The "World-Wide Web": A Distributed Hypermedia over the Internet

Paper preparred for Learntec 94 Karlsruhe

Daniel Schneider,
TECFA,
Faculté de Psychologie et des Sciences de l'Education,
Université de Geneve, Switzerland.
9 route de Drize, CH-1227 Carouge
Phone: +41 22 70 9694
Email: schneide@divsun.unige.ch

ABSTRACT

The "World Wide Web" (WWW) is a distributed hypermedia system that runs over the Internet. In a hypertext, if you want more information about a particular subject mentioned, you can usually "just click on it" to read further details. WWW documents can be linked to other documents written by different authors, in various locations. To access the web, you run a browser program. The browser reads documents, and can fetch documents from other sources. Information providers set up hypermedia servers which browsers can get documents from. The browsers can, in addition, access files by FTP, NNTP (the Internet news protocol), gopher and an ever-increasing range of other methods. Finally, the browsers permit searches of documents and databases as well as a "fill-out form" interfaces for accessing any kind of external programs running on a server.

The WWW has great potential as a distance education tool. It delivers information to the learner in an easy way and allows for more sophisticated computer mediated communication tools such as dynamic hypertext, question/answer programs, conferencing systems, etc. This text covers the main functionalities of WWW and related Internet Tools. We will show how WWW documents are produced and discuss how the WWW is currently being used in education.

1. 1. THE INTERNET, THE CLOSET THING TO THE FUTURE INFORMATION HIGHWAYS

The Internet is a wide-ranging amalgam of networks based on the TCP/IP protocol. In 1969, it started out as a single network for research sponsored by ARPA, the American Department of Defence Advanced Research Projects Agency. By the end of the eighties, TCP/IP had become the network standard for all major research networks throughout the world. Currently, the Internet has replaced almost all other research networks (like UUCP, Bitnet, CSNET, etc.) and it is no longer restricted to research and education. It now connects over 8000 networks on 7 seven continents and it is estimated that over 10 million people use it on a regular basis, with an additional 10-20 million people using it just for e-mail. The Internet carries many kinds of traffic. This article will give a short introduction for helping a user to become part of this "information traffic" and to understand its educational potential. Technical details are necessarily simplified.

In order to understand how services are provided on the Internet, let's shortly define the "server-client" concept. A server is a computer and/or program which provides resources to other machines or programs. A Client (or client program) uses resources from a server. "Traffic" on the information highway refers to information packets going from one machine to another using a certain communication protocol. Each computer on the Internet has a unique ID number (e.g. 129.194.81.44). Most computers also have a symbolic name (e.g. "tecfa.unige.ch"). The last element refers to a network (e.g. "ch" means "Swiss research net"). The second last element refers to an organisation (e.g. "unige" means "University of Geneva"). The first element usually refers to a specific machine. In principle, one can connect from any machine to any other. Special gateways allow some access to machines on other kinds of networks. Let's now examine a few protocols providing interesting services on the Internet:

"FTP" stands for File Transfer Protocol; on many systems, it's also the name of a user-level program that implements that protocol. This program allows a user to transfer files to and from a remote network site.

"Telnet" is the name of the protocol which allows you to connect to a remote machine. Normally you need access privileges, but some services, e.g. library or educational database lookups are provided freely without restrictions.

"Anonymous FTP" indicates that a user may log into the remote system as user "anonymous" with an arbitrary password. Thousands of anonymous FTP servers exist. A lot of research institutions make available research papers and software that way. Normally, each country has one or several huge and well known servers "Archie" servers index major servers and you can query them to find some on which huge amounts of programs and papers are made available. Special appropriate software (including educational) on the net.

"Usenet" is a large set of computers (not restricted to the Internet) who exchange "news". People can post articles tagged with one or more universally-recognised labels, called "news groups". Those news groups are organised in hierarchies. Most of those hierarchies are public and spread out through the entire world. Some (like "Clari News") are commercial and some are local. A number of software packages are used for reading and posting articles. On the Internet, most organisations use the Network News Transfer Protocol (NNTP) to receive a "News feed", to make articles available to client news readers of the local network and to propagate locally posted articles to other sites.

Given the huge amount of files and services available and the relatively high "noise/information level" of "Usenet News", applications have been invented to organise access to network resources in a uniform way. The first widely used system has been "Gopher", a menu based system which allows you to browse and search documents, link to resources such as Usenet, Archie, campus information systems, and even connect to other Gophers. All gophers in "gopher space" are also linked with a geographical index. Since one can get easily lost in gopher space, special "Veronica Servers" exist which index major gopher sites. Gophers are very popular in education and with institutions dealing with Education.

Many other services and protocols exist. A few of them are: WAIS (a Wide Area information Server with a a powerful full-text indexing engine); Finger and X500 (systems to provide information on people); IRC ("Internet Relay Chat", implementing real-time discussion); Hytelnet (an integrator for library access), MUDs (real-time multi-user textual virtual realities, mainly used for role-playing and adventure gaming), NFS (the most popular network file sharing system for workstations), MBone ("Multicast BackBone", an extension to the IP protocol allowing for video-conferencing on standard WAN lines).

The "World-Wide Web" (WWW) is now becoming the most popular campus and world-wide information system on the Internet. It has all of Gopher's functionalities, but is not hindered by Gopher's relatively simple menu-based organisation. The WWW allows you to explore the unlimited world-wide digital "web" of information. It's integrated hypertext system relates information in an easy to use way. Almost every document you look at provides you with pointers, or links, which you can follow to other documents on related subjects. A number of interactive search engines exist to locate information in and outside of the WWW. Some of those are based on "spiders", programs crawling through the net and picking up information. The WWW is an evolving standard and it's potential for education is now recognised by both academia and business.

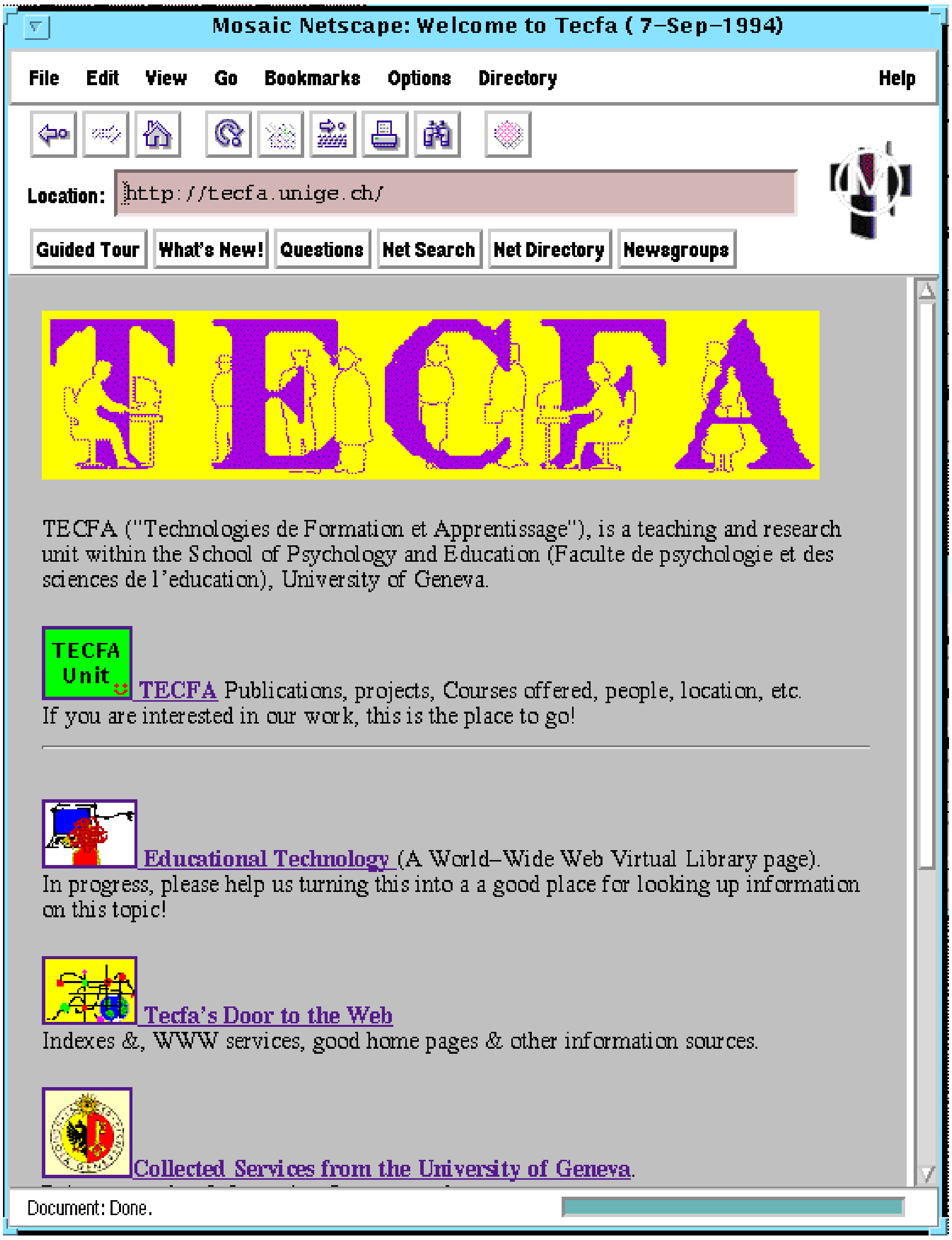

Fig 1: The example of TECFA's home page

2. TECHNICAL ASPECTS OF THE WWW, THE INTEGRATOR OF "CYBERSPACE"

The World-Wide Web has been conceived by Tim Berners-Lee at CERN in 1989 as an information integrator within which all available information on the Internet could be accessed in a simple and consistent way on every kind of machine architecture. The WWW has exponential growth rates. Currently, over 3000 WWW server sites are known. In the US Academic Network, WWW traffic (in byte counts) amounts to 7% of the total network traffic and it has a daily growth rate of 1% ! It is now among the "big three" of academic information systems, besides Gopher and "Usenet News".

A standard WWW browser (i.e. the client program for the WWW) can access at least the following communication protocols: HTTP (WWW's Hypertext Transfer Protocol), FTP, NNTP, WAIS and Gopher. Central for information retrieval is the Universal Resource Locator (URL). An URL (e.g. *"http://tecfa.unige.ch/welcome.html"*) is composed of a protocol indicator (e.g. *"http"*), an Internet machine name (e.g. *"tecfa.unige.ch"*) and a file name (e.g. *"welcome.html"*). Your can see part of this page in Figure 1. Usually the file name stands for a document to be retrieved. Sometimes, a program will be launched, e.g. *"http://www.ucc.ie/htbin/acronym"* will launch a program for looking up Internet Acronyms).

2-1 The HTML Mark-Up Language

The standard WWW information vehicle is a hypertext document, an text file encoded with the so-called "Hypertext Mark-Up Language" (HTML). An HTML file will display on your screen according to special mark-up commands and according to the possibilities of your client-program. Special high-lighted buttons in a displayed "page" allow retrieval of other pages from anywhere in the world (including any kind of file formats that your set-up can handle). When a file format is accessed that the browser does not understand (e.g. a Postscript file), a "helper" software is called, if it is available on your machine. This explains that you can configure your client to use any kind of program to display any multimedia image, sound, video and data format. Note also, that most WWW clients can display in-line images in GIF format.

Short of being able to discuss the HTML mark-up language in length, let's examine a simple example of a WWW "page" demonstrating its syntax:

```
<HTML>
<HEAD>
<TITLE>Title (Name) of this html page</TITLE>
</HEAD>

<BODY>
<H1>Big Title</H1>
<H2>Title 2</H2>Here is a list:
<UL>
  <LI> element 1
  <LI> element 2
</UL>
This is an enumerated list:
<OL>
  <LI> element 1
  <LI> element 2
</OL>
This is a simple paragraph. Text will be formatted.
<A HREF="http://tecfa.unige.ch/info-www.html">Click HERE </a> to learn more about the
World-Wide Web.<p> Here is a preformatted paragraph:
<pre>
This is a simple paragraph. Text will not be formatted.
This is a simple paragraph. Text will not be formatted.
</pre>
<A HREF="http://tecfa.unige.ch/tecfa/general/tecfa-people/schneider.html">D.S.</A>
</BODY>
```

```
</HTML>

                                Title (Name) of this html page
                  BIG TITLE
Title 2

   Here is a list:
     * element 1
     * element 2

   This is an enumerated list:
    1. element 1
    2. element 2

This is a simple paragraph. Text will be formatted. Click HERE to learn more about the
World-Wide Web.
Here is a preformatted paragraph:
This is a simple paragraph. Text will not be formatted.
This is a simple paragraph. Text will not be formatted.

D.S.
Figure 2: Sample Output using a line browser
```

HTML is an extension of the SGML language which is widely used in the editing industry for describing the logical structure of a document. An HTML element normally is delimited by tags: *"<tag_name>* text *</tag_name>"*. A tag may include a name, some attributes and some text or hypertext. For example: "*<TITLE>Title (Name) of this html page</TITLE>"* specifies the name of a page. An HTML document is composed of a single element: *"<html>...</html>"* that is, in turn, composed of head and body elements: "*<head>...</head>"* and *"<body>... </body>"*. Text displayed in a browser is part of the body and its structure is defined by tags like *"<h1>* . . . *</h1>"* (prominent header), "*<p>"* (paragraphs) or *" First item in the list, <li> Next item in the list, ...
"* (unordered lists). There are several kinds of hypertext links. Its simplest form is: *"<A HREF="*Universal Resource Locator*">D.S.</A>"* ("D.S" will display as a hypertext button). The HTML page in the example above will display with a simple line browser (like "Lynx") as shown in figure 2.There are several ways of producing HTML pages:

1. Using a simple text editor and inserting html tags manually;
2. Using a "smart" text editor (like Emacs) or an "HTML tool" (like HotMetal) which inserts the necessary HTML tags via menus or hot keys;
3. Using "filters", i.e. translation programs that translate from text processing or other formats into HTML;
4. Using programs that create HTML "on the fly", a feature frequently used to display output resulting from a data base request.

Each of these methods has its advantages. All of them are well described in the introductory literature available on the Network.

2-2 Interactive "Forms" pages

An important feature of HTML are so-called "forms". Forms allow a server to query the user with a few standard graphic user interface (GUI) widgets like push-buttons, radio-buttons, text editing window and scrolling lists. This information can then be processed by the server. It makes up a powerful query-interface to various kinds of data bases. Because HTML pages can be generated dynamically by a server, such pages can be tailored according to the needs of the user, which is an interesting feature for educational systems. An other feature are interactive maps which report to the server the position of mouse clicks. Such maps can be used to build user navigation aids.

3. 3. THE WWW AND EDUCATION

Currently, the most important educational uses of the WWW are educational information systems and dissemination of Educational Material (including text and software). The WWW has potential for distance and "just in time" open learning. However, it does not provide a full learning environment. One learns by (a) doing something and (b) by pursuing an instructional goal. Those two strategic requirements are supported by a number of elements like (a) teaching and tutoring (providing guidance to the curriculum and tasks); (b) monitoring and testing (ensuring that the learner is learning); (c) co-operation with fellow learners (in order optimally construct knowledge); (d) learning material (like exercises, simulation software, educational hypertext). The WWW can support a number of those, especially since teaching and monitoring can be done partly by the learner himself, partly by software. Today one can already find a variety of educational uses, like:

1. Information servers to look up information (manuals, books, expositions, bibliographies, programs, etc.)
2. Distribution of educational material (texts in any format, programs)
3. Curricula information & guidance to lessons
4. Educational textbooks in hypertext format, including exercises.
5. Tools for collaborative work (dynamic hypertext, "News" like conferencing systems, co-writing hypertexts)
6. Implementations of "Jigsaw puzzles", information blocks with holes that have to be filled in by the Student.
7. Question & Answering pages (e.g. tests monitoring acquired knowledge, Skinner & Bloom type of learning to some extent)
8. Interfaces to local clients (e.g. simulations, programming environments, tutors, etc.). For security reasons, those are generally restricted to local area networks.
9. Interfaces to learning environments running on a server machine (e.g. an interface that lets a user program an algorithm and shows how it is executed).

One must make a distinction between (a) person-to-many-persons" delivery of learning material and learning instructions (e.g. exercises, educational text, educational software, instructional information) and (2) person-to-person interactions. Individual tutoring and monitoring as well as collaboration between students fall into the second category and are important instructional functions. Currently, they are not the strength of the WWW, but can be provided by other asynchronous and synchronous communication software on the Internet such as Email, "virtual classrooms" and computer-supported collaborative work tools. Note however that some of those (e.g. Email, dynamic hypertext and asynchronous conferencing systems like "News") can be implemented or at least interfaced within the WWW using the "forms" interface. The fact that the WWW can't do everything is not really a problem, since at least todays Internet users have the habit to select the protocol and service that fits a task best. In the future, real-time communication tools will integrate access to the WWW. Some developers already work on this and prototypes exist.

You can visit the Globewide Network Academy ("http://uu-gna.mit.edu:8001/uu-gna/") which is a non-profit initiative for providing free Education on the Internet. Courses are entirely based on Internet tools. In addition to the WWW used for distributing curricula information, manuals, introductory texts etc., GNA uses Virtual Textual Realities (MUDs) to provide "virtual classrooms" and on-line assistance.

Now lets come back to interactive applications on the WWW, both for monitoring and proper learning. Some already do exist. However, since HTTP is a stateless protocol (meaning that the connection to a server is closed after a requested document is delivered) and since most network connections are slow due to saturations of server machines and the Internet, the practical potential of this kind of application is currently unknown. The author's WWW server points to some interesting examples. One of those is a multi-media on-line tutorial for "frog dissection" ("http://curry.edschool.virginia.edu/~insttech/frog/") by the Curry School of Education, University of Virginia. It can be used in high school biology classrooms as a valuable preparation tool or even as a useful substitute for laboratory dissection. Currently, there are many indexes on "WWW for Instructional Use". One of the best is provided by the University of Texas (http://www.utexas.edu/world/instruction/index.html)".

In May 1994 a first International Workshop on "Teaching and Learning with the Web" was held within the First International Conference on the World-Wide Web at CERN in Geneva. More information can be obtained at the author's WWW and FTP Site.

4. 4. HOW TO GET STARTED

4-1 Get connected first !

The WWW is free! However, you need access to a machine which is connected to the Internet and you should be familiar with the FTP Protocol. Throughout Europe, private customers can obtain dial-up Internet accounts by providers such as EUNET for a reasonable fee. In academia, most microcomputers have now potentially access to the Internet. Make sure that you have an appropriate TCP/IP software installed (especially under Windows). Then you can FTP a WWW client. We suggest getting Mosaic Netscape for UNIX, Mac and Windows via anonymous ftp from *"ftp.mcom.com"* or Mosaic NCSA from *"ftp.ncsa.uiuc.edu"*. If you work under MS-DOS, install Lynx (available at *"ftp2.cc.ukans.edu"*. Lynx is also available for UNIX and VMS and works great over slow dial-up connections. For a demo of Lynx, you can telnet to "ukanaix.cc.unkans.edu". Login as "www". For most WWW server and client software, binaries are provided and installation is not too difficult.

4-2 More Information

The author's WWW server (*"http://tecfa.unige.ch/"*) can be used as a starting point for general "web surfing" and obtaining technical information on the Internet and the WWW. More importantly, the provided "WWW Virtual Library on Educational Technology is a good starting point for information on educational technology on the Internet and educational services provided by the Internet.

4-3 Introductory Bibliography

GAFFIN, A. and HEITKOETTER, Jörg (1994) *Big Dummy's Guide to the Internet.* Electronically available via WWW from *"http://www.germany.eu.net/books/bgtti/bgtti.html"* or via FTP from *"ftp://ftp.germany.eu.net/pub/books/big-dummys-guide/"*. Huge, up-to-date and recommended.

GILSTER, P. (1993) *The Internet Navigator: The Essential Guide to Network exploration for the Individual Dial-Up User,* Wiley, New York, 470 pages, 1993. Recommended.

HAHN,H. (1993) *The Internet Complete Reference*, McGraw-Hill, New York, 800 pages, 1993.

KEHOE,B.P. (1992) *Zen and the Art of the Internet: A Beginner's Guide to the Internet*. Prentice Hall, Englewood Cliffs, 112 pages, 1992. The big classic, but outdated.

KROL,E. (1992) *The Whole Internet: Catalog & User's Guide*. Sebastopol, O'Reilly, 376 pages, 1992. Recommended, but somewhat outdated.

LAQUEY, T. and RYER,J.C. (1993), *The Internet companion: a beginner's guide to global networking*. Addison-Wesley, New York, 208 pages, 1993. Recommended. An on-line version is available at *"ftp://ftp.germany.eu.net/books/laquey/"*.

RANDALL, Neil (1994) *Teach yourself the Internet around the world in 21 days*, SAMS Publishing, 1994. Not too authorative, but has many pictures from many WWW sites.

WIGGINS, Richard W. (1994) *The Internet for Everyone: A Guide for Users and Providers*: McGraw-Hill, New York, 1994. Most recommended, has a good section on the WWW.