An Introduction to the Internet, Part 3: Internet Services

Jerry V. Glowniak

Nuclear Medicine Service, VA Medical Center, Portland, Oregon

In this third article in a four-part series on the Internet, Internet services are described and examples are given on how to use them. After reading this article, a technologist should be able to name the major Internet services, understand how they function and use them to obtain information from the Internet.

Key Words: Internet; computers; electronic mail; USENET; telnet; ftp; archie; gopher; veronica; World Wide Web

J Nucl Med Technol 1995; 23:231-248

The resources available on the Internet are accessed by various methods which are referred to as Internet services. Each of these services has distinctive features. Internet services are implemented by application programs or protocols, the most important of which are telnet, ftp, gopher, email, WAIS, archie and the World Wide Web. These are discussed in detail below.

The user interacts directly with these programs which are at the highest level of the protocol stacks discussed in the first part of this series. The application program formats user input and transfers it to lower levels of the protocol stack, TCP/IP software, then routes the information onto the Internet. The information is received by a remote computer and passed up to an application program that then performs some prescribed function and returns the results to the local application program.

All Internet services function in a client-server mode. As described in Part 1, this method of interaction between computers involves splitting an application program into two parts: the client program that accepts input from the user and displays results; and the server program that resides on the remote computer and performs the tasks requested by the client. Each of the common Internet services is defined by a standard protocol that describes how connections between computers are set up, what information is passed between them and how a particular service functions.

In the traditional model of the client-server architecture, a client program for a particular service can only communicate with its corresponding server. Thus, a gopher client can only communicate with a gopher server, not with a World Wide Web or ftp server[1]. While there are several reasons for making resources available by the different services, this approach complicates access to information. To access all the resources on the Internet, a user must know how to use the different services and must have the client software for each of the different services. In addition, a user must know what type of server is available on a remote computer.

To circumvent the problem of a client having access to only a portion of the resources on the Internet, two partial solutions have evolved. One method of increasing access to different types of services is to allow servers to interact through special programs called gateways that permit different protocols to communicate with one another. For example, an individual who sets up a gopher server may want to provide information that is available on an ftp server, such as to allow downloading files. One way to do this is to create a program that can translate requests from a gopher server to a format understood by an ftp server. Another method of accessing multiple services is to incorporate multiple client applications into a single client program. The most widespread application of this approach is with Web browsers. These are client programs whose primary function is to contact World Wide Web servers. The World Wide Web is a distributed system of documents written in hypertext format (see below for further details). Web browsers have built-in capabilities to access ftp, WAIS, gopher servers and, depending on the particular browsers, other services such as email, telnet or USENET.

SERVICES, PROTOCOLS, AND SOFTWARE

Before describing particular Internet services, it is important to understand the terminology. Strictly speaking, there is a distinction between an Internet service, the protocol describing that service and the software program that implements the protocol.

For correspondence or reprints contact: Jerry V. Glowniak, MD, VA Medical Center (115), PO Box 1034, Portland, OR 97207.

[1] Telnet clients are the single major exception to this rule. Using a telnet client alone a user can send email and download information from gopher, World Wide Web and other servers, but it is beyond the scope of this article to explain how to do this.

TABLE 1
Common Internet Services and Their Protocols and Port Numbers

Service	Protocol	Port number
Directory service	CSO	105
Directory service	whois	43
Electronic mail	SMTP	25
File transfer	ftp*	21, 20
Finger	finger	79
Gopher	gopher	70
IP address lookup	DNS	53
Remote login	telnet	23
USENET	nntp	119
WAIS	Z39.50	210
World Wide Web	http	80

Key: DNS = domain name service; ftp = file transfer protocol; http = hypertext transport (transfer) protocol; nntp = network news transfer protocol; SMTP = simple mail transfer protocol; and WAIS = Wide Area Information Service.

*ftp uses two ports; port 21 is the control channel for sending and receiving commands and port 20 is the data channel used for transferring files.

An Internet service is a description of some function performed between two computers. Examples are electronic mail service, remote login, transferring files between computers or accessing distributed document systems (gopher and the World Wide Web). Table 1 lists the common Internet services, their related protocols and port numbers.

Since the Internet services involve communication between computers, two computers must speak the same language if they are to understand each other. The rules of communication are specified in documents called protocols. Each protocol covers a particular aspect of computer communications, and protocols are generally written to conform to specific levels in the Internet protocol model. The documents describing Internet protocols are called RFCs (Requests for Comments) and are available at several locations around the Internet. All the RFCs can be obtained by ftp from the InterNIC (Internet Network Information Center) at ftp.internic.net. The common Internet services are specified by protocols that are described in RFCs.

A protocol describes what is to be done, not how to do it. Thus, the ftp protocol describes, among other things, what control messages are exchanged between computers when files are being transferred between them. The protocol does not state how the computer generates these messages, which is the task of software. Unlike the protocols for the Internet services, which are freely available and independent of any type of computer or operating system, the software implementing the protocols must be written for each type of computer and/or operating system. For example, although there are many different types of commercial and noncommercial email programs, they all use the simple mail transfer protocol (SMTP) for sending and receiving messages on the Internet. The Internet protocols standardize all the important Internet functions. Certain aspects of communication are not standardized, however, such as the display of files. Different software programs can produce considerably different appearances to downloaded data. Furthermore, software for the various Internet services may have options that allow a user to change parameters within the program that affect the display and execution of the program. This is especially true of Web browsers. These options increase the usefulness of the program, but they also increase its complexity. Given the large amount of Internet software, it is worthwhile to evaluate different packages in order to find one that best suits a user's needs.

The terms for describing Internet services, protocols and the software implementing these protocols are often used interchangeably which can occasionally lead to confusion. In addition, the commands used to run software programs are often the same as the protocol names, which is a common feature of UNIX systems. To make matters even more confusing, some computers are named according to the service they provide. For example, the National Institutes of Health have a computer called gopher.nih.gov that can be reached using a gopher client program. The command for accessing the gopher software on a UNIX system is simply "gopher." Thus, to reach this computer, one would give the command gopher gopher.nih.gov on a UNIX system. Finally, the names of common Internet services are often used as verbs. The statement "Telnet to liberty.uc.wlu.edu to use the netlink service" means that the user should use his telnet software to connect to the computer liberty.uc.wlu.edu which offers a service called netlink.

From the user's point of view, one of the major differences between software programs for the Internet services is the way data is displayed and commands are entered. There are two basic types of interfaces: graphical user interfaces (GUIs) and text-based system (also called command line interfaces). In general, GUIs are easier to use but require more complex computer hardware and software. GUIs use the entire computer screen to display text and graphics. Common examples of these systems are Macintosh computers, IBM's OS/2 and OS/2 Warp operating systems, Microsoft Windows and Windows 95, and the UNIX X Window system. Commands and options can be entered with a mouse by clicking on menu items or icons. With text-based systems, commands are typed at a prompt or chosen from a menu. Examples of these systems are DOS for IBM and IBM-compatible PCs, standard UNIX systems, and VMS/VAX by Digital Equipment Corporation. Software for all the common Internet services is available for either text-based or GUI systems. The visual appeal and some of the functionality of the World Wide Web, however, is lost using text-based systems. On the other hand, certain services are text oriented, such as telnet, and GUIs have few advantages over text-based systems for these services.

IN GENERAL

Internet resources are accessed by choosing an Internet service and supplying the computer name or IP address of a computer on the Internet that offers that service. When a TCP/IP packet arrives at the destination computer, there must

be some means of identifying to which application program the data is to be delivered. This is accomplished by including a number in the TCP header which identifies the required Internet service. This number is called the port or port number. An Internet computer may designate any number in the range 0 through 65535 as a port for a specific service it offers. If number assignments were completely random, it would be difficult for a user to reach services at remote sites since the user would need to know how a particular computer assigns port numbers to its services. Fortunately, the commonly used Internet services are assigned standard port numbers that are referred to as "well known ports." If a user fails to specify a port number for one of these services in a command, the standard port number for that particular service will be used. These standard port numbers are assigned to servers.

A client also functions through a port on the local system, but the local system randomly assigns port numbers to clients for each connection to a remote system. The TCP header in a TCP/IP packet contains the port number that the local client uses in addition to the server port number. The remote server, when it communicates with a client, identifies a particular connection by means of the randomly assigned port number. This mechanism allows multiple simultaneous connections to be maintained between users of a client on a particular system and a remote server. Although the users are all using the same local client program, they each have a different local port number. Table 1 lists the standard port numbers for common Internet services. Occasionally, a particular computer may assign a nonstandard port number to one of its services. In that case, the user must specify the port number in order to access the service. An example is weather information from the University of Michigan's Weather Underground on port 3000. A user can access this service by telnet at downwind.sprl.umich.edu 3000.

With some software programs, like Trumpet Winsock, a TCP/IP program for Windows systems, a user may encounter the term "socket." A socket is defined as the combination of the IP address of a computer that offers a particular service and the port number of that service. Communications on the Internet take place through connections. A connection is a pair of communicating sockets: the local socket of the client and the corresponding socket of the remote server. Thus, a connection is defined by 4 numbers, the IP addresses and port numbers of the client and server. Each data packet exchanged between the client and server contains these 4 numbers. This ensures that each client-server interaction on the Internet at any given time is unique.

The method of accessing a remote computer in a text-based system is:

command ⟨computer name⟩ [port number]

The IP address can be substituted for ⟨computer name⟩, and [port number] can usually be omitted. ⟨computer name⟩ can be entered in upper or lower case. For DOS or VMS systems, all text is case insensitive and commands can be entered as either upper or lower case. UNIX commands are case sensitive, almost all of which must be in lower case. In systems that use GUIs, point and click methods are used and destination computers can be chosen from a predefined list of computers, or the computer name can be entered manually. For UNIX systems, the command for a particular service is often the same as the protocol name. This is the case for telnet, ftp, gopher and finger; but electronic mail, USENET and the World Wide Web have different commands, depending on the particular software program used to access that service.

In the description of Internet services that follows, examples will be given in UNIX format. While most readers are probably unfamiliar with UNIX, I have chosen this approach for several reasons. UNIX is the most widely used operating system on the Internet and a user can gain more information about resources at many sites with a basic knowledge of UNIX. Secondly, although text-based systems are not as easy to use as GUIs, they are easier to describe because they have a more standardized format than GUIs. Lastly, a large proportion of Internet access providers offers UNIX-based shell accounts. Since these accounts can be purchased for very low prices, this may be a cost-effective way of accessing the Internet for users who don't require or cannot use the graphical or audio capabilities of services such as the World Wide Web. It should be pointed out, however, that many UNIX sites offer an optional menu system that allows a user to access specific services by choosing menu items so that a knowledge of UNIX is not required.

Commands and responses in the examples given below are indicated by indented text. Where computer prompts and user inputs are shown together, user input is in boldface type. The reader should enter a carriage return after typing the indicated text. Computer prompts and responses are indicated by ordinary text. Items enclosed in square brackets [ ] indicate optional user input, while items enclosed in angled brackets ⟨ ⟩ indicate mandatory user input.

ELECTRONIC MAIL (EMAIL)

One of the most widely used services on the Internet is electronic mail, or email. It is also one of the more difficult services to describe in detail because of the large number of programs in existence. Email serves such a useful purpose that several networks with no other connections to the Internet have Internet email gateways. Since these networks often employ different addressing conventions than those used on the Internet, addressing mail to users on some of these networks can be complicated. A reader having trouble sending mail to or receiving mail from one of these networks should consult with the network administrator or a text that describes mail addressing conventions. The standard Internet format for addressing mail is:

⟨username⟩@⟨computer name⟩

where ⟨username⟩ is the way an individual is identified on his local computer, and ⟨computer name⟩ is the computer running the mail server that receives and distributes mail on the remote

network[2]. No spaces are allowed anywhere within the address. As an example, the author's email address at the Veterans Affairs Medical Center in Portland, Oregon is:

glowniak@portland.va.gov

On UNIX systems, the user name is often the same as the login name.

Mail programs differ considerably. Although some mail programs come as single software packages, especially programs for personal computers, most mail programs on the Internet have a basic two-part structure: a program with which the user interacts, the mail user agent (MUA); and a program that sends and receives mail messages, the mail transfer agent (MTA). The MUA is similar to a word processor. An individual uses this program to create, modify, read and delete mail messages. Messages are stored in files in a directory that is usually referred to as a user's mailbox. Messages a user creates are stored in a subdirectory called an out box and received messages are stored in an in box. The MTA is actually responsible for sending and receiving messages. The MTA deposits received messages in the user's in box and retrieves and sends messages from the out box when instructed to do so by the MUA. MTAs employ the standard Internet mail protocol, the Simple Mail Transfer Protocol (SMTP). The most common MTAs on the Internet are the programs sendmail and smail.

The two-part structure of mail programs offers a great deal of flexibility in installing mail systems on a computer used by many individuals. Different MUAs (the programs usually identified by the user as the mail program) can be employed that all use the same MTA. This simplifies implementing and updating mail programs since only the MUAs need to be changed rather than the MTAs.

One of the major differences between mail programs is the type of information they can send and receive. Mail programs were originally developed to handle simple text. Thus, computer programs, image files and files with special nontext formatting codes (e.g., spreadsheets and many word processing files) could not be sent directly by email. One solution to this problem is to encode the contents of these files into text equivalent characters. These text equivalent versions of a file appear as ordinary characters in a mail message, but the file is completely unreadable as ordinary text. The recipient of such a file must decode the file to use it. The most common method of encoding and decoding files on a UNIX system is with the uuencode and uudecode commands.

Another method for sending nontext files is with an extended version of the standard Internet mail protocol, the Multipurpose Internet Mail Extensions (MIME) protocol. This protocol describes how various file types can be sent by email. To use this method of file transfer, both the sending and receiving mail programs must be MIME compliant. MIME-compliant mail programs can incorporate images, graphics, sound or even video clips in a mail message.

[2] In actual practice, ⟨computer name⟩ is often not the name of the computer with the mail server but, rather, an alias that is a shorter or easier to remember name than that of the mail server. This alias is frequently just the network portion of the mail server's name. The domain name system, described later, knows how to locate the actual mail server given an alias.

MAILING LISTS

A mailing list is a specialized form of email. Mailing lists define a group of individuals who correspond through a program that delivers mail to all members on the list. Mailing lists require no special software for members on the list aside from a mail program on the local system. The program responsible for maintaining the mailing list is called a mailing list manager (MLM) and is an MUA. The most common MLMs on the Internet are LISTSERV (a contraction of list server), Majordomo and ListProcessor. These programs are also referred to as mail bursters or mail exploders.

The program that runs the mailing list has two addresses. One address is used for administrative functions such as subscribing or unsubscribing to the list and providing information on how the program works. The other part of the program is the address to which members post messages and which sends the postings to all members on the list. A good example of a nuclear medicine mailing list is the nucmed mailing list at the University of Western Ontario. To subscribe to this list, a user sends a message to the address:

listserv@largnet.uwo.ca

with the message: subscribe nucmed ⟨name⟩, where ⟨name⟩ is the user's real name. To unsubscribe to the mailing list, the message should be: unsubscribe nucmed. Postings to the list are sent to the address:

nucmed@largnet.uwo.ca

A message sent to the above address is sent out to all subscribers on the list.

USENET

A service related to mailing lists is USENET (a contraction of User's Network). USENET is a service devoted entirely to discussion groups or newsgroups, as they are more commonly called. Originally, USENET was an independent network of UNIX-based computers that exchanged messages over phone lines. USENET gradually began to use the Internet for exchanging messages so that at present nearly all USENET traffic travels over the Internet. Unlike mailing lists where there is a computer to which all messages are sent, messages to USENET are sent to newsgroups which can be located on a large number of computers on the Internet. Newsgroups are devoted to specific topics and are named in a manner similar to Internet computers. Newsgroup names are strings of descriptive characters separated by periods with the most general category on the left and the most specific category on the right of the string. The major general categories are listed in Table 2. Each general category has a number of subcategories that are not as well-defined. The alt and misc categories are for newsgroups that do not precisely fall into the other categories. Specific newsgroups can be very general such as:

TABLE 2
Usenet Categories

alt	alternate
comp	computers
misc	miscellaneous
news	newsgroups
rec	recreation
sci	science
soc	society
talk	debate oriented

sci.chem

for chemistry or very specific such as:

alt.sports.baseball.ny-yankees

for a newsgroup discussing the New York Yankees.

There are thousands of newsgroups world wide. Newsgroups can be local, regional, national or international in distribution. An individual can start a newsgroup by posting a message to news.announce.newgroups, a newsgroup specifically established for setting up newsgroups. A name for the newsgroup is proposed and its function described. Individuals who are interested can then vote on whether or not to create the newsgroup. If enough interest is expressed, a new newsgroup is created.

Postings to newsgroups are stored in specific files on a computer offering USENET. Access to these files is through a special program called a newsreader. An individual uses the newsreader program to create a list of newsgroups he is interested in reading. When the individual wishes to read postings in a newsgroup, he activates the newsreader program which then displays the newsgroups to which he has subscribed. The user picks a newsgroup, and the newsreader displays the subjects of messages in that newsgroup. The user then chooses which messages he wishes to read. There are different types of newsreader programs, but all programs allow a user to subscribe and unsubscribe to newsgroups, reply to postings and save messages to his mailbox. Unlike mailing lists, postings to newsgroups are not stored in a user's mailbox, but rather in files accessed by the newsreader. USENET uses its own protocol, the network news transfer protocol (nntp), to transfer messages between computers.

Each USENET site has an administrator, often the same individual who maintains the computer on which the newsreader is located. The administrator is responsible for deciding which newsgroups are offered on the local system. He can also create a number of newsgroups for discussing issues pertinent to the local network, institution or geographical area. Messages posted on the local system are sent to other USENET sites that accept posting for the same newsgroups. An individual can specify when posting a message how far the message should be distributed, e.g. local, regional or worldwide, and to what newsgroups the message should go. USENET is a very popular service and postings can take up large amounts of space on a computer system. Few computers have the resources to make all USENET newsgroups available, and the administrator selects which groups are offered on the local system. If a newsgroup is not available on the user's system, he can request that the administrator make the group available.

TELNET

The principal method of logging onto remote computers on the Internet is by means of the telnet protocol. This versatile protocol is a built-in feature of UNIX systems. Telnet turns a user's computer into a terminal of the remote computer, and only commands understood by the remote computer can be properly interpreted. Of all the Internet services, telnet gives a user the greatest potential access to a remote system's resources. Most publicly accessible telnet sites, however, limit the resources a user can access and a user is usually presented with a menu system after logging on. Depending on the site, the user may be prompted for a login name and, occasionally, a password. As an example, entering the command:

telnet rs.internic.net

will connect the user to the registration service at the InterNIC which offers several large databases of Internet related information. As with many sites, an information screen appears showing the services available and commands for accessing those services. Telnet can also be used to access services at remote sites that may not be available on the local computer. There are public sites that offer archie clients, gopher clients and text-based Web browsers programs. An example of the latter, at the University of North Carolina, allows public access to the text-based Web browser called lynx. To access this service, a user enters:

telnet sunsite.unc.edu
login: **lynx**

FILE TRANSFER PROTOCOL (FTP)

The file transfer protocol (ftp) is the primary method for transferring files on the Internet. There are thousands of publicly accessible sites on the Internet from which files can be downloaded to the user's local computer. The types of files available cover a wide spectrum from text files and documents to computer software of every imaginable variety. There are large archives on the Internet that specialize in certain types of files, such as software archives for Macintosh, Windows and UNIX systems, reference sources for specific subjects, and government documents to name a few. Unlike public telnet sites, all ftp sites require a login and password. For public ftp sites, the login name is "anonymous" while the password is usually the user's email address. With text-based interfaces, the user sees a prompt, usually ftp>, at which a command must be entered.

The file and directory structure of the remote system will determine how the data is arranged, but with UNIX systems (by far the majority of systems on the Internet) files and directories are arranged in a hierarchical structure similar to

that of DOS or Windows systems. With a UNIX client program, the following commands are available:

dir or ls	list the contents of the current directory
cd ⟨name⟩	change current directory to ⟨name⟩
get ⟨filename⟩	download ⟨filename⟩ to the local computer
cdup	move up one directory
quit	close the connection to the remote computer
help	list available commands
binary	switch to binary mode file transfer
ascii	switch to text (ASCII) mode file transfer

A directory listing using the "dir" command will display file and directory names as well as other information while the "ls" command lists just file and directory names. For ftp servers that employ the UNIX operating system, a directory is identified by the letter "d" in the first column of an entry and a file by a "–". See examples below. There are two methods of downloading files: binary mode and text (or ASCII) mode[3]. The mode of transfer is set by typing "binary" or "ascii" at the ftp prompt. ASCII mode is the preferred method for downloading text files while binary mode is used for downloading binary files. Binary files are computer programs and any files that contain nontext characters. These latter files include image, audio and video files as well as text files that contain control or formatting codes which are commonly added to many types of spreadsheet and word processing files. A binary transfer of a text file will usually cause no problems, but a text-mode transfer of a binary file will produce a severely garbled file. When in doubt about file type, perform a binary transfer first. If this does not produce a usable file, an "ascii" transfer can be tried.

An example of how to use "ftp" is shown below. To download the very popular Web browser, Mosaic, from the University of Illinois at Urbana-Champaign, the following commands are entered:

ftp ftp.ncsa.uiuc.edu
Name > **anonymous**
password > **glowniak@portland.va.gov**

The author's email address is entered as a password. After logging in, the user is at the main, or root, directory. A "dir" command is given to display its contents:

ftp> **dir**
–rw–r––r–– 16557 Jan 3 README
drwxr–xr–x 2048 Aug 22 Web

Only two of 35 lines of output are shown, and the output has been shortened to show particular items. The first string of characters describes file attributes. The first character in this string denotes the file type. README is an ordinary file (a "–" is the first character) while Web is a directory (a "d" is the first character). This format also indicates that the remote computer is a UNIX system. The next three groups of characters are the file length in bytes, the date the file was last altered and the file name. The next step is to change directories:

ftp> **cd Web**
ftp> **dir**
drwxr–xr–x 2048 Jan 6 Mosaic

Only the relevant directory listing is shown. After changing to the Mosaic directory, another "dir" command shows the subdirectories "Windows," "Mac" and "Unix." The Windows version of Mosaic will be downloaded. Changing to the Windows and Win31x directories and giving a "dir" command shows several files. To install the Mosaic program, two files are needed: mosaic20.exe and w32sOLE.exe. Installation instructions are in the README.TXT file in the Win31x directory. To retrieve these files, first change to binary mode:

ftp> **binary**
ftp> **get mosaic20.exe**
ftp> **get w32sOLE.exe**
ftp> **quit**

ARCHIE

Archie is a search program, written at McGill University in Montreal, Canada, for locating files at public ftp sites. Approximately once a week, archie searches most public ftp sites on the Internet that can be reached by anonymous ftp and creates a complete listing of all files and directories at these sites. This listing is added to a database that can be searched for keywords. Archie functions in a typical client-server mode and, for best results, the client should be run on the user's local system. Archie, however, is not a standard Internet service and the user's local system may not have an archie client. In this case, a user can reach a public archie client by telnet. There are over a dozen sites on the Internet that offer public archie clients.

It is important to realize that archie searches for directory and file names, not for contents of a file. If a user knows the name of a file, archie will locate that file, but if a user is looking for information on a specific topic, a certain amount of guesswork will be required. For example, if a user wants information on hepatitis, he can try search terms such as "liver," "hepatitis" or "hepatic." Each file that contains the search term in the file name or in the name of a directory leading to the file will be displayed along with the directory path to the file and the computer where the file is located. The user then needs to download and examine the files to see what information they contain.

On a UNIX system with an archie client, archie is accessed in the following manner:

archie [–h archie server] ⟨search term⟩

where "archie server" is name of a computer with a server and ⟨search term⟩ is a string of characters. The "–h" characters indicate the following string of characters is a host computer name. If the [–h archie server] term is omitted, as is usually the

[3] ASCII stands for American Standard Code for Information Interchange. ASCII is the standard method of encoding letters, numbers and punctuation into seven-bit binary numbers. This method of text encoding is used by all computer systems.

case, the archie client will connect to a predefined archie server. Depending on how the server is set up, archie can search for arbitrary strings of characters or whole file names. The user can specify which type of search he wants. If a user uses a search mode that looks for arbitrary strings, the search term "hepat" will locate "hepatitis" and "hepatic." An example of an archie search with a public archie client is shown below. The search will be performed at McGill University in Montreal, Canada for a Windows program, hopchkw.exe (Windows hop check), that implements a service called traceroute:

telnet archie.mcgill.ca
login: **archie**

A password, "archie," may be required at other sites. After logging in, the archie prompt is displayed. Typing "help" activates the help program that explains how the program works. Typing "?" at the help prompt will display a list of commands. The command "prog" followed by a keyword is used to search for files:

archie> **prog hopchkw.exe**
ftp://ftp.isri.unlv.edu/private/dos/winsock/hopchkw.exe
Date: 20:00 11 May 1994 Size: 27392 bytes

Archie found the program on the computer ftp.isri.unlv.edu which is at the University of Nevada at Las Vegas. The computer, directory path and file name are given in the URL format (see below under World Wide Web for an explanation of URLs). The date the file was put on the system and its size are also specified.

GOPHER

One of the major drawbacks of telnet and ftp is that one needs to know what information is located at what sites. To be able to use the Internet effectively, a user would have to know a lot about the location of resources. To overcome this problem, the gopher program was written. This program, developed at the University of Minnesota and named after its mascot, was made freely available to the Internet community in 1991. Gopher is one solution to the problem of finding resources at widely separated sites without knowing where these resources are located.

Superficially, a text-based gopher program appears similar to telnet. Gopher operates in the typical client-server mode. A user accesses a gopher client by giving the gopher command followed by the name of a site that runs a gopher server. If the gopher command is given without specifying a server, the client will connect to a predetermined, or default, gopher server. At many institutions, the default server is the institution's own gopher server. Unlike telnet, no login or password is required. After connecting to a gopher server, a menu is returned that appears very similar to menus seen at public telnet sites. Picking a menu item retrieves another menu or a document in a manner similar to telnet. The difference, however, is that the menu item may be located anywhere on the Internet. Associated with each menu item is a location and a path to a file or other directory located either on the original gopher server or some other gopher server on the Internet. When a menu item is chosen, the local gopher client makes a connection to the associated location, downloads the file or menu and then terminates the connection. If a menu is downloaded, its items may be linked to other gopher servers. Thus, by specifying a single gopher server, a user could potentially have access to all the gopher servers on the Internet. Whether or not this is possible will depend upon the links the original gopher server has to other servers. These links are established by the individual who sets up the gopher server and creates the menus. Some gopher servers have very few links to other servers while others, such as the gopher server at the University of Minnesota, have links to thousands of other servers. A gopher client connecting to this server can access a large proportion of all the publicly available gopher servers on the Internet.

Because all gopher servers can be linked together, the total amount of information on the Internet that can be accessed by the gopher program is often referred as gopherspace. Gopherspace defines a distributed document system where information is located using descriptive terms (i.e., the menu items) rather than by stating where the information is located. One can think of gopherspace as a tree with branches and leaves. The trunk of the tree is the first gopher server accessed. The branches are menu items. If a menu item is another menu, that branch gives off further branches. If a menu item is a file, the branch ends in a leaf where information is located. The gopher client keeps track of a user's path along this tree and allows a user to retrace his steps backward to the trunk.

The types of files that gopher can access are quite general. The gopher program has built-in abilities to display files from ftp sites and can retrieve information using search programs such as WAIS and CSO. Although gopher can download image, sound and movie files, it does not have the ability to display or play these files by itself. External programs, called viewers, can be linked to the gopher client to perform these functions. The computer on which the client resides must, of course, also have the hardware capabilities for image, audio and video support. Gopher programs can access documents on the World Wide Web if a Web browser is linked to the gopher client as a viewer.

As an example of how to use the gopher program, a session will be described with the University of Minnesota gopher server. A user gives the following command:

gopher gopher.tc.umn.edu

The UNIX shell activates the gopher client which contacts the University of Minnesota gopher server. The server returns the main menu shown below and then terminates the connection:

Internet Gopher Information Client v2.1.3

Home Gopher server: gopher.tc.umn.edu

→ 1. Information About Gopher/
2. Computer Information/

3. Discussion Groups/
4. Fun & Games/
5. Internet file server (ftp) sites/
6. Libraries/
7. News/
8. Other Gopher and Information Servers/
9. Phone Books/
10. Search Gopher Titles at the University of Minnesota ⟨?⟩
11. Search lots of places at the University of Minnesota ⟨?⟩
12. University of Minnesota Campus Information/

Press ? for Help, q to Quit Page: 1/1

Some of this text is added by the client, such as the title "Internet Gopher Information Client v2.1.3" that identifies the version of the client software that was obtained from the University of Minnesota. The client identifies the gopher server in the second line and numbers the menu items returned from the server. The client adds identifying information at the end of each menu item. The following symbols are used:

/	a directory
⟨?⟩	a keyword search item
⟨Picture⟩	an image file
⟨Movie⟩	a video clip

No symbol at the end of an item indicates a text file[4]. The client also provides a help line at the bottom of the display screen. Typing "?" will display a list of commands that control how gopher functions. Gopher clients that use GUIs often provide similar functions in pull-down menus. Finally, the client lists, in the bottom right-hand corner, the current and total number of screens of text in the menu. A user chooses an item by moving the arrow on the left of the menu to a desired item and then hits return or types an item number followed by a return.

Choosing item 1, "Information About Gopher," from the main menu causes the client to reconnect to the University of Minnesota gopher and download another directory.

View item number: **1** Connecting... Retrieving Directory...

Internet Gopher Information Client v2.1.3

Information About Gopher

→ 1. About Gopher
2. Gopher News Archive/
3. GopherCON '95/
4. GopherCON '94/
5. Gopher Software Distribution/
6. Commercial Gopher Software/
7. Gopher Protocol Information/
8. University of Minnesota Gopher software licensing policy
9. Frequently Asked Questions about Gopher
10. Gopher+ example server/
11. comp.infosystems.gopher (USENET newsgroup)/
12. Adding Information to Gopher Hotel
13. Gopher T shirt on MTV movie (big) ⟨Movie⟩
14. Gopher T shirt on MTV movie (small) ⟨Movie⟩
15. Gopher T-shirt on MTV #1 ⟨Picture⟩
16. Gopher T-shirt on MTV #2 ⟨Picture⟩
17. How to get your information into Gopher
18. Reporting Problems or Feedback

Press ? for Help, q to Quit, u to go up a menu Page: 1/1

Notice that in the help line at the bottom of the display, a new command "u" appears which allows a user to return to the previous menu. Item 6, "Commercial Gopher Software/", is chosen next.

View item number: **6** Connecting... Retrieving Directory...

Internet Gopher Information Client v2.1.3

Commercial Gopher Software

→ 1. AIRGopher
2. Chameleon
3. GUIDE for Macintosh
4. GUIDE for Windows
5. WinGopher

Press ? for Help, q to Quit, u to go up a menu Page: 1/1
Really quit (y/n) ? **y**

This menu has five items, all of which are text files describing various types of proprietary gopher client programs. A user exits the program by typing "q" and then "y" to confirm his choice.

One important feature of gopher clients is the ability to make entries, called bookmarks, to a user-created list of gopher menu items. With the particular client shown, a user simply types "a" when the arrow is next to a menu item to add it to the list. The list is created automatically when the first bookmark is chosen. This bookmark list, which can be viewed at any time by typing "v", is used like any other gopher menu. The bookmark list is, in effect, a user's own personalized gopher server.

VERONICA

Although gopher allows a user to find resources throughout the Internet without needing to know where they are located, the process of finding information is not very efficient. A user has to follow links on menus, but there is no guarantee that the links lead to what the user wants or that the links are the most appropriate resources. A partial solution to this problem was the creation of a program called veronica. Veronica is to gopher as archie is to ftp. As with archie, the veronica program periodically searches all of gopherspace (approximately once per month) and creates a database of all menu items it finds. In order to do this, veronica starts at the University of Minnesota gopher and walks through the menus recording each item and following each link. When a link leads to another gopher server, veronica connects to that server and follows a similar process, eventually finding all the menu items of all gophers linked directly or indirectly to the University of Minnesota gopher.

[4] GUI clients typically identify menu items by icons placed at the left of each item.

Veronica functions in a client-server mode, but there are no specific veronica clients. A veronica server is accessed through a gopher client that links to gopher sites with a veronica server. A user chooses the veronica menu item and is asked for a keyword search term. Veronica searches its database and returns a menu whose items contain the keyword along with the links for these items. A user chooses an item from this menu and is connected to its gopher server. One can think of a veronica server as a gopher server that creates customized gopher menus. Veronica was created at the University of Nevada and there are now about two dozen veronica servers on the Internet. Any gopher server can contain links to these sites and there are several gopher servers that contain a list of all the known veronica sites. Some of these gopher servers provide information on which veronica servers are currently active. Unfortunately, veronica servers tend to crash frequently. Many gopher servers have menu items that link them to veronica servers. Two examples are the servers at the University of Minnesota, gopher.tc.umn.edu (under "Other Gopher and Information Servers/" on the main menu) and the server at the University of Nevada at Reno, futique.scs.unr.edu.

A restricted version of veronica is jughead. Jughead performs the indexing and search functions of veronica on a single gopher server. The administrator of the gopher server decides whether to make jughead searches available at his site.

WORLD WIDE WEB

The Internet service that is responsible for the explosive growth of the Internet in the last two years is the World Wide Web (WWW). Like gopher, the WWW is a distributed document system with links between documents. One of the main differences between these two services is how information is accessed. Gopher systems are strictly hierarchical. Menu items are either directories or files that are arranged in a tree-like structure with all the information stored in files. A user moves sequentially down the tree until a file is reached. If one wants more information about a topic in a file, one can retrace the path along the tree to see if there are related files on the topic of interest. This information may or may not exist at some point further up or down the tree but, in general, there is no well-defined method for finding related information on a topic. Although files on WWW servers are arranged in a fashion identical to those on gopher servers, the method of accessing WWW files is more flexible. A user can jump to any document on the WWW without first accessing the intervening directories. Unlike gopherspace where the information is contained in files that are at isolated end points, all documents in the WWW can contain links to other documents making the structure more analogous to a Web than a tree with branches and leaves. WWW documents can contain text, images, graphics, video and sound clips, or any combination of these elements. Within a document, words, phrases or even pictures are highlighted. Choosing the highlighted item, by clicking on it with a mouse, retrieves and displays a linked document that itself may contain highlighted items that are linked to other documents. Text written in this format is called hypertext, while documents containing images, videos or sound files are referred to as hypermedia, although the term hypertext is commonly used to describe both types. Documents available on the WWW are called Web pages. The main entry point at a Web site is called the home page, a term without a precise meaning since any document at a site could be so designated.

Another major difference between gopher and the WWW is the format of documents. While both services can display text, image or video files, gopher can only display one type at a time. WWW pages, on the other hand, can contain text and multimedia in the same document because Web browsers for GUI systems have built-in capabilities for displaying the most common types of image files. Images that are displayed as part of a Web page are called inline images.

The idea of hypertext did not originate with the WWW. The Hypercard program on Macintosh computers first used this system extensively, while the on-line help system of Microsoft Windows is also based on hypertext. The WWW, however, was the first Internet service to use hypertext as the basis of a distributed document system. Hypertext is a more natural way of linking documents together. For example, a document describing wildlife in the North America may have the words "cougar," "robin," and "North America" highlighted because a description of these terms may not be appropriate to the rest of the document. Choosing the highlighted word "cougar" would return another document that might describe cougars in more detail. This text itself could have highlighted words, along with pictures of cougars or a map of their range in North America.

WWW documents are written in a special format called the hypertext markup language (html). See Appendix A for a discussion of html. This language inserts special formatting codes into the documents that allow WWW client programs to display Web pages properly and permits the embedding of links to other files within the document. This contrasts with documents that can be displayed by telnet, ftp and gopher clients which are simple text. The method of transferring Web documents between computers is defined by the hypertext transport (or transfer) protocol, http, which normally functions on port 80. As with gopher, as soon as a document is transferred from a Web server, the connection to the client is terminated.

When the WWW was first introduced on the Internet in 1991, it was not initially widely used for two main reasons. While it is a straightforward procedure to make simple text documents available by telnet, ftp and gopher, the same documents would have to be reformatted in html in order to have links within the document. Web browsers can display simple text files but, without links within a document, a Web browser is not much different than a gopher client. Until a large number of sites began offering html documents through WWW servers, there was simply not a lot of information accessible through the WWW. Secondly, the first WWW clients were text-based interfaces and, in many cases, these did not have an obvious advantage over the menu-based format of gopher programs.

The number of sites on the WWW gradually increased until November 1993 when the WWW entered an explosive growth phase. The sudden increase in the size and popularity of the WWW was due to the release of a program, called Mosaic, by the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign. Mosaic is a WWW client program that can download and display not only hypertext but linked image files. When properly linked with additional programs, Mosaic can also play sound and video files. Mosaic was made freely available to the entire Internet community in Windows, Macintosh and UNIX versions. The response was enormous. Within a few months, millions of copies of Mosaic were downloaded from NCSA. The number of people using the WWW greatly accelerated the creation of Web sites, especially in the commercial sector. NCSA continues to improve Mosaic, offering new versions at regular intervals. The commercial potential of Mosaic was quickly realized and several companies licensed the technology from NCSA to create proprietary versions of Mosaic. The most popular of these proprietary products is Netscape by Netscape Corporation. Some programs related to Mosaic can be found in Appendix B.

Mosaic and its related programs are called Web browsers and they are more than just WWW client programs. Web browsers have built-in capabilities to access gopher, WAIS and ftp servers. Depending on the browser, built-in capabilities for accessing telnet, USENET and email are also available or can be linked to the program using additional software that can perform these functions. Web browsers have built-in capabilities to display the more common types of image files, such as the widely used GIF and JPEG files.

In general, Web browsers have more extensive hardware and software requirements than software for the other Internet services. To run Mosaic and related programs, a computer must have image display capabilities. These features are standard on Macintosh computers and on any system that can run Windows or IBM's OS/2 operating system. UNIX systems require a special display environment called the X Window system. For personal computers, the minimum memory requirement is 4 megabytes of random access memory (RAM), although 8 megabytes are recommended. For personal computers accessing the Internet through a modem, the TCP/IP software must be run on the user's computer rather than on the remote computer of the Internet access provider.

Web browsers use a special notation for accessing resources on the Internet called Uniform Resource Locators (URLs). Every file on the Internet can be described in terms of its access method, location on a given computer, port number, directory path to the file and the file name itself. URLs incorporate this information into a character string in a standardized format. The formal names used in URLs for the access methods are telnet, ftp, gopher, http (for the World Wide Web), wais, news (for USENET) and mailto (for email).

An example of a URL for a Web page on the WWW at the University of Utah is:

http://www-medlib.med.utah.edu:80/WebPath/webpath.html

The main components of a URL are illustrated above: (1) an access method, http, followed by a colon and two forward slashes; (2) the name of the computer on which the file resides followed by a colon, a port number and a forward slash (the colon and port number are included for illustrative purposes only; since the WWW normally resides at port 80, they could be omitted in the above example); (3) a directory path to the file of interest followed by a forward slash; and (4) a file name[5]. The computer name can be entered in upper or lower case. The access method is read by the browser while the directory paths and file names are read by the server. Some browsers, such as Netscape, will accept the access method in upper or lower case while others, such as Mosaic, accept only lower case. The directory paths and file names are case sensitive for UNIX servers but case insensitive for Macintosh (e.g., http://www.apple.com) and Windows (e.g., http://www.microsoft.com) servers. Since a user will rarely know the type of server he is accessing, it is best to enter URLs exactly as given. A UNIX server will return a message that the file cannot be located if a path or file name is entered in the wrong case. No spaces are permitted in a URL. If the file name and directory path are omitted, a Web browser will display a default document that is often the home page at the Web site. Each directory in the directory path can have a default document associated with it (called index.html on UNIX systems) that is displayed when only the directory path, or a portion of it, is entered. If no default document is associated with a directory, a directory listing is returned for the last directory listed. When a Web browser is used to access a gopher or ftp server and the directory path is not specified, the main menu at that site is displayed. Some browsers and some sites require a trailing "/" if only a computer name or a computer name and a directory path are given.

Web browsers have a host of interesting features only a few of which will be described here. Web browsers differ somewhat from each other, so it's not possible to give a detailed generic description of a Web browser. In the following discussion, the Windows version of Mosaic will be described. The Macintosh and UNIX versions are very similar. Since Mosaic is freely available, anyone can download and run this program. In addition, most commercial Web browsers are based on Mosaic and have functions similar to those described below.

An example of a Web page from Loyola University as displayed by Mosaic is shown in Figure 1. There are six main parts to this display. From the top downward, they are: (1) the title line; (2) the menu bar; (3) the toolbar; (4) the URL indicator line; (5) the display area for Web pages; and (6) the status line. The Web document title is displayed on the title line[6]. Each word in the menu bar has an associated pull-down menu. These menus can be seen by moving the pointer to an item in the menu bar and clicking the left mouse button. A menu item

[5] For email and USENET, the format of a URL is somewhat different. The URL format for email is mailto:⟨address⟩ where ⟨address⟩ is an email address. For USENET, the format is news:⟨newsgroup⟩ where ⟨newsgroup⟩ is name of a newsgroup such as sci.chem mentioned above.

[6] A title is an optional element of a Web page, although it is strongly recommended that authors include titles for Web pages.

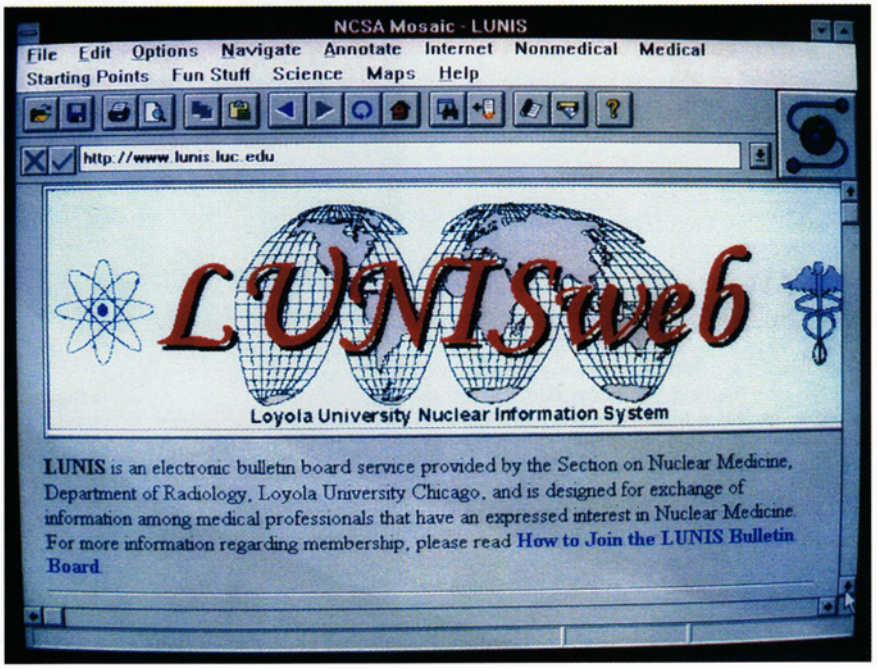

FIGURE 1. The home page of LUNIS, the Loyola University Nuclear Information System, a nuclear medicine bulletin board. The URL of this Web page, http://www.lunis.luc.edu, is shown immediately above the display area. Text highlighted in blue is linked to another document. The actual Web page is considerably longer than the displayed text and can be viewed by using the scroll bar on the right. The Web browser used here is NCSA Mosaic, final beta version 2.0 for Windows.

is selected by moving the pointer to that item and clicking on it. The toolbar contains buttons that perform various functions by clicking the mouse on the buttons. For example, the arrow buttons allow a user to move forward or backward to previously viewed Web pages. The URL indicator line shows the complete URL of the currently displayed Web page. A user can manually enter a URL on this line to access a new Web page. The main portion of the screen is used to display the Web page. The user can remove some or all of the other elements mentioned above to create a larger display area for the Web page. Many Web pages are too large to be displayed on a single screen, so the scroll bars to the right and at the bottom of the screen can be used to view different parts of the document. The last element in the display is the status bar at the bottom. As a document is being transferred, information appears in this area describing the status of the transfer, such as the size of the document and what portion of the document has been received.

One very important feature of Web browsers is the hotlist. This is a list of URLs a user creates that can be accessed through pull-down menus from the menu bar. Hotlists are the equivalent of bookmark lists with gopher clients. Figure 1 shows several hotlists I have created: Internet, Nonmedical, Medical, Fun Stuff, Science and Maps. Starting Points is also a hotlist that comes preinstalled with Mosaic. A user adds an item to a hotlist by choosing a hotlist and clicking on the "Add To Current Hotlist" item under the "Navigate" menu. The title of the document is then added to the hotlist. As an example, Figure 2 shows my Medical hotlist, opened from the menu bar. Moving the pointer to any item on the list and clicking the left mouse button transfers that Web page to the browser. In Figure 2, I have moved the pointer to "Welcome to MedSearch America", the main Web Page for a company advertising available positions in several medical related fields, including nuclear medicine technology. After clicking the left mouse button, the Web page in Figure 3 is retrieved and displayed.

When a Web browser is first started, it reads information from a start-up file. This start-up file, called mosaic.ini in Windows systems, contains a large number of parameters that control the function of the browser. These parameters can be changed by using the "Options" item on the menu bar and selecting "Preferences. . ." One piece of information a browser needs is the initial Web page to be displayed when the browser is started. This page is specified by a URL in the start-up file. Most browsers are preconfigured to display the home page of the company or institution that created the browser. A user, however, may want to display a different Web page when starting the browser, such as the Web page I have created for my own use shown in Figure 4. With Mosaic, the initial Web page can be changed using the "Preferences. . ." selection mentioned above. To use custom-made Web pages, a user creates an html document, as described in Appendix A, and stores it on the computer's hard drive. The browser is then instructed to use this document as the initial Web page.

One of the functions of a URL is to tell a Web browser which server to contact to access a file. A Web browser, however, can also directly access files on its own computer without passing the files through a server. In order to do this, any reference to a local file, such as a link in a Web page, must use the following syntax:

file:///drive/path/filename

or

file://localhost/drive/path/filename

In this syntax "drive/path/filename" is how a file is identified by the local operating system; file:/// or file://localhost tells the

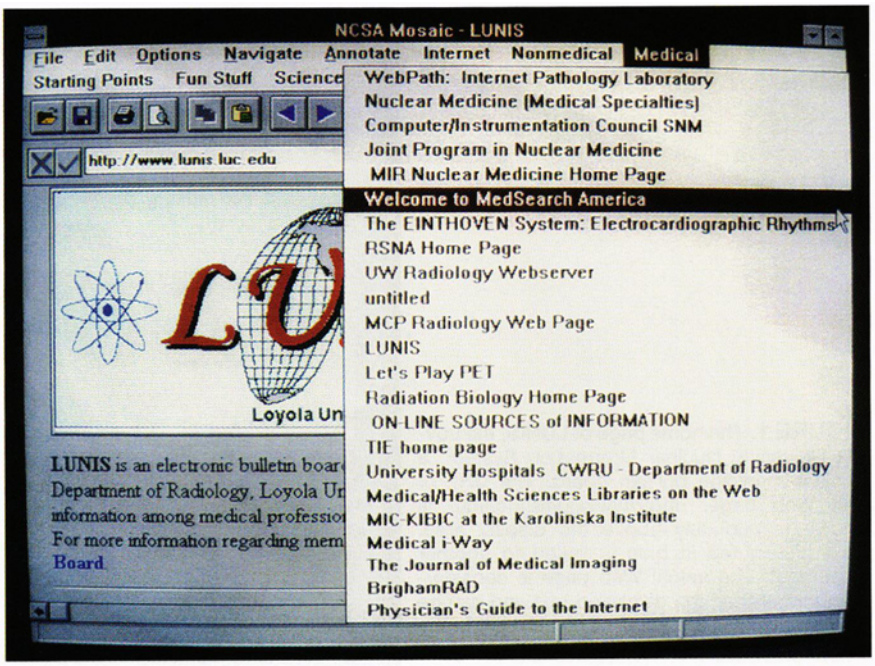

FIGURE 2. A pull-down menu associated with the hotlist "Medical." This list is displayed by placing the arrow over the word "Medicine" in the menu bar and clicking the left mouse button. The items in this list are the titles of the Web pages, not their URLs. Sliding the indicator line over an item and clicking the mouse button connects to the title's URL. In this example, Welcome to MedSearch America was chosen and is displayed in Figure 3. Notice item 10, "untitled." This Web page did not have a title.

browser the following file is on the local computer. For example, a text file called phones.txt in the windows directory on the C drive of a PC could contain a list of phone numbers. If a user wanted to link this file to a local Web page, the file would be referred to as:

file:///C:/windows/phones.txt

or

file://localhost/C:/windows/phones.txt

Some browsers, such as Mosaic, do not require "drive/" if the file is located on the same drive as the browser. One could use Mosaic to display files and images located on the local computer without having a connection to the Internet or any other network. This could be done by creating Web pages with links only to local files. In fact, most Web browsers have options that allow a user to select and display files directly from local disk drives even if the files are not linked to any Web pages.

For any Web browser to display image or video files or to play audio files, it must have appropriate software for handling these files. These multimedia files are written in multiple different formats and corresponding software programs are needed for handling each type of format. Most Web browsers have built-in capabilities to display the more common types of

FIGURE 3. The home page of MedSearch America, URL http://www.medsearch.com. In order to show as much of the page as possible, presentation mode is used which removes from the screen all graphic elements of the Mosaic program except for the scroll bar on the right. Presentation mode is activated and deactivated by pressing the Alt and P keys simultaneously. MedSearch America is an information service for the allied health professions. The circled text in the image is linked to other Web pages. As in Figure 1, this page is too long to be displayed in its entirety.

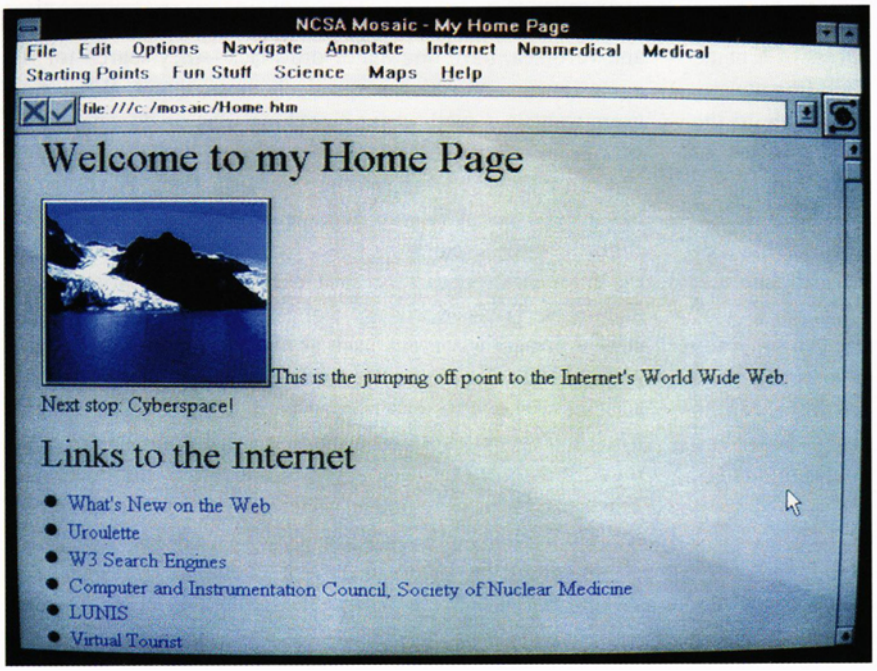

FIGURE 4. A Web page created for use as the initially displayed document when Mosaic is started. The Options item from the menu bar has been used to turn off the display of the status bar and the tool bar. The six lines at the bottom are links to other sites. Note that since this page is read directly from the computer's hard disk, its format starts with file:///. The html document for this page is shown in Figure 5.

image files: GIF and JPEG. There are, however, many other image formats that cannot be directly handled by Mosaic and most video and audio formats are also not supported. In order to display these other types of files, Web browsers can be linked to programs, generically referred to as external viewers, that can display these files. When a browser downloads a file it cannot directly display, it checks its start-up file to see if there is a viewer on the local system to which it can send the file. The user must specify the name of the external viewer and its location on the local system in the start-up file. These viewers may or may not be available as part of the software that comes with the operating system. If they are not, a user has to find and install the viewers he desires. External viewers are required for more than just multimedia files. For example, some text files, such as word processing files, are also written in special formats and, for these types of files to be displayed, the word processing program must be linked to the browser through the start-up file. Fortunately, the procedure for linking a viewer is now a fairly automated process with most browsers, but it is still one of the more difficult tasks in making a browser a highly versatile display tool.

OTHER INTERNET SERVICES

There are a large number of other Internet services. Some of these are defined by standard Internet protocols and others are implemented by various commercial and noncommercial software programs. A few of these services are described below.

Finger is a standard Internet protocol that provides information about users on a system. Finger is a built-in feature of all UNIX systems. The format of the command is:

finger [user name]@[computer name]

If the finger command is given without arguments, a list of users who are currently signed onto the local system and the programs they are running are displayed. If only [username] is given as an argument of the finger command, finger returns more complete information about a user on the local system. On UNIX systems, this includes the contents of a file, the .plan file, that can contain any information a user may wish to make available. The file may contain simple text or a list of commands that are performed whenever that user name is accessed. If the user is on a remote system, [computer name] must be specified. If only @[computer name] is given, users who are currently signed onto the remote system are listed. As an example of how finger can be used to disseminate information, the command:

finger copi@oddjob.uchicago.edu

returns the .plan file created by the user "copi" on the computer oddjob.uchicago.edu at the University of Chicago. The file lists significant events that have occurred on the current date. Because finger can provide important information about a system and its users, many sites have disabled this service for remote users.

The domain name system (DNS) in one of the most important standard Internet protocols. The primary function of DNS is to provide IP addresses for computer names. As discussed in the first article in this series, communication on the Internet is done in terms of IP addresses; computer names are a convenience for humans who find it difficult to remember IP addresses. DNS clients obtain IP addresses from DNS servers that are usually referred to as name servers. These are specifically defined computers that maintain lists of computer names and their associated IP addresses. The servers return the IP addresses of computer names. Name service is such an important and widely used function, that DNS clients are set up to automatically query name servers and return IP addresses whenever computer names are given in a command. Some

software programs, such as TCP/IP software for personal computers, may require a user to enter the IP address of a name server when installing the program. For most TCP/IP software, the DNS client functions automatically and is invisible to the user. Name servers provide more information than just IP addresses for computer names. Some systems have commands that allow a user to directly access a name server for information about the computers and the computer networks that they serve. On UNIX systems, these commands are nslookup and dig.

Ping is one of the simplest services a computer can run. Ping sends a single, small data packet to a remote computer. The remote system sends back a reply that the packet was received. The purpose of ping is to determine if a computer is active and connected to the Internet. The structure of a ping command is:

ping ⟨computer name⟩

The response is either that the computer is active or that the computer cannot be contacted, or no response is received. If a remote computer does not respond to ping, it will usually not respond to any other service request from the local computer. As an example, if a user cannot ping a remote computer name but can ping the computer's IP address, either the user has typed an incorrect computer name or there is a problem accessing the name server of the local computer.

Traceroute is a nonstandard Internet service that is a sophisticated use of ping. The traceroute command is similar to ping:

traceroute ⟨computer name⟩

Traceroute displays a listing of all the computers, routers and networks through which a packet travels on its way to the remote computer. It performs this function by sending out a series of ping commands to the intervening computers and routers. Equivalent commands on other systems are hopchkw on Windows and tracerte on OS/2 Warp.

One of the most frequently asked questions is how to find the email address of an individual on the Internet or other information such as his business address or phone number. Although efforts are being made to provide the equivalent of phone books for individuals on the Internet, any comprehensive listing is still a long way off. Nevertheless, a variety of protocols and indexing methods are available, and only a superficial discussion is presented here. One of the first attempts to provide directory service at local institutions was through CSO servers, sometimes called qi servers, named after the Computing Systems Office of the University of Illinois. This service was introduced as a type of electronic phone book. It is commonly found on gopher servers at many universities and allows keyword searching. The problem with this service is that a user must know where an individual is located. More comprehensive directory services are provided by a number of protocols such as the whois and X.500 protocols. A good example of the whois service is available at the interNIC which can be reached by telnet (rs.internic.net, login: whois), gopher (gopher.internic.net) and the WWW (http://www.internic.net).

The main Internet search programs described so far, archie and veronica, have one major drawback—they search for file names and menu items, respectively, not content. What most users want is a method for searching files or documents for specific information. This is accomplished by a service called WAIS (Wide Area Information Service). Each WAIS server has a database of related documents, such as files on Asian culture. WAIS creates an index for each document that lists the number of occurrences of each nontrivial word in the document. Trivial words are words that have little or no specific information content, such as and, or, the, not, which, etc. A WAIS client accepts keywords and sends them to a user selected server. The server returns not only a list of documents with the keyword, but also a ranking based upon several criteria such as how often the keyword appears in the document. This ranking is a method of determining the relevance of a specific keyword in a particular document. A user can pick a document in the list to display its contents. The value of a WAIS client is dependent upon the number of WAIS servers it can contact, each of which is dedicated to a specific topic. There are now about 1,000 such servers on the Internet. It would be impossible for a user to be familiar with all these servers and a database of WAIS servers, the directory of servers, has been created to help search for a particular server. A public WAIS client is available on the WWW at http://www.wais.com.

WAIS is one of the most important search functions on the Internet, but it is also one of the most difficult to use. Many keyword search capabilities on telnet, gopher and WWW servers access WAIS servers through gateways in a manner that is transparent to a user. For users who need more in-depth search capabilities of these WAIS databases, commercial WAIS clients are available.

A large number of Internet services are available as homegrown programs on specific computers that can be reached by telnet. There are too many of these programs to describe even superficially. Two examples have been listed previously, the University of Michigan's Underground Weather Service and netlink. The WWW provides a rich source of search programs and other services that will be described in more detail in the last article in this series, "Resources on the Internet."

APPENDIX A

Creating Hypertext Markup Language (html) Documents

One of the characteristics of html documents that makes them easy to create is that they are simple text files. Although they contain special formatting codes, tags, these tags are themselves text. This contrasts with most word processing programs that add nonprintable, proprietary formatting codes to documents they create. These codes can be read and have meaning only to the word processing program or printers. A user can create an html document using one of these word processors, but the program must be instructed to save the document as a simple text (or ASCII) file, in other words without the proprietary formatting codes. The word processing

```
<TITLE>My Home Page</TITLE>
<H1>Welcome to my Home Page</H1>
<P>
<IMG SRC="file:///c:/internet/gif/glacier__.gif">
This is the jumping off point to the Internet's World Wide Web.<BR>
Next stop: Cyberspace!
<H2>Links to the Internet</H2>
<UL>
<L1><A HREF="http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/whats-
new.html">
What's New on the Web</A>
<L1><A HREF="http://www.ukans.edu/uroulette.html">
Uroulette</A>
<L1><A HREF="http://cuiwww.unige.ch/meta-index.html">
We Search Engines</A>
<L1><A HREF="http://gamma.wustl.edu:80/tf/caic.html">
Computer and Instrumentation Council, Society of Nuclear Medicine</A>
<L1><A HREF="http://www.lunis.luc.edu">
LUNIS</A>
<L1><A HREF="http://wings.buffalo.edu/world">
Virtual Tourist</A>
</UL>
```

FIGURE 5. The html file, home.htm, for the Web page shown in Figure 4. See Appendix A for an explanation of symbols. For Web pages stored on Windows systems, html files are identified by the file extension "htm."

program Notepad with Windows creates simple text files. Figure 5 is an example of a Web document that I created and Figure 4 shows how this document is displayed by Mosaic. Tags are enclosed within angled brackets ⟨ ⟩.

The tags in Figure 5 are of two general types. Tags that perform a simple function, such as indicating where a paragraph begins ⟨P⟩ or where a carriage return ⟨BR⟩ should be inserted within text, are single symbols called standalone tags. Most tags, however, occur in pairs such as ⟨UL⟩ and ⟨/UL⟩ indicating where an unordered list of items, each identified by ⟨LI⟩, begins and ends. An ending tag always contains a forward slash. The most important of these tags are the anchor tags. The beginning anchor tag requires another piece of information to define the anchor—a reference to linked information. Only one reference type is used in Figure 5, the hypertext reference HREF. This indicates that the following text in quotes is a URL or a local file. The text between the anchor tags is displayed by the Web browser as a highlighted item. In Figure 4, the text is highlighted in blue. The wording of the text is arbitrary but descriptive of the underlying link. In fact, the text could be replaced by the name of an image file, in which case a picture is displayed on the Web page. When the user clicks on the picture, the underlying link in the anchor tag is retrieved. See below for a further explanation. An item in a list does not need to be linked to a URL. Text after an ⟨LI⟩ tag without a link is displayed as ordinary, nonhighlighted text.

The tags in Figure 5 are as follows:

⟨TITLE⟩ Name ⟨/TITLE⟩	Document title
⟨H1⟩ Heading ⟨/H1⟩	Main document heading
⟨P⟩	Begin a paragraph
⟨IMG SRC="file"⟩	Insert an image specified by "file" (file is a URL or a local image file)
⟨BR⟩	Break; insert a carriage return
⟨H2⟩ Subheading ⟨/H2⟩	Secondary document heading
⟨UL⟩ Items ⟨/UL⟩	Unordered list of items
⟨LI⟩	List item
⟨A reference⟩ text ⟨/A⟩	Anchor

The title tag identifies the enclosed text as the title of the document. The title is different from the URL of the document and is a readable version of what the document contains. Mosaic and other browsers display the document title in different ways and use the title to identify documents in hotlists. The ⟨H⟩ tags (header tags) are a method of creating subdivisions with a document. Note that simple text, such as "Welcome to Cyberspace," does not have any tags. It is displayed

sequentially after the previous tag or text. The important part of this document is the list of links or anchors. For brevity, I have inserted only six links in this document, but many more could be added. Each beginning anchor tag contains a reference. In this document only one type, the hypertext reference, is used. Its format is HREF and a following URL, such as HREF="http://cuiwww.unige.ch/meta-index.html". The immediately following text is highlighted and displayed by the browser.

In this document, one image is included, a GIF file showing the scene of a glacier. This image could have been linked to another document if the reference to the image file had been placed between anchor tags. For example, assume there is an explanation of glaciers in the file glaciers.txt on the computer earth.geo.edu and that the URL of this file is http://earth.geo.edu/land/glaciers.txt. Then, if line 4 in Figure 5

⟨IMG SRC = "file:///c:/internet/gif/glacier_.gif⟩

was replaced with

⟨A HREF = "http://earth.geo.edu/land/glaciers.txt"⟩
⟨IMG SRC = "file:///c:/internet/gif/glacier_.gif⟩⟨/A⟩

the resulting document would be displayed exactly like the original one except for a blue border around the image. This would indicate that the image is linked to another document, and clicking on the image would retrieve and display the file glaciers.txt from the computer earth.geo.edu.

For individuals who do not want to learn html, there are several html editors available on the Internet, as well as commercial software that will create Web documents. These editors are similar to word processing programs and automatically insert appropriate tags in the document when a specific function is requested.

APPENDIX B

Software for Internet Services

There is a large amount of software for the various Internet services. This software can be obtained from both commercial and noncommercial sources. Many software programs are freely available from several sites around the Internet. Although freely available, these programs usually have some restrictions on their use. Freeware refers to programs that can be downloaded and used free of charge. Shareware is freely available with the request that any individual who uses the program registers with the developers and pays a small fee, usually less than $50.00. Some programs by commercial developers are offered for free use with the restriction that they are used only by individuals or on a small scale by educational institutions. These programs are often early versions of software that will be offered for sale at a later date when a more mature product is developed. Finally, some programs are freely available but can only be used in conjunction with a specific on-line access provider who charges for the time the user is on-line. Freeware and shareware are usually made available with a requirement that they will not used or sold for commercial purposes without the express consent of the developers.

Software is available for both clients and servers for all the common Internet services for all types of operating systems. Since most servers on the Internet use the UNIX operating system, the majority of server software is written for UNIX systems. This appendix does not discuss server software, but those who wish to set up servers on their personal computers should be aware that such software exists.

Freeware and shareware usually consist of individual programs for each of the Internet services. There are many ftp sites on the Internet that offer these programs. A listing of a few of these sites is given below. The user should read the various README or other documentation files at each site to get an idea of what programs are available and where they are located. Programs are often in subdirectories listed in the pub directory on the main menu. The most widely used TCP/IP shareware for Windows is Trumpet Winsock (tcpman.exe is name of the program). Specific programs can be found by archie searches. The program Winsock FTP for Windows (ws_ftp.exe) contains a list of sites with Internet software. A user can download this program and review this list even if he does not actually connect the program to the Internet. More general topics can be found with veronica searches. For example, a search term "macintosh archives" will return a list of sites where Macintosh programs can be found.

Many commercial packages are bundled software, in other words they contain multiple software programs for Internet services and TCP/IP software. Some are preconfigured to access a given Internet access provider. All the major online services offer Internet access. See Part 2 of this series for a description of these services. A few of the better known commercial software products are listed below. The reader should be aware that software products change often and more up-to-date information is available in many computer-oriented magazines and publications.

Software Available by ftp on the Internet:

Address	Comments
ftp.ncsa.uiuc.edu	National Center for Supercomputing Applications, University of Illinois. Many programs available including viewers. Mosaic is in the Web/Mosaic directory.
fatty.law.cornell.edu	Cornell Law School. Cello, a Windows Web browser not based on Mosaic, is found in the pub/LII/Cello directory.
ftp.netscape.com	Netscape, the most popular commercial version of Mosaic, is found in the netscape directory. This version is free subject to licensing restrictions.
ftp.iastate.edu	Iowa State University. Change to the pub directory after logging on. Many Internet application programs are found here.

wuarchive.wustl.edu oak.oakland.edu	Washington University at St. Louis and Oakland University in Rochester, Michigan contain some of the largest software archives on the Internet.

Commercial Software:

Product	Comments
Internet in a Box 2.0	A complete set of Internet application programs for Windows from Spry, Inc. Included is the Web browser, Air Mosaic and TCP/IP software. The package will automatically connect to an Internet access provider, SprintLink, or let a user choose another provider. For more information, contact http://www.compuserve.com/prod_services/consumer/consumer.html.
Mosaic in a Box	Internet in a Box adapted for Windows 95. Contact: as above.
MacWeb WinWeb	MacWeb and WinWeb are Web browsers available from EINet galaxy. Alpha versions of these products are available as shareware by ftp from ftp.einet.net in the einet subdirectory. Contact http://www.einet.net.
NetCruiser	From NetCom Online Communications. This is a fully integrated package of Internet application programs for Windows including a Web browser. The software can be downloaded by modem. This package connects to Netcom's own proprietary network. New versions of the software can be obtained for free online. Contact http://www.netcom.com.
OS/2 Warp	IBM's OS/2 latest operating system includes a complete package of Internet software, including the Web Explorer. The TCP/IP software is preconfigured to connect to the Internet through the IBM Global Network. Contact http://www.austin.ibm.com/pspinfo.
Chameleon	Netmanage, Inc. Provides the most complete set of Internet and networking tools for Windows including the ability to run ftp server software. Contact http://www.netmanage.com.
Netscape	From Netscape Corporation. The commercial product provides full user support and a more enhanced version of the product than the shareware mentioned above. Available in Windows and Macintosh version. Contact http://www.netscape.com.
Windows 95	The operating system comes preconfigured to connect to the Microsoft Network. TCP/IP and Internet application software are available. Contact http://www.windows95.com.
Macintosh	Macintosh does not have a fully integrated set of Internet tools. The System 7.5 operating system has built-in TCP/IP software. There are application programs for all the individual services such as Turbogopher and Fetch (for ftp) as well as Web browsers. Contact http://www.apple.com.
Pipeline	A complete set of Internet applications with preconfigured access to PSINet, the proprietary network of PSI, Inc. (Performance Systems International). Both Macintosh and Windows packages are available. Contact http://www.pipeline.com.

SUGGESTED READING

There are hundreds of books about the Internet. Some are devoted to specific topics such as Mosaic, Netscape, html, USENET, email, etc. The books listed below are recommended because they have been reviewed by the author or have received favorable reviews in the computer literature.

Beginning Level

Levine JR, Young ML. *The Internet for dummies quick reference*. Foster City, CA: IDG Books Worldwide, Inc.; 1994.

Seiter, C. *The Internet for Macs for dummies quick reference*. Foster City, CA: IDG Books Worldwide, Inc.; 1994.

Duntemann J, Pronk R, Vincent P. *Web explorer pocket companion*. Scottsdale, AZ: The Coriolis Group, Inc.; 1995.

Kent P. *10-minute guide to the Internet*. Indianapolis, IN: Alpha Books; 1994.

Intermediate Level

Krol E. *The whole Internet: user's guide and catalog*, 2nd ed. Sebastopol, CA: O'Reilly and Associates, Inc.; 1994.

Gilster P. *Finding it on the Internet: the essential guide to archie, veronica, gopher, WAIS, WWW, (including Mosaic), and other search and browsing tools*. New York, NY: John Wiley and Sons, Inc.; 1994.

Kehoe BP. *Zen and the art of the Internet. A beginner's guide*. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1994.

Marine A, Kirkpatrick S, Neou V, Ward C. *Internet: getting started*. Englewood Cliffs, NJ: Prentice-Hall, Inc.; 1993.

Advanced Level

Comer DE. *Internetworking with TCP/IP: principles, protocols, and architecture*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, Inc.; 1995.

Lynch DC, Rose MT, eds. *Internet system handbook*. Reading, MA: Addison-Wesley Publishing Company, Inc.; 1993.

Liu C, Peek J, Jones R, Buus B, Nye A. *Managing Internet information services*. Sebastopol, CA: O'Reilly and Associates, Inc.; 1994.

Hunt C. *TCP/IP network administration*. Sebastopol, CA: O'Reilly and Associates, Inc.; 1992.

Dyson P. *Novell's dictionary of networking*. San Jose, CA: Novell Press; 1994.

Graham IS. *HTML sourcebook: a complete guide to HTML*. New York, NY: John Wiley and Sons, Inc.; 1995.