# Internet Discovery Tools: The Basics

## Greg A. Wade
Southern Illinois University at Carbondale

### Introduction

An Internet discovery tool is a software package or protocol which allows a user to seek out and retrieve information from computers attached to the Internet. There are many such tools. Some of which perform a very specific task and others which are more general purpose. However, to be successful at "net surfing" you do not need an in-depth knowledge of them all. You only need to be slightly familiar with most, and proficient with a gopher client or World Wide Web (WWW) browser.

### Accessing Remote Computers

Telnet allows you to remotely logon to another computer anywhere on the Internet. Once logged in, you can operate the remote computer as if you were at its keyboard or at one of its terminals.

To use a remote machine you will usually need an account consisting of a userid and password to access the system. Some computers allow access to guest accounts. These accounts, for example, are usually restricted to accessing bulletin boards or searching library catalogs. However, most systems only allow access to local users. The need for publicly accessible accounts on computers has decreased sharply since the introduction of other Internet services. The task of finding public systems that still exist has been eased by the introduction of gopher and the WWW.

### Finding and Downloading Files

A computer may not allow public access via a guest account, but it may have an archive of files which may be downloaded by any Internet user. These archives contain software, papers, statistical data, and almost anything else you can imagine. The main task is finding what you need.

Archie was developed by McGill University to aid in this task. Deriving its name from the word archive, archie helps you to find where particular files are located on the Internet. Archie servers maintain databases of the names of files stored at known public archive sites. Besides filename databases, some archie servers keep databases of file descriptions. This is helpful when you know what you are looking for but you don't know the name of the file. There are many ways to access archie servers, but the simplest is through gopher or WWW archie gateways.

After a file is located at an archive site, it then must be transferred or copied to your computer. This can be done using the file transfer protocol (ftp). With ftp you can anonymously retrieve files from public archive sites. Thus, the development of the term "anonymous ftp". Most networking packages provide a standalone ftp client, which is used by many to transfer files. Recently most archive sites have become accessible via gopher and WWW.

### Searching Databases

In addition to files, there are many searchable databases on the Internet. Many of these databases are searchable with WAIS. WAIS stands for Wide Area Information Servers. This technology was created by the Thinking Machines Corporation to provide a common protocol for indexing databases and extracting information from them. Although there is client software available for performing WAIS searches, many gopher and WWW sites offer gateways to perform these. However, some databases indexed with WAIS technology are not directly accessible by client software. These must be accessed through gopher or WWW.

**Surfing the Net**

Besides providing the sole access method for some databases, gopher and WWW programs are the best way to explore the Internet. These programs are similar but very different. Both attempt to provide information in an organized fashion. However, they each do this quite differently.

Gopher clients allow you to access computers running special server software that allows access to collections of documents and services organized into very easy to use menus. Besides being easy to use, gopher clients provide many additional features. Most allow you to save or print locally anything you can view on the screen. Many also allow you to create a list of bookmarks. These bookmarks help you quickly return to information of interest.

Gopher client software is available for most computers. The archive site located at the University of Minnesota, where gopher was created, is the best place to find it. The site is located at boombox.micro.umn.edu and accessible by anonymous ftp. Once connected look in the /pub/gopher directory for the name of your computer or operating system.

Like gopher clients, WWW browsers allow you to access computers running special server software. However unlike gopher, WWW browsers are not restricted to presenting information as textual menus. They present information as hypertext, full color images, audio clips, and full motion video. Many WWW browsers also support imagemaps and fill-out forms.

Imagemaps allow you to "click" on a portion of a graphic image to retrieve information. An example of this would be a United States map. When a state is "clicked" on, more detailed information about that state would be presented. Fill-out forms allow for greater interaction between the user and the WWW site. Their uses range from providing very user-friendly interfaces for performing WAIS searches to conducting online surveys.

Currently, there are two very popular WWW browsers, Mosaic and the Netscape Navigator. Mosaic was created by the National Center for Supercomputing Applications (NCSA) and the Netscape Navigator is a commercial product of Netscape Communications Corporation. They may be obtained by anonymous ftp from ftp.ncsa.uiuc.edu in the directory /pub/Mosaic and ftp.netscape.com in the directory /netscape1.1b3 respectively. Both can be used freely by academic institutions and non-profit organizations, but others should first check the terms of the license agreement accompanying each software package.

The latest versions of Mosaic and the Netscape Navigator have very comparable features with one notable exception. Netscape Communications has created a secure sockets layer (SSL). Their browser uses SSL to perform secure data transfers with commercial WWW sites running the Netscape Commerce Server [Netscape]. This technology allows you to safely transmit confidential information, such as credit card numbers, to WWW sites.

Besides Mosaic and the Netscape Navigator, there are many other less popular WWW browsers. These include WinWeb, Cello, IBM OS/2 WebExplorer, MacWeb, lynx, and many others [Boutel]. Complete or demonstration versions of most WWW browsers can be found at anonymous ftp sites which are located easily by Archie.

**Finding the Waves**

Once you begin to surf the Internet, you will ultimately ask the question: How do I find information on...? Although there is no central card catalog for the Internet, there are ways to search for information by keywords.

Gopher sites can be searched by a very easy rodent-oriented net-wide index to computerized data, commonly known as Veronica. This index contains the titles of 99% of all gopher menu items [Foster]. Veronica can be searched by pointing your gopher client at veronica.scs.unr.edu. Similarly, WWW can be searched with Lycos which gets its name from the first five letters of the Latin name for the Wolf Spider. Lycos is being developed at Carnegie Mellon University, and has the ability to search the text of 2.5 million documents in a matter of minutes [Mauldin].

For those who would rather browse than search for information, there are subject catalogs for both gopher and WWW. A listing of gopher sites by subject can be found at belden.lib.depaul.edu and the Yahoo subject catalog of

WWW sites can be found at  http://www.yahoo.com.  However, these indices and subject catalogs should not be considered definitive.

New information resources are being added daily.  To see a listing of the newest WWW sites routinely check NCSA's What's New located at http:// www.ncsa.uiuc.edu/SDG/ Software/Mosaic/Docs/whats-new.html.  This listing is arguably the most read WWW document.  It is seen by as many as 2 million readers per month [Welz].  Yahoo also maintains a listing to the sites most recently added to their subject catalog.

**Conclusion**

Although there is a large amount of information available on the Internet, accessing it doesn't have to be painful.  It can actually be a quite enjoyable experience with gopher or WWW.  You never know what may lie just beyond the next click of your mouse.

**References**

Anderson, Faye and Greg Wade.  1994.  *The UWIN Guide to the Internet*, http://www.uwin.siu.edu/ welcome/guide.html.

Boutell, Thomas.  1995.  *World Wide Web Frequently Asked Questions*, http://sunsite.unc.edu/boutell/faq/ www_faq.html.

Foster, Steven.  1995.  *Common Questions and Answers about Veronica, a title search and retrieval system for use with the Internet Gopher.* gopher:// veronica.scs.unr.edu/00/veronica/veronica-faq.

Netscape Communications Corporation.  1995. *Netscape Data Security*, http://www.netscape.com/newsref/ ref/netscape-security.html, 1995.

Mauldin, Michael L.  1995.  *CMU Lycos, Frequently Asked Questions*, http://lycos.cs.cmu.edu/lycos-faq.html.

Welz, Gary.  1995.  Cyber Spiels, *Internet World*,  6(5):48-50.

**APPENDIX B: GLOSSARY OF USEFUL TERMS**

Archie: a system for locating files that are publicly available by anonymous FTP. Archie tells you where a file is based on the name you provide.

BBS (Bulletin Board System): this service typically has several features that may include read-only information, boards where users can post messages, mail, chat rooms where several users can interact, etc. You must log onto a bbs to participate in its capabilities.

Bookmark: a facility within gopher that remembers places you use often; for example, you can place a bookmark on the UWIN MainMenu. The bookmark allows your software to remember the site and go there immediately.

Browser: a World Wide Web client such as Netscape or Mosaic; an information retrieval tool.

Client: a software application that runs on your machine andextracts some service from a network server. Gopher, Mosaic, and Netscape are examples of such client applications. In short, client software brings some of the Internet's capabilities directly to your computer.

E-mail (Electronic Mail): the most commonly used application on the Internet that allows you to write and send messages. E-mail is passed along from machine to machine until it finally arrives. This mail service critically depends on the address used so you need to maintain a list of the e-mail addresses of those you want to correspond with.

FAQ (Frequently Asked Questions): A list of questions and their answers that are provided so that basic questions are covered and do not have to be addressed over and over again.

FTP (File Transfer Protocol): a protocol that defines how to transfer files from one machine to another. Ftp also refers to the application program which moves files using the File Transfer Protocol.

GIF: graphics interchange format; an image file format

Gopher: a client-server application that provides a menu-driven system for exploring Internet resources (created at the University of Minnesota).

History List: a list of document titles and URLs kept in your client's memory that represents the sites visited during an on-line session.

Home Page: the graphical door to the information a Web server provides (analogous to a main menu); a top level document

Hotlist: a list of specified URLs to a set of preferred Internet information resources; e.g. the UWIN Wetlist.

HTML (HyperText Markup Language): the markup system that is used
to create hypertext documents that are available on Web servers.

HTTP (HyperText Transfer Protocol): the protocol used by WWW browsers and servers to communicate. This protocol supports the transmission of hypertext and hypermedia documents. The UWIN web server, that you can access via Netscape or Mosaic, is an HTTP server.

Hyperlink: a link in a document to information contained within another document. These links are usually represented by highlighted words or images.

Hypermedia: refers to documents that contain a variety of information types such as text, image, audio, and video. These media are accessed through hyperlinks

Hypertext: hypertext is data that contains links to other data. The program that you use to read a hypertext document is called a browser.

Internet: the Net is the worldwide network of networks and is nothing more than a collection of computers, telephone cables and satellite transmission systems that relay data to and from thousands of points across the world.

Internet Service Provider: an organization that provides access to the Internet (or parts of the Internet) for free or for a fee.

IP: Internet Protocol is an important protocol underlying the Internet. It allows a packet to travel across multiple networks on the way to its ultimate destination.

Jughead: Jughead performs similar tasks as Veronica only for a specifically defined set of gopher menus.

Listserver: a machine that runs an automatic discussion list. You must subscribe to such a list to receive all the correspondences of the list. Typically, such lists are organized around particular discussion topics.

Mosaic: a graphical WWW browser which acts as a client to HTTP servers (created by the National Center for Supercomputer Applications).

Netiquette: the term used to refer to the informal rules of behavoir for the Internet

Netscape: a graphical WWW browser which acts as a client to HTTP servers (created by Netscape Communications, Inc.).

Newsgroups: similar to a listserver with one major difference. Instead of the messages coming directly to you via your e-mail address, users must use special client software to access ("point at" ) newsservers; e.g. Netscape can point at newsservers.

PPP: Point to Point Protocol is a version of the Internet software that runs over normal phone lines using standard high-speed modems. PPP allows a computer to use TCP/IP protocols and is thus a direct connection to the Internet. PPP is a newer standard that is replacing SLIP. Both bring the power and flexibility of the Internet to your computer over an ordinary phone line.

Protocol: a set of rules outlining how computers can talk to one another.

Public Domain: refers to software that can be used free of charge.

Server: a software application that allows a computer to offer a service to other computers over the Internet. These other computers contact the server program by means of the matching client software. Server also refers to the computer on which the server software runs.

SLIP: Serial Line IP is a protocol that allows a computer to use the Internet protocols using a standard phone line and a high-speed modem. Like PPP, it allows for a direct Internet connection. It is commonly used but is being replaced by PPP.

TCP: Transfer Control Protocol is one of the protocols that the Internet is based on.

Telnet: a terminal emulation protocol that allows you to log in to other computer systems on the Internet. Telnet also can refer to the application program that allows you to log in to another computer system using the telnet protocol.

Terminal Emulation: different types of computer terminals understand different control codes for such features as clearing the screen, cursor movemnts, etc. Terminal emulation refers to which type of a well known terminal your computer behaves like, i.e., emulates. When you log in to another system, you will be asked about your terminal type. This is to enable the machines to communicate with one another so that you can use the other machine's

services smoothly.  Typically, you can say uou are emulating a "vt100" terminal.  When you create an account on WaterTalk, you will be asked your terminal type.

UNIX:  an important operating system in the development of the Internet.  However, YOU do NOT have to know anything about UNIX to use the Internet.

URL:  URL stands for Uniform Resource Locator which is a draft standard for specifying an address on the Internet.  The URL contains four distinct parts:  the protocol type, the machine name, the direct path, and the file name.

Veronica:  this is a tool that keeps track of many gopher menus around the world.  It performs a search and looks for all the menu items that contain the keywords you desire.

WAIS:  Wide-Area Information Servers are powerful systems for looking up information on the Internet.  WAIS helps you search for information based upon what's in documents located on servers.

WWW (World Wide Web):  the Web is a relatively new browsing and searching system on the Net using hypertext and hypermedia (sometimes refered to as just the Web).

The following is the UWIN listing of water related listservers and mailing lists that you can sign up for and receive posts at your e-mail address. Most listservers accept the command "subscribe listname your e-mail address" to sign you on. Do not send subscribe commands to the post address for the list.
This listing is posted on the UWIN system where it is updated frequently. Please send any additions to faye@uwin.siu.edu.


AQUIFER:        Pollution and Groundwater Discharge
Subscribe: LISTSERV@IBACSATA.BITNET
Post: AQUIFER@IBACSATA.BITNET


AQUA-L:        Aquaculture List; discussions on the science, technologies and business of aquaculture.
Subscribe: LISTSERV@VM.UOGUELPH.CA
Post: AQUA-L@VM.UOGUELPH.CA


BPWSP-L:        Bureau of Public Water Supply Protection discussion list.
Post: BPWSP-L@ALBNYDH2


COASTGIS:        Coastal GIS Distribution List
Subscribe: LISTSERV@IRLEARN.UCD.IE
Post: COASTGIS@IRLEARN.UCD.IE


COASTNET:        Coastal Management Conference; a forum for planners, managers, researchers and others to discuss national and international coastal management issues.
Subscribe: LISTSERV@URIACC.URI.EDU
Post: COASTNET@URIACC.URI.EDU


CTURTLE:        Sea Turtle Biology and Conservation
Subscribe: LISTSERV@NERVM.NERDC.UFL.EDU
Post: CTURTLE@NERVM.NERDC.UFL.EDU


DEEPSEA:        Deep Sea and Hydrothermal Vent Biology; serves the global community of deep sea hydrothermal vent and seep biologists working in the areas of evolution, ecology, biogeography, paleontology, systematics, phylogenetics, population genetics and zoology.
Subscribe: LISTSERV@UVVM.UVIC.CA
Post: DEEPSEA@UVVM.UVIC.CA


DIALOG-AGUA-L: information pipeline for the Inter-American Resources Network of the Inter-American Dialogue on Water Management.
Subscribe: MAILSERV@ACC.FAU.EDU
Post: DIALOG-AGUA-L@ACC.FAU.EDU


FISH-ECOLOGY:        Network for professionals on fisheries ecology and related topics; international computer conference for academic and other personnel to exchange views on and approaches to fisheries ecology issues.
Subscribe: LISTSERV@SEARN.SUNET.SE
Post: FISH-ECOLOGY@SEARN.SUNET.SE


FISHFOLK:        Fisheries Social Science Network. Note: subscription is not automatic, request is forwarded to list owner.
Subscribe: LISTSERV@MITVMA.MIT.EDU
Post: FISHFOLK@MITVMA.MIT.EDU

FISH-JUNIOR:   Marine Science and Children/Youth List; purpose of list is to enable youth to interact with scientists and scientific issues related to fisheries ecology.
Subscribe: LISTSERV@SEARN.SUNET.SE
Post: FISH-JUNIOR@SEARN.SUNET.SE

GLIN-ANNOUNCE:      Great Lakes Information Network Announcements; postings of various announcements with relevance to Great Lakes issues.
Subscribe: GLIN-MAJORDOMO@GREAT-LAKES.NET
Post: GLIN-ANNOUNCE@GREAT-LAKES.NET

GLIN-EDUCATION: Great Lakes Education Forum; discussion of educational issues and materials related to the Great Lakes.
Subscribe: GLIN-MAJORDOMO@GREAT-LAKES.NET
Post: GLIN-EDUCATION@GREAT-LAKES.NET

GLIN-GIS:       Geographic Information System Forum for Great Lakes Resource Managers, Planners and Others.
Subscribe: GLIN-MAJORDOMO@GREAT-LAKES.NET
Post: GLIN-GIS@GREAT-LAKES.NET

GLIN-HEALTH:         Human Health and Toxic Pollution Research; managed by the Great Lakes Human Health Research and Information Exchange Network; up-to-date news, funding opportunities, general information and publication citations on toxic pollution and its effects on human health.
Subscrbe: GLIN-MAJORDOMO@GREAT-LAKES.NET
Post: GLIN-HEALTH@GREAT-LAKES.NET

GLIN-TALK:     Great Lakes Discussion Forum; discussion of Great Lakes' issues.
Subscribe: GLIN-MAJORDOMO@GREAT-LAKES.NET
Post: GLIN-TALK@GREAT-LAKES.NET

GWM-L:          Discussion of Groundwater Modeling
Subscribe: LISTSERV@GWRP.CCIW.CA
Post: GWM-L@GWRP.CCIW.CA

HYDROLOGY:  A mailing list for those interested in hydrology - the science of water in the environment. It is relevant to those outside of traditional engineering hydrology, and emphasises multidisciplinary, fundamental, scientific research in hydrology. The list is intended to have an Australian focus, but those in countries with similar hydrological conditions are also encouraged To subscribe.
Subscribe: MAJORDOMO@ENG.MONASH.EDU.AU
Post: HYDROLOGY@ENG.MONASH.EDU.AU

ICAM-L:          Integrated Coastal Area Management
Subscribe: LISTSERV@IRMFAO01.BITNET
Post: ICAM-L@IRMFAO01.BITNET

IRRIGATION-L:          List for irrigation theory and practice; discussions related to any type of irrigation system; format is typically question-and-answer but postings covering research projects and results, new technologies and products and other topics related to irrigation are welcome; list covers not only technical aspects of irrigation such as design, hydraulics, system layouts but also topics involving irrigation management and socio-cultural, political and economic aspects of irrigation.
Subscribe: LISTSERV@VM.GMD.DE
Post: IRRIGATION-L@VM.GMD.DE

LAKES-L:      Discussion of lake management and ecology; for professionals and others who are concerned with the health of lakes and their watersheds.
Subscribe: MAJORDOMO@BADGER.STATE.WI.US
Post : LAKES-LO@BADGER.STATE.WI.US

MAR-FACIL:    Marine Facilities List; discussion of technical and business topics related to marine research facilities, aquaculture operations, public aquaria and other facilities supplying seawater for the support of marine life.
Subscribe: MAILSERV@AC.DAL.CA
Post: MAR-FACIL@AC.DAL.CA

MARINE-L:     Marine Studies/Shipboard Education Discussion.
Subscribe: LISTSERV@UOGUELPH.BITNET
Post: marine-l@@UOGUELPH.BITNET

MARMAM: Marine Mammals Research and Conservation Discussion.
Subscribe: LISTSERV@UVVM.UVIC.CA
Post: MARMAM@UVVM.UVIC.CA

MEDSEA-L: Marine Biology of the Adriatic Sea. Note: subscription is not automatic, request is forwarded to list owner.
Subscribe: LISTSERV@AEARN.BITNET
Post: MEDSEA-L@AEARN.BITNET

MEH2O-L: Middle East Water List; discussion group for information and research relating to water in the Middle East.
Subscribe: LISTSERV@VM.TAU.AC.IL
Post: MEH2O-L@VM.TAU.AC.IL

RZWQM-L: Root Zone Water Quality Model Discussion List.
Subscribe: RZWQM-L-REQUEST@GPSRV1.GPSR.COLOSTATE.EDU
Post: RZWQM-L-@GPSRV1.GPSR.COLOSTATE.EDU

SALMON: Discussions of the life cycle of salmon and the depletion of salmon runs in western America and other parts of the world.
Subscribe: MAISER@RIVERDALE.K12.OR.US
Post: SALMON@RIVERDALE.K12.OR.US

SEWER-LIST: Sewerage and Waste Treatment Discussion List; for people involved in design,  installation, permitting, operation and maintenance of sanitary sewers, storm sewers, combined systems, lift stations and publicly-owned treatment works; focus is on practical matters rather than academic, theoretical or policy concerns.
Subscribe: LISTPROC@MCFEELEY.CC.UTEXAS.EDU
Post: SEWER-LIST@MCFEELEY.CC.UTEXAS.EDU

SWMM-USERS: Stormwater Management Model Users List; Discussion of stormwater and sanitary sewer design software.
Subscribe: LISTSERV@UOGUELPH.CA
Post: SWMM-USERS@UOGUELPH.CA

TRICKLE-L: Trickle Irrigation Discussion List; Unmoderated discussion of all aspects of trickle or drip irrigation.
Subscribe: LISTSERV@UNL.EDU
Post: TRICKLE-L@UNL.EDU

WATER-L: Water Quality Discussion List; oriented toward water quality issues addressed by cooperative extension.

Subscribe: LISTPROC@LISTPROC.WSU.EDU
Post: WATER-L@LISTPROC.WSU.EDU

WATER-ON-LINE: Water-on-Line Discussion List; this list covers water-related information found on Internet.
Subscribe: LISTPROC@UCDAVIS.EDU
Post: WATER-ON-LINE@UCDAVIS.EDU

*WTEDUC: WaterTalk discussion forum on water resources education issues.
Subscribe: WTlists@uwin.siu.edu
Post:

*WTGIS: WaterTalk discussion forum on GIS and water resources.
Subscribe: WTlists@uwin.siu.edu
POst:

*WTGLOBAL: WaterTalk discussion forum on global water issues.
Subscribe: WTlists@uwin.siu.edu
Post:

*WTGWQ: WaterTalk discussion forum on groundwater quality issues.
Subscribe: WTlists@uwin.siu.edu
Post:

*WTHYDROLOGY: WaterTalk discussion forum on hydrology topics.
Subscribe: WTlists@uwin.siu.edu
Post:

*WTPOLICY: WaterTalk discussion forum on water policy, planning and management issues.
Subscribe: WTlists@uwin.siu.edu
Post:

* These are WaterTalk mailing lists. To subscribe, send the following command in the body of an e-mail message to WTlists@uwin.siu.edu:

<div align="center">subscribe forumname</div>