

THE INTERNET — AN ELECTRONIC TREASURE TROVE

Gerhard Obenaus
Translation Laboratory, University of Iowa
gerhard-obenaus@uiowa.edu

The Internet, a network of computers with distinctive software and hardware, interconnects millions of people world wide and offers tremendous amounts of information. Translators may benefit from the Internet throughout the translation process. The Internet is growing so fast that trying to find the right information is like looking for a needle in a haystack. The following paper will discuss some of the benefits of the Internet for the translator, and then point out various tools and guides which can be used to get the most out of the Internet.

INTRODUCTION

Anything written or said about the Internet should start with a disclaimer. A self-organizing network of networks, the Internet is so volatile that whatever is said about it today is bound to be obsolete tomorrow. If you use the Internet to sharpen your hacking skills, you will appreciate its fast pace, the thrill of the hunt, as you try new tools and tricks to delve even deeper into cyberspace. The hacker's thrill is the novice's nightmare. Like a tourist in a foreign land where people speak a strange language and road signs make no sense, the novice will feel utterly lost. Without a good guide, that is. This paper cannot provide such a guide. There are many of them already, serving different people with different needs. It will attempt to provide enough information on guides, sources, and the possibilities of the Internet, however, to allow translators eager to exploit its potential to get off to a good start.

The Internet received its name from the fact that it connects a variety of computers with distinctive software and hardware. These computers are inter-networked to allow them to communicate by translating messages into a mutually understandable language referred to as communications protocol. The protocol used by the Internet is called TCP/IP. The first true Internet was Arpanet, a military network established in the United States in the late 60ies. The following figures illustrate the phenomenal growth the Internet has experienced in the past decade. In 1981, 213 computers participated in the Internet. As of August 1993, the Internet comprised 15,160 different networks supporting at least 1,776,000 host computers in 60 countries [Merit (1)]. The number of users is estimated to grow by a million every few months. A conference on the Internet which attracted some 200 people in 1983 recently had between 50,000-60,000 visitors in San Francisco.

This vast interconnection of computers and the great popularity the Internet enjoys

provide an ideal infrastructure for sharing resources. The Internet allows users to transfer files between incompatible computers, send messages across the globe, and log into databases thousands of miles away within the blink of an eye. There is no central agency to coordinate the Internet. Each of its subnetworks is administered separately, and support tasks are accomplished by cooperative arrangements. Major Network Information Centers, NICs, provide support to users by making documents, other information and various services available to them [Farley (2)]. Other agencies coordinate Internet design, engineering and management. Internet networks within each country are organized in hierarchical order into national, regional, and local networks. Universities frequently act as local hubs to provide Internet access to secondary and primary schools as well as the public in general.

INTERNET ACCESS

Before going on to discuss the uses of the Internet for the language professional, a word on getting access to the Internet is in order. You may distinguish between direct and indirect access. Indirect access refers to electronic mail which may be sent to and from the Internet from other networks through inter-connections, called gateways. Users of CompuServe, FidoNet, and national networks may send messages to the Internet, participate in discussion groups and, to a limited extent, access databases. To take full advantage of the possibilities offered by the Internet, direct access to a computer connected to it is required. Universities, organizations, and large companies frequently offer access to the Internet to their affiliates. In the United States, some institutions have made it their policy to offer Internet access to anybody who asks for it. Other institutions make accounts available to the public for a moderate fee.

In Europe, finding convenient access to the Internet is a little more involved. If accounts are not available through the computer center of a university near your home, access is possible through national Videotex providers. Videotex access still carries a hefty surcharge of \$ 15.00 or more per hour in addition to telephone and other charges which may accrue. Complicated and expensive access procedures, along with a lack of awareness, seem to be the main reasons why the Internet has not really caught on with translators. As the Internet continues to grow and gain in popularity, the number of private companies providing competitive Internet accounts will increase and facilitate access for those not affiliated with large institutions. Please refer to the appendix for a short list of commercial Internet service providers.

SERVICES AVAILABLE ON THE INTERNET

The following services are available on the Internet: electronic mail, file transfer protocol, and telnet.

Electronic Mail

Electronic mail, hence e-mail, is used to send messages to other people or programs on other computers. It is the most extensively used service on the Internet. Through gateways, it is possible to communicate with the Internet from other networks such as CompuServe, FidoNet, or Bitnet. Please refer to the appendix for a short discussion of addressing conventions.

Electronic mail is useful throughout the various stages of the translation process. Finding work is one of the many things e-mail facilitates. Job notices for translators appear regularly on several discussion groups. Unsolicited advertisement of your own services on special interest groups (SIGs) violates e-mail etiquette, however, and may result in a strong reaction of other group members. Some translators have started to use the Internet to communicate with their clients and receive and send translations. As the commercial side of the Internet develops, it may turn into a medium of choice for translator-client communication. At this point, using accounts provided by universities and other non-profit institutions to carry out private business is considered unethical and may result in the loss of user privileges.

The most valuable resources available to translators through e-mail are special interest groups. Whatever the topic, there is a good chance that it will be discussed at one or more SIGs. On Bitnet, these groups are referred to as LISTSERVers, on UNIX platforms they are called USENET newsgroups. LISTSERVers, named after a computer program designed to mass distribute mail, are more easily accessible from outside the Internet than USENET newsgroups. The program keeps a list of all members of the group. When a member posts a message, it is redistributed to the mailbox of every subscriber. Messages posted to USENET newsgroups, on the other hand, are not redistributed to all members. Only one copy of the message is stored on the host computer. To read a message, a user must have access to the host where it is stored. Because they are more easily available, LISTSERV groups will be discussed in greater detail. Most groups archive their messages in a database, and it is possible to query the archives via e-mail to retrieve old messages by topic, date, keyword, or author. One group, Lantra-L, is devoted to translation. Intel-L deals with software localization issues. There are many other groups across a wide spectrum of fields that the translator may find useful as a source of information. To receive a list of available LISTSERV groups, send an e-mail message to LISTSERV@BITNIC.BITNET with the following text: LIST GLOBAL. You will receive a rather large file in your mailbox listing all discussion groups of which the LISTSERV program is aware. To receive help for using LISTSERVers, send an e-mail message to the address indicated above with HELP as text.

SIGs may be very helpful in locating information. If a translator has difficulty understanding a term or a concept, a message sent to the right group may provide a speedy answer. Most groups unite experts in the field from various countries, and since many of them rely on the Internet for help, they are usually most willing to provide it to others. Without the help of the computer it may even be impossible to find an expert. In addition, it is much easier to post a note requesting help on a discussion group than to pick up the phone book and call somebody whom you have never met to ask for a favor. SIGs also provide a good forum to exchange information with colleagues around the world and to stay abreast of the latest developments in the field.

File Transfer Protocol

File Transfer Protocol, hence ftp, requires full access to the Internet. It makes it possible to transfer files between two Internet computers easily and quickly. Depending on network traffic, it may only take a few seconds to transfer a Mb of information from a computer in London to a computer in Taiwan. Two kinds of ftp access are possible: anonymous ftp and full service ftp. Anonymous ftp allows anyone on the Internet to access computers which contain public file archives. The user simply logs on as *anonymous* and transfers files located in the archives to his/her computer. There are thousands of public ftp-sites around the world, providing a wealth of information ranging from documents and software for personal computers and mainframes to computer graphics, and a variety of other data. Full service ftp requires a logon name and a password and provides access to files outside the public archive section of the host computer.

The main problem with using ftp is finding out where a file is located. Archie is a tool designed to facilitate file searches on ftp sites. Without it, finding the right file is next to impossible unless you know on which ftp site it is located. Archie scans all ftp-sites periodically and updates its catalog of files. Archie can either look for a particular set of characters contained in a file name or produce a list of files associated with a certain subject. To find ISO standards, for example, type *prog iso* to produce a list of files whose name contains the string "iso". *What is iso*, on the other hand, produces a list of subjects with which the string "iso" is associated. Although it is possible to query Archie using e-mail, full Internet access is required to retrieve the actual files. To try Archie, telnet to one of the Archie hosts, such as *archie.doc.ic.ac.uk* in the United Kingdom or *archie.mcgill.ca* in Canada. Refer to the appendix for an example of a typical output produced by an Archie query.

Telnet

Telnet allows Internet users to log into other Internet computers to access online databases, electronic library catalogs, Internet information services, or to access their account [Kochmer (3)]. While ftp is available only for users with full access to the Internet, many sites with telnet access can also be reached through packet switching networks, such as CompuServe or Tymnet. Similar to ftp, two levels of access are available through telnet. Many sites offer guest logins which provide limited, yet useful, capabilities. To gain full access to databases and other information services available through telnet, a password and a login name have to be obtained from the host organization.

OPACs — Online Public Access Catalogs. Hundreds of university libraries all over the world have made their catalogues available online. The ability to search catalogues at a number of institutions from your desktop offers tremendous potential in identifying resources. Keyword searches may produce items not available at your local library. In addition to locating resources, OPACs may also be very helpful in terminology work. The dictionary offers *venous thrombosis* and *thrombosis of the veins* for the German *Venenthrombose*. A computer search produced several books with *venous thrombosis* in the title, but none with *thrombosis*

of the veins, thus providing a good indication as to which equivalent to use. In addition, remote catalogues may offer more detailed cataloging information for certain types of publications, such as government reports or periodicals. Many OPACs also offer access to specialized databases, ranging from manuscript and statistical collections to full text documents, such as newspapers, periodicals, or literary classics [Obenaus (4)]. Using these capabilities, translation projects can be researched in a fraction of the normal time. Several tools are available to find out which libraries provide their catalogues online. One of them is Hytelnet, a hypertext guide to online libraries and other Internet sources. Hytelnet resides in memory and waits to be activated with a hotkey. In addition, Hytelnet also provides information on FreeNets, Campus Wide Information Networks (CWIS) and other telnet information sites. Please refer to the appendix for a list of ftp sites for Hytelnet. The other tool is a document compiled by Art St. George and Ron Larsen, entitled "Internet Accessible Library Catalogs & Databases". Please refer to the appendix to see where it is available.

Text Archives. There are various projects around the world which collect language corpora for research and other purposes. Some of these corpora are made available to the general public. Corpora are very useful to the lexicographer and terminologist looking for usage examples. In many cases, the corpora are distributed with specialized access or concordance software. At Georgetown University, a Catalogue of Projects in Electronic Texts was established which can be searched online. It provides detailed information on the various electronic text projects, such as Project Gutenberg at the University of Illinois, ARTFL at the University of Chicago, or the Oxford Text Archives. The database is available via telnet or modem. Please refer to the appendix for a contact address.

Specialized Databases. I would like to mention the databases provided by the European Community as an example of specialized databases accessible through the Internet. ECHO, the European Community Host Organization, offers a wide range of unique information services online. They are of special interest to the translator because of their multilingual nature. The databases range from information on research projects, EC documents, to a multilingual terminology database. The latter, EURODICATUM, is of special interest to translators and terminologists. It contains technical and scientific terms as well as contextual phrases. It allows searches limited to certain fields as well as free text searches. Many of the terms contained in EURODICATUM are not yet available in printed form, increasing the value of the database. To use ECHO databases, telnet into ECHO.LU. When prompted for a user code, simply type GUEST. Users of the national VIDEOTEX services, such as the German Btx, the French Teletel, or the British Prestel Service, also have immediate access to ECHO. A contact address for ECHO customer service is provided in the appendix.

The databases listed above are only a small sample. The databases available through the Internet would fill a book. The next section will discuss some of the online tools which help the user navigate through the Internet by providing a more uniform and user-friendly interface.

NAVIGATING THROUGH THE INTERNET

WAIS — Wide Area Information Servers

One of the problems with using diverse databases is that the user is confronted with a new and unfamiliar interface for each database. WAIS addresses this problem by providing a uniform, user-friendly interface on the user's desktop computer which translates the query into a language the database can understand. Using WAIS, it is possible to read newspapers, get information on any country in the world, scan many specialized databases, and obtain information about the Internet. WAIS is particularly suited for retrieving textual information. It can search through many large documents for the occurrence of keywords and provide a scored list of the documents which are most relevant to the query. Using a technique called "relevance feedback" the user can customize the scoring process to make WAIS even more effective [Kochmer (3)]. A feature of particular interest to translators is the capability of WAIS to collect all items of interest automatically on a daily, weekly, or monthly basis. Thus, if you are interested in the use of a certain financial term, you can retrieve all documents in which it occurs in the Wall Street Journal and instruct ways to send you an updated list with documents containing the term on a regular basis. To try WAIS, telnet to quake.think.com and log in as WAIS.

Gopher

Gopher servers provide a menu driven interface to allow users to browse through the network without knowing telnet addresses and *ftp* sites. If the user finds a menu that looks interesting, gopher will get the information associated with it if you press a mouse button or the ENTER key. Gopher provides easy access to Archie servers, online library guides, ftp sites and numerous other Internet information sources. The number of Gopher systems is increasing rapidly, and the software is becoming more sophisticated. All Gopher sites connect to one another, allowing the user to explore virtually any part of the Internet by following a simple menu system. Gophers do not provide anything that is not available via telnet or ftp [Raish (5)]. They just allow the inexperienced user to gain access to information without having to learn telnet and ftp commands and browse through numerous guides to find out where the information is located. To experiment with a Gopher server, telnet into gopher.cic.net or any of the other gopher sites.

One of the drawbacks of Gopher is that the user has to negotiate several layers of menus before locating the right information. VERONICA, which stands for "Very Easy Rodent-Oriented Netwide Index to Computerized Archives" is an attempt to streamline Gopher searches by permitting the user to search gopher menus by keywords to find relevant menu entries more quickly. The number of VERONICA systems is still small. Thus, the ones that do exist are used heavily, resulting in slow response times. The popularity of VERONICA is bound to increase as more people learn about it, and within a year, most major Internet sites will probably have VERONICA installed [Raish (5)].

World Wide Web

World Wide Web, short WWW, is similar to Gopher and WAIS. However, instead of a menu driven design, it uses hypertext. WWW is not as widely available as Gopher, nor is it as fully developed. Since hypertext is a very powerful organizational tool, WWW holds great promise. To experiment with WWW, telnet into info.cern.ch located in Zurich, Switzerland.

OUTLOOK ON THE FUTURE

The Internet is growing at a phenomenal pace. The inter-connectivity it provides between different computer platforms has attracted the attention of a large number of commercial users. Several years ago, most of the users of the Internet were part of an elite who used high-powered computers to transfer massive amounts of research data across the continent. Today, a growing number of users is interested in the Internet for commercial reasons, pushing the research elite into the background. Thus the Internet is going through a transition from a research and educational to an increasingly commercial network.

One of the reasons that commercial applications have not appeared sooner is that the Internet was started for research purposes with government seed money. No provisions were built in for billing users. It was of no concern at the time. As more and more people realized the commercial potential of the Internet, companies have begun to think seriously about finding ways to charge users for online services they would like to offer. Cable companies, such as Continental Cable or the Hybrid Corporation in San Francisco, are beginning to offer high-speed access to the Internet through cable-TV lines, enabling everybody with cable access to participate in this global network. Since most people do not care to learn UNIX, the operating system used on most Internet sites, user-friendly interfaces have been developed which allow the users to hop across continents with a few mouse clicks. These interfaces make it possible to access text, graphics, sound, and full-motion video, turning the Internet into a true multi-media environment. These capabilities hold tremendous potential for language professionals. When doing a translation on arthroscopy, for example, a click on a camera icon will show an animated sequence of the insertion of the cannula into the knee. Click on any word in the text of an article, and the program will instantly look it up in its online dictionary. The technology to do it is here today. What is needed now is a concerted international effort on the part of the profession to cooperate in the compilation and dissemination of terminological and other data. Some of these efforts, such as the Text Encoding Initiative, are alive and well. But more are needed. Online research techniques should be incorporated in translator training, as some institutions have begun to do. Gopher servers should be set up to distribute materials for case studies for translator training, terminology work, and other areas of interest to the profession.

It is important that more translators get actively involved in the Internet for several reasons. The commercialization of the Internet, while providing some advantages, also involves many risks. Private information providers may attempt to restrict information access to corporate users who can afford to pay several hundred dollars an hour for online tune. One of the great things about the Internet is its flexibility, and if enough translators get involved, either personally or through professional groups, we may be able to influence its

development and keep it a place where information is available to everybody.

In addition, if more translators get involved, the amount of useful information available to the profession through the Internet will increase. Although the Internet is very helpful, it has yet to live up to its potential for the language professions. Thus we should not wait and see what the Internet can do for us, but think what we can do for it. If only half of the terminological papers written at various translator training institutes throughout Europe were put online, it would be a big step forward. I hope that this paper has aroused enough of your curiosity to venture out into the Internet and explore. Only through our active participation will the Internet become a true treasure trove.

List of References

1. Merit Network, Inc. (Statistics are available via FTP from nic.merit.edu; directory: nsfnet/statistics; filename: history.netcount)
2. Farley, Laine (ed.), August 1991, "Library Resources On The Internet: Strategies For Selection And Use." (available via FTP from dla.ucop.edu; directory: pub/internet; filename: libcat-guide)
3. Kochmer, Jonathan, 1991, "NUSIRG. User Services and Internet Resource Guide." Online Version, NorthWestNet Academic Computing Consortium, (available via FTP from ftphost.nwnet.net; directory: nic/nwent/user-guide; nusirg.whole-guide.ps.tar.Z)
4. Obenaus, Gerhard, 1993, "The Use of Computer Networks for Translators." Language International 5.1. 6-10.
5. Raish, Martin, 31 March 1993, "Network Knowledge for the Neophyte, Version 3.0". (available from listserv@bingumb.bitnet; File: NETKNOW NEOPHYTE)
6. Dillon, Martin; Jul, Erik; Burge, Mark; Hickey, Carol, 1993, "Assessing Information on the Internet. Toward Providing Library Services for Computer-Mediated Communication." OCLC, Dublin, Ohio.

Acknowledgements

I would like to thank Prof. James Marchand at the University of Illinois for making his vast collection of Internet documents available to me to write this paper.

Appendix

The appendix provides a brief list of public access points to the Internet, sources of information, and tools.

COMMERCIAL INTERNET PROVIDERS AND OTHER CONTACT ADDRESSES

AlterNet USA and International
Tel: +(703) 876-5050
info@uunet.uu.net

ANS USA and International
Joel Maloff
Tel: +(313) 663-7610
maloff@nis.ans.net

- PSInet PSI, Inc.
 Tel: +(703) 620-6651
 Offers Global Dialup Service for USA and international access. It costs \$39.00 a month and is available through local dialup in about 160 cities throughout the USA. You can reach PSI through X.25 PAD access at 311060740013602 or through TELENET/SPRINTNET by typing "c psinet, ACCOUNTNAME". On the Internet, send mail to info@psi.net for more information.
- Georgetown Paul U. Mangiafico, Project Assistant
 Catalogue of The Center for Text and Technology
 Projects in Academic Computer Center
 Electronic 238 Reiss Science Building
 Text Georgetown University
 Washington, DC 20057
 Tel:+(202) 687-6096; Fax:+(202) 687-6003
 Internet: pmangiafico@guvax.georgetown.edu
- ECHO ECHO Customer Service
 BP 2373
 L-1023 Luxembourg
 Tel:+352 34 98 11; Fax:+352 37 98 1234

ADDRESSING PROTOCOLS

The addressing protocols for the major networks are listed below. A good online source of information about gateways is "The Inter-Network Mail Guide", by John Chew. It is available via ftp from ra.msstate.edu; directory: pub/docs; file: internetwork-mail-guide.

<i>Network</i>	<i>To Internet</i>	<i>From Internet</i>
Applelink	user@domain@DASNET#	user@applelink.apple.com
CompuServe	Internet:user@host.domain	user@compuserve.com Note, the comma in the user-id in CompuServe changes to a period when sending mail from the Internet. xxxxx,yyy - > xxxxx.yyy@compuserve.com
FidoNet	send to UUCP at nearest gateway site and set first line of message to "To: user@domain"	user@Note.Net.Zone.fidonet.org

ARCHIE LISTING OF HYTELNET FTP SITES

Host animal-farm.nevada.edu

Location: /pub/ibm.pc

FILE -rw-r-r- 613480 Jun 24 09:37 hyteln65.zip

Location: /pub/mac

FILE -rw-r--r~ 600064 Feb 9 00:00 Hytelnet6.4.sea.bin

FILE -rw-r-r-- 733154 Jun 24 09:42 Hytelnet6.5.sit.hqx

FILE -rw-r-r- 1820 Jul 29 1992 hytelnet.readme

Host askhp.ask.uni-karlsruhe.de

Location: /pub/infosystems

DIRECTORY drwxr-xr-x 1024 Aug 25 17:37 hytelnet

Location: /pub/infosystems/hytelnet/amiga

FILE -rwxr~r~ 84309 Jul 14 1992 Ami-HyTelnet.lha

FILE -rwxr-r- 275 Jul 14 1992 Ami-HyTelnet.readme

Location: /pub/infosystems/hytelnet/mac

FILE -rwxr-r- 733154 Jun 24 05:21 Hytelnet6.5.sit.hqx

Location: /pub/infosystems/hytelnet/pc

FILE -rwxr-r- 613480 Jun 21 04:45 hyteln65.zip

Location: /pub/infosystems/hytelnet/unix

FILE -rwxr-r- 21125 Jun 21 07:14 hytelnet.tar.Z

OTHER INFORMATION SOURCES AND BOOKS ON THE INTERNET

Library Catalogs

Art St. George and Ron Larson, "Internet Accessible Library Catalogs & Databases." available via FTP from: nic.cerf.net; directory: cerfnet/cerfnet_info; file: Internet-catalogs-XX-XX.txt where XX-XX stands for the month and year of the most recent version.

Internet Resources

Judi Harris and the students of TEB 8000. Internet Resources Directory, Part 3: File Archives (FTP Sites) of Interest to Educators. Available via ftp from ftp.virginia.edu (128.143.2.7); directory: /pub/IRD/; file: IRD-FTP-sites.txt.

Judi Harris and the students of TEB 8000. Internet Resources Directory, Part 4: Ideas for Curricular Infusion of Telecomputing Tools and Resources. (July 1992)

J. Paul Holbrook and Christine S. Pruess, editors. CICNet Resource Guide. (June 1992). Available via ftp from nic.cic.net (192.131.22.2); directory: pub; file: resourceguide.

Introductory Books

Kehoe, Brendan P, 1993, "Zen and the Art of the Internet." Prentice Hall. ISBN 0-13-010778-6, 112p \$ 22.50

Kroll, Ed., 1992, "The Whole Internet User's Guide & Catalogue." O'Reilly and Associates. Sebastopol, CA. ISBN 1-56592-02502, 376pp, \$ 24.95

Advanced Books

Frey, Donnalyn & Adams, Rick, 1991, "!% @:: A Directory of Electronic Mail Addressing and Networks." O'Reilley & Associates, Sebastopol, CA. ISBN 0-937175-15-3.