Navigating the Internet*

By Seth M. Powsner, M.D.
Associate Professor, Department of Psychiatry
Member, Center for Medical Informatics

Nancy K. Roderer, M.L.S.
Director, Harvey Cushing/John Hay Whitney Medical Library
Member, Center for Medical Informatics

Yale University School of Medicine
New Haven, Connecticut 06510

Navigating any complex set of information resources requires tools for both browsing and searching. A number of tools are available today for using Internet resources, and more are being developed. This article reviews existing navigational tools, including two developed at the Yale University School of Medicine, and points out their strengths and weaknesses. A major shortcoming of the present Internet navigation methods is the lack of controlled descriptions of the available resources. As a result, navigating the Internet is very difficult.

Trying to find a reference, text file, data file, or anything else on the Internet is much like trying to find an item in a large department store. The item may be available, perhaps even a bargain, but where is it? In department stores, you search for something specific by asking a clerk or, if you are not sure exactly what you need, by browsing. In libraries, you either can search for a specific item in the catalog or browse the stacks. But the Internet, unlike libraries or department stores, was designed for neither searching nor browsing, and at this stage both are difficult. This article contrasts library and Internet organization to provide a context for a review of approaches to navigating the Internet. The focus is not on technical details but rather on support for conventional searching and browsing.

BACKGROUND

It is taken for granted that libraries arrange their holdings to facilitate browsing. Dewey Decimal and Library of Congress numbering systems keep all the holdings on infectious disease on the same shelf. In department stores, less formal rules keep the bowls, knives, and tea kettles in the same vicinity. Other arrangements are possible; indeed, practical considerations do influence arrangements. Oversized books are stored together in libraries. Rapid turnover in periodicals creates pressure to put them in separate periodical rooms. However, organization by subject is important for modern hunting and gathering [1–2].

The Internet is not "arranged" in the usual sense of the word. It evolved along communication lines from the ARPANET of the late 1960s. ARPANET was an experiment in distributed communications, with distributed control meant to survive nuclear attack; it was not constructed as an information repository [3].

Internet information is not "selected" in the usual sense of the word, in the way that a librarian would choose the best and most relevant from among available references to add to the library's collection. In the 1970s, programmers working on experimental computer networks tended to exchange software files. There was a wider range of discourse once e-mail was delivered reliably. By the early 1980s, individual e-mail exchanges had evolved into news groups, later listservs, like electronic reincarnations of the colonial circular letters spreading news and ideas up and down the Boston Post Road. Improvements in disk storage during the 1980s then made it reasonable to keep file archives continuously available on the Internet. Archive contents were dictated by the computer center that provided disk space.

An individual now can offer his or her own Internet file archives and databases. Many universities, along with some government and private organiza-

* This work was supported in part by NIH Contract N01 LM13537 from the National Library of Medicine.

tions, are well enough connected for staff members with spare disk space and computer processing cycles to publish an Internet database strictly on their own initiative. Communications technology and basic file-transfer protocols are widely available (and often included with Unix-based computers). Such personal databases appear on the Internet alongside the offerings of traditional computer information service providers, such as DIALOG Information Services or the National Library of Medicine (NLM).

Duplicated files are common. Internet band width far exceeds available bits of scholarly information. One Internet link can transmit more than three million journal articles a year, assuming a common 56,000-bit-per-second link and thirty double-spaced pages per article (50,000 characters per article, six articles per minute). Few institutions boast three million original publications a year. Reruns are inevitable.

SOFTWARE FOR NAVIGATING THE INTERNET

Software for Internet access and navigation permits some conventional searching and browsing. The following sections review commonly available Internet software that runs on almost any desktop computer, whether a Macintosh or a PC-compatible running Windows, that has an Internet connection. Detailed technical descriptions are available from many other authors, of whom we cite only a few [4–9].

The focus here is on conventional, not conceptual, searching and browsing [10]. In the future, increasing numbers of library patrons may want to truly navigate information sources. The current experience of the authors is that patrons expect to be able to search an index or browse a table of contents.

Telnet

Telnet provides teletype-style communications to other computers connected to the Internet as an alternative to telephone communication using modems. Telnet really offers video-display-terminal–style communication. An Internet address is entered rather than a phone number. Software makes the connection through intermediate computers along the network. Typing and printing may be slow or fast, depending on network data loads. However, having telnet access is no more useful for searching or browsing than having library borrowing privileges: both are useful only to users who already know where to find what they want.

Online public access catalogs

Libraries around the world are allowing online public use of their card catalogs. Telnet provides adequate access for most of these catalogs. Online public access catalog (OPAC) software running on a library computer or a computer nearby accepts telnet connections and prompts the user to enter an author's name, a book title, or a subject keyword. In return, OPAC software sends back a list of holdings.

OPAC software does not help users search or browse other Internet sources, although Internet access to OPAC is a boon for distant and local library patrons alike. Patrons must know in advance which library's OPAC to check and its Internet address. Then they can connect and search or browse.

Client-server software

Advanced Internet software packages are separated into "client" and "server" components. Server software usually runs on medium- or large-scale computers to service requests for information. Server software is distinguished from traditional online retrieval programs by its inability to respond directly to requests from terminals. Servers only respond to properly packaged requests arriving over the network. Users must run suitable client software on their own computers to submit proper requests.

Client software usually is customized to run on a small desktop computer and to take advantage of specific computer features, such as graphical interfaces, fast color displays, and disk storage. Client software absorbs the burden of conversational user interaction, relaying only fully formed requests to the server software. GRATEFUL MED can be considered client software; however, it usually is connected over a telephone circuit or telnet to MEDLARS, which does respond to traditional terminals.

The success of client-server–style software stems from the combination of responsive personal-computer-system interaction with large-computer-system breadth of retrieval. Simpler client programs may handle terminal or telnet requests from users without appropriately equipped personal computers.

File-transfer protocol and Archie

File-transfer protocol (FTP) facilitates file retrieval from file archive (i.e., server) sites. Individual text files, data files, or even software can be moved quickly between Internet FTP clients and servers. Speeds of 5,000 to 10,000 bytes per second to sites halfway around the world are common (varying with network traffic loads and specific communication links). If the authors added the text of this article to an FTP file system, a reader at the University of Queensland in Australia could retrieve the text in less than a minute. FTP actually can handle megabytes of information.

Unfortunately, FTP, like telnet, requires that the user know the Internet address, file name, and file type (e.g., text, binary data, program) they want to retrieve. Knowing an FTP server system's address may

permit browsing of file directories; no searching or browsing capability is built in.

Archie software, developed at McGill University, provides some search capability for international FTP file archives. Archie server software periodically checks the holdings in a list of FTP archives and constructs an index. (Index updates can be done at night to reduce Internet traffic.) A user can contact an Archie server system and request a list of files matching some specification. A list of matches, including the address of the FTP archive sites holding the files, is returned. FTP software then is used to retrieve individual files.

Archie can offer only limited browsing. It primarily indexes by file name. There has been some attempt to catalogue FTP archive holdings by file content; however, file content coding requires voluntary manual work that is not essential for basic FTP operation.

Wide Area Information Service

Wide Area Information Service (WAIS) goes a few steps further than Archie and FTP by indexing the content of text files stored on a networked computer. WAIS assumes a collection of articles or reports stored on hard disk. WAIS indexing software reads through the files, creating a word index. WAIS server software then uses the index to answer search requests.

WAIS client software communicates a user's search request over the network to a number of WAIS servers. Each server sends back a response for inclusion in a more complete list than one server alone would provide. WAIS client software then handles specific file requests to specific WAIS servers, much like an FTP interaction, making the file available for review.

WAIS permits searching and browsing something like a subject card catalog does. Automated word indexing lacks the discipline of a professionally maintained catalog, but in fields where terminology is consistent, a search may yield a reasonable batch of files to "thumb through."

Gopher and Veronica

Gopher server and client software takes advantage of hierarchical hard disk file storage to provide outline-style access to files. This software also takes advantage of Internet connections to allow the outline to be maintained and stored by different people using different computers. Gopher originally was designed to handle campus bulletins and directories at the University of Minnesota. Its use has spread to many other campuses and other institutions. Advanced versions of the software can handle many types of files, including graphics, and connect to other types of servers (e.g., FTP, WAIS).

Cross-references can be entered into a gopher server to direct an individual automatically to a server outside the institution. A campus library gopher server might direct a student to a medical center library computer for holdings in the history of medicine and hours of operation.

Veronica software was developed at the University of Nevada to provide a networkwide index of gopher entries, just as Archie provides an index of FTP offerings. Like Archie, Veronica indexes entry names.

Gopher and Veronica are more conducive to browsing and searching than are FTP and Archie. Gopher's outline presentation offers information in a form reasonably like a table of contents. Unfortunately, browsing efficiency is compromised by the absence of any overall organization. Browsing efficiency is compromised further by duplicate intergopher links, which repeatedly lead the user to the same basic data. Veronica searches are helpful but have the same drawbacks as Archie searches.

World Wide Web and Mosaic

World Wide Web (WWW) software offers hypertext documents that pull together information from sources around the world. Instead of merely citing supporting references, WWW authors can include active links directing the reader to the WWW server holding the supporting information. Mosaic extends WWW documents to allow for use of graphics and forms and to permit access to other types of servers, such as FTP and gopher. The combination of graphics and free-form documents, as opposed to a strict outline approach, has made Mosaic very popular.

Mosaic documents offer very interesting possibilities for scholarly or business publication. Practical sound and video software add-ons would offer truly electronic publication of multimedia documents. The challenge is to provide sufficiently powerful viewing (client) computers with graphical user interfaces and network connections of adequate speed.

WWW and Mosaic software systems alone do not facilitate browsing and searching. The user must know in advance where to find a document of interest. However, WAIS access is possible from a Mosaic document, and, if WWW and Mosaic documents are indexed reliably by WAIS servers, the combination of these software packages would provide reasonable network browsing and searching.

Yale UMLS network navigation software

NetMenu and the Information Sources Directory (ISD) were developed at Yale University's Center for Medical Informatics (YCMI) as part of NLM's Unified Medical Language System (UMLS) project [11-12]. NetMenu permits network browsing. ISD provides network searching.

NetMenu. NetMenu permits reference-room-style browsing. It permits access to a carefully reviewed selection of popular network resources, such as the MEDLINE and Current Contents databases, and to services such as e-mail. NetMenu was designed primarily for quick access to the most commonly used information sources. Although prototyped in HyperCard for Macs, a portable implementation runs on Macs and PC-compatibles, making possible a uniform front end on library, hospital, and desktop machines. More than seventy machines make approximately 16,000 network connections per month.

NetMenu was designed as a simple, two-level menu system for network access by inexperienced users. The first menu groups selections into broad categories as shown in Figure 1, and the second level offers specific choices. Descriptions of specific choices are available by clicking the "information" button. Clicking the "connect" button launches the appropriate communication program or other application. The ability to access resources quickly without any Internet knowledge makes browsing possible. A number of other institutions, such as the University of Colorado, have developed menu systems similar to NetMenu [13].

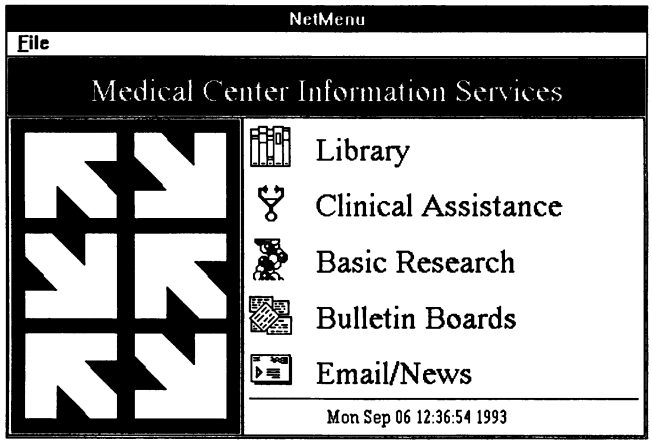
NetMenu offerings are reviewed by a committee run by medical library staff, with physicians and network personnel as members. NetMenu configuration files and network connection scripts are maintained centrally under library control. Centralized authority and maintenance responsibility lead to a reference-room-style collection of network resources.

Centralized configuration and script maintenance has been critical to ensuring routine network use. Network connection failures are indistinguishable from NetMenu program failures from a user's point of view. Changes in network address or log-on protocol frustrate users just as much as a back hoe cutting the network cable. However, with NetMenu configuration and communication script files on one file server, it is possible for one staff member to isolate the cause of the failure and fix it for all.

Information Sources Map/Directory. The NLM's UMLS included the development of an Information Sources Map (ISM) to guide biomedical researchers and clinicians towards networked information sources. YCMI produced the ISD as part of the ISM development [14]. The ISD is available for networked PC-compatible computers on the Yale campus and is being adapted for use at another university.

A database of information sources is at the heart of the ISD server. The ISD database is built from the ISM files included with NLM's UMLS Knowledge Sources CD-ROM. The content of each network information source is indexed by subject, using Medical Subject Heading (MeSH) terms, by type (e.g., bibliographic citations, textual data, organizational addresses) and by typical use (e.g., research or clinical care). Campus information sources are added to the server as is information on access. Information sources not found on any network can be included if desired.

Figure 1
Initial screen of Yale NetMenu



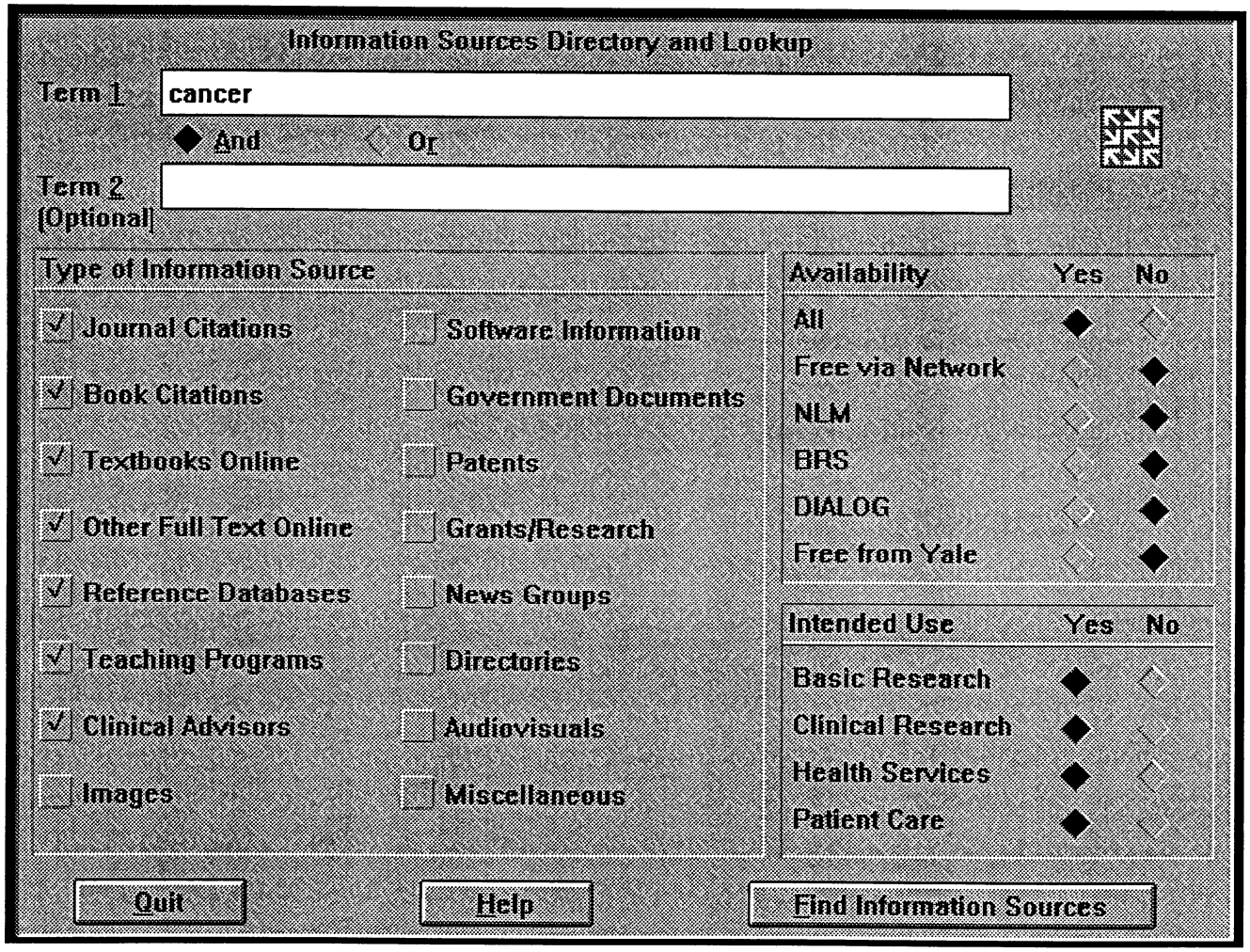
The ISD is designed to allow for both browsing and searching. It accepts one or two biomedical terms indicating the subject of interest and allows the user to check boxes indicating what sort of databases or resource materials might be of interest and whether the results should be limited to local or free resources (Figure 2). Non-MeSH terms are looked up in the UMLS Metathesaurus. The searcher is offered a list of related terms as alternatives.

Based on the subject terms and the user's indication of materials of interest, a list of information sources is presented to the searcher. The list is ranked from "highly likely to be relevant" to "possibly relevant" (Figure 3). The list can be restricted to only those most likely to be relevant or can be reorganized according to other factors (e.g., access restrictions, type of research). The searcher can browse the list and obtain more detailed descriptions of specific databases.

The user selects one database and clicks the "connect" button. Like NetMenu, the ISD includes communication scripts to make network connections. (The user is prompted for a password if one will be needed.) The user then interacts directly with the database system; the ISD does not attempt to formulate or translate search requests.

The Yale ISD is implemented as a client communications script running on a PC-compatible computer, which interacts over the network with a server program running on a Sun workstation. The PC communication program (Dyna Comm for Windows) does double duty by providing a graphical client for the

Figure 2
Yale Information Sources Directory search screen



ISD server and also providing terminal emulation for connection to a biomedical database.

DISCUSSION

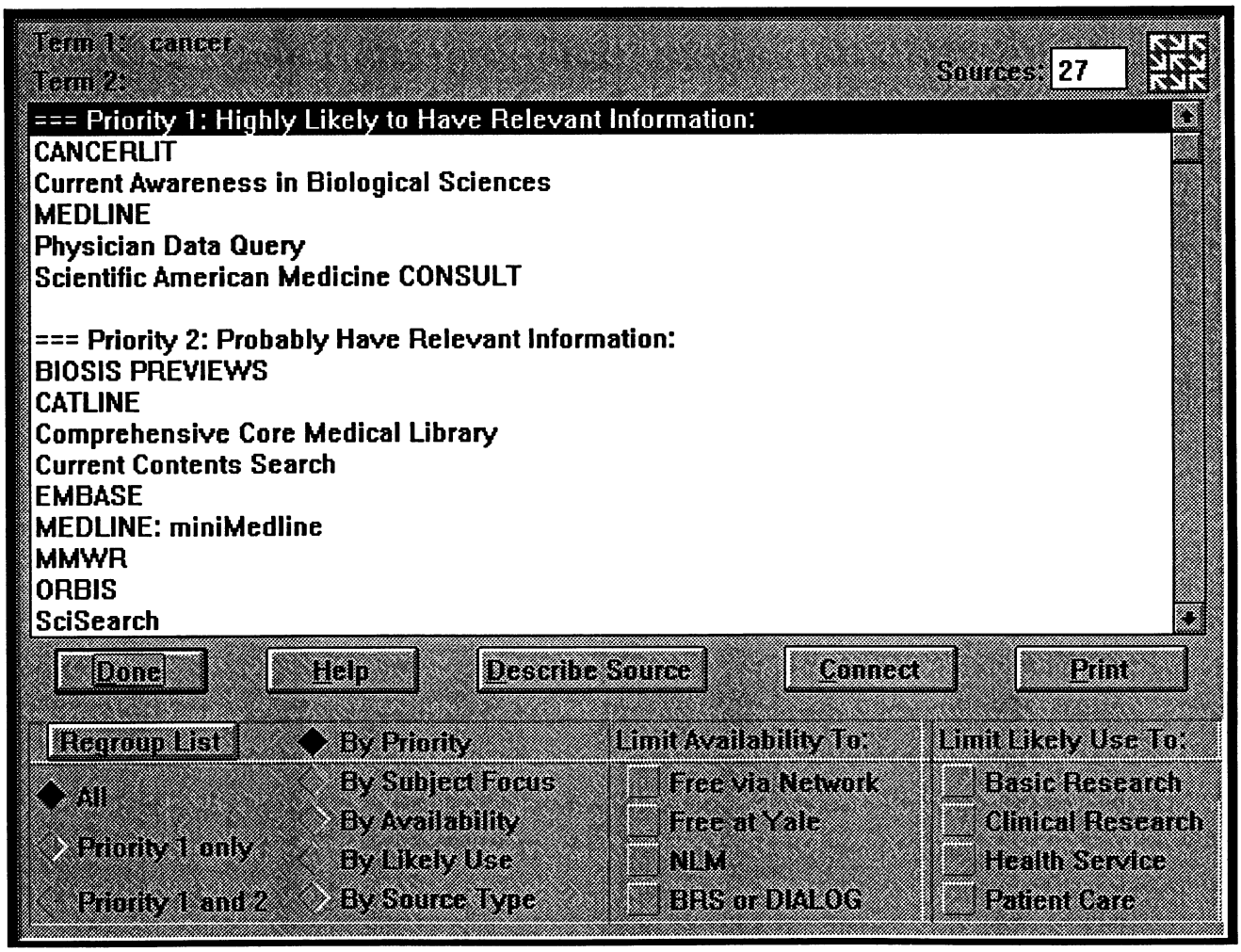
Only WAIS software offers something akin to conventional browsing through Internet information. Gopher software provides its own form of network browsing, which has a certain appeal. NetMenu offers something akin to browsing a reference room. ISD offers browsing as a means of narrowing a request.

Archie, ISD, Veronica, and WAIS software all accept search requests and return references to Internet resources. Archie handles only file names automatically, and Veronica handles only Gopher menu entries, making searching entirely dependent on the precision of an author-assigned file name. The ISD depends on MEDLINE MeSH headings meant for indexing individual citations to categorize entire databases. WAIS indexes contents of files by text words. Each approach has distinct limitations and may or may not be understood easily by library patrons.

None of these systems is exhaustive; there is no requirement that useful Internet information be registered with any searching or browsing system. For now, an exhaustive attempt to find information on the Internet must include perusal of news groups and personal contacts by e-mail. The ISD does use a centrally controlled list of resources, so there is one place to register information or complaints.

The distributed, decentralized nature of the Internet is one of its main strengths; however, these characteristics are a fundamental problem when searching for a reference source. There is no oversight.

Figure 3
Yale Information Sources Directory results screen



Information on the Internet is not peer reviewed or even grouped by subject. There may be many versions of a particular document, including some with errors.

Internet browsing and searching would be facilitated by features found in the scholarly journal system: peer review, indexing, and editorial commentary. Network data replication requires descriptions that distinguish truly unique information from different routes to reorganized forms of the same information. Procedures mapping out access routes to online databases would be helpful.

CONCLUSION

What should it mean to "search" or "browse" the Internet? Must users navigate the Internet the way *les voyageurs* navigated the waterways of North America? Science fiction author William Gibson imagined network browsing with the help of a personal computer and direct brain electrodes serving as the user interface. This arrangement would allow survey flights over an information landscape for which he coined the name *cyberspace*. One of Gibson's characters reenters cyberspace:

> Disk beginning to rotate, faster, becoming a sphere of paler gray. Expanding—And flowed, flowered for him, fluid neon origami trick, the unfolding of his distanceless home, his country, transparent 3D chessboard extending to infinity. Inner eye opening to the stepped scarlet pyramid of the Eastern Seaboard Fission Authority burning beyond the green cubes of Mitsubishi Bank of America, and high and very far away he saw the spiral arms of military systems, forever beyond his reach. . . . Molly was gone when he took

the trodes off, and the loft was dark. He checked the time. He'd been in cyberspace for five hours [15].

Present-day Internet access is not as engaging as Gibson's cyberspace, but neither does it require surgically implanted electrodes. Any desktop computer, a Macintosh or PC-compatible running Windows, can run the programs described in this article. The Internet, soon to become "the Information Superhighway" or "Infobahn" or simply "I-way" into cyberspace [16], beckons library users and librarians pursuing scholarly pots of gold at the end of some telecommunication link. Very few maps are available, and many files that look like pots are just spittoons or pot-shaped signs marked "detour."

ACKNOWLEDGMENTS

NetMenu and the ISD are primarily the work of colleagues in the Center for Medical Informatics and the Harvey Cushing/John Hay Whitney Medical Library: Kei H. Cheung, Jeffrey I. Clyman, Sandra Frawley, Jan Glover, Majlen Helenius, Naomi Ikeda, Xia Liu, Perry L. Miller, Kenney Marone, John A. Paton, Mark A. Shifman, and Lawrence W. Wright.

REFERENCES

1. MARCHIONINI G, SHNEIDERMAN B. Finding facts and browsing knowledge in hypertext systems. Computer 1988 Jan;21(1):70–80.
2. SHNEIDERMAN B. Information exploration tools. In: Designing the user interface: strategies for effective human-computer interaction. 2d ed. Reading, MA: Addison-Wesley, 1992:395–437.
3. KANTOR A. Internet: the undiscovered country. PC Magazine 1994 Mar 15;13(5):116–8.
4. FRISSE ME, KELLY EA, METCALFE ES. An Internet primer: resources and responsibilities. Acad Med 1994 Jan;69(1): 20–4.
5. GIBBS M, SMITH R. Navigating the Internet. Carmel, IN: Sams Publishing, 1993.
6. GILSTER P. The Internet navigator. New York: Wiley, 1993.
7. METCALFE ES, FRISSE ME, HASSAN SW, SCHNASE JL. Academic networks: Mosaic and World Wide Web. Acad Med 1994 Apr;69(4):270–3.
8. OBRACZKA K, DANZIG PB. LIS: Internet resource discovery services. Computer 1993 Sep;26(9):8–22.
9. SCHWARTZ MF. Internet resource discovery at the University of Colorado. Computer 1993 Sep;26(9):25–35.
10. CIMINO JJ, ELKIN PL, BARNETT GO. As we may think: the concept space and medical hypertext. Comput Biomed Res 1992 Jun;25(3):238–63.
11. CLYMAN JI, POWSNER SM, PATON JA, MILLER PL. Using a network menu and the UMLS Information Sources Map to facilitate access to online reference material. Bull Med Libr Assoc 1993 Apr;81(2):207–16.
12. SHIFMAN MA, CLYMAN JI, PATON JA, POWSNER SM ET AL. NetMenu experience in the implementation of an institutional menu of information sources. In: Safran C, ed. Seventeenth Annual Symposium on Computer Applications in Medical Care: patient-centered computing. New York: McGraw-Hill, 1994:554–8.
13. SCHWARTZ, op. cit.
14. CLYMAN, op. cit.
15. GIBSON W. Neuromancer. New York: Berkeley, 1984:52.
16. GLEICK J. The information future out of control. New York Times Magazine 1994 May 1:54–7.

Received March 1994; accepted May 1994