by individuals or by groups. Such collections are often not complete, but have the virtue that the entries are being subjectively rated by their value. As opposed to pure search tools (discussed below), catalogues are always browsable. Since they can be quite large (100 kilobyte and more), catalogues are often also searchable – a feature that tends to blur the distinction between catalogues and search tools.

3.1 Astronomy-Oriented Catalogues

The comprehensive *AstroWeb* [48] and *Star*s Family* [52], [53], [54] of databases have already been mentioned. (AstroWeb was recently merged with the *Astronomy and Astrophysics* [49] collection in the WWW Virtual Library.) In addition there is the review paper *Network Resources for Astronomers* [50] by Andernach, Hanisch & Murtagh; the searchable on-line version [51] is updated every now and then. Finally, André Heck's Star*s family of astronomical databases deserves to be mentioned. This high-quality resource collection includes three distinct parts: *StarWorlds* [52] listing astronomical and related organisations, *StarBits* [53] with abbreviations, acronyms, contractions and symbols, and *StarHeads* [54] with URLs to WWW pages of astronomers and space scientists.

3.2 Library-Oriented Catalogues

Librarians might find the *University of Waterloo Electronic Library* [55] a good starting point for Internet explorations. It provides access to resources organised by discipline, walking tours of the libraries at the University of Waterloo, a WWW interface to the Library's on-line catalogue, links to several web search engines, and more (e-books, e-journals). The *Library Section of Buzzard's Nest* [56] is a personal collection of library resources, whereas *HyperMedia Zines on the Net* [57] is an index to electronic journals and magazines.

3.3 General Catalogues

Among the many general resource directories *Yahoo – A Guide to WWW* [58] with its almost 40,000 searchable entries certainly stands out. It has been created by two dedicated individuals, David Filo and Jerry Yang, originally at Stanford University. Yahoo, which apparently went commercial recently, sports e.g. the *Commercial Directories* [59] with more than 50 entries. End of April 1995 Yahoo was ranked first in "The Whole Internet Catalog Top 25".

The *WWW Virtual Library* [60] is a distributed subject-oriented database (i.e. different subjects handled by different sites), which is also available in a format conforming to US Library of Congress standards [61]. The WWW Virtual Library contains the very useful directory of *Virtual Libraries* [62].

The *GNN (Global Network Navigator)* [63] from O'Reilly & Associates is accessible by free subscription; it sports the searchable *Whole Internet Catalog* [64]. The *Internet Resources Guides, Tools and Services* [65] by the US Library of Congress is another valuable resource repository.

The *Internet Resources Meta-Index* [68] at NCSA is a loosely categorised meta-index of the various resource directories and indices. The searchable *EINet Galaxy* [69] is a guide to world-wide information and services. *Nova-Links* [70] is an extensive Internet navigator from Nova Southeastern University; it includes pointers to listservers, software libraries, and library catalogs [71]. The *Mother-of-all BBS* [72] is a huge file maintained by Oliver McBryan at Univ. of Colorado. Relying on external contributions it offers pointers to WWW home pages of companies, universities, research centers, government agencies, research projects, hardware or software announcements, etc.

The *UofS – Search the Internet* [73] at University of Saskatchewan is a very commendable site. Worth mentioning are also the *Internet Resource Discovery Tools* [74] by the Emory University Health Sciences Center Library; the *LLNL List of Lists* [75] directory at Lawrence Livermore National Laboratory featuring a special science list; the *InterNIC Directory and Database Services* [76] providing access to the Internet White Pages information using whois, netfind, and X.500 systems; *IntIndex – An Internet Index* [77] by SilverPlatter World, a contribution-based world-wide library of electronic information resources; and the searchable *Shops and Services on the Net* [78] by Inter-Market.

Richard P. Bocker's *Planet Earth Home Page (PEHP)* [66] collection is a commendable virtual library featuring the searchable *PEHP Virtual Library Comprehensive Image Map* [67].

Those interested in the commercial side of things will find the *Commercial Sites Index by Open Markets* [79] useful. It is a directory of commercial services, products, and information on the Internet; ANDed and ORed Boolean searches are permitted. Other directories worth mentioning are the *Interactive Yellow Pages* [80] by NetCenter; the *DA-CLOD (Distributedly Administered Categorical List Of Documents)* [81], a searchable database where anybody can add URLs; and the *Web of Wonders* [82] by Weicomp Consulting Inc.

Among the many personal collections one finds frequent references to the *Internet Resources* [83], maintained by Neil Enns, Brandon University, and nominated "Cool Site of the Day" in February 1995; *Todd's Links to Other Great Network Resources* [84] by Todd Ellis Van Hoosear, Michigan State University; *Yanoff's Special Internet Connections* [85] maintained by Scott Yanoff, Univ. of Wisconsin-Milwaukee; *Internet Resources* [86] by Michael Fisk, New Mexico Tech.; *Joel's Hierarchical Subject Index* [87] by Joel Jones, Univ. of Illinois; the *Scott List* [88] (huge file) by Michael Louis Scott, listing 1,325 WWW sites in 40 categories; the searchable *Internet Sleuth* [89] by Internet Business Connection; *Brent's Internet Jumpstation* [90] by Brent Hunter; and Patrick Debois' *URL Heaven* [91] (which however is likely to stop soon).

4 Search Tools

Search tools permit seeking information on the Internet. For FTP archives there is Archie and Martijn Koster's *ArchiePlex* [151], a forms based Archie gateway for the WWW; for Telnet resources there is *HyTelnet* [92]; for Listserv's and Newsgroups there is *List search* [93]; for Gopher space there are *Veronica* [94], *Jughead* [95], *BUBL* [96], and *Subject Trees* [97]. And for the WWW there are crawlers, nomads, robots, spiders, wanderers, and worms [98].

Due to its size and distributed nature the Internet as such cannot be searched in near real-time. Therefore all search engines operate on a database which, similarly to the operation of Archie, is assembled prior to the query by automatically exploring the web in a piecemeal fashion, and collecting URLs together with auxiliary descriptive information. The databases underlying such "search engines" are different from the directories reviewed above, in the sense that they are assembled automatically and tend to be much more complete and therefore larger to the extent that they cannot be browsed and inspected interactively.

4.1 Directories of Search Engines

Currently the Internet hosts at least two dozen different search engines (see below). Therefore several collections of (URLs to) search engines have been assembled which can be used to keep abreast of new developments. Some such collections can be used directly by allowing immediate entry of a query.

At present the presumably most comprehensive list of search tools is *Searching the Web* [99] at Yahoo. Other useful collections are the *World Wide Web Robots, Wanderers, and Spiders* [100] at NEXOR, and the *W3 search engines* at CERN [101], mirrored at North Carolina State University [102].

CUSI (Configurable Unified Search Engine) [104] by NEXOR, is a commendable single WWW form permitting to use one of several different distributed WWW search engines. The service, which originated in a personal initiative of Martijn Koster in 1993, is mirrored at multiple sites, and can be imported to run locally.

The very valuable *Multithreaded query page* [103] permits to issue a query simultaneously to several distributed search engines which then run in parallel. The response takes as long as the slowest search engine selected.

CityScape Internet Services [105] collects search engines geared towards specific topics such as information servers, software, people, publications, News/FAQs (frequently asked questions), documentation, and other interesting things.

The *OTIS Index* [109] at Otis Systems permits to search for people via Netfind, via USENET authors, via X.500 UFN search, and via Internet Domains. *LinkSearch* [108] by WHERE.COM offers canned searches for software, people, dictionaries etc.

Among institutional collections of search engines we find additionally: the *Web Indexes* [110] at Univ. of Washington; the *Search Utilities* [111]; the *Internet Search* [112] by Netscape; the *Internet Search* [113] by DCA; and the *Search Sites* [114] page by Cera.

Other institutional collections are the *Internet Search Engines* by BBC [115] which features the *Web Jungle* [116] listing search tools for information servers, software, people, News/FAQs, and Internet documentation; the *LINKS, Search Tools & Libraries* [117] by the Internet Commerce Center; the *Charm Net's Pointers To Information Search Services* [118] by Charm which is a nicely commented list; the *Links to Search Tools, Virtual Libraries and other Index pages* [119]; and the *Virtual Library/CyberWeb Search* [120] at CyberWeb, a large, extensively commented collection.

The *PEHP's Search Engine Room* [106] features the *Search Engines* [107] collection. Other personal collections include the *World Wide Web Search Engines* [121] list by Andy Black, Univ. of York, and the *Internet Search Engines* [122] directory by Joe Kung, MIT.

4.2 Search Engines

Today the most prominent search engine is presumably *Lycos* [123] by Carnegie Mellon University, a "GNN Best of the Net" [124] nominee for 1995, and recently non-exclusively licensed to Microsoft. In November 1994 altogether 862,858 documents had been indexed, a number which had climbed to 3.6 million in May 1995. Lycos allows a search by document title, headings, links, and keywords. It also offers a list of 6,898 distinct *HTTP servers* [125].

Its direct competitor, the *Harvest Information Discovery and Access System* [126], consists of an integrated set of tools to gather, extract, organise, search, cache, and replicate relevant information across the Internet. It features the *WWW Home Pages* [127], a searchable index of reportedly over 35,500 home pages [128].

InfoSeek [131] is reported to be a comprehensive and accurate commercial WWW search engine (query by plain English, key words and phrases, or special query operators). It claims to index the entire contents of a month of over 10,000 Newsgroups at once (over 3 million articles), and to cover more than 200,000 fully indexed Web pages. The Internet Magazine called InfoSeek the best search tool on the Internet. Free access is permitted for 1 month.

The *CUI W3 Catalog* [134] is a fairly comprehensive, semi-automated, high-quality global index permitting search via word/pattern or Perl regular expressions. It is mirrored at various other sites.

The *JumpStation II* [133], which supersedes the older *JumpStation* [132], allows interactive searching of documents by title, header, and subject. Furthermore it also offers server search by partial server address with wildcards.

The *WebCrawler* [135] at Univ. of Washington allows ANDed or ORed keyword search; it is equipped with document relevance ranking.

The *RBSE's (Repository Based Software Engineering)* URL database [139] at NASA Johnson Space Center & University of Houston, is a spider building an Oracle database. It uses freeWAIS indexing/searching of more than 36,000 full-text documents.

AliWeb [140], a public service at NEXOR, is a semi-distributed, slowly growing special purpose global index for the Web. It regularly retrieves specially formatted files, and combines them into a searchable database. AliWeb is mirrored at Indiana University [141].

The *GOLD (Global On-Line Directory)* [143] by CityScape aims to be a complete guide to the Internet. It reportedly indexes several wire services (including AP, Reuters, Newsbytes, PR Newswire, and BusinessWire) as well as the Computerworld and InfoWorld magazines.

The *WWW Worm (WWWW)* [129] was designated the best navigational aid in 1994. It allows the search for words embedded in document titles or URLs, and permits using so-called "regular expressions" for search refinement (a la Unix's egrep). Unfortunately WWWW was last run in September 1994, so it is pretty much outdated by now.

The *GNA (Global Network Academy) Meta-Library* [142] allow field-by-field Boolean search. It is maintained manually and thus is likely to become outdated too.

SG-Scout [147], developed for the Xerox Palo Alto Research Center; the robot was last run in August 1994, when it discovered over 7,250 WWW servers and 250,000 HTML and text pages. Again if such a tool is not continuously operated it becomes useless after a while.

Other noteworthy general search engines are the *Search the Web* [136] engine at Galaxy EINet; *NIKOS* [137] at Rockwell Telecommunications (formerly WWW Nomad) permitting ANDed search; and the good looking *Internet Wizard* [138] at Spry.

Several search engines internally use the well-established WAIS search mechanism, such as *Internet Services* [144], a WAIS database of Internet resources; *MESCH (Multi-WAIS Engine for Searching Commercial Hosts)* [145] by the Internet Presence & Publishing Corporation; and *Discover* [130], a system for finding WAIS databases on over 500 WAIS servers. The *Nordic WAIS/World Wide Web Project* [146] automatically indexes and classifies WAIS databases found on the Internet.

The *FOUR11* [150] engine may be used for finding e-mail addresses among over 1 million entries.

Then there is the very interesting *Fish-Search* [148] engine which offers user-controllable information fishing starting from an arbitrary user-specified URL. Thus WWW servers may be searched which do not offer a search facility by themselves.

And finally *URouLette* [149] at the University of Kansas allows serendipitous "discoveries" by carrying the user (virtually, of course) to a random site on the web.

5 Conclusion

The Internet offers a multitude of directories and at least two dozen search engines for locating desired information resources. Using the information compiled in this overview will enable an Internet user to find the resource discovery tool which best suits his/her needs. However, it is currently not easy to obtain reliable information for these tools about web coverage, the relevance ranking mechanisms used, and the overall quality. Search engines differ widely in the means used to precollect data, in admissible queries, and in the presentation of results. It is hoped that in the future more search engines will use the WAIS/Z39.50 standards. Unfortunately some of the databases underlying the search engines are growing out of date, and some doubts can be raised about the longevity of personal collections. Still the amazing fact is that information seeking in an anarchic system such as the Internet is possible on a global scale!

References

[1] http://www.nw.com/zone/WWW/report.html

[2] http://www.netgen.com/info/growth.html

[3] http://ecf.hq.eso.org/astroweb/yp_astro_resources.html

[4] http://www.igd.fhg.de/www/www95/workshops/ref-a.html

[5] http://guinan.gsfc.nasa.gov/WebStars/AST.html

[6] http://axp2.ast.man.ac.uk8000/astro.html#Astronomy

[7] http://www.ast.cam.ac.uk/Astro/HomePages.html

[8] http://duke.usask.ca/~lowey/encyclopedia/a/astronomy.html

[9] http://duke.usask.ca/~lowey/encyclopedia/index.html

[10] http://www.w3.org/hypertext/DataSources/bySubject/Libraries.html

[11] http://www.lib.ncsu.edu/staff/morgan/alcuin/wwwed-catalogs.html

[12] http://www.lib.washington.edu/~tdowling/libweb.html

[13] http://sjcpl.lib.in.us/homepage/PublicLibraries/PublicLibraryServers.html

[14] http://fiat.gslis.utexas.edu/internet/internet.html

[15] gopher://gopher.tc.umn.edu/11/Libraries

[16] http://liberty.uc.wlu.edu/~hblackme/iii.html

[17] http://www.ub2.lu.se/resbyloc/Nordic_lib.html

[18] http://munin.ub2.lu.se/ub2.html

[19] http://www.bookshop.co.uk/

[20] http://www.bookshop.co.uk/V3/SEARCH3.HTM

[21] http://lcweb.loc.gov/global/internet/guides.html

[22] http://www.stat.ucla.edu/cruising/

[23] http://www.mecklerweb.com/mags/iw/books/books.htm

[24] http://www.brandonu.ca/∼ennsnr/Resources/guides.html

[25] http://nearnet.gnn.com/gnn/bus/ora/catalog/twi2.html

[26] http://www.hep.net/documents/eegtti/eegtti.html

[27] http://www.rpi.edu/Internet/Guides/decemj/icmc/top.html

[28] http://login.eunet.no/∼presno/index.html

[29] ftp://ftp.internic.net/pub/internet-doc/zen.txt

[30] http://www.eit.com/web/www.guide/

[31] gopher://ds.internic.net/11/resource-guide

[32] http://www.cs.indiana.edu/internet/internet.rsrc.html

[33] http://info.cern.ch/hypertext/DataSources/WWW/Geographical.html

[34] http://www.netgen.com/cgi/comprehensive

[35] http://www.netgen.com/cgi/wandex

[36] http://http2.sils.umich.edu/∼lou/chhome.html

[37] gopher://una.hh.lib.umich.edu70/77/inetdirsstacks/.waisindex/index

[38] http://www.mecklerweb.com/webguide/resource.htm

[39] http://web.city.ac.uk/citylive/pages.html

[40] http://web.city.ac.uk/citylive/whosearch.html

[41] http://www.cs.indiana.edu/internet/internet.html

[42] gopher://wings.buffalo.edu/hh/internet/

[43] http://wings.buffalo.edu/world/

[44] http://www.tue.nl/maps.html

[45] http://s700.uminho.pt/europa.html

[46] http://s700.uminho.pt/ec.html

[47] http://www.gold.net/

[48] http://ecf.hq.eso.org/astroweb/yp_astro_resources.html

[49] http://www.w3.org/hypertext/DataSources/bySubject/astro/astro.html

[50] http://www.eso.org/online-resources-paper/rrn.html

[51] http://www.eso.org/online-resources.html

[52] http://cdsweb.u-strasbg.fr/∼heck/sfworlds.htm

[53] http://cdsweb.u-strasbg.fr/∼heck/sfbits.htm

[54] http://cdsweb.u-strasbg.fr/∼heck/sfheads.htm

[55] http://www.lib.uwaterloo.ca/

[56] http://www.infi.net/∼cwt/lib.html

[57] http://www.nwu.edu80/ezines/

[58] http://www.yahoo.com/

[59] http://www.yahoo.com/Business/Directories/

[60] http://info.cern.ch/hypertext/DataSources/bySubject/Overview.html

[61] http://info.cern.ch/hypertext/DataSources/bySubject/LibraryOfCongress.html

[62] http://info.cern.ch/hypertext/DataSources/bySubject/Virtual_libraries/Overview.html

[63] http://gnn.com/gnn/gnn.html

[64] http://nearnet.gnn.com/wic/newrescat.toc.html

[65] http://lcweb.loc.gov/global/internet/internet.html

[66] http://godric.nosc.mil/planet_earth/info.html

[67] http://godric.nosc.mil/planet_earth/everything.html

[68] http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/MetaIndex.html

[69] http://www.einet.net/galaxy.html

[70] http://alpha.acast.nova.edu/start.html

[71] http://kufacts.cc.ukans.edu/hytelnet_html/SITES1.html

[72] http://www.cs.colorado.edu/homes/mcbryan/public_html/bb/summary.html

[73] http://www.usask.ca/search/index.shtml

[74] http://www.cc.emory.edu/WHSCL/resource.discovery.html

[75] http://www.llnl.gov/llnl/lists/listsl.html

[76] http://www.internic.net/ds/dspg01.html

[77] http://www.silverplatter.com/intindex/intro.html

[78] http://www.gold.net/inter-market/index2.html

[79] http://www.directory.net/

[80] http://netcenter.com/lite/cgi-bin/yellowsindex.pl

[81] http://schiller.wustl.edu/DACLOD/daclod

[82] http://www.digimark.net/wow/

[83] http://www.brandonu.ca/~ennsnr/Resources/

[84] http://clunix.msu.edu/~vanhoose/link.html

[85] http://www.uwm.edu/Mirror/inet.services.html

[86] http://nmt.edu/~mfisk/inet.html

[87] http://www.cen.uiuc.edu/~jj9544/index.html

[88] http://www.picosof.com8080/html/scott.html

[89] http://www.charm.net/~ibc/sleuth.html

[90] http://www.catalog.com/bhunter/welcome.htm

[91] http://hobbes.rug.ac.be/~pdebois/

[92] http://galaxy.einet.net/hytelnet/HYTELNET.html

[93] http://alpha.acast.nova.edu/cgi-bin/news-lists.pl

[94] gopher://veronica.scs.unr.edu/11/veronica

[95] gopher://liberty.uc.wlu.edu3002/7

[96] gopher://ukoln.bath.ac.uk7070/11/Link

[97] http://alpha.acast.nova.edu/subjects.html

[98] http://www.ics.uci.edu/WebSoft/MOMspider/WWW94/paper.html

[99] http://www.yahoo.com/Reference/Searching_the_Web/

[100] http://web.nexor.co.uk/mak/doc/robots/robots.html

[101] http://cuiwww.unige.ch/meta-index.html

[102] http://www.lib.ncsu.edu/meta-index.html

[103] http://www.sun.fi/mtq/mtquery.html

[104] http://pubweb.nexor.co.uk/public/cusi/doc/list.html

[105] http://www.gold.net/internet-access/index.html

[106] http://corpinex.nosc.mil80/planet_earth/Library/sei_room.html

[107] http://godric.nosc.mil/planet_earth/pehp_2.html?search+engines#first_hit

[108] http://where.com/ls/LinkSearch.html

[109] http://www.interlog.com/~gordo/otis_index.html

[110] http://webcrawler.cs.washington.edu/WebCrawler/WebIndexes.html

[111] http://www.tiig.ist.ucf.edu/search.html

[112] http://home.mcom.com/home/internet-search.html

[113] http://www.dca.net/Internet-Search.html

[114] http://www.cera.com/search.htm

[115] http://auntie.bbcnc.org.uk/babbage/iap.html

[116] http://www.bbcnc.org.uk/babbage/search.html

[117] http://www.save.com/links.html

[118] http://www.charm.net/info.search.html

[119] http://info1.vub.ac.be8080/search_tools/index.html

[120] http://www.charm.net/~web/Vlib/Users/Search.html

[121] http://www.york.ac.uk/~apb104/engines.htm

[122] http://cecelia.media.mit.edu8080/htdocs/search.html

[123] http://lycos.cs.cmu.edu

[124] http://gnn.com/gnn/wic/best.toc.html

[125] http://lycos.cs.cmu.edu/servers94nov.html

[126] http://rd.cs.colorado.edu/harvest/

[127] http://harvest.cs.colorado.edu/brokers/www-home-pages/query.html

[128] http://www.town.hall.org/brokers/www-home-pages/query.html

[129] http://www.cs.colorado.edu/home/mcbryan/WWWW.html

[130] http://www.bocklabs.wisc.edu/www95/papers/82/html-files/discover.html

[131] http://www.infoseek.com

[132] http://js.stir.ac.uk/jsbin/js

[133] http://js.stir.ac.uk/jsbin/jsii

[134] http://cuiwww.unige.ch/w3catalog

[135] http://webcrawler.cs.washington.edu/WebCrawler/WebQuery.html

[136] http://galaxy.einet.net/www/www.html

[137] http://www.rns.com/cgi-bin/nikos

[138] http://wizard.spry.com/

[139] http://rbse.jsc.nasa.gov/eichmann/urlsearch.html

[140] http://web.nexor.co.uk/public/aliweb/aliweb.html

[141] http://www.cs.indiana.edu/aliweb/form.html

[142] http://uu-gna.mit.edu:8001/uu-gna/meta-library/index.html

[143] http://www.gold.net/gold/

[144] wais://munin.ub2.lu.se:210/internet_services

[145] http://www.ip.net:80/cgi-bin/mesch/

[146] http://www.ub2.lu.se/W4.html

[147] http://www-swiss.ai.mit.edu/~ptbb/SG-Scout/SG-Scout.html

[148] http://www.win.tue.nl/bin/fish-search

[149] http://www.ukans.edu/uroulette.html

[150] http://www.Four11.com/

[151] http://cuiwww.unige.ch/./archieplexform.html