

PENNPRINTOUT

The [University of Pennsylvania's](#) Online [Computing](#) Magazine



February 1993 - Volume 9:4

[[Printout](#) | [Contents](#) | [Search](#)]

Navigating the Internet: Tools for discovery

By Judy Smith and Daniel Updegrove

Franklin, Penn's online library catalog, PennInfo, our campus-wide information system, and Whois, the online electronic mail directory, are three popular resources on PennNet, the campus data network. Since passwords are not required for these systems, many users are, in fact, affiliates of other universities, government agencies, or industrial firms. Their link to PennNet is the Internet, a network of more than one million computers in forty countries.

Penn students, faculty, and staff, in turn, can access thousands of resources around the world: library catalogs, campus information systems, directories, databases, and archives. The number of institutions joining the Internet, the number of individuals with access, and the number of resources being contributed to this public domain continue to grow rapidly. But how does one navigate in such a vast sea of information?

Until recently, only intrepid researchers and networking gurus understood enough about network addressing, user command interfaces, and technical tricks to use the Internet for more than electronic mail. New navigation tools--easy to use, widely available, and free--have dramatically changed this. Now anyone can, in a matter of minutes, learn to explore from Sweden to Singapore in search of scholarly, technical, or avocational treasure. (See the "Internet hunt," sidebar for examples.)

Five Internet navigation tools--Archie, Gopher, Veronica, WAIS, and World-Wide Web--are introduced in this article, followed by instructions on accessing Gopher, which provides links to the other four tools, as well as to PennInfo. As with PennInfo, these tools are usable via any computer that can emulate a VT100 terminal, but additional power and ease of use are available with versions that operate as clients on Macintoshes and other workstations with IP/ethernet connections.

Archie

The Internet community has been amassing text, image, software, and database resources for over twenty years. Historically, these resources have been stored in public repositories known as anonymous FTP servers. FTP is the Internet-standard high-speed file transfer protocol, used for exchange of private information by trusted parties with passwords as well as for publishing information without passwords, i.e., anonymously.

Hundreds of archives now exist but, up until a year ago, no one tracked them. Archie (ARCHIvE server) was developed at McGill University to index the contents of all FTP servers and provide keyword searching of the index. Its approach is simple but powerful: Every night it re-indexes roughly one thirtieth of the servers; the result is a database that is completely refreshed each month.

Although Archie enables you to locate information, it does not allow you to view or retrieve the information. To do that, you need FTP software on an IP-connected workstation or host (see Penn Printout, March 1992).

WAIS

Wide Area Information System (a joint project of Apple Computer, Dow Jones, KPMG Peat Marwick, and Thinking Machines Corporation) provides a uniform interface to many full-text databases, together with a sophisticated "relevance search" capability. You can search any WAIS database using any word or phrase and the system will return a menu of documents, ordered from more to less relevant. WAIS databases are commonly collections of related data (The Bryn Mawr Classical Review), primary source documents (Clinton speeches), or reference works (CIA World Fact Book, Roget's Thesaurus). There are currently almost 400 WAIS databases, and new ones appear frequently.

Since it can be difficult to determine the focus of a WAIS database from its name, a Directory of Servers, itself a WAIS database, was developed. You can search this directory for topics that interest you, and it will suggest WAIS databases for you to explore. For example, you could search the directory using the keyword "religion," and you would be referred to three WAIS databases: the Book of Mormon, the Qur'an, and

the Bible.

Gopher

Gopher began as the University of Minnesota's version of PennInfo, a menu-driven campus-wide information system (CWIS). Gopher's simplicity as a distributed, client/server CWIS led to its rapid adoption by other institutions, some of which developed new client or server software for desktop or host computers and contributed them to the Gopher software archive (accessible via anonymous FTP, naturally). Soon thereafter, Minnesota offered to provide a menu of all Gopher servers that any other Gopher could access. The result was what networkers have been talking about for years: an interoperating set of information systems linking several hundred organizations around the world, all with a common user interface!

The next step in Gopher's evolution was addition of gateways to FTP, Telnet (the Internet standard remote terminal protocol), Archie, WAIS, and WWW. Gopher was thus transformed from an integrated set of CWIS programs into the most successful Internet navigation tool. But success became problematic: As the worldwide menu structure grew, locating information became increasingly tedious. Something like Archie was needed to help researchers locate information quickly in this new, ever expanding "Gopherspace."

Veronica

In November, 1992 a search tool, Veronica, was contributed to Gopher by a team from University of Nevada at Reno. The original Veronica ("Very Easy, Rodent-Oriented, Net-wide Index to Computerized Archives," a comic acronym if ever there was one) provides a search through all menus using a single keyword. The result is a dynamically created menu of all Gopher resources that contain the keyword in their menus. Now, a second Veronica search tool has appeared--an indexed WAIS database extended to allow Boolean searches of menu documents. Although both searches are limited to words in menus (as opposed to the full text of documents), the combination of Veronica and Gopher results in a powerful capability to search for and retrieve information from all over the Internet, with the location of the information effectively irrelevant.

World-wide Web (WWW)

WWW was developed as a hypertext system at the Center for Nuclear Energy Research (CERN) in Geneva. It allows links with and between WWW documents and, like Gopher, provides access to other Internet resources and navigation tools. Although much admired by many in the Internet community for the elegance of its design, WWW has not proliferated as has Gopher, in part because WWW services are more complex to create and maintain, and in part because security restrictions at CERN restricted Internet access.

Access to Gopher

From the PennNet annex: prompt, issue the command telnet gopher and you will be presented with a main menu including local and remote gopher servers, the Gopher-PennInfo gateway (that is, PennInfo menus and documents accessible via the Gopher user interface), as well as Archie, Veronica, WAIS, and WWW. Alternatively, use the "worldwide" command from within PennInfo. It should be noted that these navigation tools currently qualify for only "best effort" support from Data Communications and Computing Services (DCCS); no formal training, local documentation, or CRC help is planned. DCCS seeks feedback about the use of these tools; send comments to Al D'Souza, Director of Program Managemet, at 898-2429 or via e-mail to dsouza@pobox.upenn.edu.

To obtain Gopher client (or server) software, FTP to <ftp.upenn.edu> or <boombox.micro.umn.edu>. Log in as "anonymous," use "guest" as your password, and change your directory to `pub/gopher`. Gopher clients may be set to "point to" the server at Minnesota; you are encour-aged to reset your client to point to <gopher.upenn.edu>.

To obtain Archie, WAIS, or WWW client software for an IP/ethernet- connected computer, simply use Veronica and Gopher. (Alternatively, search PennInfo by the keyword "navigation" to determine FTP archive addresses, and then use FTP.) From a WWW client, point to <http://www-penninfo.upenn.edu:1962/> for access to PennInfo.

Home again

The joint development effort, client/server paradigm, and

no-cost distribution of these navigation tools is characteristic of the Internet. Consider:

- In 1991 Penn "imported" software developed at MIT for TechInfo, made minor modifications, and introduced PennInfo.
- A few months ago, MIT, in turn, installed the Gopher gateway to TechInfo, developed by Linda Murphy of DCCS.
- Penn's recommended communications software for IP-connected Macintoshes is NCSA Telnet, developed and distributed by the University of Illinois' National Center for Supercomputing Applications.
- Fetch, an elegant Macintosh implementation of FTP developed at Dartmouth, is now supported here.
- Eudora, a Mac-based electronic mail program from the University of Illinois, is being used in several offices at Penn.

Also characteristic of the Internet are a strong international flavor, creative programmers ("hackers" in the best sense), droll humor, and a growing number of enthusiastic and productive scholars navigating the net--without leaving their desks.

Sidebar 1: Internet Hunt

Below is a sampling of questions from the December and January Internet Hunts. This monthly treasure hunt was created by Rick Gates, Director of Library Automation at the University of California at Santa Barbara. For more information use the keywords "internet hunt" to search PennInfo.

- What is the atomic weight of boron?
- I'm trying to find a new book on the Internet by an author named Krol. Are there any local bookstores that might carry this? I live in Albuquerque, New Mexico.
- Early last month, U.S. president-elect Bill Clinton proposed a new technology policy. Where can I find the text of this proposed policy?
- I'm going to be in Denver, Colorado on the nights of Jan 22-25. Will the Denver Nuggets basketball team be playing at home on any of those nights?

- What was the total amount of sales in liquor stores in the United States in September of this year? Was this more than last year?
- I'm volunteering some time for a local hiking association. I'd like to know if anything's been written on the development of trails for the handicapped?
- I'm going to London next February. Is there a place that I can ask about some of the different pubs that might help take the chill away?
- Where is the ACM's SIGGRAPH '93 Conference being held next August?
- How does one say "Merry Christmas and a Happy New Year" in Czech?
- Is the Toyota Motor Corporation connected to the Internet?
- I read in an electronic journal somewhere that a conference was held in Padova, Italy on models of musical signals. I wrote down the name of a contact, "Giovanni De Poli." Can you find his e-mail address for me?
- What is the primary religion in Somalia?
- I understand that the Net is being put to use distributing information and pictures of missing children. Where can I find out more, and where can I find the pictures?
- Where can I find tables listing the nutritive values of different foods?
- What is the text of the 1st Amendment to the Constitution of the United States?

Sidebar 2: Further Reading

Any one of the following books on the Internet is recommended for further reading:

- Kehoe, Brendan P. (1993) *Zen and the Art of the Internet: A Beginner's Guide to the Internet*, (second edition). Prentice Hall, Englewood Cliffs, NJ. (\$22.00) The first edition is also available via FTP at [ftp.upenn.edu](ftp://ftp.upenn.edu) in the directory `pub/DCCS`.
- Krol, Ed. (1992) *The Whole Internet User's Guide and Catalog*. O'Reilly & Assoc., Inc. Sebastopol, CA.

(\$24.95)

- LaQuey, Tracy, and Jeanne C. Ryer. (1992) *The Internet Companion: A Beginner's Guide to Global Networking*. Addison-Wesley, Reading, MA. (\$10.95)
- Tennant, Roy; John Ober; and Anne G. Lipon. (1993) *"Crossing the Internet Threshold: an instructional handbook*. Library Solutions Press. (\$45.00)

Also note that PennInfo contains detailed information about these navigation tools--see the "Internet Navigation" topic in the "Computing" menu or use a keyword search to find information.

JUDY SMITH is a Technical Writer for ISC Communications Group. DANIEL UPDEGROVE, Associate Vice Provost for Information Systems and Computing, is Executive Director of DCCS.
