World Wide Web Primer

Nathan Torkington

September 16, 1993

Introduction

This document is an introduction to the World Wide Web. It is intended to be a gentle primer for users who have heard of the Web and wish to learn more. It explains the concepts underlying the Web, and explains how to try it out for yourself. It is not intended to be a guide to providing information on the Web.

This document is available on the Web, as well as being posted fortnightly to the Usenet newsgroups **comp.infosystems.www**, **alt.hypertext**, **news.answers**, **comp.answers** and **alt.answers**. It is available as LaTeX, plain ASCII, DVI and Postscript files via anonymous ftp. For instructions on retrieving the latest version of this document, consult the last section, called "How to obtain this document".

This document was last revised on *Wed Sep 15 15:26:45 NZT 1993* by *Nathan Torkington.*

Table of Contents

6. Providing Information
7. See Also
8. How to obtain this document

The Vision of the Web

The World Wide Web is the vision of programs that can understand the numerous different information-retrieval protocols (FTP, Telnet, NNTP, WAIS, gopher, ...) in use on the Internet today as well as the data formats of those protocols (ASCII, GIF, Postscript, DVI, TeXinfo, ...) and provide a single consistent user-interface to them all. In addition, these programs would understand a new protocol (HTTP) and a new data format (HTML) both geared toward hypermedia.

The programs already exist — "Lynx", "Mosaic", "Cello" and CERN's "LineMode Browser" are in use at hundreds, if not thousands, of sites around the Internet today. The ability of the programs to understand existing protocols means that they can access the huge body of gopherspace, FTP files, WAIS databases and news articles already extant. In addition to this, large amounts of new hypertext is being introduced through HTTP and HTML.

What is in the Web

gopher is a similar system to the Web, but not as powerful. The gopher software implements its own protocol, with limited access to other protocols. **gopherspace**, as the information accessable through gopher is called, consists of menus which can contain text files, binary files, images, keyword-search items, or more menus. The principle limitation of gopher is that it can't exploit hypertext. See the entry on "gopher" in the section "See Also" for information on obtaining the gopher software.

Hypertext is a term first coined by Ted Nelson. It is the logical combination of computers and text — a computer interface to text which allows cross-references to be followed. In a graphical situation, the user can follow cross-references by clicking with their mouse on the phrase cross-referenced. This would bring up the document at the "other end" of the cross-reference. **Hypermedia** is the extension of this to include graphics and audio as things which can be selected or viewed.

WAIS is a full-text database system produced by Thinking Machines Corp, and placed in the public domain. Full-text databases allow retreival of documents by specifying any of the words which occur in them. WAIS also gives document ranking and (with the appropriate extensions) boolean searches. WAIS servers communicate with users' programs via the ANSI standard Z39.50 protocol. See the entries on "WAIS" and "Z39.50" in the section "See Also" for information on obtaining the WAIS software.

FTP is the standard Internet protocol for copying files between computers. A very large amount of information is available via **anonymous FTP**, a variant of FTP where a set of files is made available for public access. See the entry on "FTP" in the section "See Also" for information on obtaining the source code to an FTP server. See the entry "FTP by Mail" for instructions on doing FTP through e-mail.

NNTP is a protocol used for moving around **Usenet News**. This is like the bulletin-board of the Internet (although plenty of non-Internet users also contribute), with articles being contributed on a wide variety of subjects. The articles are grouped into **newsgroups** depending on their content — the author of an article specifies which newsgroup(s) it is to go in. See the entry on "NNTP" in the section "See Also" for information on obtaining the source code to an NNTP server.

Documents on the Web are referred to using URLs (Uniform Resource Locators). A URL looks like **http://www.vuw.ac.nz/campus/home.html**. It consists of three parts — the method of retrieving the document (**http**), an option machine name (**www.vuw.ac.nz**) and a pathname (**/campus/home.html**). The URL format is nearly an Internet standard — see the entry "URL" in the section "See Also" to find more information on URLs.

How to See More

Several computers on the Internet have public-access World Wide Web clients accessable by **telnet**. Here is the current list:

info.cern.ch You will be connected directly (no username or password required). This is in Switzerland, however, so non-European users might be better off using a closer browser.

ukanaix.cc.ukans.edu A full screen browser "Lynx" which requires a vt100 terminal. Log in as **www**.

www.njit.edu (or telnet 128.235.163.2) Log in as **www**. This is a full-screen browser in New Jersey Institute of Technology, USA.

www.huji.ac.il A dual-language Hebrew/English database, with links to the rest of the world. The line mode browser, plus extra features. Log in as **www**. Hebrew University of Jerusalem, Israel.

sun.uakom.cs Slovakia. Has a slow link, only use from nearby.

If you are interested in this, consider compiling or FTPing one of the browsers so you can browse from your own machine. Browsers exist for IBM PCs, Unix, VMS and Macintoshes, and there are at least two for X Window System users as well. The entry on "Browsers" in the section "See Also" has a list of which browsers are available for which computers.

Providing Information

To add information to the Web, you will need either a HTTP server, a gopher server, an FTP server, or a WAIS database server. These are all available in source-code via anonymous FTP (see the relevant entries in the section "See Also" for information on obtaining the source code for these servers). Which server you choose depends on your needs.

If you are only wanting to serve plain-ASCII databases, then install a WAIS server. If you want to serve unformatted text, with the option for WAIS searching, install a gopher server and WAIS.

If you want to deliver hypertext, and speed is unimportant, use an FTP server (beware, though — FTP is very slow for this). If you want to deliver hypertext with reasonable speed, use an HTTP server.

A thorough discussion of the merits and disadvantages of the three main HTTP servers appears in the companion document "An Information Provider's Guide to Web Servers". See the section "How to obtain this document" for more information. What follows is a summarised version of that document.

There are three HTTP servers around, all available without restriction to academic users. They are:

The CERN Server This has mapping (ability to redirect requests), a security filter, and can act as a gateway to most things.

The NCSA Server This is a small and simple server, with the ability to act as an annotation server as well. It can also understands the gopher setup, and can run on top of the same data.

Plexus This is written in Perl (see the entry on "Perl" in the section "See Also" for more information on Perl) by Tony Sanders (sanders@bsdi.com). It comes with ArchiePlex, an archie gateway (see the entry on "archie" in the section "See Also" for more information on archie) and various calendar, manual page and finger gateways. It even has a converter from setext to HTML (see the entry on "setext" in the section "See Also" for more information on setext).

See the entry on "Servers" in the section "See Also" for instructions on obtaining these HTTP servers. The newsgroup comp.infosystems.www is a good place to ask questions for help on compilation and setup.

If you are serving hypertext to the Web, you will need to know about HTML (the HyperText Markup Language) and the converters that exist between HTML and RTF, LaTeX and others. See the document "An Information Provider's Guide to HTML", posted fortnightly to comp.infosystems.www, for more information on HTML, and HTTP servers. The entry on "HTML" in the section "See Also" has more information on obtaining this document. Note that **you**

don't need to know about HTML if you're not serving hypertext.

See Also

Archie Archie is a database of files available via anonymous ftp. You can specify a filename, part of a filename, or a regular expression, and archie will give you the name of the computers that have the filename you asked for available via anonymous FTP. For more information on archie, see the file README available through anonymous FTP in the directory pub/archie/doc on archie.ans.net.

Browsers MS-DOS users have several choices, depending on their software installation. Windows users, with an appropriate Winsock-compliant TCP/IP stack, should use Cello or NCSA Mosaic. PC-NFS users should try the CERN LineMode Browser.

Macintosh users should try MacMosaic (currently in alpha-test) or MacWWW.

X Window System users should try XMosaic — XMosaic requires the Motif libraries to compile, but precompiled binaries are available for many platforms. A similar interface is provided by TkWWW — TkWWW uses the tcl/tk language and graphics libraries.

Unix users can obtain CERN's simple LineMode Browser. The browser Lynx, is harder to compile but looks better.

VMS users can use the LineMode browser or Lynx, or the VMS WWW browser (see "VMS" for information on obtaining this browser).

Cello Cello is available via anonymous FTP from fatty.law.cornell.edu in the directory /pub/LII/Cello/.

CERN Server The CERN server is available via anonymous FTP from info.cern.ch, in the directory /pub/www/src/ as WWW-LineMode_XXX.tar.Z where XXX is a version number.

gopher The gopher software is available via anonymous FTP from boombox.micro.umn.edu in the /pub/gopher/ directory. The gopher protocol is documented in RFC 1436 (see the entry "RFC" to find out how to obtain copies of RFCs).

FTP The source for a reliable and useful FTP server is available via anonymous FTP from ftp.uu.net in the directory /networking/ftp/wuarchive-ftpd/

FTP By Mail Send e-mail to mail-server@rtfm.mit.edu with "send usenet/news.answers/finding-sources" in the body.

HTML HTML, the HyperText Markup Language, is document in the file html-spec.txt.Z, available via anonymous FTP from info.cern.ch in the directory /pub/www/doc/ or via anonymous FTP from ftp.uu.net in the directory /networking/info-service/www/doc/. The document "An Information Provider's Guide to HTML" is posted fortnightly to the same Usenet newsgroups as this document, and is available via FTP from the same places (see the section "How to obtain this document" for more information).

LineMode Browser The PC-NFS version is available via anonymous FTP from info.cern.ch in the directory /pub/www/bin/pc-nfs/wwwpcnfs.zip. The source-code (which compiles under Unix and VMS) is available via anonymous FTP from info.cern.ch in the directory /pub/www/src/ as WWWLineMode_XXX.tar.Z, where XXX is a version number.

Lynx Lynx requires the "curses" full-screen library, and is available via anonymous FTP from ftp2.cc.ukans.edu.

MacWWW MacWWW is available via anonymous FTP from info.cern.ch in the directory /pub/www/bin/mac/.

Mosaic WinMosaic is currently in alpha-test and is not available to the public. MacMosaic is also in alpha-test, and is available via anonymous FTP from ftp.ncsa.uiuc.edu in the directory /Web/MacMosaic/. XMosaic is in wide release and is available via anonymous FTP from ftp.ncsa.uiuc.edu in source form in the directory /Web/xmosaic-source/ and in binary form in the directory /Web/xmosaic-binaries/.

NCSA Server The NCSA server is available via anonymous FTP from ftp.ncsa.uiuc.edu in the directory /Web/ncsa_httpd/ as ncsa-httpd-0.4.tar.Z — the 0.4 is the version number, and will change if new versions are released.

NNTP The Net-News Transfer Protocol (NNTP) is described in RFC 977 (see the subsection "RFC" for information on obtaining RFCs). Several implementations are available, the latest and most efficient is INN (available via anonymous FTP from ftp.uu.net in the directory /networking/news/nntp/inn/).

Obtaining the servers The CERN server is available via anonymous FTP from info.cern.ch, in the directory /pub/www/src/ as WWWLineMode_XXX.tar.Z where XXX is a version number. The NCSA server is available via anonymous FTP from ftp.ncsa.uiuc.edu in the directory /Web/ncsa_httpd/ as ncsa-httpd-0.4.tar.Z — the 0.4 is the version number, and will change if new versions are released. Plexus is [where].

Perl Perl is an interpreted language, especially good for text handling. It is available for anonymous FTP from ftp.uu.net in the directory /pub/languages/perl/ as perl.tar.gz.

Plexus Plexus is [where?].

Provider's Guide The document "An Information Provider's Guide to Web Servers" is posted fortnightly to the same Usenet newsgroups as this document, and is available via FTP from the same places (see the section "How to obtain this document" for more information).

RFC RFC stands for "Request for Comments". Internet RFCs are documentation of protocols, proposals, pipe-dreams and plans. They are numbered sequentially from 1, and are available for anonymous FTP from nic.ddn.mil in the directory /rfc/.

setext setext stands for Structure Enhanced Text, and is a markup system that provides a way to format ASCII documents with visually unobtrusive anchors to parts of it above the paragraph level. More information is available via anonymous FTP from garbo.uwasa.fi in the directory /mac/tidbits/setext/

TkWWW TkWWW is available via anonymous FTP from any X11 site in the contrib/ directory — TkWWW uses the tcl/tk language and graphics libraries.

URL The draft URL specification is available via anonymous FTP from info.cern.ch in the directory /pub/www/doc/ as urlX.txt where X is a version number.

VMS The Hebrew University of Jerusalem have a VMS browser tested under UCX/Multinet and UCX_APX (Alpha). It uses the VMS/SMG screen routines and is available via anonymous FTP from www.huji.ac.il in the directory /www/vms_client/.

WAIS The Thinking Machines release of WAIS is available via anonymous FTP from ftp.uu.net in the directory /networking/info-service/wais/ as wais-8-b5.1.tar.Z. The CNIDR release of free-WAIS is available via anonymous FTP from ftp.cnidr.org in the directory /pub/NIDR.tools/ as freeWAIS-0.1.tar

Z39.50 ANSI standard Z39.50 is a standard for communication for information retrieval. The draft specification is available via anonymous FTP in the same place as either version of the WAIS source. The file is probably called z3950-spec.txt. ANSI charge for paper copies of the real standard.

How to obtain this document

The latest version of this document is always available on the Web as **http://www.vuw.ac.nz/non-local/gnat/www-primer.html**, and the most recently posted ASCII version will be available via anonymous FTP from rtfm.mit.edu in the directory /pub/usenet/news.answers/www as **primer**. The ASCII, LaTeX, DVI, and PostScript versions will be available via anonymous FTP from wuarchive.wustl.edu in the directory /doc/misc/www/.

This document is part of a series: "World Wide Web Primer", "An Information Provider's Guide to HTML", and "An Information

Provider's Guide to Web Servers". The other documents in the series are available from the archives above.

Please send feedback to the author, Nathan Torkington, at the e-mail address Nathan.Torkington@vuw.ac.nz — all discussion will be treated as public domain and may be used in future versions of this document.