Entering the World−Wide Web: A Guide to Cyberspace

Kevin Hughes
Honolulu Community College
September 1993

Table of Contents

What is the World−Wide Web?

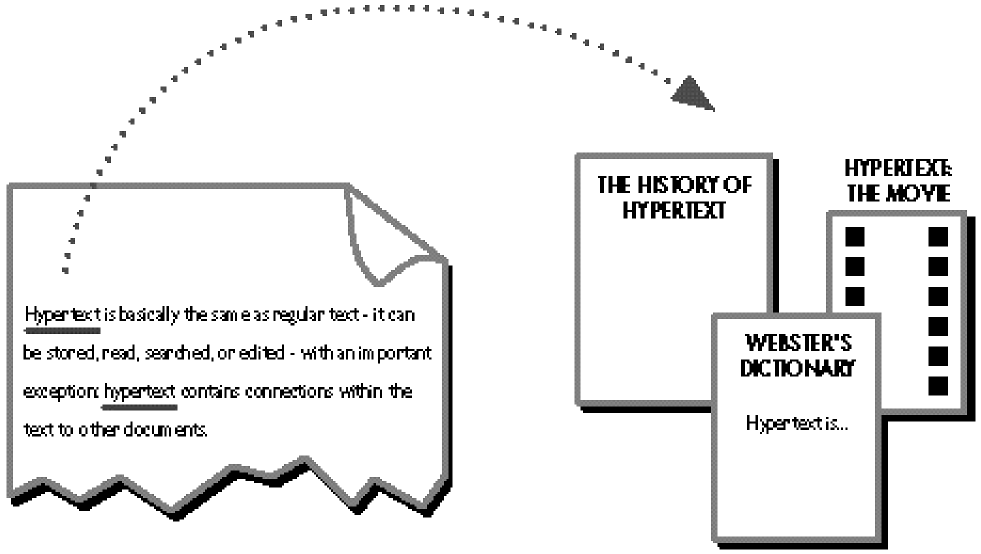
For fifty years, people have dreamt of the concept of a universal information database – data that would not only be accessible to people around the world, but information that would link easily to other pieces of information so that only the most important data would be quickly found by a user. It was 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 until now has the technology caught up with these dreams, making it possible to implement them on a global scale.

The official description describes the **World−Wide Web** as 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 hypertext and hypermedia?

The operation of the Web relies on **hypertext** as its means of interacting with users. Hypertext is basically the same as regular text – it can be stored, read, searched, or edited – with an important exception: hypertext contains connections within the text to other documents.

For instance, suppose you were able to somehow select (with a mouse or with your finger) the word "hypertext" in the sentence before this one. In a hypertext system, you would then have one or more documents related to hypertext appear before you – a history of hypertext, for example, or the Webster's definition of hypertext. These new texts would themselves have links and connections to other documents – continually selecting text would take you on a free−associative tour of information. In this way, hypertext links, called **hyperlinks**, can create a complex virtual web of connections.



Selecting a piece of hypertext... **...can take you to more hypertext, books, movies, sounds, and Images.**

Hypermedia is hypertext with a difference – hypermedia documents contain links not only to other pieces of text, but also to other forms of media – sounds, images, and movies. Images themselves can be selected to link to sounds or documents. Here are some simple examples of hypermedia:

- You are reading a text on the Hawaiian language. You select a Hawaiian phrase, then hear the phrase as spoken in the native tongue.

- You are a law student studying the Hawaii Revised Statutes. By selecting a passage, you find precedents from a 1920 Supreme Court ruling stored at Cornell. Cross–referenced hyperlinks allow you to view any one of 520 related cases with audio annotations.

- Looking at a company's floorplan, you are able to select an office by touching a room. The employee's name and picture appears with a list of their current projects.

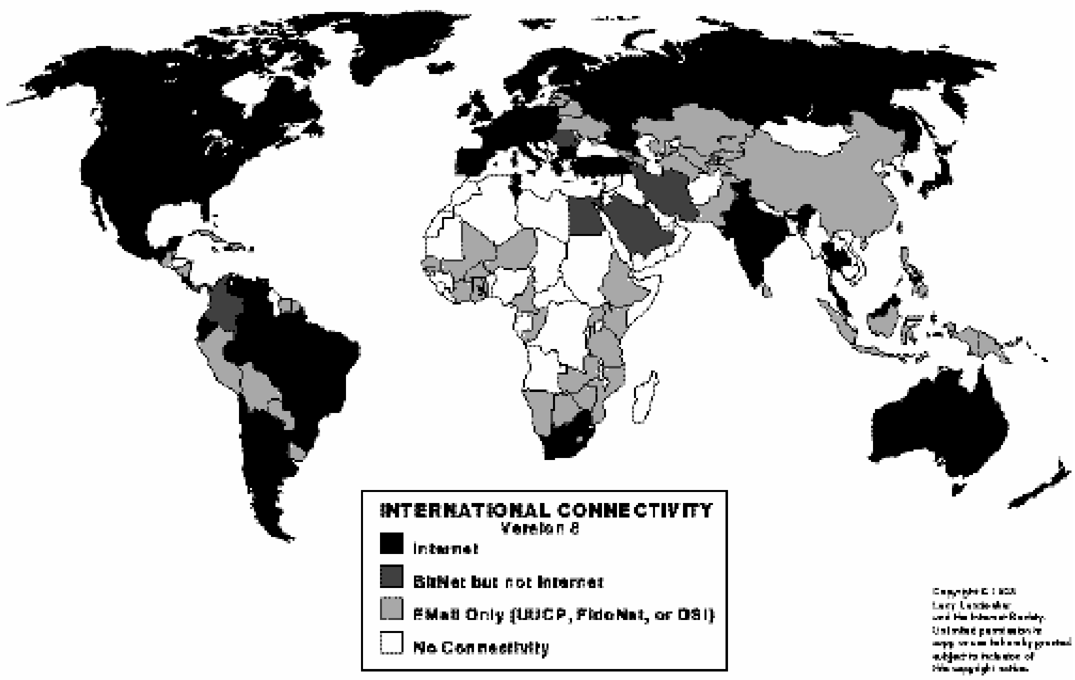
- You are a scientist doing work on the cooling of steel springs. By selecting text in a research paper, you are able to view a computer–generated movie of a cooling spring. By selecting a button you are able to receive a program which will perform thermodynamic calculations.

- A student reading a digital version of an art magazine can select a work to print or display in full. If the piece is a sculpture, she can request to see a movie of the sculpture rotating. By interactively controlling the movie, she can zoom in to see more detail.

The Web, although still in its early years, allows many of these examples to work in real life. It facilitates the easy exchange of hypermedia through networked environments from anything as small as two Macintoshes connected together to something as large as the global Internet.

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.



The countries in black are connected to the Internet.

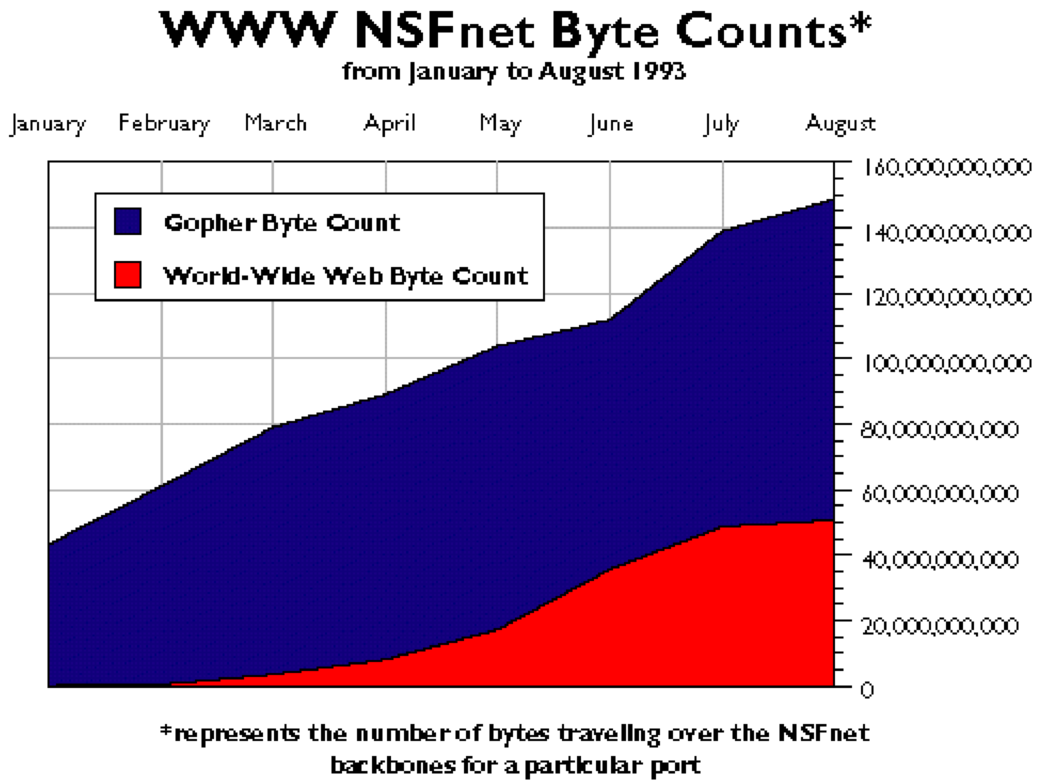
How was the Web created?

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 a goal of CERNs for many years, as its members were located in a number of countries.

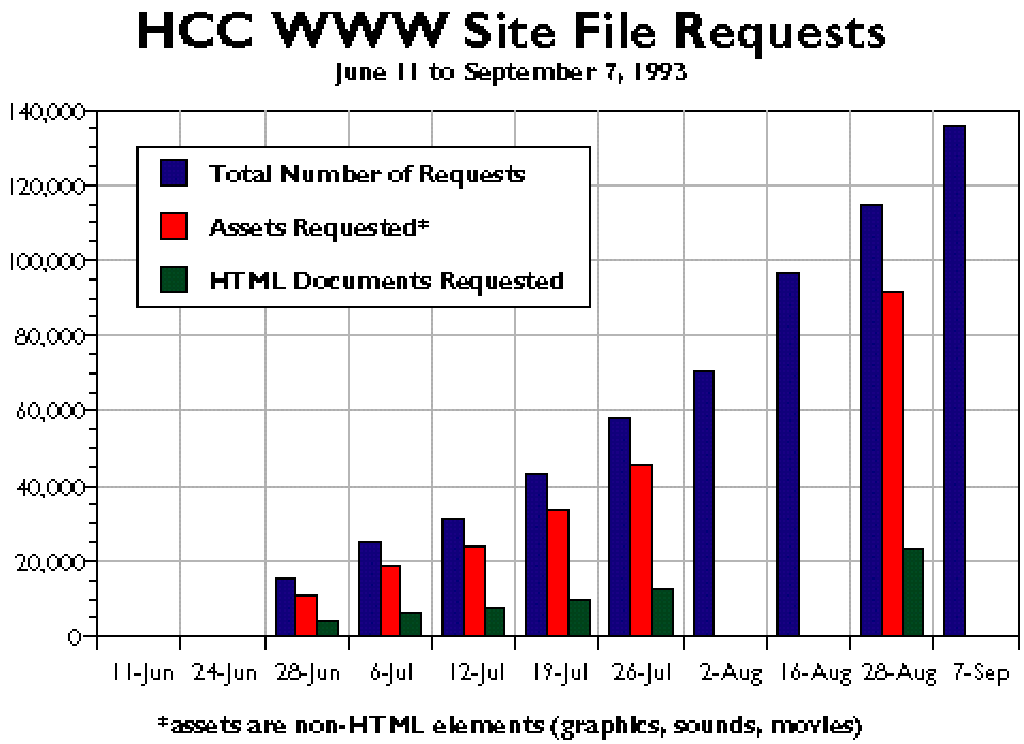
How popular is the Web?

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. The Web is now ranked 13th of all network services in terms of sheer byte traffic. In January its rank was 127. 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.



World–Wide Web growth.
Statistics available by FTP from nic.merit.edu.

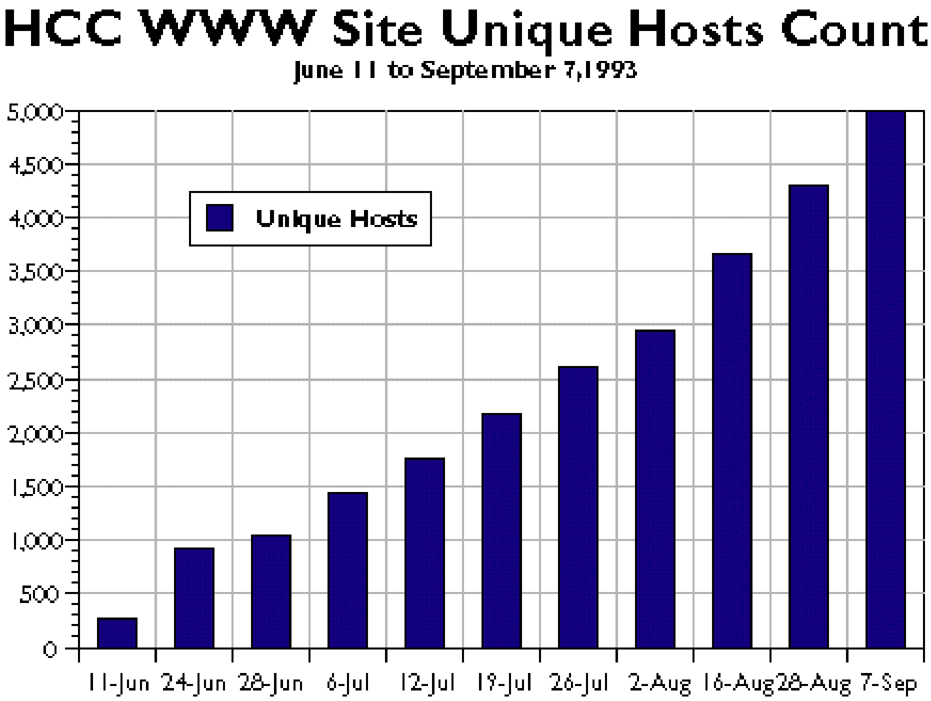
Honolulu Community College officially announced their opening of their hypermedia server – the first Web server in Hawaii – at the end of May 1993. By September of that year (after 105 days of service), they had received over 23,000 requests for documents and over 112,000 requests for assets from nearly 5,000 separate hosts on the network. From September 1 to 7 they received traffic from over 600 separate hosts, an all–time high. It is expected that traffic will increase further as the school year begins and student involvement in the Web increases.



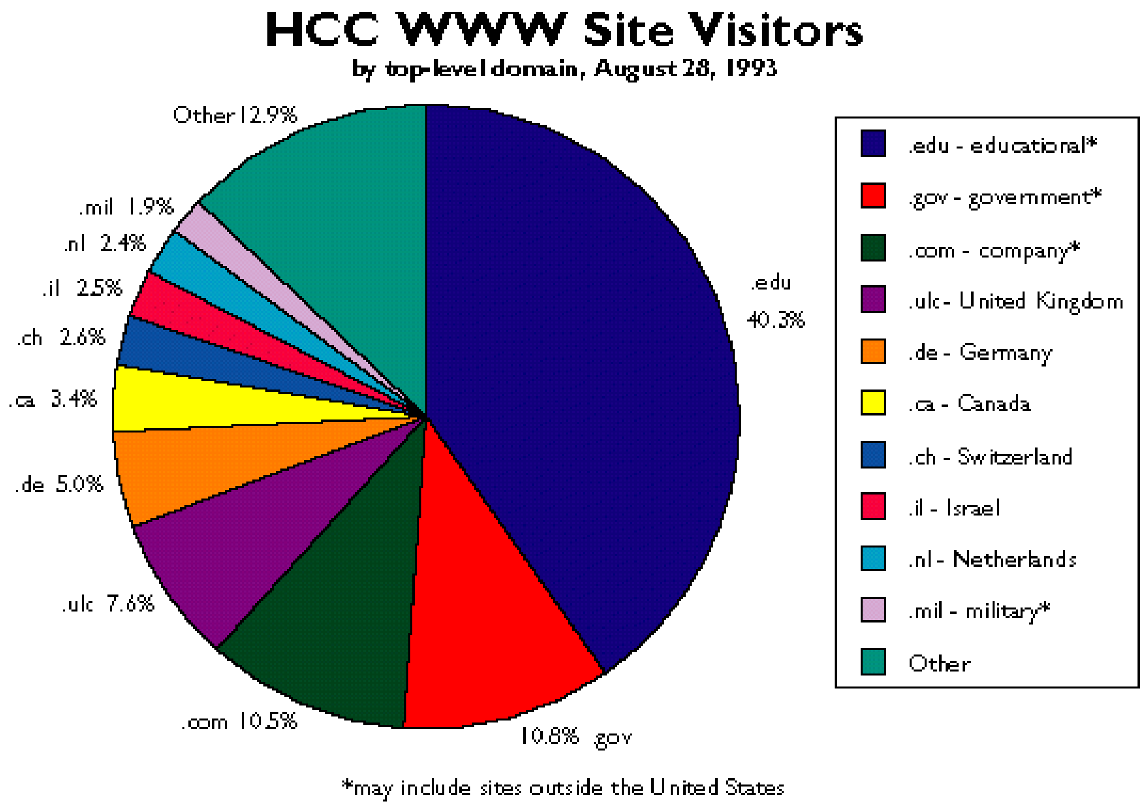
Since the site's opening, HCC has received virtual visitors from Xerox, Digital Equipment Corporation, Apple Computer, Cray, IBM, MIT's Media Lab, NEC, Sony, Fujitsu, Intel, Rockwell, Boeing, Honeywell, and AT&T (which has been one of the most frequent visitors), among hundreds of other corporate sites on the Internet.

Collegiate visitors have originated from campuses such as Stanford, Harvard, Carnegie–Mellon, Cornell, MIT, Michigan State, Rutgers, Purdue, Rice, Georgia Tech, Columbia, University of Texas, and Washington University, as well as other campuses in the United Kingdom, Germany, and Denmark, to name but a few.

Governmental visitors have come from various departments in NASA, including their Jet Propulsion Laboratories, Lawrence Livermore National Laboratories, the National Institute of Health, the Superconducting Supercollider project, and the USDA, as well as government sites in Singapore and Australia. A few dozen Army and Navy sites throughout the world have browsed around as well.



Because HCC's server began operation when there were relatively few such sites in the world, and in part due to its popularity, the growth in traffic has closely reflected the growth of the Web. Further analysis of HCC's server logs indicate the following breakdown in classifications:



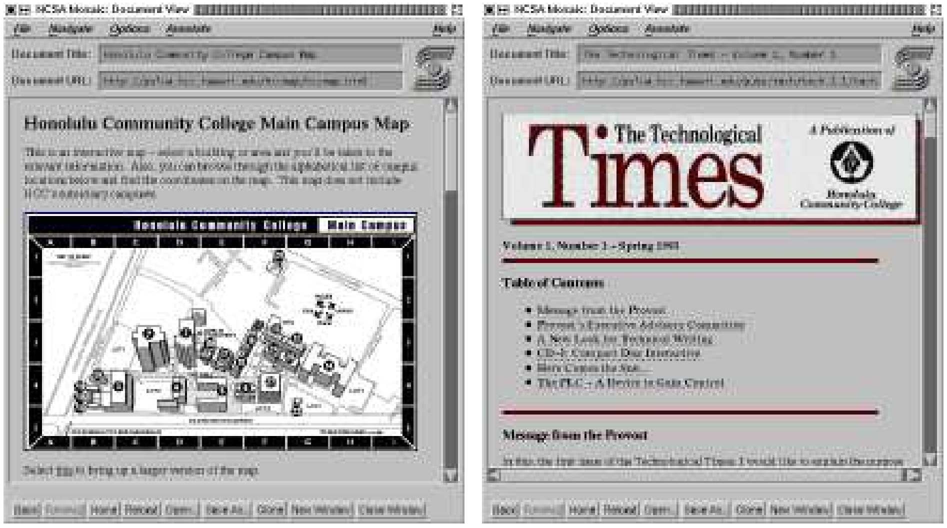
Although it is impossible to know for sure, it can be guessed that the largest segment roaming the World–Wide Web consists of four–year campus populations within the United States.

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 earlier beta versions were distributed, Mosaic had developed a strong yet small following by the time it was officially released.

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.



NCSA's Mosaic for X windows.

What can Mosaic do?

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

- 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 travelled hyperlinks.
- The ability to store a list and retrieve a list of URLs for future use.

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.

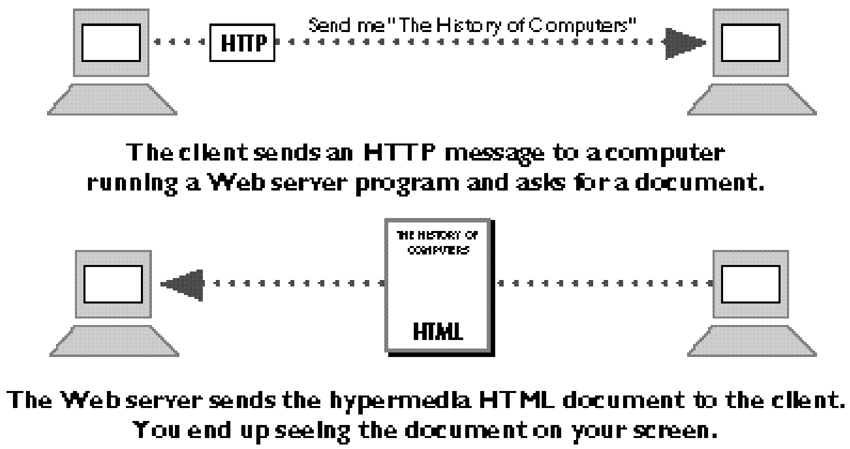
Here's an example of how the process works:

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 computers 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 users screen.

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.

Future Web servers will include encryption and client authentication abilities − they will be able to send and receive secure data and be more selective as to which clients receive information. This will allow freer communications among Web users and will make sure that sensitive data is kept private. It will be harder to compromise the security of commercial servers and educational servers which wish to keep information local. Improvements in security will facilitate the idea of "pay−per−view" hypermedia, a concept which many commercial interests are currently pursuing.



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.

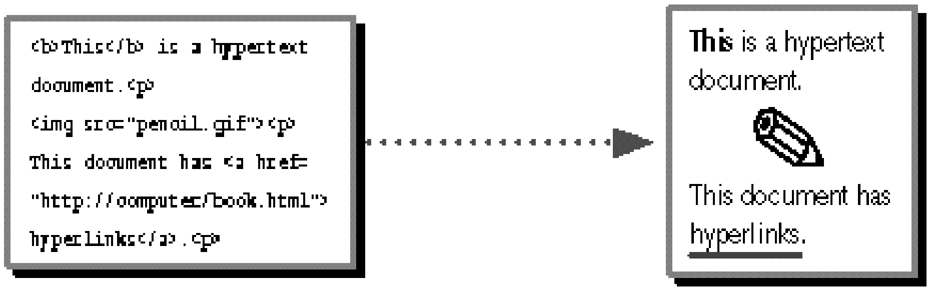
The phrase "World−Wide Web" is often used to refer to the collective network of servers speaking HTTP as well as the global body of information available using the protocol.

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 convertors 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 1993. HTML+ will support interactive forms, defined "hotspots" in images, more versatile layout and formatting options and styles, and formatted tables, among many other improvements.



What an HTML document looks like... ...what you see on the screen.

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 the data or service is located. Further parts may specify the names of files, 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.
- **file://pulua.hcc.hawaii.edu/directory/** – Displays a directorys contents.
- **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.

HTML+ will include an email URL, so hyperlinks can be made to send email automatically. For instance, selecting an email address in a piece of hypertext would open a mail program, ready to send email to that address.

What software is available?

World–Wide Web **clients** (browsers) are available for the following platforms and environments:

- Text–only (dumb) terminal, 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
- Browsers written in perl are available.
- Browsers written for the emacs environment are available.

World–Wide Web **servers** are available for the following platforms and environments:

- UNIX
- Perl
- Macintosh
- VM, VMS

For details on how to obtain Web client and server software, refer to the section "How can I get more information?"

How can I get more information?

Most of this information is available on the Internet. In order to access resources specified by 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
    mail2html, converts electronic mailboxes to HTML documents
    ftp://info.cern.ch/pub/www/dev
    Word Perfect 5.1 to HTML convertor
    http://journal.biology.carleton.ca:8001/Journal/background/ftp.sites.html
    rtf2html, converts Rich Text Format (RTF) documents to HTML
    file://oac.hsc.uth.tmc.edu/public/unix/WWW
    latex2html, converts LaTeX documents to HTML
    http://cbl.leeds.ac.uk/nikos/tex2html/doc/latex2html/latex2html.html

HTML+ Document Type Definition (DTD)
ftp://info.cern.ch/pub/www/dev/htmlplus.dtd

Information/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.

Browsers Accessible by Telnet

A comprehensive list of telnet−accessible clients
http://info.cern.ch/hypertext/WWW/FAQ/Bootstrap.html

telnet info.cern.ch
The simplest line mode browser.

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

telnet eies2.njit.edu
Log in as "www". A full−screen browser.

telnet vms.huji.ac.il
Log in as "www". A line−mode browser.

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

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

telnet info.funet.fi

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 aixtest.cc.ukans.edu, in directory /pub
    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.
ftp oac.hsc.uth.tmc.edu. in directory /public/Mac
    Macintosh server.
ftp fatty.law.cornell.edu, in directory /pub/LII/cello
    Browser for Microsoft Windows.

Note: The alpha versions of Mosaic for Windows and for Macintosh are not being distributed. To be on their alpha testers mailing lists, send mail to mosaic−win@ncsa.uiuc.edu or mosaic−mac@ncsa.uiuc.edu. Completion for these browsers is scheduled for November 1993.

About the Author

For the last two years Kevin Hughes has been working as a student systems programmer with Dr. Ken Hensarling, Honolulu Community College's Director of Academic Computing. He designed and implemented HCC's World−Wide Web site and is currently doing freelance graphics and programming work for various companies and organizations in Hawaii. He can be reached through the Internet as **kevinh@pulua.hcc.hawaii.edu**.

Index/Glossary

A

Archie
    A network service that searches FTP sites for files.

B

browser
    Software that provides an interface to the World−Wide Web.

C

CERN
    The European collective of high−energy physics researchers (European

Organization for Nuclear Research).

client
: A computer or program requests a service of another computer or program.

client−server model
: A structure in which programs use and provide distributed services.

Collage
: Collaborative (shared whiteboard) software developed by the NCSA.

CSO
: Central Services Organization. A service which facilitates user and address lookup in databases.

D

Doug Engelbart
: The inventor of many common devices and ideas used in computing today, including the mouse.

F

finger
: A service that responds to queries and retrieves user information remotely.

FTP
: File Transfer Protocol. A common method of transferring files across networks.

G

Gopher
: A versatile menu−driven information service.

H

Honolulu Community College

HTML+
: The latest version of HTML.

hyper−g
: A distributed hypertext system mostly popular in Europe.

HyperCard
: A personal hypermedia/multimedia creation system for use on Apple Computers.

hyperlinks
: Connections between hypermedia or hypertext documents and other media.

hypermedia
: Hypertext that includes or links to other forms of media.

hypertext
: Text that, when selected, has the ability to present connected documents.

HyperText Markup Language (HTML)
: The standard language used for creating hypermedia documents within the World–Wide Web.

HyperText Transmission Protocol (HTTP)
: The standard language that World–Wide Web clients and servers use to communicate.

hytelnet
: A hypertext interface to telnet.

I

Internet
: The global collective of computer networks.

M

Mosaic
: A mouse–driven interface to the World–Wide Web developed by the NCSA.

N

National Center for Supercomputing Applications (NCSA)
: A federally–funded organization whose mission is to develop and research high–technology resources for the scientific community.

National Science Foundation (NSF)
: A federally–funded organization that manages the NSFnet, which connects every major research institution and campus in the United States.

NNTP
: News Network Transfer Protocol. A common method by which articles over Usenet are transferred.

P

PALS
: A standard library database interface.

S

server
: A program which provides a service to other client programs.

SGML
: Standard Generalized Markup Language. A generic language for

representing documents.

Software Design Group
: The group within NCSA that is responsible for designing computer applications.

T

techinfo
: A common campus−wide information system developed at MIT.

Ted Nelson
: The inventor of many common ideas related to hypertext, including the word "hypertext" itself.

telnet
: A program which allows users to remotely use computers across networks.

texinfo
: A common campus−wide information system.

Tim Berners−Lee
: The inventor of the World−Wide Web.

U

Uniform Resource Locators (URLs)
: Standardized formatted entities within HTML documents which specify a network service or document to link to.

Usenet
: The global news−reading network.

V

Vannevar Bush
: Originator of the concept of hypertext.

Veronica
: A network service that allows users to search Gopher systems for documents.

W

WAIS
: Wide−Area Information Service. A service which allows users to intelligently search for information among databases distributed throughout the Internet.

whois
: A name lookup service.

World−Wide Web
: The initiative to create a universal, hypermedia−based method of access to

information. Also used to refer to the Internet.

X

X.500
: A standard which defines electronic mail directory services. Mostly used in Europe.

Thanks to Tim Berners−Lee for a better definition of the Web!

Fourth Edition: September 21, 1993