CHAPTER 6

Searching the Internet for Earth Science Resources

The Internet is often described as a huge, distributed library of information. A library is an *archive* of information, be it documents, books, video or audio tapes. A distributed library or archive refers to the fact that the information about a particular topic is distributed across several computers rather than being all on the same one. In some ways, you might already use a "distributed library." On many larger campuses the library collection cannot be housed in one building. Instead, books are distributed to smaller libraries scattered around campus. When preparing to do research you might visit several of these branch libraries to get the information you need. You might even find that the information you require is located off-campus. This isn't a problem because you get this information by requesting it through interlibrary loan, and within a few days it arrives at your desk. The problem in maintaining a distributed library system is organizing the location of the resources and providing a way for library patrons to get to the information. The distributed nature of the Internet is no exception. The rapid expansion of the Internet brings megabytes of information on to the Internet each day. This fact makes organizing the location of resources extremely difficult. Not only that, system administrators will move the files to new locations on a server or to a completely different server. Updating new locations becomes a problem too.

From previous chapters you have learned that information may be accessible via different services (Gopher, FTP, the Web). Not all the information located on the Internet can be accessed by one service, with the possible exception of the ever evolving capabilities of World Wide Web browser software. This means that you need to know how to use several kinds of search tools to effectively mine information from the Internet. Not only that, you need to understand a little about how each service searches the Internet. This will help you choose the right place to go to seek out the information you desire. You need patience too, for the explosive growth of Internet users has not kept up with the diffusion of high-speed network connections. The rapidity at which information and data are made available makes updating search tool databases difficult. What you might find today will be different tomorrow. What you find with one tool will be different from another. The fastest way of working yourself into a corner is to limit your searches to a single service.

Choosing the right tools and services for your needs is important to effectively locate information. Before starting a search you should ask yourself several questions. Where do you think your information will be located? Are you looking for a data file? Then FTP

searches would be a good place to start. Are you looking for documents about a particular topic? Then a document search engine would be a logical place to start. Maybe the information you need is not online but located in a library. You can still use the Internet to find your information: Telnet to an online library card catalog. How much time do you have to conduct your search? Do you need your information now or do you have time to browse? These are the kinds of questions you need to answer before starting your Internet search. Preparing for your search will make it a more efficient and less frustrating experience.

The purpose of this chapter is to introduce you to several tools available for searching out information on the Internet. It is not intended to be an exhaustive treatment of the development of the tools and or all the ways to use them. There are many good books (e.g., Gilster, 1994) that fully describe Internet search tools and search methodologies. Frequently-asked-question files, Usenet groups and archived help files contain a great deal of useful information about searching the Internet. Instead, I'll pass along the basics of using many of the common search tools and services to get at earth science information in the most efficient way. In this chapter we'll examine the syntax of searching and ways

- to find earth-science-oriented World Wide Web sites using search engines and subject-oriented guides
- to search for files using the Archie service
- to seek out Gopher information using Veronica
- to locate information about people connected the Internet using Netfind, WHOIS and Finger

The Syntax of Searching

Searching the resources of a remote earth science server is much like creating a program. You have to know the syntax for the language that the server's search software speaks. Depending on the search software interface, search engines will accept everything from keyword search terms strung together with Boolean operators to natural language word strings. When searching out resources on the Internet, you have to use the appropriate language syntax for the search engine your using.

Boolean Searching

Most search services recognize Boolean operators for searching their resources. Boolean operators are designed to guide the search engine's method of retrieving information. Various systems and search engines recognize different kinds of Boolean operators. Your search query is interpreted from the right-hand side to the left-hand side of the words entered. Parentheses can be used to force a particular order on to the query. Be aware that some searches are case-sensitive. To minimize unwanted returns or problems, you should check the search system help documentation before creating a search.

The common Boolean operators are AND, OR, NOT and *.

OR Search terms joined with the OR operator represent a logical union between documents. For instance, queries like “rivers OR streams” will return documents with either of the two search terms located in the document. Use the OR operator when combining synonyms or similar concepts.

AND Search terms combined with the Boolean AND operator represent an intersection of those terms. A search query like “rivers AND streams” will turn up every document that contains both search terms. The AND operator is best used when combining dissimilar terms like

astronomy AND education
geology AND journals
climate AND data

If you simply enter two words together, most search engines will recognize it as an “AND” query. For example, if you entered

mississippi river

it would be recognized as “mississippi” AND “river.” The restrictions imposed by the AND operator means that fewer “hits” or documents are returned.

NOT The NOT operator is used to exclude documents from a set based on the search terms identified. In this case, using “rivers NOT streams” would exclude all documents in which the search terms are found. The NOT operator works best when you are trying the eliminate a subset of documents from your search. The restrictions imposed by the NOT operator reduce the number of returned documents.

Right-hand truncation (*) Known as *stemming*, right-hand truncation uses an asterisk (*) to create a group of documents containing terms with common roots but different suffixes. The asterisk is used like a wild card when working with directories of files on your system. It does not work within, or to the left of, any characters. For example, the search “geo*” would produce results like geography, geology, and geophysics. The number of documents returned increases, in some cases dramatically, when you use right-hand truncation.

Searching the World Wide Web

The hypertext environment of the World Wide Web makes it both a challenge and a delight to navigate, depending on your immediate need for information and the time you have to find it. Be prepared for some disappointment if you need a quick answer to a question,

because the Web can be a tangled place to find it. Organization of information on the Web has come a long way in the last few years. A number of search engines have been created to

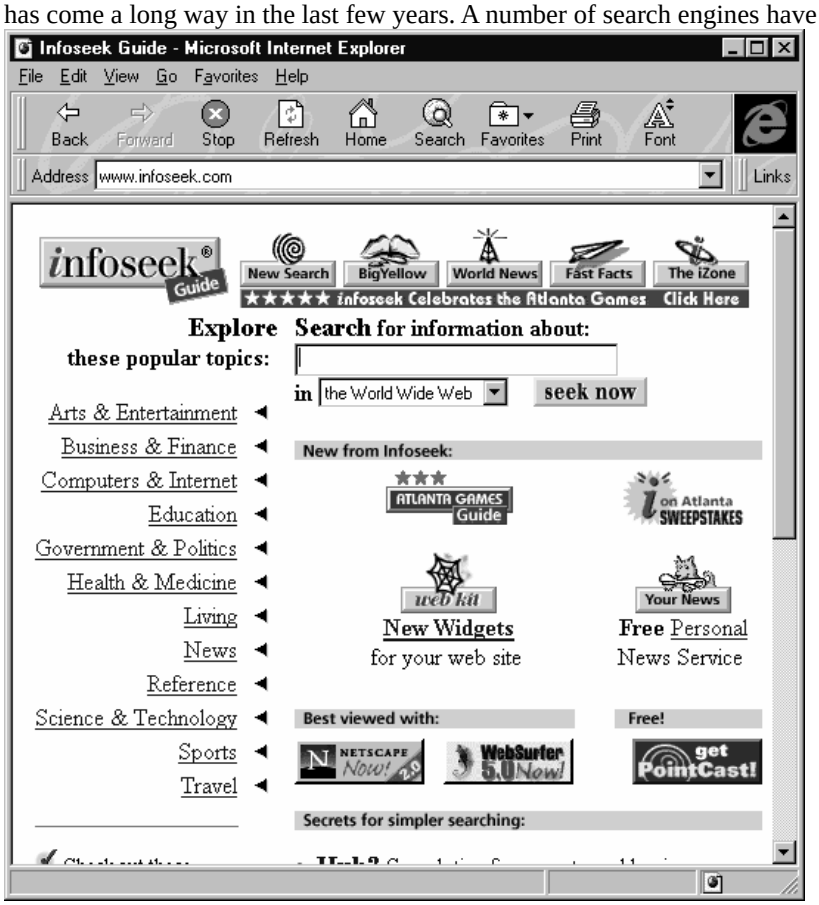

Figure 6.1 The Infoseek Guide search service home page

roam the Internet and seek out online information. Searchable databases of World Wide Web home page resources are coming online, and their holdings, though varying between services, are quite extensive.

How much of the World Wide Web can you search with a Web search service? This depends on what service you choose and how it operates. Archive services for Web pages come as hierarchically organized directories like the popular Yahoo! Web site or searchable Web page databases that use special programs, called “robots and spiders,” to search the World Wide Web for home pages.

Either archive service has its advantages and disadvantages. Most search services offer

Boolean searching of their Web page databases. Many let you customize your search. Search engines like the Infoseek Guide **(URL - http://www.infoseek.com/)** (Figure 6.1) let

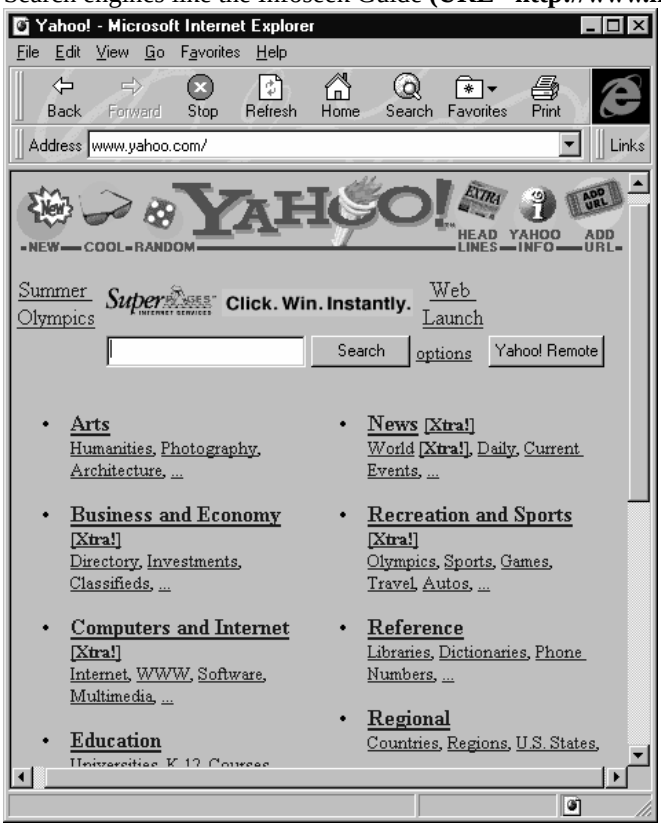

Figure 6.2 Portion of the Yahoo! internet guide

you search through all World Wide Web pages archived in their database. You can limit your search to reviewed Web pages or search through articles posted to Usenet newsgroups. Internet frequently-asked-question files (FAQs) can be searched from Infoseek Guide too. But due to the enormous growth in the popularity of the Web, sites like these are often quite busy.

Best described as a “Web robot,” the WebCrawler search engine processes documents one at a time as it navigates the World Wide Web. WebCrawler maintains an index of the Web for searching. WebCrawler creates its index of Web sites by an incomplete first traverse, subsequently relying on auto-navigational mechanisms to fill in the gaps (Pinkerton, 1994). WebCrawler has a second searching component, that of being able to search the Web on demand.

Yahoo! **(URL - http://www.yahoo.com)** is an example of a hierarchically organized directory of Web pages (Figure 6.2). A hierarchical directory is good for those who know

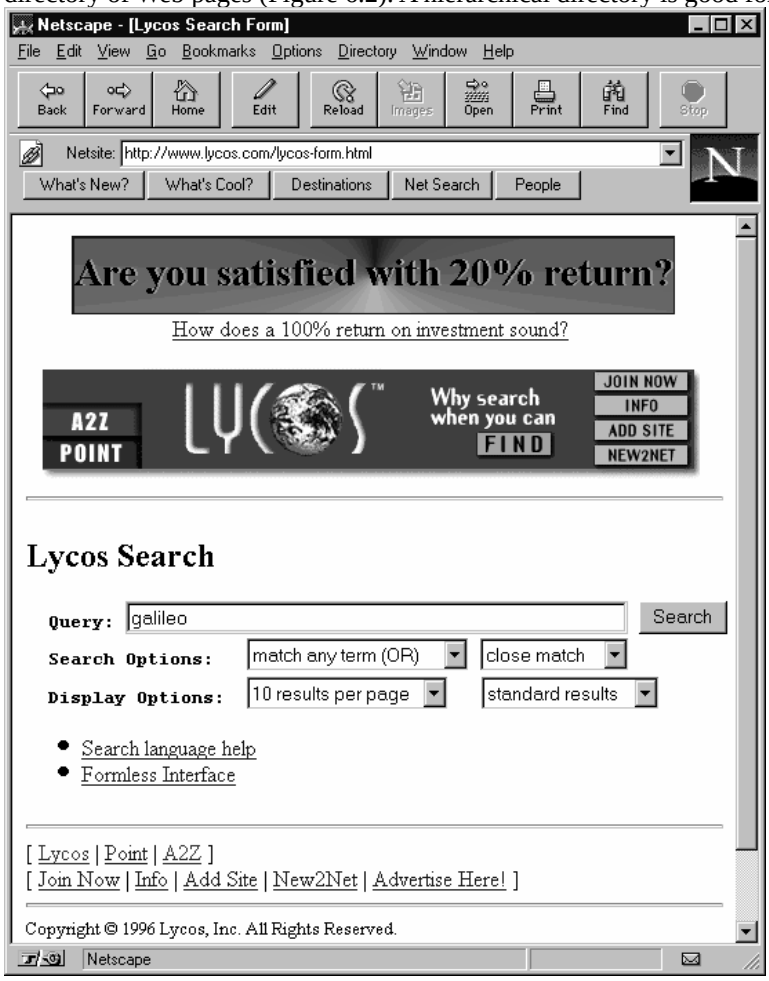

Figure 6.3 Lycos search service

what they're looking for and what category to look for the information in. These kinds of directories generally do not let you view an abstract describing the contents of a Web page. They also require more time to find the information you are looking for by digging through layers of menus and links. Most hierarchical directories permit you to search their holdings too.

If you choose a service that uses robot or spider programs to roam the World Wide Web, your search will likely cover a substantial portion of it. The Lycos **(URL -**

http://www.lycos.com/) search service advertises that it encompasses over 95% of the entire World Wide Web (Lycos, 1995) (Figure 6.3). These spider programs are searching not only the World Wide Web but also Gopher and FTP sites each day, continually adding new resources and updating those that have changed their name or address. The addition of new resources to their site database means that subsequent searches will likely reveal new information. Each address in the database or catalog is associated with information about what the site contains, an outline of the information, and the number of times the site is referenced by other addresses.

A search engine is the software that queries the search service's database of sites and returns a list of "hits." Search services vary in how they search their databases and the Internet. Some allow searching of the entire text of Web documents in their database. Some permit you to search through URLs or titles. I suggest avoiding title searches. Not all Web authors accurately describe the contents of their Web document in the titles, and some titles can be misleading. Search engines like Lycos sort your hits in order of relevance. The Lycos search engine catalogs three kinds of files: HTTP files, Gopher files and FTP files. Information about the file type allows you to find image, full-motion video, sound and other nontext files.

Each search result has a "relevancy" score ranging from 0 to 100. The scores for the search results are determined by the words and phrases that you enter in the search text box; the number of occurrences of the words and phrases on a page determines the score. Common words generate lower scores because they are found on many pages and cannot be easily used to distinguish pages. Uncommon words generate higher scores because the words do not appear on many pages. Higher scores also result from words combined into phrases because the combinations of words are not as common as the individual words. Relevancy scores should be used as a guide. Just because a site ends up with a low relevancy score doesn't mean that it hasn't anything to offer.

I'll use the Lycos search engine to demonstrate a typical search for information on the World Wide Web. One of the big events of 1995 was the *Galileo* visit to Jupiter. To see what I can find out about the project I open a session to Lycos and type the search term "galileo" (without the quotes) into the text field and click the search button beside it. After a few seconds Lycos returns a new screen (Figure 6.4).

The search results indicate that 11,088 documents were found matching my keywords. That is three times as many as I had when running the same search just six months prior. Of these only 83 documents score above a .010 relevancy score. The documents are arranged from highest to lowest score. The underlined words indicate that they are links. If I choose them, Lycos will return a page or more of documents containing the search word I chose. Next I'll scroll down the list of "hits," or documents that Lycos has in its catalog that match my search term (Figure 6.4)

GREAT! I'm shouting (that's what the capitals mean on the Internet). Among the documents is the Project Galileo home page created by NASA's Jet Propulsion Lab. The title is highlighted and underlined, meaning that it is a hot link to the Project Galileo home

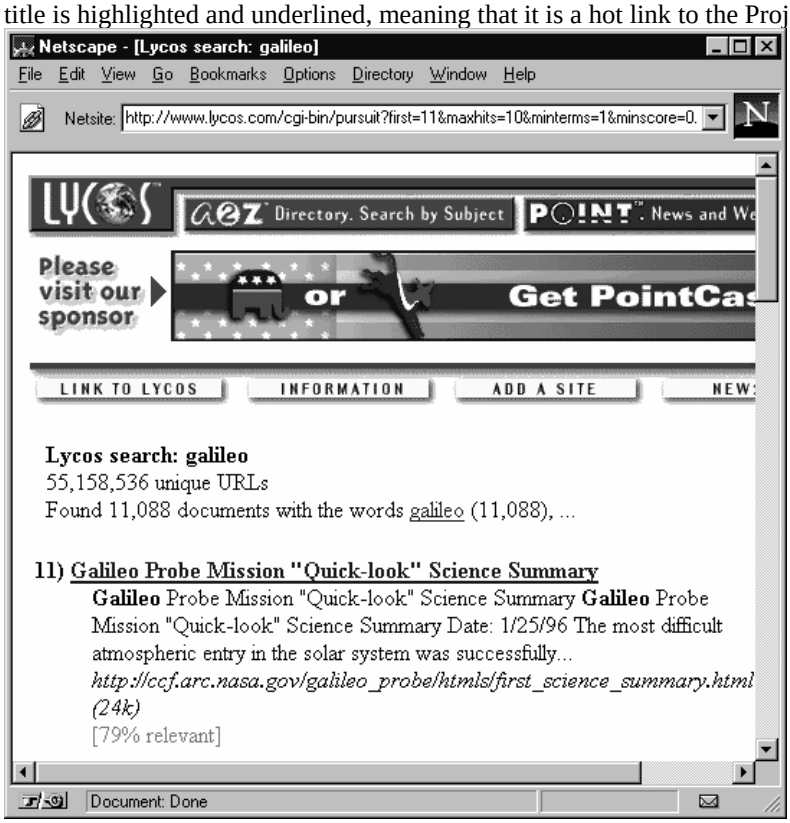

Figure 6.4 Typical Lycos home page abstract

page. The abstract gives me a notion of what the home page is about. I can continue to look through the rest of the hits or go directly to the Project Galileo home page.

There are a number of World Wide Web search engines that you can use. Table 6.1 lists several search engines and their addresses.

Searching for Files with Archie

In Chapter 4, "File Transfer over the Internet," you were introduced to the amazing amount of archived information located on computers connected to the Internet. The inevitable question is, How do I locate a file among the gigabytes of programs and data scattered all across the world? The answer lies with Archie. A play on the word "archive," and not the

comic book character, Archie was developed at McGill University in Montreal. Archie servers automatically gather a list of files, along with associated directory information, from public FTP sites and place the list into an indexed database. When you connect to an Archie

Table 6.1 World Wide Web search engines

Service	URL
WebCrawler	URL - http://webcrawler.com
Yahoo	URL - http://www.yahoo.com
Lycos	URL - http://www.lycos.com
InfoSeek	URL - http://www2.infoseek.com
Excite	URL - http://www.excite.com
The Whole Internet Catalog	URL - http://nearnet.gnn.com/gnn/wic/index.html
TradeWave Galaxy	URL - http://galaxy.einet.net/galaxy.html
Inter-Links	URL - http://www.nova.edu/Inter-Links
The World Wide Web Worm	URL - http://wwww.cs.colorado.edu/home/mcbryan/WWWW.html
The Clearinghouse for Subject-Oriented Internet Resource Guides	URL - http://www.lib.umich.edu/chhome.html
Special Internet Connections	URL - http://www.uwm.edu/Mirror/inet.services.html
The Awesome Lists	URL - http://www.clark.net/pub/journalism/awesome.html
Planet Earth Home Page Virtual Library	URL - http://www.nosc.mil/planet_earth/everything.html
Internet Resources Meta-Map	URL - http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Demo/metamap.html
The WWW Virtual Library	URL - http://www.w3.org/hypertext/DataSources/bySubject/Overview.html

server you can search the database by keywords. The server will return a list of all file names and FTP site locations that match your keyword. Archie does not make a comprehensive sweep across *all* Internet FTP sites. Instead, it targets specific FTP servers whose system administrators have given their permission to search. Within the Archie system, over 230 gigabytes of files are available for searching, and this is only a subset of all the Internet sites that have made their resources available to the public. Archie updates its database by regularly searching through a site's holdings. Be aware that your Archie search may turn up a file that has been removed, or a new file may have been uploaded after the last Archie sweep of a site.

There are several ways to access an Archie service:

- Use Telnet to contact an Archie server. Connect to Archie by issuing the Telnet <server-name> command at your system prompt. The server responds with the Archie prompt, and you're ready to start searching.

- Access Archie via electronic mail. To do this you send your search commands in the body of the message and address the email as archie@<server name>.
- Use an Internet browser like Gopher or a Web client program like Netscape or Mosaic. Archie appears as a menu choice on some Gopher servers. World Wide Web browsers have been specifically engineered to connect to many kinds of server, Archie being one of them.
- Use a client program on your system if one is available. To do this simply type:

  archie <return>

  and the Archie program will start up.

Each approach has its own advantages. Archie servers are usually busy places to access. Quite often you will have difficulty logging on to one, especially during prime workday hours. If the Archie server is busy, it will display a list of alternative Archie servers that you can try. Table 6.2 gives you a list of potential Archie sites. If you have access, use your own system's client program. This approach will save you time and network resources. If you're not in a great hurry to get results, then search Archie via electronic mail. Searching via electronic mail may take longer, but a printout of the results is sent to you. You can save the search results in a file for later use.

Accessing Archie Using Electronic Mail

When demand for an Archie service gets too high, consider using electronic mail to conduct your searches. Archie servers usually process interactive logins first, and when demand decreases, will process electronic mail requests. To access Archie through electronic mail, a user must compose an electronic mail message that contains the necessary search commands and send it to an Archie server. If I wanted to search for the popular file decompression program pkunzip.exe, I would compose this message:

From:mritter@uwspmail.uwsp.edu
To:archie@archie.unl.edu
Subject:
Date:October 20, 1995

find pkunzip.exe

Once the message is processed, the Archie server responds by sending a list of sites with subdirectory information to you. There are some guidelines you should follow and restrictions you should note when using electronic mail to search for files with Archie:

- Each command must be on a line of its own.
- Each command should begin in the first column of a line.

Table 6.2 ***Selected Archie servers***

Server Address	Location	Country
ds.internic.net	AT&T InterNic Directory and Database	USA
archie.sura.net	SURAnet, Baltimore, MD	USA
archie.unl.edu	University of Nebraska, Lincoln	USA
archie.rutgers.edu	Rutgers University, N.J.	USA
archie.au	University of Melbourne	Australia
archie.uquam,ca	University of Quebec	Canada
archie.doc.ic.ac.uk	Imperial College, London	United Kingdom
archie.funet.fi	Finnish University and Research Network	Finland
archie.th-darmstadt.de	Technische Hochschule, Darmstadt	Germany
archie.ac.il	Hebrew University of Jerusalem	Israel
archie.unipi.it	University of Pisa	Italy
archie.wide.ad.jp	WIDE Project, Tokyo	Japan
archie.nz	Victoria University, Wellington	New Zealand
archie.luth.se	University of Lulea	Sweden

- Archie servers consider any message with the word “help” in it as a request for help information. As such, the return message will only contain help information.
- Archie servers will ignore all lines that contain an invalid command.
- Archie servers will treat the subject line as part of the body of the message, so it’s best to leave it blank. Many electronic mail systems have message length limitations. The Archie software will automatically split a long message into multiple 45-kilobyte-long messages to safeguard against any possible problems with message length.

Accessing Archie with Telnet

Most people access Archie via Telnet. Recall that Telnet is the Internet’s interactive remote login service. To perform an Archie search, initiate the Telnet client software and connect to the Archie server. If you are at the system prompt type:

>**telnet** <*server address*>

For example:

>telnet archie.unl.edu

Or at the Telnet prompt type:

telnet>**open** <*server address*>

telnet>**open archie.unl.edu**

When prompted for a login name, you enter “Archie.” No password is required. After login the server sends a few messages about the server and then lets you begin entering commands. At this point you can enter “help” to get information on Archie commands, or you can start searching.

To search for a file type:

archie>**prog** <*filename*>

or

archie>**prog pkunzip.exe**

The server will respond with the status of your search, in particular your position in the queue of other ongoing searches and time to process your request. If your request is successful a list of sites and the subdirectory location will be printed to screen. The list of sites will scroll across your screen at a rapid pace. If you don’t have a rollback feature, the previous screens will probably be gone. A good thing to do is to have your results electronically mailed to you by sending the command

archie> **mail** <*email address*>

For example:

archie> **mail mritter@uwspmail.uwsp.edu**

You’ve undoubtedly noticed that I used both “find” and “prog” as commands to search out files with Archie. The command “prog,” standing for program, was used by earlier versions of Archie and now has been replaced by “find.” However, many systems still recognize the “prog” command.

The Archie program functions by means of a set of environment parameters that are set up upon loading the program. You can see the default settings by issuing the “show” command:

archie>**show**

You have the opportunity to change several parameters to suit your search by typing

archie> **set** <*parameter*>

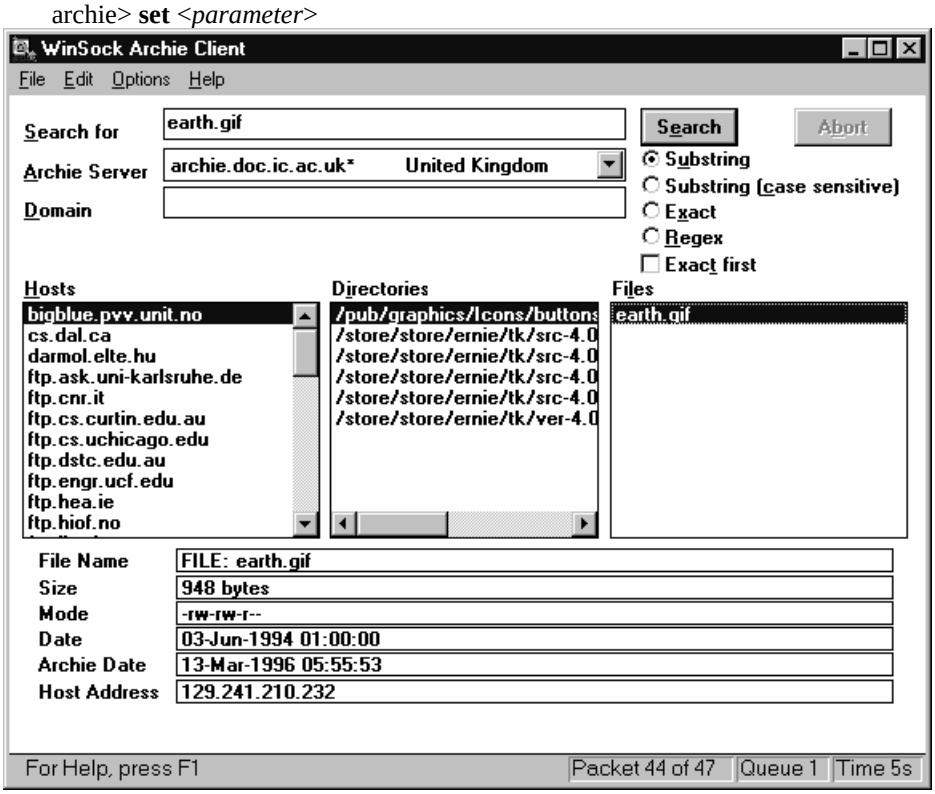

Figure 6.5 WinSock Archie GUI client program

For instance:

archie> **set search sub**

The search will retrieve all files and directory names containing your search term.

Accessing Archie with WinSock Archie

Graphical user interface applications make Archie searches a snap. The WinSock Archie application window is shown in Figure 6.5. A file can be searched for by simply typing the file name in the “Search for” form field, selecting a Archie server and clicking the “Search” button. The search results are displayed in the middle of the window. All FTP servers found to have a file with the same name that I used in my search are displayed in the Hosts box. The location of the file within the directory structure of the server is displayed in the Directories box. The file name is displayed in the Files box. File details (e.g., size, date,

Archie date) are retrieved and displayed at the bottom of the application window. The WinSock Archie client software can be configured to initiate an FTP session with the WS_FTP client program. By double-clicking on the file name, you call up a dialog box that sends an FTP command through to WS_FTP to retrieve the wanted file.

Searching Gopher Space with Veronica

One of the nice things about Gopher, like the World Wide Web, is that the software removes the complexities of accessing information and lets us focus on what we are looking for. The hierarchical menu structure is easy to learn and helps organize information for retrieval. Using Gopher by itself to find needed information can be a time-consuming process of working through menu after menu, only to find yourself at the end of a long journey with not much to show except a lot of wasted time. What is needed is a way to retrieve the information you want without having to stop at every intermediary menu along the way. A search engine that could act like Archie, finding the location of the requested information amongst the barely charted environment of Gopher space and making it accessible to the user, was needed. And so Veronica was created, in Archie's image so to speak.

Veronica is an index and retrieval system that can locate items on most of the Gopher servers connected to the Internet. Veronica finds your information by searching for words in the title of the resource as it appears in the menu of its home Gopher server. It does not do a full-text search of the contents of the resources; it finds resources whose titles contain your specified search word(s).

Veronica servers will give you a choice of how to search Gopher space. You can search Gopher space *by keywords in titles*. This type of search will find all types of Gopher resources (e.g., ASCII documents, image files, Gopher directories, etc.) whose titles contain the search terms specified. You can search Gopher directories only by keywords in titles. This search will return only Gopher directories that contain the search terms requested. You connect to Veronica from a Gopher menu pick; the program will display menu choices (Figure 6.6).

Directory searches are for broad topics, such as geomorphology. If you are looking for something more specific, like dunes, you'll want to search for title words. You can narrow your search by telling Veronica to locate particular types of Gopher resources. You specify the type of resource by adding the "-tX" option to your query, where the X is a letter or number code for the Gopher resource type. It doesn't matter where you place the "-t" flag; it can be in front of or behind the search word. For instance, the query

ozone -t1

or

-t1 ozone

will yield the same results.

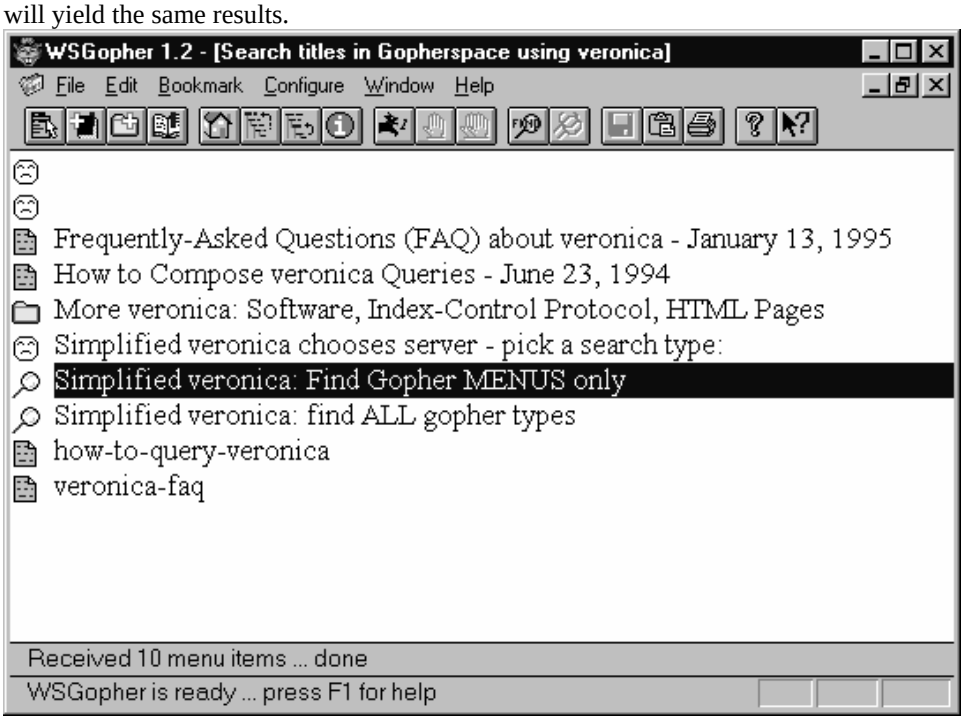

Figure 6.6 Veronica menu

You can specify more than one type in a search. For instance, if you used

Jupiter -t1g

Veronica would return all directory and GIF images with the word “Jupiter” in the titles.

Veronica will permit you to specify the number of items to find (X) by using the “-mX” flag. The “-l” flag is used to create a “log” file of links for the retrieved items. The log file will be displayed as the first “pick” on the Veronica results menu. You must use a single hyphen for each search option. For example:

ozone -t1 -m100

request 100 directory items containing the word “ozone.”

Let’s look a sample search. I’m going to use “galileo” again as the search term. It is

broad enough that I'll search for directories under Galileo. Notice the question marks at the end of all the menus. A question mark indicates that the menu choice is a searchable menu element. I'll move my arrow cursor down to item 8 and use the right arrow to get a box to

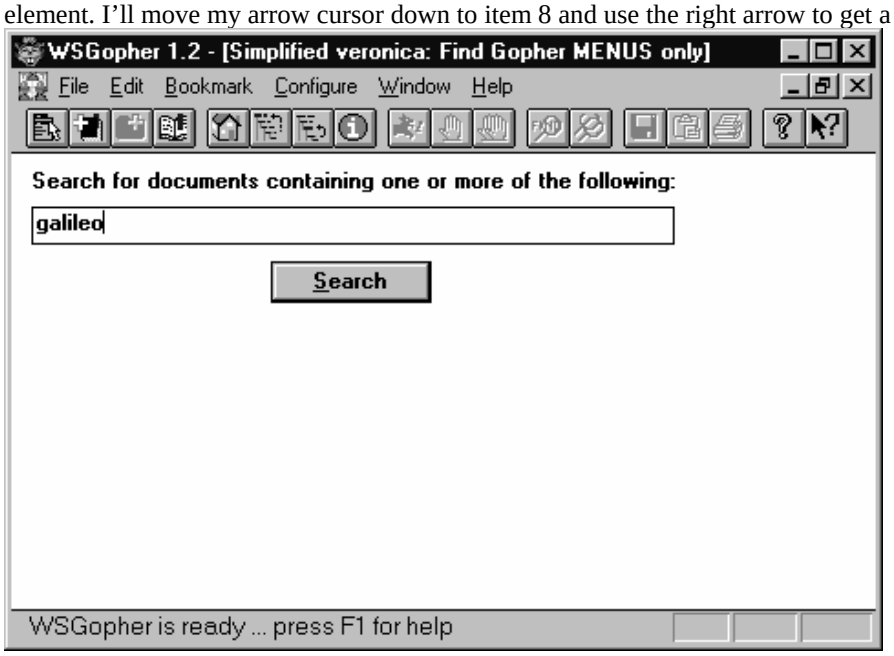

Figure 6.7 Initiating a Gopher search with Veronica

type in the search term. After typing the search term in I press the return key to initiate the query (Figure 6.7).

The query returns another long menu of items (directories) that contain the word "galileo." Notice that some are about the great scientist Galileo, while others are about the NASA *Galileo* project (Figure 6.8).

Wide Area Information Server (WAIS)

Archie and Veronica searches are basically term searches; you are matching keywords to the titles of files or directory menus. They do not search through the text of a document to find your search term. Wide Area Information Server (WAIS, pronounced "ways") enables you to search and retrieve text or multimedia documents stored on the servers. WAIS searches through full text indexes for the information requested and returns a list of documents. You can request WAIS to send a copy of any of the documents found. There are a number of WAIS searchers for nearly 500 databases containing everything from technical documents and bibliographies to newsgroup archives. You can access WAIS in several ways:

- Run a WAIS search in the World Wide Web. This is probably the easiest way.
- Run a WAIS client on your local host computer. This is the most efficient way to do a WAIS search.

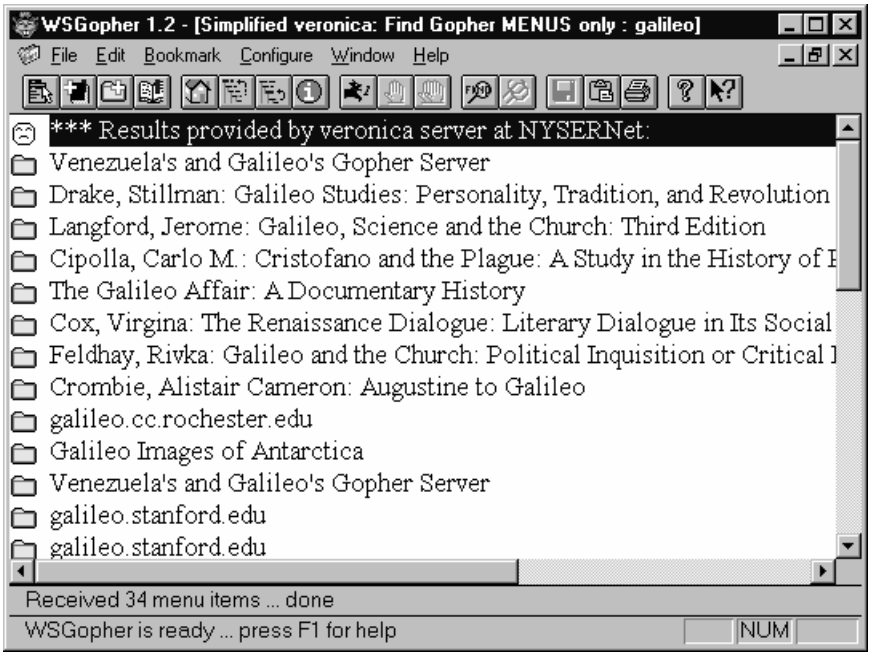

Figure 6.8 Menu of "hits" from Veronica Gopher search

- Telnet to a remote computer that offers access to a public WAIS client. WAIS searching during a Telnet session can be difficult if not impossible. The WAIS interface is complicated and depends on the terminal emulation software you are using.
- Run a WAIS session from a Gopher menu option or via the World Wide Web. The Gopher version of WAIS is easier to use than those accessible via Telnet but still not as easy as a local client. The main drawback to using a Gopher interface to WAIS is that you can only search one database at a time. This provides relevance feedback, which indicates how close a match your search produced. Table 6.3 shows how to access WAIS via the World Wide Web.

Some WAIS searchers specialize in environmental topics, physics, astronomy, and so on. This means that you will have to know where to go to get to the index you wish to

Table 6.3 Sample WAIS sites

WAIS Sites	Login

quake.think.com	WAIS
sunsite.unc.edu	WAIS
WAIS.com	WAIS

search. Because WAIS allows text searching, anything archived on the Internet can be described in a database: Usenet articles, email discussion list archives, binary data. and so on. Programs or images can be downloaded so long as they can be identified by a text descriptor in a WAIS database.

The significant difference between WAIS and other search services is WAIS's ability to respond to search frames in plain English. For instance, you might enter:

```
give me all documents related to the ozone hole
```

Don't think that WAIS can actually understand the entire sentence. What it does is extract individual words to search. WAIS ignores the more common words like "related," "or," "to," and "the" and searches for the less common words like "ozone hole." You can just as easily omit all the unnecessary words and simply type:

```
ozone hole
```

and achieve the same search results as with the longer phrase. Once the search is complete a list of matching items is retrieved. Most WAIS interfaces permit you to retrieve and download the information to your desktop.

Looking for People on the Internet

WHOIS

WHOIS enables you to search out people connected to the Internet. Like other search services, WHOIS searches a database of names. Given that there are as many as 20 million people connected to the Internet, there is no way that WHOIS can find any particular person. The typical WHOIS server lets you search a database of names within your own organization. You can move between organizations and their WHOIS servers with relative ease. However, before you get started you need to know some information about the person you're trying to find. You should know the name of course, the organization the person is affiliated with (e.g., a company, university, or government agency) and the city of residence. WHOIS servers can be accessed by:

- opening a Telnet connection to a WHOIS server with the command

  telnet <*server name*>

- using a local WHOIS client
- sending an electronic mail WHOIS search to **mailserv@ds.internic.net**

WHOIS databases can be searched by name, handle (a unique identifier), or mailbox. Table 6.4 lists the addresses and locations of selected WHOIS servers.

Table 6.4 Selected WHOIS servers

Address	Location
whois.pacbell.com	Pacific Bell (US)
whois.sunquest.com	Sunquest Information Systems (US)
zippy.telcom.arizona.edu	University of Arizona (US)
whois.bates.edu	Bates College (US)
whois.berkeley.edu	University of California at Berkeley (US)
whois.cwru.edu	Case Western Reserve University (US)
whois.dfci.harvard.edu	Dana-Farber Cancer Institute (US)
whois.messiah.edu	Messiah College (US)
whois.msstate.edu	Mississippi State University (US)
whois.ncsu.edu	North Carolina State University (US)
whois.oxy.edu	Occidental College (US)
whois.cc.rochester.edu	University of Rochester (US)
whois.sdsu.edu	San Diego State University (US)
whois.slac.stanford.edu	Stanford Linear Accelerator Center (US)
whois.bcm.tmc.edu	Baylor College of Medicine (US)
whois.ubalt.edu	University of Baltimore (US)

Earth Online Tip: WHOIS servers are for queries about specific information. Extended queries intended to obtain large sections of the directory represent not only an excessive use of server resources and bandwidth but also an unfair use of directory information that belongs to individuals.

Netfind

Netfind is like the white pages of your telephone directory. Give Netfind a person's name and description of where he or she works, and it will attempt to locate the telephone number and electronic mailbox information. Recall that to conduct a WHOIS search you need to know the right server to search on. With Netfind you can log on to any Netfind server and provide it with the person's name and work location, and it will conduct a search over many servers. Once supplied, Netfind searches its seed database to find Internet domains that match the specified keywords. If there is more than one matching domain,

Netfind displays the list of matching domains and asks you to select up to three to search.

When the search is completed, Netfind returns a summary of problems encountered during the search through remote domains, information about the most promising electronic mail address for the person being sought (if available), and information about when the person most recently logged on to the Internet (if available). If more than one person is located by a search, the summary does not include information about email targets and most recent/current logins.

Table 6.5 Netfind servers

Host Address	Country
bruno.cs.colorado.edu	USA
ds.internic.net	USA
eis.calstate.edu	USA
krnic.net	Korea
monolith.cc.ic.ac.uk	England
mudhoney.micro.umn.edu	USA
netfind.anu.edu.au	Australia
netfind.ee.mcgill.ca	Canada
netfind.if.usp.br	Brazil

There are two ways to run a Netfind search:

- Use a client program on your system. Type “netfind” at your system prompt to see if you have access to Netfind.
- Connect to a Netfind server via remote login with Telnet. Telnet to **bruno.cs.colorado.edu.**

You can use the Netfind software at your site, or you can Telnet to one of the hosts listed in Table 6.5.

Netfind requires the name of a person, with keywords to indicate where that person works. I’ll Telnet into the University of Colorado Netfind service and search for information about myself. After I log on to my Internet service provider I’ll type:

```
telnet> open bruno.cs.colorado.edu
Trying 128.138.243.150...
Connected to bruno.cs.colorado.edu.
Escape character is '^]'.

SunOS UNIX (bruno)
```

```
Login as 'netfind' to access netfind server
```

At the login prompt I'll do as the server instructs and type:

```
login: netfind



======================================================

Welcome to the University of Colorado Netfind server.

======================================================
```

A long list of alternatives sites are given in case the present Netfind server is too busy.

```
Alternate Netfind servers:
    archie.au (AARNet, Melbourne, Australia)
    bruno.cs.colorado.edu (University of Colorado, Boulder, USA)
```

I think that your terminal can display 24 lines. If this is wrong, please enter the "Options" menu and set the correct number of lines.

```
Top level choices:
    1. Help
    2. Search
    3. Seed database lookup
    4. Options
    5. Quit (exit server)
```

I'll choose 2 to search for myself:

```
Enter person and keys (blank to exit) --> ritter uwsp
```

The Netfind server finds three potential domains to look for my profile.

```
Please select at most 3 of the following domains to search:
    0. uwsp.edu (university of wisconsin, stevens point)
    1. lib.uwsp.edu (library, university of wisconsin, stevens
           point)
    2. me.uwsp.edu (mechanical engineering department,
           university of wisconsin, stevens point)
Enter selection (e.g., 2 0 1) -->
```

If there are more than 100 matching domains, Netfind will list some of the matching domains/organizations and ask you to form a more specific search. You can use any of the parts of an organization's name (or any of the components of its domain name) as keywords in searches. Using more than one keyword implies the logical AND of the keywords. Specifying too many keywords may cause searches to fail.

Item “0 uwsp.edu” looks like a reasonable choice. Netfind returns with:

```
Enter selection (e.g., 2 0 1) --> 0
( 1) got nameserver spu1.uwsp.edu
( 1) got nameserver spdns1.uwsp.edu
( 1) SMTP_Finger_Search: checking domain uwsp.edu
( 1) do_connect: Finger service not available on host uwsp.edu->
cannot do user lookup
------
Domain search completed. Proceeding to host search.
------
( 1) SMTP_Finger_Search: checking host uwspmail.uwsp.edu
( 2) SMTP_Finger_Search: checking host sknapp.uwsp.edu
( 3) SMTP_Finger_Search: checking host sis.uwsp.edu
( 4) SMTP_Finger_Search: checking host gbernd.uwsp.edu
( 5) SMTP_Finger_Search: checking host afranz.uwsp.edu
( 3) do_connect: Finger service not available on host
sis.uwsp.edu -> cannot do user lookup
( 3) SMTP_Finger_Search: checking host ariel.uwsp.edu
( 1) SMTP_Finger_Search: checking host worf.uwsp.edu
SYSTEM: worf.uwsp.edu
   Login: mritter                  Name: Mike Ritter
   Directory: /usr/fac/mritter     Shell: /bin/csh
   Office: Geog/Geol, 715-346-4449  Home Phone: n/a
   Last login Sun Mar 10 13:38 (CST) on ttyp0 from
         198.150.193.52
   No Plan.
   Login: critter                  Name: Christoph Ritter
   Directory: /usr/stu/critter     Shell: /bin/csh
   Office: Magdeburg Project, 715-346-4127 Home Phone: n/a
   Last login Fri Mar 15 06:00 (CST) on ttyp1 from 141.44.27.16
   No Plan.
```

Two hits were made, and mine is the first one.

Finger

The Finger program is used to retrieve a file or information about a person by using an electronic mail address. To use Finger you simply type “finger” at your system prompt and enter the electronic mail address:

```
worf-6> finger mritter@uwspmail.uwsp.edu
Login: mritter                        Name: Mike Ritter
Directory: /usr/fac/mritter           Shell: /bin/csh
Office: Geog/Geol, 715-346-4449       Home Phone: n/a
On since Mon Dec 18 13:22 (CST) on ttyp0 from 143.236.26.121
Plan:
Geography/Geology World Wide Web Page
```

Basically the same information has been retrieved in a much shorter time and with less hassle than the Netfind service. However, Netfind can use much less specific information to run a search than Finger. Finger can also be used to retrieve other files. For instance, to

retrieve an update in earthquake activity you can Finger to **quake@geophys.washington.edu**.

```
worf-8> finger quake@geophys.washington.edu

[geophys.washington.edu]
Login name: quake       In real life: Earthquake Information
Directory: /u0/quake    Shell: /u0/quake/run_quake
Last login Mon Dec 18 10:47 on ttyp6 from 164.157.206.13
Mail last read Mon Oct 23 03:46:40 1995
Plan:
The following catalog is for earthquakes (M>2) in Washington and
Oregon produced by the Pacific Northwest Seismograph Network, a
member of the Council of the National Seismic System. PNSN
support comes from the US Geological Survey, Department of
Energy, and Washington State.
Catalogs for various regions of the country can be obtained by
using the program 'finger quake@machine' where the following are
machines for different regions.
gldfs.cr.usgs.gov (USGS NEIC/NEIS world-wide),
andreas.wr.usgs.gov (Northern Cal.),
scec.gps.caltech.edu (Southern Cal.),
fm.gi.alaska.edu (Alaska),
seismo.unr.edu (Nevada),
mbmgsun.mtech.edu (Montana),
eqinfo.seis.utah.edu (Utah),
sisyphus.idbsu.edu (Idaho),
slueas.slu.edu (Central US),
tako.wr.usgs.gov (Hawaii)

Additional catalogs and information for the PNSN (as well as
other networks) are available on the World-Wide-Web at URL:
'http://www.geophys.washington.edu/'
DATE-TIME is in Universal Time (UTC) which is PST + 8 hours.
Magnitudes are reported as local magnitude (Ml). QUAL is
location quality A-good, D-poor,
Z-from automatic system and may be in error.
DATE-(UTC)-TIME LAT(N) LON(W) DEP MAG QUAL COMMENTS
yy/mm/dd hh:mm:ss deg. deg. km Ml
95/11/16 18:08:06 45.00N 122.58W 26.2 2.1 C 25.6 km SE of
Woodburn, OR
95/11/21 03:01:05 47.71N 120.30W 5.6 2.8 B 8.7 km NW of Entiat
95/11/28 14:42:53 49.18N 123.60W 10.0 2.3 C 40.4 km W of
Vancouver,BC
95/12/09 17:32:53 48.40N 122.21W 0.0 2.1 B 9.0 km E of Mount
Vernon
95/12/17 15:01:47 47.58N 120.21W 12.4 3.1 B 7.4 km S of Entiat
```

Notice that the earthquake information contained in the report is a part of the "Plan" file. The report delivered to our desktop provides us with a wealth of information. The document points us to other possibilities to finger for information in different parts of North America and on other kinds of servers that such information can be obtained from. At the document is the information we are looking for: dates, times, magnitudes, and locations of earthquake activity in northwestern North America.

You can keep up to date with NASA by "fingering" NASA News at **nasanews@space.mit.edu** (Figure 6.9). Users can retrieve the latest information on such things as the schedule for upcoming space shuttle missions, including the crew's activity schedule for that day.

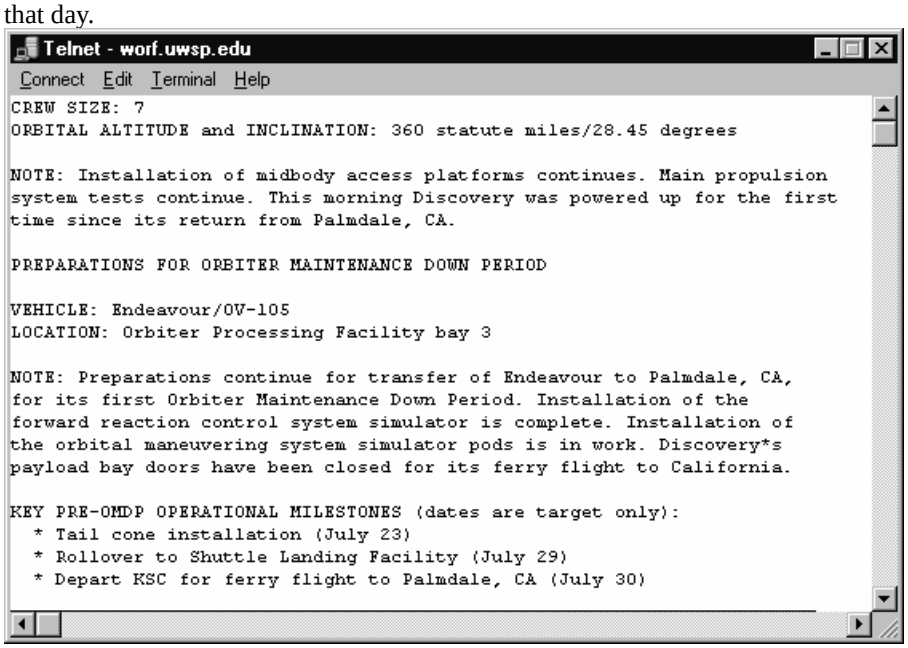

Figure 6.9 Finger response from NASA News

What You Have Learned

- Search engines use Boolean operators to frame a query for their database.
- Database and network searches can be accomplished with robot-like programs that scour the Internet for resources.
- Subject-oriented database services permit a user to look for information by navigating through a hierarchy of menus.
- Archie is used to search for the location of a particular file on the Internet.
- Veronica is employed to search Gopher space for menu titles and directories.
- WAIS is a database index search engine for archives.
- WHOIS, Finger and Netfind are used to search for information about Internet users.

Apply It!

Searching for resources on the Internet can be a time-consuming job. In previous Apply It!

sections, we have relied on knowing the address of a site to start our pursuit of climate change Internet resources. We turn to the search services discussed in this chapter when we need more help finding our way.

Figure 6.10 Starting an "advanced" Lycos search

In this Apply It! section I'm going to search for information related to the impact of climate change on stream flow. Earlier in this chapter we examined the Lycos search engine. Let's return to Lycos to conduct our new search. This time I'm going to use the advanced features of Lycos, which enable me to decide how matches are executed and what type and amount of information will be returned to me about my query. First, I'll connect to the Lycos search form at **URL - http://www.lycos.com/lycos-form.html** (Figure 6.10). The search form provides the usual keyword field and search button to initiate the query. Below these items are additional options to shape my search. First is the "Search Options," the default of which is to use the OR Boolean operator. The OR operator will search out all documents that contain any of the terms. The drop-down "Search Options" list box lets me use a Boolean AND operator to match various combinations of terms in my search list. This is especially handy if you're unsure of the spelling of one your keywords. For instance, say you wanted to search for information about Shishaldin volcano in Alaska but were unsure of its spelling. In the Query field you might enter "Shisaldin Shisalden Alaska." Lycos searches through its holdings to match both of the spellings to Alaska. In this way you are sure to find the right one. (Hopefully *one* of the search terms is spelled correctly!) The

Lycos search lets me determine the selectivity of the match as well. Lycos will use its relevancy score to determine the strength of the match and whether to return the match to you. The stronger the match, the fewer returns I'll receive, but this also eliminates irrelevant

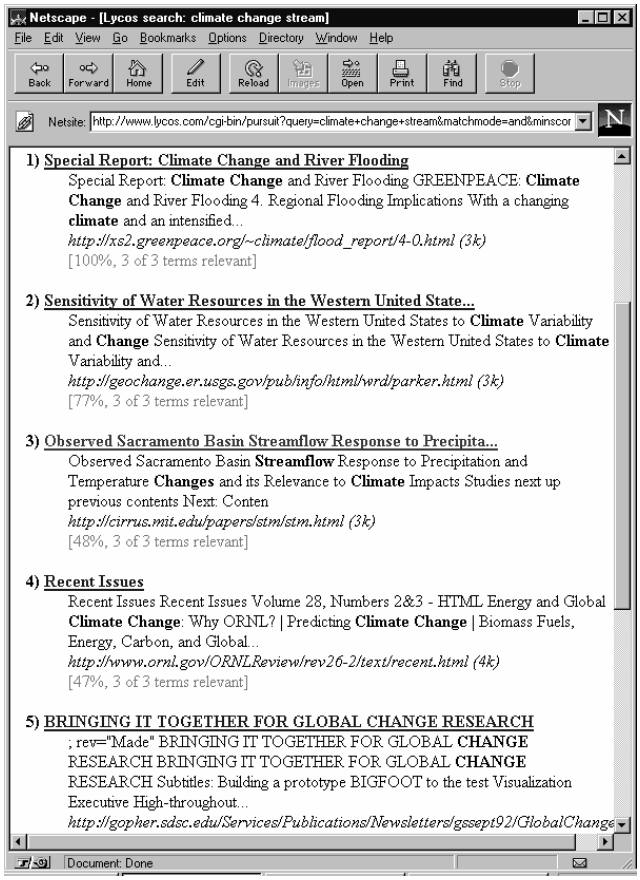

Figure 6.11 Lycos advanced search query results

documents. To save yourself some time, use a strong match for common keywords and a loose match for less common ones.

Lycos lets you determine how your query results are displayed. Two options are given, one for the number of results ("hits") displayed on a page and one for the amount of information returned. There are three levels of detail: standard (the default), summary (the minimum amount of information is displayed) and detailed (all information is displayed).

To begin my search I'll enter the keywords "climate change stream" (Figure 6.10). I

purposely left the word “stream” in singular form because Lycos might match “stream” to larger words like “streamflow. ” I’ll set the Search Options to “match all terms (AND)” first but keep the selectivity at “loose.” My query is started by clicking the “Search” button.

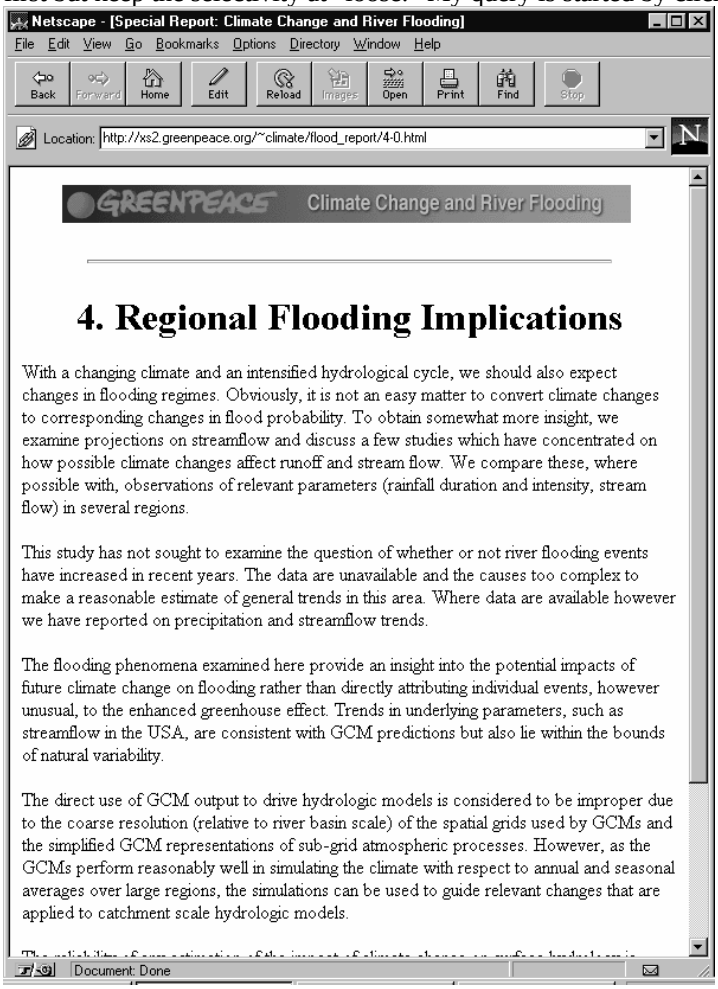

Figure 6.12 Greenpeace page on Regional Flooding Implications and climate change

The results of my search, shown in Figure 6.11, contain some good prospects, particularly the first document from the environmental organization Greenpeace, which has a relevancy score of 100%. Organizations like Greenpeace contribute an enormous amount of material to the online community.

This page discusses the validity of general circulation models to forecast regional flooding due to changes in climate (Figure 6.12). It is a portion of a special report furnished

by Greenpeace focusing on climate change and river flooding **(URL - http://xs2.greenpeace.org/~climate/flood_report/)**. It's a good addition to my growing list of resources so I'll bookmark it for later reference.

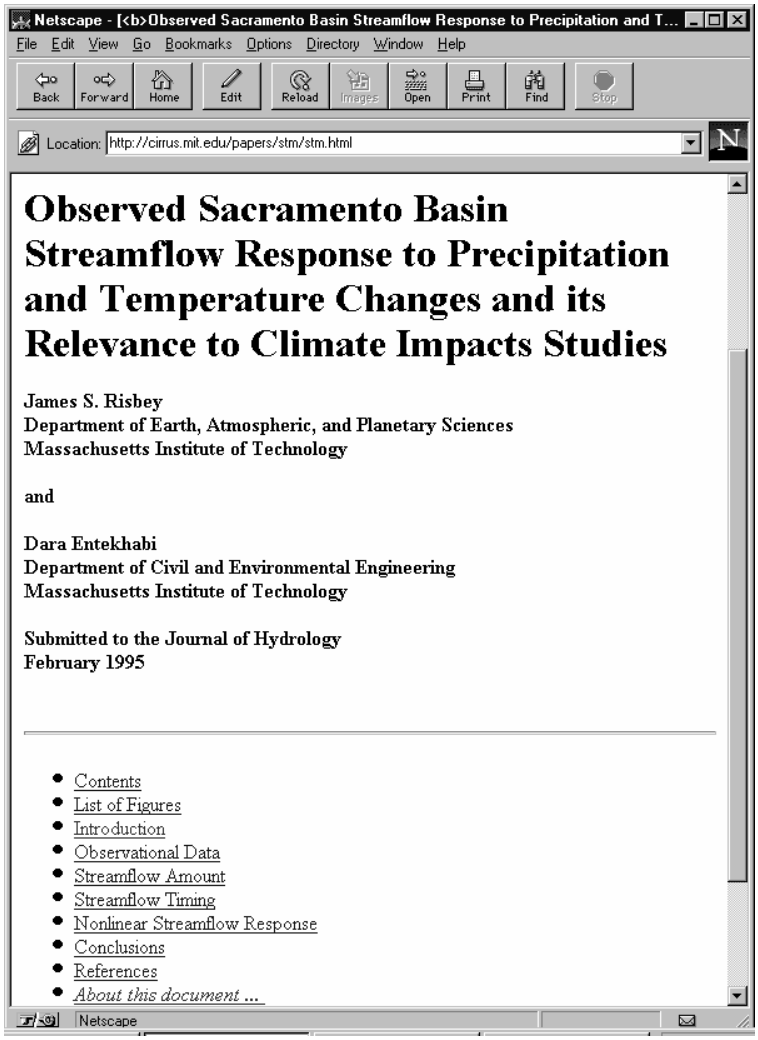

Figure 6.13 Research paper about the streamflow response to climate change

Returning to my search results I'll scroll down to item 3, "Observed Sacramento Basin Streamflow Response to Precipitation," and click on it to see how useful it is. It has a low relevancy score, 48% but it did match the three terms. This is another good find (Figure 6.13) as it is an actual research paper investigating "streamflow response from the

historical record rather than from hydrological models of the basin, and [comparing] observational results with model results" **(URL - http://cirrus.mit.edu/papers/stm/stm.html).** Note that the paper was submitted to a professional journal and so at this time has not undergone peer review of its research methods or results. Nevertheless, useful information is provided.

This Apply It! has generated a few new materials to add to our climate change resource list. Be aware that if you apply this same search you'll probably end up with different results because Lycos, like other Internet search services, is adding new resources every day. The search I make tomorrow may yield different results than the one I ran yesterday. It can be frustrating to keep up with, but the addition of new, up-to-date online materials is worth it.

Try It Out!

1. Each of the following items can be found on the Internet. Use the appropriate search service to locate the resources below. How many can you find?

    - CIA World Factbook
    - world maps
    - Roget's Thesaurus
    - periodic table of the elements
    - geographic name server
    - an image of Jupiter
    - the weather forecast for Minneapolis, Minnesota
    - a weather map of Europe

2. Finger the latest earthquakes at **quake@gldfs.cr.usgs.gov**. Plot the location of the earthquakes on a base map of the world. Where do most of the reported earthquakes originate? Is there any pattern? Compare the location of the earthquakes to a tectonic map of the earth. What were the most recent earthquakes to strike the northwestern portion of the United States?

3. Find out what the author's latest plans are by fingering him at **mritter@fsmail.uwsp.edu**.