

12.0 Navigational Enhancements

Gopher provides an effective mechanism to allow an administrator to provide a hierarchical browsing list, for a campus, for the Internet, or both. But the organizational scheme carefully designed during months of deliberation by a campus menu design team may bear no relation whatsoever to the organization of a given user's brain. Even when the organizational scheme is suited to the user, the process of searching for information across many Gophers can be tedious. Recent developments have addressed both concerns.

12.1 "Road Map" and TS/TB

In July 1992, Dennis Boone of Michigan State University implemented a "Road Map"--a representation of an entire Gopher's directory structure in a single text document. This allows users to read the map and get a feel for the lay of the land. Also that month, Boone implemented a Gopher title search tool. Known as "ts/tb," the title search allows the administrator to prepare keyword search indexes for the entire Gopher tree or for any part thereof. The user wanting to find the schedule for the campus opera society can do a keyword search on "opera," and all titles including that word will be served up as a menu. The title search tool has been adopted for use at a number of sites. One shortcoming is that document titles are pulled up from wherever they appear in the menu hierarchy, and they are presented without putting them in the context of the menu hierarchy. However, this usually does not lead to confusion.

12.2 Veronica

In November 1992, Steven Foster and Fred Barrie of the University of Nevada announced a service that could search document titles across many Gopher servers. Their original code was a modification of Boone's title search. Where Boone's index deliberately limits itself to one Gopher server, they devised a mechanism to provide a single title index that could span many Gopher servers. The Nevada team christened their tool Veronica, which they say is an acronym for "Very Easy Rodent-Oriented Netwide Index to Computerized Archives." Just as Archie surveys anonymous FTP servers for their list of holdings, Veronica must go out and poll its target servers periodically in order to build a database. A user connecting to Veronica specifies a keyword to search for, say for the word "biology," and gets a list of document titles from throughout Gopherspace that match.

At this point, Veronica is not without its problems. The most noticeable problem is performance. As of this writing, there are two Veronica servers for the entire Internet and these servers are often quite busy. It can take several minutes to get a response to a Veronica query at busy times. Of course, if the search eliminates many minutes (or hours) of manual browsing, a wait of several minutes may well be worth it. Steve Foster observes that users have told him they are willing to wait for the results of a search of all of Gopherspace, assuming the results are accurate.

Another concern is discerning the meaning of titles listed as the result of a Veronica search. As with the single-Gopher title search, the document titles are shown stripped of their hierarchical context. Within one Gopher server this is not necessarily daunting. Do a Veronica search on "presidential," however, and you'll find items ranging from the 1992 Presidential debates to presidential searches at various campuses. The labels may not provide enough context to make clear the location or purpose of a given document. The user ends up having to select unwanted documents just to determine what they contain, and even then there may not be sufficient context within the document. Initially, Veronica did not resolve multiple references to a given document down to one listed item. Popular items are pointed to by many Gophers, so a Veronica search might yield 100 pointers to the same document. The Veronica developers modified the software in December 1992 to fix this problem: all references to a particular selector string/host are collapsed to one listing. This does not solve the problem of sites that make separate local copies of documents they find on the network; each such copy is indexed separately by Veronica. The only hope for solving that problem is widespread adoption of a Uniform Resource Number scheme. (The Internet Engineering Task Force is working to define such a numbering scheme, which, with companion Uniform Resource Locators, would allow for agreed-upon ways to number and to locate Internet documents and services.)

Finally, Veronica is not currently consistent in the search results it delivers to the user. Part of the problem is that Veronica, like most Gopher indexes, relies on the public domain WAIS search code, which returns a "relevance score" rather than an absolute answer as to whether something was found. At one point, the two existing Veronica servers yielded differing results for the same searches. Inconsistent results may also arise from the fact that Veronica may be unable to do its periodic survey of a given server. One proposal for addressing this problem calls for a new Gopher document type, the Veronica index, which would be delivered to a polling Veronica server in one operation. This would be a much simpler, more efficient transaction than having the Veronica server manually traverse the entire directory tree of each Gopher server whose titles it wants to survey.

+ Page 38 +

Despite these shortcomings, Veronica represents an exciting development. Already users are finding it far more efficient than browsing innumerable menus when searching for specific resources in Gopherspace. Proof comes from the results of recent Internet Hunts. These monthly quizzes, issued by Rick Gates of the University of California at Santa Barbara, challenge users to find out what information servers on the Net can answer particularly arcane questions. (For example, "Who is the author of the only book held by Victoria University of Wellington on the training of sheep dogs?") Increasingly, the winners of the Hunt have exploited the power of Veronica to enhance their burrowing through Gopherspace.

Many Gopher servers now provide a pointer to Veronica under "Other Gophers" or an equivalent folder. (The Nevada server is located at veronica.scs.unr.edu, port 70; use the selector string "1/veronica" to access the server.) The Veronica developers are exploring setting up several servers around the Internet, to balance out load factors. With the number of Gopher servers

growing daily, the list of servers in production has become a management issue in its own right--there are currently over 1,100 servers. Recently the staff at the University of Minnesota reorganized the "list of all Gopher servers in North America," breaking it up by state. This makes browsing easier for some purposes, but if you are looking for, say, Johns Hopkins University, and you do not know it is located in Maryland, you may want a different path. (Indeed, for most purposes, the geographical location of the server is irrelevant to the user's query.) Dennis Boone has implemented a global keyword index of Gopher site names to allow a user to quickly find a known site regardless of where it is located. (This service is available as "7/internet/others/ts" on the gopher.msu.edu server. Gopher administrators may want to add pointers to this index.)

12.3 Jughead

In March 1993, Rhett Jones of the University of Utah announced a Gopher title search tool called "Jughead." Jughead can index a single Gopher or all of Gopherspace, as Veronica does.

John Doyle of Washington and Lee University has used Jughead to create several indexes of Gopherspace by document type (e.g., Telnet session, CSO phone book, and index). His Jughead server is found at liberty.uc.wlu.edu, port 3002.

13.0 Quality of Service Concerns

Gopher has come a very long way in a very short time. Because servers have been deployed around the Net for only about sixteen months, many services are still experimental. A user may stumble upon a document that sounds just right using Veronica, only to discover in dismay that the needed server is consistently down.

+ Page 39 +

To some extent, this is inherent in networked information. And perhaps users should learn to expect to find resources that are unavailable from time to time. As Nancy John of the University of Illinois at Chicago has observed, patrons of libraries have long been used to the idea of locating a particular book they want, only to find that book has been checked out by another patron. [6]

Still, administrators confront choices that may affect reliability of service. The Gopher model encourages distributed deployment of services across many servers. It is tempting to place documents on a wide variety of Gopher servers, across a campus and even within departments. But as Joel Cooper of Notre Dame University points out, some forms of information are vital "corporate data"--data that must be reliably provided for the good of the enterprise. [7] For instance, Gopher presents real opportunities for saving trees by replacing printed matter. But if the phone book server is located in a locked office that is inaccessible at night, who will reboot it when it crashes? Gopher administrators may well want to put documents that are vital to their institutions' business on a central server, perhaps in an area staffed by full-time computer operators. By using automated document posting processes (either via e-mail or FTP), the administrator can make posting easy for information providers, while utilizing carefully managed servers.

Even the best maintained server will experience outages. Gopher clients determine that a server is down simply by "time-

out" criteria--they wait for a predetermined period of time and display an error message if the Gopher server doesn't respond. Some clients allow interruption of a query if the user does not want to wait for the entire time-out interval. It would be desirable if all clients allowed the user to configure the time-out interval.

The Gopher protocol provides a mechanism whereby a server can deliver two selector strings for a given document. In this way, a site could offer a primary and an alternate server for critical information. So far, this feature of Gopher has not been widely exploited, but it is worth noting that the University of Minnesota now runs a server that replicates the documents offered at Gopher Central. (The backup is gopher2.tc.umn.edu, port 70.)

Prudent Gopher administrators will also provide for regular backup of the server software and data files. Any production data resource demands regularly scheduled incremental and full backups. Again, centrally located servers may be in the best position to be backed up, perhaps by staff who already perform such backups for other production hosts on a daily basis.

+ Page 40 +

Reliability of service is one issue; quality of content is another. In many cases, Gopher administrators view themselves as publishers, and they rely on the original providers of documents to provide high-quality information. Of course, the quality of documents delivered can vary wildly. Administrators can try to set standards for quality of content and for currency of information, but these can be hard to enforce. In some cases, removal of documents, while a blunt instrument, is the only answer to this problem.

Again, the ability of Gopher to point to other documents becomes an issue. Just about anyone can declare an electronic serial to be an e-journal. By what standard do we decide whether to point to it in a Gopher? Include all journals? Those that manage to produce two issues? Those with ISSNs? Those that appear to be "scholarly"? Those that are peer reviewed? Over time, librarians will probably have to apply the same sort of collection development procedures to online information as they do for their print holdings. Although most e-journals are currently delivered without a fee, there are costs to including substandard material--the human and machine costs of mounting the material as well as the time the patron spends avoiding the chaff.

Librarians face the interesting question of deciding whether to list networked information resources in general, and Gopher-based services in particular, in their online catalogs. Already some libraries have included selected e-journals in their online catalogs. With the proliferation of networked resources, each library undertaking to catalog the Internet faces the same issues of what ought to be included in the virtual collection that confront Gopher administrators. Moreover, the effort involved in preparing a catalog record exceeds by a considerable margin the cost of adding a Gopher link. In fact, Martin Dillon, Director of Research at OCLC, has proposed a four-tier approach to cataloging Internet resources, with the effort expended proportional to the scholarly value of the item being cataloged. In brief, these levels are:

1. Cataloging is equivalent to that used for scholarly

materials today.

2. "Brief record" cataloging as used by libraries for materials not of the highest rank.

3. Cataloging by automated means, supplemented by some human editing.

4. Catalog record is supplied by item's creator and collected automatically. [8]

+ Page 41 +

Should libraries undertake to catalog networked resources, they may wish to adopt a scheme such as Dillon proposes. Moreover, they may decide to use consortia and specialization so that not every library takes on the task of cataloging the entire Internet.

14.0 Privacy and Security Issues

Like many networked information servers, Gopher servers maintain logs of what documents have been accessed at what time from what computer. These log records contain the Internet address of the originating user's computer. These days, many workstations have only one regular user. Therefore, an Internet address can correspond to a single human being. Just as libraries have long had policies to avoid tracking information about an individual patron's reading habits, Gopher administrators need to take steps to ensure that users' privacy is not violated. Some institutions have issued formal Gopher privacy policies. Two of these institutions are Rice University and Michigan State. (At the latter, logs are periodically "sanitized"--information is summarized so that individual accesses are no longer discernible.)

Some applications in Gopher could best be handled given a scripting facility. Such a tool could allow a server to cause a program to execute on the user's workstation. A very simple example would be to have the workstation's Telnet program type essential login information after a session is opened (e.g., a command all users must type to connect to a service on a given host; for instance, "DIAL VTAM"). However this sort of facility poses risks: an unscrupulous Gopher administrator could devise a script that is dangerous or even malicious in execution. Even a PostScript interpreter could be dangerous, as PostScript is a language in its own right, able to open and modify files on disk; the only protection would be a PostScript browser known not to be able to write to disk.

+ Page 42 +

Administrators also face problems with users who try to use Gopher as a pathway into machines where they are not welcome. Any Gopher administrator can trivially add a link to a host that may prefer not to be made visible. There is a need for a registry of what sites are willing to accept Telnet sessions from random users across the Internet--these sites would be considered fair game for pointers in Gopherspace. (In practice this has not been a source of widespread concern so far.) Conversely, Gopher administrators may find that with tools like Veronica, their documents may take on greater visibility than desired. For

instance a departmental Gopher might offer information intended for local consumption (e.g., minutes of a faculty meeting) that is sensitive if viewed from other sites. Where necessary, it is possible for an administrator to restrict access to readers on a local network. This is essential for delivery of restricted licensed information, such as electronic newspapers.

15.0 Electronic Publishing

Gopher is sometimes labeled an Internet document delivery system. Yet, we are currently forced, for the most part, to deliver Gopher documents as flat ASCII files. This is because we cannot assume that there is software on the user's desktop to allow display of anything more sophisticated. As a result, Gopher administrators must spend considerable effort "undesktop publishing" documents that have been carefully made ornate by information providers. Not only does this slow the posting of information, it also denies the end user the benefits of well-designed material. These are serious concerns for Gopher administrators.

Some sites have begun to experiment with delivering documents in PostScript form. A public domain viewer for PostScript is available from the GNU project. It is called Ghostscript. At this point, this public domain tool has not been widely deployed, and it does not appear to be a practical solution for all computing platforms.

+ Page 43 +

Early in 1993, the Adobe Corporation announced a product called Acrobat, which will be a multi-platform PostScript viewer. Acrobat revolves around a special "distilled" PostScript file type that is said to yield documents only 25% larger than equivalent ASCII files. Additionally, Acrobat will include the ability to display embedded JPEG images. Acrobat is a promising tool for moving the online information world into true electronic publishing, with font changes and embedded images efficiently delivered. It would make an ideal browser for Gopher-delivered documents. The key question is whether Acrobat (or competitive tools) will be priced low enough so as to allow near-universal distribution.

16.0 Multimedia

The Gopher community has already begun providing networked multimedia. Early examples include the ability to deliver audio over the Internet. This has been demonstrated, but it is currently limited to NeXT or Sun workstations and it requires fast Internet paths (at least 64 kilobytes per second). Michigan State University has posted samples from its extensive Vincent Voice Library, such as Nixon's resignation speech and Kennedy's inaugural. Most recently, Michigan State mounted audio of the entire 1992 Presidential debate that was held on its campus, along with photos of the event, a transcript, and coverage from the student newspaper.

Many sites have begun using Gopher to deliver still images. The Fine Arts School of the University of Victoria uses Gopher to deliver art images to its students. Jeremy Kargon, a consultant at Johns Hopkins University, has mounted a series of architectural images via Gopher. The Instituto Tecnológico y de Estudios Superiores de Monterrey delivers Mexican scenes via

Gopher. Ohio State University will make available a set of building diagrams for use by their campus community, for instance to assist in wiring buildings for the campus computer network. Medical schools are exploring the idea of using Gopher-delivered images for instructional purposes. The University of Virginia has "Gopherized" two still-image collections from the Library of Congress--a set of recent Soviet documents and a collection of historical items from the Vatican.

+ Page 44 +

Two formats are most commonly used for still image delivery over computer networks: GIF and JPEG. GIF is the Graphics Interchange Format popularized by the information utility CompuServe. JPEG (Joint Photographic Experts Group) is another standard file format for digital still images, which allows efficient storage and fairly accurate reproduction. GIF viewers have been implemented in many commercial, shareware, and public domain programs on all common computer platforms. JPEG is a newer standard that is catching up to GIF in terms of prevalence of viewers. Under current usage, Gopher type "I" refers to GIF or JPEG files; type "g" refers specifically to a GIF file.

There are two factors to consider when evaluating whether delivery of still images via Gopher is practical: disk space consumed and time required to display the image. The GIF format typically requires 100 KB to 200 KB of storage per image, but the cost of disk storage is plummeting--large SCSI drives can now be obtained for less than \$1.50 per megabyte. However, even if cost of disk storage can be borne, one would probably not want to use GIF or other bitmapped image formats as a medium for delivery of multiple page documents: GIF viewers are usually too slow to allow for rapid paging through image files. Depending on the hardware and the GIF viewer, it can take 5 to 30 seconds to paint an image. (All other things being equal, JPEG files generally take more time to interpret and display.) In addition to the time it takes a viewer to decode an image for display on the screen, an administrator must also consider the network overhead of moving 100 KB files for each end-user selection. With favorable trends in the cost of disk, the speed of processors, and the bandwidth of networks, Gopher-mediated delivery of still images will become more viable over time.

Recently, there have been efforts to support network delivery of multimedia documents. David Lacey of the University of Iowa has integrated a browser that implements the MIME protocol. Sample documents mounted at Iowa include photographs of artifacts from a campus museum, along with descriptive text. Viewed on color NeXT workstations, these multimedia documents are rich and crisp. In the PC environment, the UGOPHER client is capable of launching software that can play audio, and it can also play MIME files if the user has the Metamail tool installed. Experiments with compressed motion video are beginning; one site is experimenting with delivery of Macintosh QuickTime files via Gopher. Apple has announced a QuickTime viewer for Microsoft Windows.

+ Page 45 +

The biggest challenge in delivering multimedia information is the proliferation of formats. In the still-image arena, there are more formats than there are 35mm camera lens mounts. Fortunately the GIF format has been popular enough for viewers to

become available for all common workstation platforms. The JPEG format may soon become deployed as widely. A Gopher administrator wanting to provide images to the broadest audience will use GIF at this point.

There is a similar proliferation of audio file formats. The 8 kilohertz mulaw format currently used for Gopher audio delivery is readily supported on Sun and NeXT workstations. Other platforms, such as the Macintosh or the PC with Sound Blaster cards, assume other formats. Some sites are offering files in Sound Blaster format for PCs so equipped, but these are not usable on machines that lack that hardware. Gopher+ may make it possible for a server to hold a file in one form but deliver it in the format best suited for the user's client software. For instance, translation among various audio formats is possible with software such as the public domain Sox utility. It may be possible with Gopher+ (and a little effort) to devise ways to archive audio in one canonical form, but deliver it in the appropriate machine-specific format.

Large, complex documents in Gopher reveal one shortcoming of the current technology: Gopher needs a mechanism for delivering a document in "chunks." Currently an entire document is delivered when the title is selected by the user. Consequently, a book or other large text must be artificially split into separate chapter documents (or further). An audio document must be split either by arbitrary time slice or based on content. It would be better if large, complicated documents could be treated as a single document, or, at the user's choice, viewed based on underlying structure. Some sort of table of contents field, analogous to the index markers on compact discs, might be used. With the deployment of Gopher+, there is the potential for adding this sort of support.

+ Page 46 +

17.0 Non-Gopher Client Technologies

We saw earlier that Gopher is best utilized when the user takes advantage of specialized client software. Recent developments in the area of non-Gopher client technology are worth noting.

At the Fall 1992 meeting of the Coalition for Networked Information, VTLS demonstrated a prototype online catalog that could retrieve information from Internet systems. The prototype system added information resembling the Gopher document descriptor to standard MARC records. This is a promising concept.

As we saw earlier, the Gopher community has already struggled with the question of how to organize information about Internet systems, including online catalogs. At the same time, some libraries have begun adding networked information resources to their online catalogs, and integrated library system vendors are providing links to remote systems from online catalogs. It is only logical to expect that someday we will see a merging of these two streams of activity. Ideally, the user would interact with a single user interface that searches for networked information: the user does not care whether it is a Gopher linked to an OPAC, or an OPAC linked to a Gopher.

Another alternative client development is a tool for the X Windows environment called NCSA Mosaic. Written by the National Center for Supercomputing Applications, Mosaic is a sort of "super client" that supports connections to a variety of servers: World-Wide Web, Gopher, USENET News, FTP, and others. Mosaic has

been deployed as a beta release as of this writing. Mosaic supports numerous document types including various image formats, audio, and PostScript. Mosaic and similar clients could be an important advance in networked information, freeing the user from having to obtain a separate client for each kind of primary server he or she wants to connect to. Mosaic aspires to provide the user with a customizable hypertext view of the networked universe.

+ Page 47 +

Naturally, with its generality, Mosaic presents the user with considerably more options to consider than does the typical Gopher client. Moreover, the X platform is used on comparatively few computers (compared to PCs and Macintoshes), even at universities. However, this client has generated considerable enthusiasm among those who have seen it. Besides serving as a unifying client, Mosaic also provides helpful navigation aids such as a complete history of where a user has been in her wanderings through the Internet; it even uses color to highlight menu items selected earlier in a session. When Mosaic is ported to more common computing platforms, it could become an important unified client. NCSA intends to deploy Mosaic on Macintoshes and perhaps under Microsoft Windows.

Another attempt at a unified client is under development at the Cornell Law School. A tool called Cello is being developed for Microsoft Windows.

Not all clients have the straightforward goal of simple retrieval of documents for the user to read. In May 1992, Steve Ludtke of Rice University announced a 3-D Gopher client for NeXT workstations. Dubbed "A Gopher in a Forest," the client presents a 3-D rendering of IP address space, and it updates the image in real time as the user moves from server to server or down the menu hierarchy. This tool generated considerable excitement, but no one was able to identify a practical use for it. Because IP address space doesn't correspond to geography, it was not a way to visualize where in the world documents were coming from, but it did inspire the idea of a client that could depict a geographical view of Gopherspace. (Such a client is yet to be written.)

Perhaps the most exotic alternative Gopher client is the MOO Gopher. An interactive, multiuser game called Mud--a sort of network-based Dungeons and Dragons--has been popular on the Internet for some time now. The Xerox Palo Alto Research Center has created a variant of Mud called MOO, which provides an object-oriented language for Mud. MOO is rich enough to support implementation of a Gopher client, and this has been done. In some MOO games, the player uses the MOO Gopher to browse the Internet for answers to questions in order to advance in play. Opinion is divided as to whether this is of research/educational value or a waste of time and bandwidth.

+ Page 48 +

18.0 Related Networked Information Server Technologies

Gopher is one of several emerging technologies that help people to navigate the Internet. It overlaps with the other tools in two ways: (1) Gopher can serve the same functions as some of the other technologies, and (2) Gopher can serve as a gateway into the other services. For instance, a Gopher can offer a menu of

WAIS services. Each pointer to a WAIS server can be presented as a Gopher index search item.

As a distributed menu system, Gopher is strong in its support of "resource discovery." Users can browse through a Gopher and learn about resources available on the Internet they never dreamed existed. This is an advantage over a tool like Archie. With Archie, you have to know the file name before you do your search. When a site "Gopherizes" its FTP service, it makes the list of files far easier for users to browse.

Like Gopher, HYTELNET is a tool that helps users identify needed Internet systems. Several years ago, Billy Barron (then of the University of North Texas) compiled a list of online catalogs available on the Internet. Building on this list, Peter Scott of the University of Saskatchewan created HYTELNET, a PC-based hypertext directory. Over time, the scope of HYTELNET was expanded to include a wide variety of Internet systems, such as CWISes. A colleague of Mr. Scott's, Earl Fogel, created UNIX and VMS versions of HYTELNET that permitted Telnet sessions to Internet systems. Mr. Scott issues frequent updates about new Internet systems on a mailing list. By letting users browse through a list of Internet systems, HYTELNET fills the role of a resource discovery tool. However, HYTELNET only identifies and links users to Internet systems; it does not retrieve files or perform index searches like Gopher.

Another tool, WAIS (Wide Area Information Servers), is a distributed networked database system. It is based upon an extension of the NISO Z39.50-1988 standard, which describes a model for an "originating application" (a client) to query databases at one or more "targets" (servers). WAIS allows a user to do keyword searches for documents, scanning a single server or multiple servers. WAIS responds to a search with a list of documents sorted by a weighted score--the higher the score, the better the match to the search. WAIS is strong in its ability to allow users to sift through a large body of documents stored in servers across the Internet. It is more appropriate for finding a small set of documents among a large set of candidates than for serving as a menuing or browsing tool. WAIS has enjoyed a fair amount of attention in the general press lately, because its developers have launched a commercial firm to market it.

+ Page 49 +

Yet another tool, World-Wide Web, evolved in the physics community, and it is being increasingly recognized as a general networked information retrieval tool. WWW is a distributed hypertext system. Information providers can deliver documents defined in a type of markup language (a subset of the Standard Generalized Markup Language, SGML, called HTML) that can be perused by users who need not know the physical location of the parts of the document they are reading. As a hypertext system, WWW frees the user from the confines of a rigid hierarchy. Advocates of WWW claim that this is an essential feature--the only effective way to deliver documents chosen from a vast universe of information is via hypertext. They point out that providing pointers within the text of documents can make it far easier for users to find and peruse materials of interest. The WWW model lets users seamlessly cruise from server to server and from topic to topic. It allows the user to retrace steps as well. Like Gopher, WWW can launch Telnet sessions to connect to online services.

WWW can serve as a browsing tool much as Gopher does. Under

Gopher, it is common for a document to include pointers such as "Look in the 'Campus Eateries' folder for more information." To follow that advice, the user must leave the current document and open the folder in question. Under WWW, the pointer text could highlight "Campus Eateries" within every document where it would be helpful to mention it; a click on the embedded title would open the referenced document. Such multiple links are unobtrusive and do not require the user to hunt through other folders. It is hard to deny that embedded links are easier for the user to navigate.

Hypertext services need not require a great deal of editorial investment. For instance, one site has built a cross-referenced set of UNIX "man" pages under World Wide-Web. ("Man" pages are brief descriptions of UNIX commands and their options, one per command.) Every mention of another "man" page within a document is shown as a link; the user can jump from page to page at will. This sort of process can be implemented by means of an automated script, without a great deal of programming or editorial time required.

+ Page 50 +

These disparate technologies have been exploited at varying rates. Although numerous subject-specific WAIS databases have been made available and World-Wide Web has been deployed at quite a few sites, Gopher seems to be the dominant tool for campus-wide information systems and for some subject-specific services. In terms of traffic generated, Gopher is leading the pack. For instance, counts of packet load on the NSFNET backbone reveal Gopher is high on the list of network traffic generators (see Table 3).

Table 3. NSFNET Backbone Traffic Distribution by Service, March 1993. (Extracted and summarized from
ftp://nic.merit.edu/statistics/nsf-9303.ports)

Packet Total: 34,874,064,400
Byte Total: 6,502,203,065,800

Service Name	Port	Packet Count	% Pkts	Byte Count	% Bytes
Gopher	70	327717650	0.940	79023945150	1.215
Z39.50	210	19506350	0.056	5415741150	0.083
WWW	80	11294550	0.032	3613584700	0.056

Gopher has grown dramatically in its ranking in these statistics over the last several months. Of course, WAIS and WWW are also growing rapidly as new users come to rely on these tools. To some extent, these figures overstate Gopher's dominance. For instance, the use of WAIS to deliver data via Gopher is masked. Also, since WWW has enjoyed more initial success in Europe, the use of these U.S. numbers understates its impact somewhat.

Not surprisingly, the relative number of servers deployed for the major tools shows Gopher with a wide lead as well. Whereas there are over 1,100 Gopher servers on the Internet, there are about 118 WAIS servers (supporting over 425 databases) and about 50 WWW servers. (Of course, these numbers are not

directly comparable. Just as the Library of Congress holds more books than a small public library, it could be the case that a given WAIS server holds more data than several Gopher servers.)

+ Page 51 +

Why is Gopher so popular? One reason is the relative ease with which an information provider can set up a server. As Steve Worona of Cornell University observes, "Compared to other Internet navigation tools, a Gopher server is quite easy to set up. As a result of the low 'entry costs' for setting up a Gopher service, it has become a very popular tool across the Internet in a very short time." [9] Another reason is that Gopher clients have been deployed on the most popular computing platforms--PC compatibles and Macintoshes. And for those users who lack client software, Gopher is fully accessible via dial-up and Telnet paths; WWW, for instance, lacked a good curses client until recently. (The University of Kansas offers a public curses client for WWW in support of their CWIS, Lynx.)

The distinctions among these technologies may blur over time. Groups within the IETF are working to define standards that will allow an information provider to offer standard abstract information in order to describe the purpose and content of a network resource as well as standardized location descriptions so that users or automated processes can connect to that resource. For instance, an FTP site could offer descriptions of all its files, and Archie could allow users to do keyword searches of those abstracts. Already, the Gopher developers have announced their intent to support these new standards. Gopher could thus take on some of the role currently filled by Archie. The Gopher team have also said they intend to embrace the Z39.50 standard. As Mark McCahill of the Gopher team notes, Gopher and other tools may become much more similar with these enhancements and similar ones.

It is important to bear in mind that these technologies can interoperate today. Therefore, a site need not use one to the exclusion of another. For instance, a site might mount a WAIS database that meets a specific need, while using Gopher as a campus-wide information system. The Gopher CWIS can include the WAIS database as a document. In fact, the combination of Gopher as a CWIS and WAIS as the server for one or more databases is becoming quite common.

+ Page 52 +

Nor is WAIS the only "back-end" database technology deployed behind a Gopher "front-end." For example, Tim Howes of the University of Michigan has implemented a gateway between Gopher and X.500 directory servers, making it easy for users to browse a list of X.500 "phone books" from around the world. Similarly, Ohio State University had an existing telephone directory server on their campus network; they implemented a gateway from Gopher to that server rather than building a separate CSO database. (Some users find CSO telephone searches overly complex compared to a Gopher index.) Gopher administrators are implementing databases using a variety of other engines, with Gopher as a primary or secondary interface to the data. The UNIX Gopher server also makes it easy for a Gopher administrator to point a folder to an FTP site or even to a collection of mail messages.

Gopher has existed just about two years, its developers are only able to work on it part-time, and documentation on how to build a Gopher service is somewhat spotty. Yet, it has grown into a production service at dozens of sites around the world. Clearly it meets a real need--it provides a distributed menu system that allows users to browse the Internet. Although Gopher faces competition from rival technologies such as World-Wide Web, its future seems assured. The Gopher+ protocol promises powerful extensions to the basic Gopher concept. The Gopher clients and servers that we will see next year may have significantly greater capabilities than today's software.

True, Gopher is not unique among Internet technologies in its ability to deliver networked information. What makes Gopher special is that it can deliver a browsing list of such documents and services in a package that's easy to administer and accessible to users. Via Gopher, users can move from server to server, quickly seeing what's available and learning about documents and services they'd never dreamed of. With Veronica searches, the user can search for documents across multiple Gopher servers. The user has the choice of browsing or doing a focused search.

Gopher's most serious competitor is probably the World-Wide Web. Some critics of WWW point out that there is the threat that a hypertext service can degenerate into a "twisty maze of passages all alike" and that users will get lost. WWW advocates respond that the hierarchical model of a Gopher is mathematically a subset of the WWW hypertext; an administrator can design a web to look just like a Gopher hierarchy. (However, a WWW adherent might ask why an information provider would choose to limit a service to a hierarchical view.)

+ Page 53 +

As WWW clients become available on more platforms and as superclients such as Mosaic and Cello are deployed, WWW will become accessible and attractive to many more users. Tim Berners-Lee, the main architect of WWW, has stated plans to make it easier to install a WWW server and to "join the Web." With broader user access and more servers online, WWW could become more competitive with Gopher. The University of Minnesota and the Gopher community could face pressure to provide hypertext support within Gopher. The question is whether Gopher can retain its ease of setup and use if it embraces hypertext.

Gopher/WWW interoperation is already common. The UNIX Gopher server can offer a Gopher hierarchy to the Web. The developers of Mosaic and NCSA have stated their intent to deliver a unified Gopher/WWW server that is capable of serving a Gopher hierarchy directly to WWW clients. These developments could encourage merging of the tools. The Internet Engineering Task Force has formed a working group that is looking at integration of resource discovery tools such as Gopher and WWW.

Recent statements from the University of Minnesota indicate that, if Gopher is used for commercial purposes, such use will require a license agreement. The University of Minnesota intends to continue to make Gopher available free of charge for nonprofit academic use. Much of the development of Gopher has been possible because of the keen interest of a supportive community. Will that community continue to contribute to a corpus for someone else's financial benefit? Already the announcement of licensing fees has spurred one site to offer a free Gopher server

implemented in a few hundred lines of code (written in Perl). Another site is offering a "GNU Gopher" server. (The GNU project is a movement dedicated to providing free versions of UNIX and related tools.) Note that related tools have also "gone commercial" recently. This is the case with Archie, now to be developed by Bunyip Information Systems, and WAIS, to be enhanced and marketed by a private company. (Note that as of this writing WWW is offered under similar terms to the new Gopher license; commercial users are expected to contribute development effort or pay a fee.) It will be interesting to watch the effect of commercialization on all of these products.

+ Page 54 +

For the present, Gopher enjoys overwhelmingly superior market penetration compared to competitive tools. It is impossible to gauge the impact of all the economic and technical developments in the Gopher community and in the networked information field at large. Over the long run, Gopher might remain a dominant player, or it might merge with another technology. It is a safe bet that it will be an important tool for the next several years at least. In the short run, no single networked information retrieval scheme will meet all of the needs of the fast-growing body of Internet users.

Notes

1. Steve Worona, personal e-mail message, 25 February 1993.
2. Billy Barron, personal e-mail message, 10 March 1993.
3. Marty Courtois, personal e-mail message, 26 February 1993.
4. Ibid.
5. Marie-Christine Mahe, personal e-mail message, 9 March 1993.
6. Nancy John, personal e-mail message, 2 March 1993.
7. Joel Cooper, personal e-mail message, 1 March 1993.
8. Martin Dillon, personal e-mail message, 9 March 1993.
9. Steve Worona, personal e-mail message, 25 February 1993.

Bibliography

Deutsch, Peter. "Resource Discovery in an Internet Environment--The Archie Approach." *Electronic Networking: Research, Applications and Policy* 2, no. 1 (1992): 45-51.

Berners-Lee, Tim et al. "World-Wide Web: The Information Universe." *Electronic Networking: Research, Applications and Policy* 2, no. 1 (1992): 52-58.

Kahle, Brewster et al. "Wide Area Information Servers: An Executive Information System for Unstructured Files." *Electronic Networking: Research, Applications and Policy* 2, no. 1 (1992): 59-68.

Lynch, Clifford A. "Information Retrieval as a Network Application." Library Hi Tech 8, no. 4 (1990): 57-72.

+ Page 55 +

Scott, Peter. "Using HYTELNET to Access Internet Resources." The Public-Access Computer Systems Review 3, no. 4 (1992): 15-21. To retrieve this file, send the following e-mail message to `LISTSERV@UHUPVM1.UH.EDU`: `GET SCOTT PRV3N4 F=MAIL`.

Glossary

Anonymous FTP. Anonymous FTP servers allow users to retrieve files without the need for assigned user IDs (literally the word "anonymous" is used as the login ID). This service is offered by many sites on the Internet.

Archie. A network service that allows users to discover which anonymous FTP sites house particular files of interest. Developed at McGill University (and now Bunyip Information Systems), future versions of Archie may allow searching for resources by abstract, author name, and other criteria.

ASCII file. A file encoded in the 128-character ASCII (American Standard Code for Information Interchange) character set. The term "flat ASCII file" is often used to refer to a simple text file, with no embedded special formatting codes or binary data. In FTP transfers, "text" and "ASCII" are synonymous.

Binary file. Binary files consist of streams of bytes whose meaning is defined by some external format standard. For example, executable computer programs are binary files. In FTP transfers, a binary file is specified by "bin" or "image" settings.

Client/Server. A model for distributing system functions between client software, residing on a user's workstation, and server software, residing on a host. The host could be a UNIX workstation, a mainframe, or another type of computer. The client handles most of the information presentation functions, and the server handles most of the database functions. A protocol specifies how the two should communicate. The client/server model is growing in popularity as workstations and networks grow in power.

CWIS (Campus-Wide Information System). A university or college information system that offers integrated access to documents (e.g., course schedules, lists of current events, and academic job openings) and systems (e.g., online catalog). CWISes began on mainframes and are now commonly implemented under the client/server model. Gopher and WWW are commonly used for CWISes; other CWISes exist (for instance, the TechInfo software from MIT).

+ Page 56 +

CSO. A protocol that allows client/server searching of simple databases such as phone books. Named after the Computing

Services Organization at the University of Illinois, the CSO protocol enjoys widespread usage in the Gopher community, despite more elaborate standards for such "white pages" services. (Sometimes called "CCSO.")

Curses. A software feature under UNIX that allows a programmer to support full-screen terminal sessions. Said to be a play on words referring to the cursor keys.

FAQ (Frequently Asked Question). Documents that list such questions and their answers are referred to as FAQs. Much of the documentation in the USENET world resides in FAQ files.

FTP (File Transfer Protocol). A standard protocol for sending files from one computer to another on TCP/IP networks, such as the Internet. This is also the command the user usually types to start an FTP session.

GIF (Graphics Interchange Format). A still-image file format promoted by CompuServe. Software to view GIF images is commonly available for most computing environments in the form of public domain, shareware, or commercial products.

GNU. A project to deliver UNIX operating system clones and related tools as freely available software. (Stands for "GNU is Not UNIX.")

Hypertext. A scheme for supporting embedded links within documents. While browsing a document with hypertext links, a user can select one of those links and quickly move to the document it points to. Popularized by the HyperCard Macintosh program.

HYTELNET. A hypertext database of Internet systems, such as online catalogs and CWISes. The PC version of the program is available from Peter Scott at the University of Saskatchewan. The UNIX and VMS versions can make Internet connections to listed systems; the PC version cannot.

IANA (Internet Assigned Numbers Authority). Internet protocol writers must agree upon standard values for fields used within the protocols. IANA is the organization that registers these standard numbers. For instance, IANA assigned TCP port 70 to Gopher. IANA will register MIME content types.

+ Page 57 +

IETF (Internet Engineering Task Force). IETF devises new and updated protocols to be used on the Internet. It is an informal body composed of Working Groups that carry on discussions over the Internet and in periodic meetings.

IP address. A unique number assigned to a computer on a TCP/IP network. Conventions have been set up to assign IP numbers in a systematic fashion across the Internet. Whether a machine is a large server or a small PC, if it is to be on the Internet, it must have an assigned IP address. IP addresses look like "35.8.2.61." There are always four numeric values, separated by periods, in IP addresses.

JPEG (Joint Photographic Experts Group). A still-image file

format that allows efficient compression. The JPEG compression scheme is "lossy"; it does not preserve all of the data, but it can yield significant space savings with little perceptible loss. JPEG viewers are relatively slow compared to GIF viewers.

MIME (Multipurpose Internet Mail Extensions). Initially, an extension of Internet mail standards to allow binary data to be embedded in SMTP mail. Since its introduction in 1992, MIME has been implemented on several computer platforms (often under the name "Metamail"), and it is increasingly viewed as the appropriate standard for handling multimedia information to be moved across dissimilar computing environments, whether sent via e-mail or otherwise.

Mosaic. An integrated client program developed by the National Center for Supercomputing Applications. Mosaic acts as a client for multiple Internet services, including Gopher, WAIS, WWW, USENET News, and FTP. Currently implemented on UNIX systems supporting X Windows and Motif. Versions of Mosaic for the Macintosh and Microsoft Windows are expected.

PostScript. A printer description language from Adobe. PostScript is the dominant format used for desktop publishing. Documents in PostScript format are commonly shared across the Internet and printed on laser printers after retrieval from a remote archive.

+ Page 58 +

PPP (Point-to-Point Protocol). A relatively new protocol that allows dial-up users to connect to the Internet. With PPP (or similar functionality) a user can use Internet client software, such as a Gopher client, in a dial-up session. Without PPP, the user must dial into a host or public client service via a terminal session (usually VT 100).

RFC (Request for Comments). Documents that define both proposed and adopted Internet protocol standards. RFCs are numbered in an ordinal fashion.

SGML (Standard Generalized Markup Language). SGML is a scheme (and an ISO standard) for embedding structural information within a document. SGML is popular in scholarly and electronic publishing as a way to support multiple views of a document. An SGML-compliant set of structures called HTML is used by World-Wide Web.

SLIP (Serial Line IP). A protocol that allows dial-up users to connect to the Internet. Being supplanted by PPP.

SMTP (Simple Mail Transfer Protocol). A protocol for sending e-mail messages between computers on TCP/IP networks, such as the Internet. The user does not run a program named SMTP; instead, various e-mail packages know how to utilize this protocol.

TCP/IP (Transmission Control Protocol/Internet Protocol). Technically, TCP and IP are separate protocols; together they allow computers on the Internet to communicate by providing a reliable way for bytes to be delivered in order over a network connection. Connections are made to TCP "ports," allowing multiple connections per machine. A port is described by a

number (e.g., Gopher servers typically use port 70).

Telnet. The TCP/IP protocol for remote terminal sessions. Usually implemented as a command of the same name.

TN3270. A variant of Telnet that allows TCP/IP connections to IBM mainframes that utilize 3270 terminal conventions.

Uniform Resource Identifier. An umbrella term for standards that describe Internet resources in a uniform way. The IETF is considering a Uniform Resource Locator, which will be a standard way to name a particular document on a particular network server. Another proposed standard, the Uniform Resource Number, will be a unique number (analogous to an ISBN) assigned to a document or resource regardless of its location and other resource descriptors.

+ Page 59 +

USENET News. A distributed discussion list system, much of whose traffic is carried over the Internet. USENET News services consist of news "feeds" typically provided by one or more servers on a campus network, and news "readers" (i.e., client software that communicates with the server over a news delivery protocol).

Veronica. A service that provides a Internet-wide index of Gopher document titles. Developed at the University of Nevada, Veronica servers periodically poll all the Gopher servers they can connect to and build a title index that can, in turn, be pointed to as a standard Gopher index.

VT 100. The dominant communications protocol for full-screen terminal sessions. The VT 100 standard was defined by the Digital Equipment Corporation in the 70s. Most terminal emulation software packages (e.g., Kermit and PROCOMM) implement VT 100 or its descendants.

WAIS (Wide Area Information Servers). Based on an extension the Z39.50-1988 standard, WAIS allows searches of documents on one or more WAIS servers. Originally promulgated by Thinking Machines Corporation, WAIS is now offered in a commercial version (by WAIS, Inc.) and a public domain version (by the Clearinghouse for Networked Information Discovery and Retrieval). The WAIS search engine is commonly used by Gopher servers. An index document under the UNIX Gopher server can point to a stand-alone WAIS server.

World-Wide Web. A network-based hypertext document delivery system. Developed at CERN in Switzerland, WWW is growing in popularity. WWW supports embedded links to documents.

Z39.50. A NISO standard defining a client/server model for searching bibliographic and other databases.

Z39.58. A NISO standard describing a Common Command Language for online catalogs.

+ Page 60 +

Note From the Author

After this document has been distributed via The Public-Access Computer Systems Review, it will be made available on the central Gopher at Michigan State University (gopher.msu.edu, port 70). It will be placed under the "More About Gopher" folder.

About the Author

Rich Wiggins, Gopher Coordinator, Michigan State University
Computer Laboratory. Voice: (517) 353-4955. Internet:
RWWMAINT@MSU.EDU.

The Public-Access Computer Systems Review is an electronic journal that is distributed on BITNET, Internet, and other computer networks. There is no subscription fee.

To subscribe, send an e-mail message to LISTSERV@UHUPVM1 (BITNET) or LISTSERV@UHUPVM1.UH.EDU (Internet) that says: SUBSCRIBE PACS-P First Name Last Name. PACS-P subscribers also receive two electronic newsletters: Current Cites and Public-Access Computer Systems News.

This article is Copyright (C) 1993 by Rich Wiggins. All Rights Reserved.

The Public-Access Computer Systems Review is Copyright (C) 1993 by the University Libraries, University of Houston. All Rights Reserved.

Copying is permitted for noncommercial use by academic computer centers, computer conferences, individual scholars, and libraries. Libraries are authorized to add the journal to their collection, in electronic or printed form, at no charge. This message must appear on all copied material. All commercial use requires permission.
