The Internet for Librarians

Uta Grothkopf

European Southern Observatory (ESO), Library, Garching near Munich, Germany

Abstract

Librarianship is currently undergoing major changes. New information sources, accessible via the "network of networks", the Internet, offer opportunities which were previously unknown, but which require continuous ongoing learning. The Internet seems to be organized badly or not at all. The poor appearance might lead to an underestimation of its value.

In the following, an introduction to the main functions will be given in order to facilitate understanding and use of the Internet. E-Mail, FTP (File Transfer Protocol) and Telnet will be covered, as well as Mailing lists, Newsgroups and the tools Archie, Gopher, Veronica, WAIS (Wide Area Information Server) and the World Wide Web (WWW).

Examples will be given to show possible applications for library services.

1 Introduction

The Internet is a means of communication, allowing the user to access documents, indexes, and other information which is available in electronic format. In addition, we can get hold of information also accessible via "traditional" methods such as "snailmail", telephone or fax, but often access via the Internet is quicker.

The Internet is the successor of the so-called ARPAnet, a network developed by the U.S. Defense Department. ARPAnet was designed in order to furnish military institutions with a decentralized network. In 1972, it was presented to the public. Now, also scholarly institutions were able to connect to the network. Especially for universities and research institutions it seemed to be an attractive instrument since it allowed communication between computers at comparably low costs.

Today, the Internet consists of a large number of smaller networks (so-called Local Area Networks or LANs) as well as individual participants from all over

the world. Because of it's decentralized structure, the Internet is also called the "Network of Networks". One of the characteristics of the Internet is the fact that participants can be users as well as providers of information and services. The membership varies continuously, a fact that influences the quality and continuity of information available on the Internet remarkably.

The technical requirements a computer must meet if it shall be able to communicate on the Internet are defined by the TCP/IP (Transmission Control Protocol/Internet Protocol).

2 Basic Functionality

The basic services offered by the Internet can be subdivided into three groups: Mail, Telnet, and FTP (File Transfer Protocol). To begin with, simple electronic (e-) mail means that one user sends a message to another user. But "Mail" goes beyond that. It also comprises Mailing Lists and electronic Bulletin Boards or Newsgroups. The Telnet program creates a link between two computers. FTP enables users to exchange files.

2.1 Electronic (E-) Mail

The easiest and most obvious application of electronic mail is sending a message from one Internet user to another. An Internet address typically consists of the name of the user (maybe in an abbreviated version), followed by the @-sign as well as one or several so-called subdomains, i.e. computers within an institution, and finally the domain. The domain indicates the country where the user (respectively his or her institute) resides, e.g. uk represents the United Kingdom, fi Finland etc. Instead of a country code, U.S. e-mail accounts show an abbreviation of the organisational area to which they belong, e.g. edu for educational institutions, mil for military, gov for governmental bodies etc. Here's an example of an Internet address: president@whitehouse.gov.

The easiest way to obtain the correct e-mail address of somebody you want to contact is to look at the address of the sender ("from"-line) if you already received a message from him or her. Of course a range of electronic and printed guides to e-mail addresses are available, but especially for printed ones it is difficult to keep them up-to-date. If you know the general e-mail address of the institute (i.e. everything following the @-sign), but not the personal account of the person you want to reach, try to send a message to the "postmaster" who is in charge of the local mailing network and ask him or her to forward your message to the correct address, e.g. postmaster@springer.de. On the World Wide

Web (see below), other names like Helpdesk, Questions and Webmaster (e.g. webmaster@www.dra.com or question@simbad.u-strasbg.fr) have appeared.

Nowadays, many different mail handling systems are available, but even without these fancy pieces of software you can send and receive messages. Only few commands are essential, e.g. mail (+ address) to send, inc to receive (incorporate) messages, as well as show, print, forward and reply.

Why should librarians use e-mail? First of all, to ensure a good, seamless, and rapid exchange of information, to enhance communication among colleagues, and to improve delivery of information and documents to the library users. Finally, these efforts will result in a considerable improvement of the library service in general.

2.1.1 Mailing Lists

An electronic discussion group or mailing list can be described as an open forum for everybody interested in a particular topic. Participants subscribe to a list by sending an e-mail message to the service address of the list. The body of the message should contain the command "subscribe <listname> <YourFirstName> <YourLastName>". The service address of the list usually is slightly different from the address to which messages meant for all subscribers have to be sent. This fact sometimes leads to confusion, and you may find "subscribe" or "unsubscribe" messages in your mailbox instead of being sent to the service address.

After subscribing to a list, participants will receive a welcoming message from the owner of the listserv, i.e. the one on whose computer the list resides. The welcoming message describes the most important commands and contains some other advice. From now on, every message sent to the whole list will also drop into the new subscriber's mailbox. Many list subscribers behave rather passively and just observe the discussions of the more active participants. However, everybody is invited to join the ongoing discussion at any time.

We distinguish between unmoderated and moderated mailing lists. Whereas everybody can send any message to the members of an unmoderated list, the latter type has a moderator to whom the messages are sent first. S/he collects and/or reviews them before mailing them to the other subscribers. Although this looks like censorship at first glance, a moderator can help very much in ensuring a high quality regarding the content of messages. It is obvious that on moderated lists angry discussions and "flames" which we may find on unmoderated lists can be avoided.

Mailing lists are the medium of choice whenever you want to contact many colleagues in one go. They are a vast source of tips and tricks, answers to ques-

tions, announcements and experiences. Colleagues who are posting a question to a mailing list can be pretty sure that other subscribers can help them solve the problem.

2.1.2 Bulletin Boards

The Newsgroups or electronic Bulletin Boards are similar to mailing lists, but the postings are not delivered to the subscriber's personal mailbox. They have to be read using a newsreader software. Which newsgroup can be accessed by the members of a particular institute will be decided by the system administrator. Nowadays, Bulletin Boards cover every possible (and certainly not only scientific) subject area. A newsgroup's name typically is composed of several parts, the parts being divided from each other by dots. The name should reflect the contents of the list in abbreviated form. The first part of the name, indicating the broadest area to which this newsgroup pertains, might be e.g. comp for computer-related newsgroups, sci for scientific groups, rec stands for recreation etc. Many of these groups are of interest to librarians (e.g. comp.internet.library or comp.infosystems.www.users).

The so-called FAQ files (Frequently Asked Questions) list many of those questions that have been posted to the list repeatedly and are therefore of high interest especially for new subscribers. They are kindly asked to read the FAQ file when joining a new list.

2.1.3 Netiquette

All users of the Internet are requested to behave according to the "Network Etiquette" or Netiquette. Some of the most prominent items of the Netiquette are:

- The access to and the use of the Internet are a privilege, not a right. Politeness and also tolerance regarding colleagues with less knowledge than yourself are a must.
- Messages should be rather short and cover only one topic at a time. This helps colleagues who want to archive messages to identify them according to the subject field.
- Communication via networks lacks gesture and mimicry which are important elements of human communication. This can lead to misunderstanding, since nothing except the mere words helps to recognize the real meaning of what has been written, be it humorous, angry, or sarcastic. Several signs have been invented to show what kind of feeling the writer wants to express. The most popular one in order to express a not-too-serious remark is the little smiley lying on its side :-) or, with the opposite meaning, the sad equivalent :-(. However, these cartoons do not fully replace means of

expression available during personal communication. Therefore, we should be especially careful with irony and sarcasm when communicating on the Internet.
- Subscribers to mailing lists who want to ask colleagues for help should mention the efforts they made already in order to answer the question themselves. Thus, colleagues will not check the same reference tools without success again and, quite importantly, they will know their good will is not being abused (it is so much easier to send a request to a mailing list instead of trying to find a solution yourself). Finally, after a request has been fulfilled, a second message should be posted in order to inform colleagues and prevent them from spending more time on this problem.
- Communication via the Internet is never private. Always be aware that messages can go astray due to technological or human errors and might be received by somebody the message was not meant for.

2.2 Telnet (Remote Login)

Telnet establishes a link between computers that enables users to use a service or software installed on the remote computer just as if it was installed on their own machine. Telnet is based on the so-called Client-Server principle. The computer requesting a service is called the Client, the supplying computer is the Server. During such an "interactive session" the Client submits commands, the Server receives and answers them.

Obvious applications of Telnet are the library catalogs available online. During day-to-day work they will help librarians in a variety of tasks – for reference and bibliographic purposes, cataloging and for collection development (e.g. by checking the lists of new acquisitions of institutes working in a similar subject area). Online library catalogs enable librarians to quickly check whether or not a particular item is available in a library and help decide whom to contact for Inter-Library Loan (ILL).

2.3 FTP (File Transfer Protocol)

The File Transfer Protocol allows users to move files via the Internet. The contents of the files need not be text, but can also consist of pictures, sound, computer programs etc.

Anonymous ftp is a frequently used method to make information publicly available, since users can log into the public areas of computers, move around on these machines and read and copy files. In order to start an ftp-session, the user types the command ftp plus the address of an ftp-server, e.g. ftp

ftp.sura.net. The login typically will be anonymous, no password is required, but as a matter of politeness, users are asked to provide their full e-mail address, so that providers of anonymous ftp services can keep track of their actual users.

A lot of different programs are available for ftp sessions, but users can also get along with a few basic commands. For instance, cd is used to change the current directory, ls to list the files in a directory, get <oldfilename> <newfilename> to copy a particular file from the remote computer onto your own machine, and finally quit to cease a session.

For librarians, knowledge about how to use ftp is of growing importance. Many institutions offer their information on anonymous ftp servers, this information being about the institute or its staff, new acquisitions or preprints lists from the local library, full articles or even books. The providing institute will save money on shipping cost and can easily update information.

3 Information Retrieval Tools

The basic functionalities explained above require that user know the exact e-mail, telnet or ftp address of the service they want to use. In order to make information retrieval on the Internet a bit easier, several tools have been developed. All of them apply the basic functionalities "behind the scenes", but the user is guided by the information retrieval tool and need not worry about the address.

3.1 Archie

Archie was designed in order to locate files in publicly accessible directories. The user can send requests for file names (or parts of them). Archie's answer will be a list of locations from where the required file can be copied using ftp. Public Archie servers are installed around the world; it is recommended to use a service geographically close to the user. Example: telnet archie.funet.fi, login: archie, password: your e-mail address.

3.2 Gopher

A Gopher is a hierarchical, menu-driven information system. Originally developed at the University of Minnesota (from where it is still available via

anonymous ftp to boombox.microumn.edu, directory pub/gopher), many institutes have installed the software on their own computers and modified the pre-designed menu pages according to their needs, thus reflecting the search area of their institutes. A user who wishes to use a public Gopher system therefore will telnet into the Gopher information system of another institute with similar research topics. Selecting one of the menu items offered by the Gopher system will lead the user to a menu on the next level. Starting from one particular Gopher system, users can also have access to full texts, various lists (e.g. the so-called phone books, compilations of e-mail addresses) and searchable indexes to databases (e.g. the Library of Congress catalog: telnet locis.loc.gov). These resources need not reside on the computers of the institution that provides the Gopher, but anywhere else in the world. The Gopher just offers a link to this service.

3.3 Veronica

Whereas Archie enables users to locate ftp addresses of individual files, Veronica (Very Easy Rodent-Oriented Net-Wide Index to Computerized Archives) is a tool to find Gopher systems which are specialized in a subject area by checking the titles (not the individual options on a menu!) of the Gopher system menus.

In order to query Veronica, the system need not be installed on the user's own computer, but can be accessed via other Gopher systems by selecting the option "Other Gopher and Information Servers/Veronica" and entering a keyword. Veronica's answer will be a list of Gopher titles that contain the search term.

4 Searchable Indexes: WAIS (Wide Area Information Servers)

The large number of databases and files which are accessible via the Internet require a tool that allows the users to search for keywords in order to find documents of interest. Some years ago, the text-searching system WAIS (Wide Area Information Servers) was developed. It is based on the American standard Z39.50. Documents indexed according to this standard can be made available through a WAIS server. In this context, "documents" can be e.g. entire texts, the archive of a newsgroup, a database or any other index.

Queries can contain all kinds of search terms. The retrieved hits will be ranked according to the relevance the WAIS-system assumes – the higher the assumed relevance of the retrieved document is, the higher the document will be listed

in the list of results ("relevance ranking"). Contrary to Gopher-systems, WAIS does not oblige users to follow a given structure, but allows them to start for instance with a search for very general terms in the "Directory-of-Servers", the meta-index of WAIS-documents. Having chosen relevant documents from the list of hits to this first query, users can enter more specific search terms in a second search, thus getting closer to the most relevant documents.

Up to now, only few library catalogs are indexed according to Z39.50. The typical access still is via a telnet link into the OPAC of the library. One of the reasons might be that until very recently the free-of-charge versions of WAIS (which can be copied via anonymous ftp e.g. from think.com, directory: wais), did not cater for searches with Boolean operators (AND, OR, NOT), nor did they allow truncation of search terms or the application of context operators.

Searching a WAIS-server is of special interest to those users who need first entry points on the Internet regarding a new subject field. Unlike Gopher-systems that provide access to information via one particular institute specialized in this subject area, WAIS does not necessarily require any knowledge about the topic. Even searching for broad search terms can guide users not only to relevant documents, but maybe also to an existing mailing list of interest as well as names of scientists and institutions working in this field.

5 World Wide Web (WWW)

Without doubt, the World Wide Web (WWW) is the most convenient Internet tool currently available. Developed at CERN in Geneva, it can be accessed using so-called browsers (i.e. user interfaces). Clients for these browsers can be obtained via anonymous ftp e.g. from ftp.ncsa.uiuc.edu, login: anonymous. WWW documents can consist of texts as well as static or moving images, graphics, sounds etc. Users need not follow hierarchical menus as they have to when using Gopher systems, but may access Web pages by clicking on highlighted words or clickable images ("icons") in the text, so-called Hypertext links. Each Web page has its Universal Resource Locator (URL) that indicates on which computer and in which directory the document is stored. Any page on the Web can be accessed directly if the user knows its URL. Another way to find relevant Web documents is to use either the subject oriented compilations of Web addresses (e.g. Yahoo at URL http://www.yahoo.com/) or one of the many search engines that allow free-text searching for titles of Web pages (e.g. WebCrawler at URL http://webcrawler.cs.washington.edu/WebCrawler/WebQuery.html).

Various compilations of library catalogs or library information servers that can be accessed via the WWW are available, e.g Libweb at URL http://www.lib.washington.edu/~tdowling/libweb.html

or the Yale list at URL gopher://libgopher.yale.edu/11/. Many institutions also started to set up information servers with full texts of library-related text, like the electronic text collection "Libraries and Internet" which is maintained at the University of Lund in Sweden (URL http://www.ub2.lu.se/UB2proj/LIS_collection/collection_top.html). Further, archives of Bulletin Boards can be searched (e.g. BUBL, the Bulletin Board for Librarians at URL gopher://ukoln.bath.ac.uk:7070/ and PACS-L (Public-Access Computer Systems List), located at URL gopher://info.lib.uh.edu/11/articles/e-journals/uhlibrary). Announcements and proceedings of library-oriented conferences can also be found on the Internet, for instance the LISA-II (Library and Information Services in Astronomy II) conference (URL http://www.eso.org/lisa-ii.html) and Digital Libraries '95 (URL http://bush.cs.tamu.edu/dl95/README.html).

6 A Few Recommendations

The Internet offers a huge number of reference sources, but often they are difficult to find. Once located, nobody guarantees that access to these sources will be continuously available. As soon as the service provider decides to withdraw his/her service, the link won't work anymore.

It takes time and patience to explore the Internet. After navigating through various documents, users might find themselves exactly where they started. Due to the hypertext links, services can be accessed via various ways, so that users might feel as if they were moving in circles. Therefore, especially new users should concentrate on a few selected entry points. Once they are more familiar with these resources, they may want to try other sources in addition.

On the other hand, we must admit that no single Gopher, WAIS or World Wide Web service will provide access to absolutely all available information. Lack of completeness is one of the disadvantages of Internet searches. For advanced users, checking various services (e.g. several search engines) in parallel might help.

Depending on the subject, Internet searches often do not replace searches in commercial databases. Especially reliable bibliographic information is not yet available for all subject areas. Although several hosts claim their databases can be queried via the Internet, this refers only to the telecommunication link itself (e.g. Internet access instead of dial-up access). Nevertheless, the user will have to pay the usual charges for the search s/he conducts.

In any case, librarians should not start searching the Internet for the first time the very moment an impatient user is waiting for results. It is nearly impossible to locate relevant information immediately if you never before had a chance to become familiar with Internet searches.

Despite these drawbacks, we should not underestimate the vast resources available on the Internet. If we take a little time to practise "learning by doing", it will turn out to be a research tool that offers access to information of still unknown dimensions.