Distributed Systems

Kevin Hughes
Enterprise Integration Technologies
459 Hamilton Avenue
Palo Alto, CA 94301
Email: kevinh@eit.com
WWW: http://www.eit.com/

Entering the World-Wide Web: A Guide to Cyberspace

What is the World-Wide Web?

For fifty years, people have dreamt of the concept of a universal information database—data that would not only be accessible to people around the world, but information that would link easily to other pieces of information so that only the most important data would be quickly found by a user. It was in the 1960's when this idea was explored further, giving rise to visions of a "docuverse" that people could swim through, revolutionizing all aspects of human-information interaction, particularly in the educational field. Only now has the technology caught up with these dreams, making it possible to implement them on a global scale.

The official description of the World-Wide Web is a "wide-area hypermedia information retrieval initiative aiming to give universal access to a large universe of documents." What the World-Wide Web (WWW, W3) project has done is provide users on computer networks with a consistent means to access a variety of media in a simplified fashion. Using a popular software interface to the Web called Mosaic, the Web project has changed the way people view and create information—it has created the first true global hypermedia network.

What is the Internet?

The Internet is the catch-all word used to describe the massive world-wide network of computers. The word "internet" literally means "network of networks." In itself, the Internet is comprised of thousands of smaller regional networks scattered throughout the globe. On any given day it connects roughly 15 million users in over 50 countries. The World-Wide Web is mostly used on the Internet; they do not mean the same thing. The Web refers to a body of information—an abstract space of knowledge, while the Internet refers to the physical side of the global network, a giant mass of cables and computers.

How was the Web created?

The Web began in March 1989, when Tim Berners-Lee of CERN (a collective of European high-energy physics researchers) proposed the project to be used as a means of transporting research and ideas effectively throughout the organization. Effective communications was one of CERN's goals of for many years, as its members were located in a number of countries.

How popular is the Web?

From January to August 1993, the amount of network traffic (in bytes) across the National Science Foundation's (NSF's) North American network attributed to Web use multiplied by 414 times. By September 1993, the Web was ranked 13th of all network services in terms of sheer byte traffic, up from 127th in January. Today there are at least 100 hypertext Web servers in use throughout the world. Since its inception, the CERN Web server traffic has doubled every four months—twice the rate of Internet expansion.

What is Mosaic?

Months after CERN's original proposal, the National Center for Supercomputing Applications (NCSA) began a project to create an interface to the World-Wide Web. One of NCSA's missions is to aid the scientific research community by producing widely available, non-commercial software. Another of its goals is to investigate new research technologies in the hope that commercial interests will be able to profit from them. In these ways, the Web project was quite appropriate. The NCSA's Software Design Group began work on a versatile, multi-platform interface to the World-Wide Web, and called it Mosaic.

In the first half of 1993, the first version of NCSA's Web browser was made available to the Internet community. Because of the number of traditional services it could handle, and due to its easy, point-and-click hypermedia interface, Mosaic soon became the most popular interface to the Web. Currently versions of Mosaic can run on Suns, Silicon Graphics workstations, IBM-compatibles running Microsoft Windows, Macintoshes, and computers running other various forms of UNIX.

What can Mosaic do?

Mosaic running on every supported computer should have the following features:

EDITOR'S NOTE

This article is a trimmed-down version of Kevin's report that was distributed at Hypertext '93. The full version is available on the World-Wide Web at `http://www.hcc.hawaii.edu/guide/www.guide.html`. Updates to this guide will be made available at `ftp.eit.com`. Kevin is also working on "The Webmaster's Cookbook," an in-depth guide to setting up and maintaining WWW sites.

REFERENCE

See *SIGLINK Newsletter*, Vol II No 2, pp. 8-9, for a quick tour of the World-Wide Web, including Lynx.

EDITOR'S NOTE

World-Wide Web statistics are available by FTP from `nic.merit.edu`. Also, check out "Measuring the Size and Growth of the Web," `http://www.mit.edu:8001/afs/sipb/user/mkgray/ht/web-growth.html`.

REFERENCE

See page 8 for a screen snapshot of Macintosh Mosaic.

A consistent mouse-driven graphical interface.
The ability to display hypertext and hypermedia documents.
The ability to display electronic text in a variety of fonts.
The ability to display text in bold, italic, or strikethrough styles.
The ability to display layout elements such as paragraphs, lists, numbered and bulleted lists, and quoted paragraphs.
Support for sounds (Macintosh, Sun audio format, and others).
Support for movies (MPEG-1 and QuickTime).
The ability to display characters as defined in the ISO 8859 set (it can display languages such as French, German, and Hawaiian).
Interactive electronic forms support, with a variety of basic forms elements, such as fields, check boxes, and radio buttons.
Support for interactive graphics (in GIF or XBM format) of up to 256 colors within documents.
The ability to make basic hypermedia links to and support for the following network services: ftp, gopher, telnet, nntp, WAIS.
The ability to extend its functionality by creating custom servers (comparable to XCMDs in HyperCard).
The ability to have other applications control its display remotely.
The ability to broadcast its contents to a network of users running multiplatform groupware such as NCSA's Collage.
Support for the current standards of HTTP and HTML.
The ability to keep a history of traveled hyperlinks.
The ability to store a list and retrieve a list of URLs for future use.

The Mosaic browsers for X and Windows now support interactive forms, which allows people to interact with forms on the screen that they can fill out and submit. This has sparked the development of on-line ordering forms and catalogs and interactive multimedia applications of all kinds within the past six months. Forms support will be available on the Macintosh Mosaic browser probably within six months.



The Web server sends the hypermedia HTML document to the client. The client interprets the HTML tags, formats the document, and displays it on the screen.

What is available on the Web?

Currently the Web offers the following through a hypertext, and in some cases, hypermedia, interface:

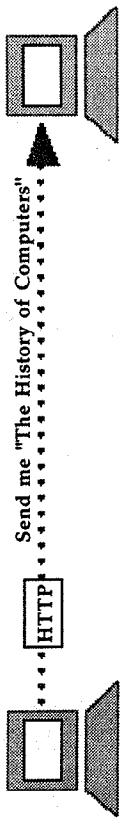
Anything served through Gopher
Anything served through WAIS (Wide-Area Information Service)
Anything served through anonymous FTP sites
Full Archie services (a FTP search service)
Full Veronica services (a Gopher search service)
Full CSO, X.500, and whois services (Internet phone book services)
Full finger services (an Internet user lookup program)
Any library system using PALS (a library database standard)
Anything on Usenet
Anything accessible through telnet
Anything in hytelnet (a hypertext interface to telnet)
Anything in techinfo or texinfo (forms of campus-wide information services)
Anything in hyper-g (a networked hypertext system in use throughout Europe)
Anything in the form of man pages
HTML-formatted hypertext and hypermedia documents

How does the Web work?

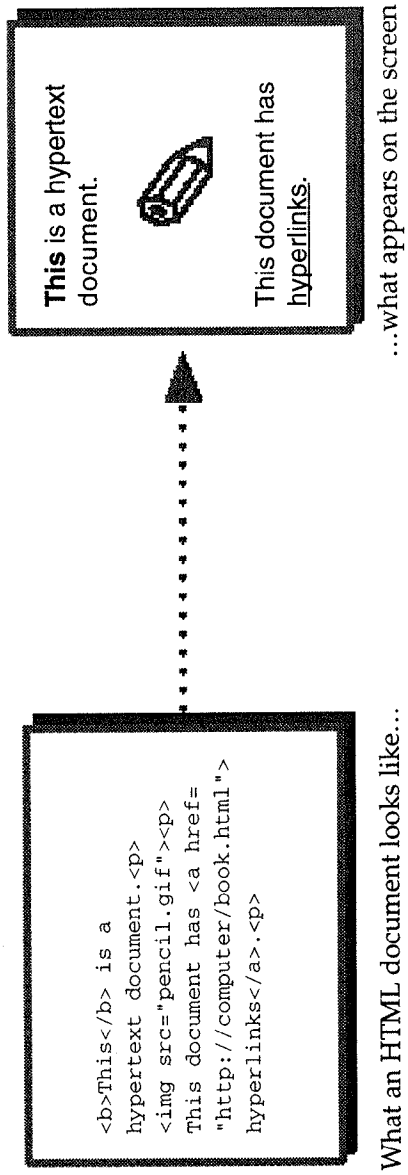
The Web works under the popular client-server model. A Web server is a program running on a computer whose only purpose is to serve documents to other computers when asked to. A Web client is a program that interfaces with the user and requests documents from a server as the user asks for them. Because the server does a minimal amount of work (it does not perform any calculations) and only operates when a document is requested, it puts a minimal amount of workload on the computer running it.

Here's an example of how the process works (see the figures at the right):

1. Running a Web client (also called a browser), the user selects a piece of hypertext connected to another text—"The History of Computers."
2. The Web client connects to a computer specified by a network address somewhere on the Internet and asks that computer's Web server for "The History of Computers."
3. The server responds by sending the text and any other media within that text (pictures, sounds, or movies) to the user's screen.



The client sends an HTTP message to a computer running a Web server program and asks for a specific document.



The World-Wide Web is composed of thousands of these virtual transactions taking place per hour throughout the world, creating a web of information flow.

The language that Web clients and servers use to communicate with each other is called the HyperText Transmission Protocol (HTTP). All Web clients and servers must be able to speak HTTP in order to send and receive hypermedia documents. For this reason, Web servers are often called HTTP servers.

The phrase "World-Wide Web" is often used to refer to the collective network of servers speaking HTTP as well as the global body of information available using the protocol.

The standard language the Web uses for creating and recognizing hypermedia documents is the HyperText Markup Language (HTML). It is loosely related to, but technically not a subset of, the Standard Generalized Markup Language (SGML), a document formatting language used widely in some computing circles.

HTML is widely praised for its ease of use. Web documents are typically written in HTML and are usually named with the suffix ".html". HTML documents are nothing more than standard 7-bit ASCII files with formatting codes that contain information about layout (text styles, document titles, paragraphs, lists) and hyperlinks. Many free software converters are available for translating documents in foreign formats to HTML.

The current HTML standard (HTML) supports basic hypermedia document creation and layout, but for current use it is still limited. The latest version of HTML, called HTML+, is still under development but will probably be completely defined by the end of 1994. HTML+ will support interactive forms, defined "hotspots" in images, more versatile layout and formatting options and styles, and formatted tables, among many other improvements.

HTML uses what are called Uniform Resource Locators (URLs) to represent hypermedia links and links to network services within documents. It is possible to represent nearly any file or service on the Internet with a URL.

The first part of the URL (before the two slashes) specifies the method of access. The second is typically the address of the computer where the data or service is located. Further parts may specify the name of a file, the port to connect to, or the text to search for in a database.

Here are some examples of URLs:

`file://pulua.hcc.hawaii.edu/sound.au`—Retrieves a sound file and plays it.

`file://pulua.hcc.hawaii.edu/picture.gif`—Retrieves a picture and displays it, either in a separate program or within a hypermedia document.

`http://pulua.hcc.hawaii.edu/directory/book.html`—Connects to an HTTP server and retrieves an HTML file.

`ftp://pulua.hcc.hawaii.edu/pub/file.txt`—Opens an FTP connection to pulua.hcc.hawaii.edu and retrieves a text file.

`gopher://pulua.hcc.hawaii.edu`—Connects to the Gopher at pulua.hcc.hawaii.edu.

`telnet://pulua.hcc.hawaii.edu:1234`—Telnets to pulua.hcc.hawaii.edu at port 1234.

`news:alt.hypertext`—Reads the latest Usenet news by connecting to a user-specified news (NNTP) host and returns the articles in hypermedia format.

Most Web browsers allow the user to specify a URL and connect to that document or service. When selecting hypertext in an HTML document, the user is actually sending a request to open a URL. In this way, hyperlinks can be made not only to other texts and media, but also to other network services. Web browsers are not simply Web clients, but are also full-featured FTP, Gopher, and telnet clients.

What software is available?

World-Wide Web clients (browsers) are available for the following platforms and environments: **Text-only (dumb) terminals**, nearly any platform; **UNIX**, text-only using curses, for SunOS 4, AIX, Alpha, Ultrix; **VMS**; **X11/Motif**, for IRIX (Silicon Graphics), SunOS 4, RS/6000, DEC Alpha/OSF 1, DEC Ultrix; **NeXT**, for NeXTStep 3.0; **IBM compatibles**, 386 and above, under Microsoft Windows; **Macintosh** computers, Classic and above; **Commodore Amiga**; **Perl**; **Emacs**.

World-Wide Web servers are available for the following platforms and environments: **UNIX**, **Perl**, **Macintosh**, **VM** and **VMS**.

How can I get more information?

Most of this information is available on the Internet. In order to access resources specified in URL format, you may need to use a Web browser or connect to a telnet site that provides a public-access browser.

General web information

Main CERN World-Wide Web page
`http://info.cern.ch/hypertext/WWW/TheProject.html`

Main NCSA Mosaic page
`http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/mosaic-docs.html`

Information on WWW
`http://www.bsdi.com/server/doc/web-info.html`

The World-Wide Web FAQ (Frequently Asked Questions) file by Nathan Torkington
`http://www.vuw.ac.nz:80/non-local/gnat/www-faq.html`

A list of World-Wide Web clients at CERN
`http://info.cern.ch/hypertext/WWW/Clients.html`

The "official" list of World-Wide Web servers at CERN
`http://info.cern.ch/hypertext/DataSources/WWW/Servers.html`

World-Wide Web newsgroup
`comp.infosystems.www`

World-Wide Web mailing lists
For general discussion: send email to `listserv@info.cern.ch`, with "add www-announce" as the body.
For developers and technical discussion: send email to `listserv@info.cern.ch`, with "add www-talk" as the body.

How to write HTML
`http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html`

How to write Web gateways and servers
`http://info.cern.ch/hypertext/WWW/Daemon/Overview.html`

HTML official specifications
`http://info.cern.ch/pub/www/doc/html-spec.multi`

HTML convertors - list of tools and convertors at CERN and NCSA
`http://info.cern.ch/hypertext/WWW/Tools/Overview.html`
`http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/faq-software.html`

HTML+ Document Type Definition (DTD)
`ftp://15.254.100.100/pub/htmlplus.dtd.txt`

HTML+ Draft
`ftp://ds.internic.net/internet-drafts/draft-raggett-www-html-00.*`

Information and reports on multimedia and hypermedia

Index to multimedia resources
`http://cui_www.unige.ch/Chloe/MultimediaInfo/index.html`

"Network Access to Multimedia Information", June 1993
`ftp ftp.ed.ac.uk`, in directory `/pub/mmaccess`
This report summarizes the requirements of academic and research users for network access to multimedia information.

"Computer Supported Cooperative Work Report", July 1993
`ftp gorgon.tft.tele.no`, in directory `/pub/groupware`
This is a comprehensive list of all known collaborative software packages and projects currently in use or under development.

"Hypermedia and Higher Education", April 1993
`gopher lewsun.idlw.ucl.ac.be`, the `/digests/IPCT` menu.
IPCT, Interpersonal Computing and Technology, is an excellent journal exploring the boundaries of education and high technology.

alt.hypertext Frequently Asked Questions list
`gopher ftp.cs.berkeley.edu`, on many other Gophers.
This list contains dozens of pointers to mailing lists, people, Internet sites, groups, books, periodicals, bibliographies, and software related to hypertext.

The author's Hypertext '93 notes, with an emphasis on WWW-related events
`http://www.eit.com/reports/ht93/ht93.report.html`

Browsers accessible by telnet

This list is from the comprehensive list of telnet-accessible clients at:
`http://info.cern.ch/hypertext/WWW/FAQ/Bootstrap.html`

`telnet info.cern.ch (or telnet 128.141.201.74)`
The simplest line mode browser. This server is in Geneva, Switzerland.

`telnet ukanaix.cc.ukans.edu`
A full screen browser "Lynx" which requires a vt100 terminal. Log in as www. University of Kansas.

`telnet www.njit.edu`
Log in as www. A full-screen browser in New Jersey Institute of Technology. USA.

`telnet vms.huji.ac.il (or telnet 128.139.4.3)`
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.

`telnet sun.uakom.cs`

EDITOR'S NOTE

This section of WWW pointers is *a lot* more useful when you are browsing the guide on the web, because each reference is just a button or key press away. Still, I found this section to be a good reference when teaching my hypertext class about the World-Wide Web this semester.

EDITOR'S NOTE

Note that Kevin's report on web activities at HT 93 has moved from where it was earlier reported (*SIGLINK Newsletter*, Vol II, No 3, p. 5).

Slovakia. Has a slow link, use from nearby.

`telnet fserv.kfki.hu`
Hungary. Has slow link, use from nearby. Login is as www.

`telnet info.funet.fi (or telnet 128.214.6.100)`
Finland.

Obtaining web browsers and servers

`ftp info.cern.ch`, in directory `/pub/www`
Simple text-only browser, as well as the CERN HTTP server.

`ftp ftp2.cc.ukans.edu`, in directory `/pub/WWW/lynx`
Distribution for Lynx, a line-mode curses-based browser.

`ftp ftp.ncsa.uiuc.edu`, in directory `/Mosaic`
Mosaic distribution, as well as the NCSA HTTP server. This site contains the Mosaic browser for Macintosh, Windows, and X.

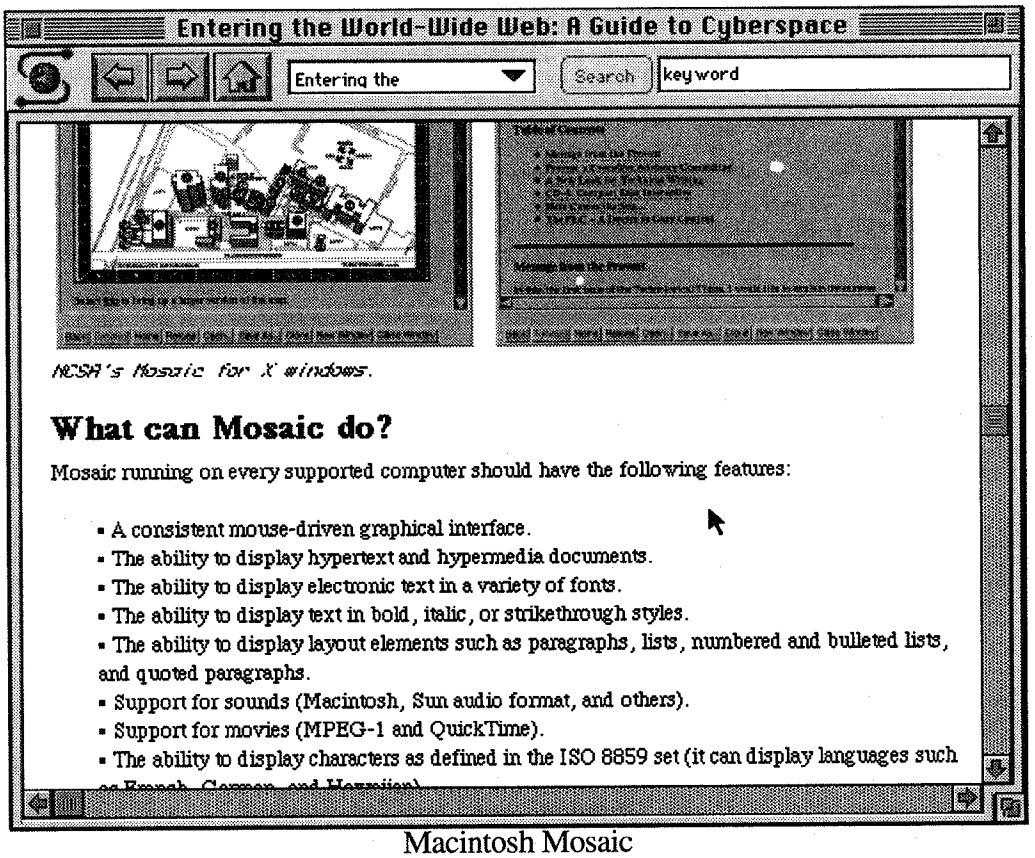
`ftp max.physics.sunysb.edu`, in `/pub/amosaic`
Distribution for Amiga Mosaic.

`ftp oac.hsc.uth.tmc.edu`, in `/public/mac/MacHTTP/machttp.sit.hqx`
Macintosh server software.

Interesting places on the web

NCSA's Demonstration Page	`http://www.ncsa.uiuc.edu/demoweb/demo.html`
Honolulu Community College	`http://pulua.hcc.hawaii.edu/`
Guide to Australia	`http://life.anu.edu.au/education/australia.html`
MSU Weather Movies	`http://rs560.cl.msu.edu/weather`
Xerox PARC (Palo Alto)	`http://pubweb.parc.xerox.com:80/`
Wired Magazine	`http://wired.com/`
English Server at CMU	`http://english-server.hss.cmu.edu/`
Electronic Frontier Found.	`http://www.eff.org/`
Legal Information at Cornell	`http://www.law.cornell.edu/lii.table.html`
Collaborative Art	`http://cui_www.unige.ch/Chloe/OtisCrosswire/`
San Francisco's Exploratorium	`http://www.exploratorium.edu/`
Global Network Navigator	`http://nearnet.gnn.com/GNN-ORA.html`
World-Wide Web EXPO	`http://sunsite.unc.edu/expo/ticket_office.html`
World Map Interface	`http://wings.buffalo.edu/world`

Kevin was a student systems administrator and World-Wide Web lecturer at Honolulu Community College when he originally wrote this paper. He is now a Software Engineer at Enterprise Intergration Technologies (EIT), an R&D and consulting company specializing in information technology for electronic commerce, collaborative engineering and agile manufacturing.



Macintosh Mosaic