

The World Wide Web has many names: WWW, the Web, W3, and so on. In this booklet, you'll see it called the Web. Some people, especially the media, even consider the Web to be the Internet. This is entirely untrue, however--Web is just one of the many technologies that the Internet makes possible.

The Web as you know it started out in 1980 at CERN (the European Laboratory for Particle Physics). A crafty fellow named **Tim Berners-Lee** created a program called *Enquire Within Upon Everything* (Enquire) to keep track of randomly associated documents by linking them together. As he put it in a 1994 interview with Internet World, he wanted to represent information in a "brain-like" way. In 1989, Berners-Lee proposed a global extension of his ideas that would link hypertext documents all over the world. The content would be put on servers, and people would use browsers to view it. Since then, the Web has grown explosively. It surpassed Gopher in usage within its first few months. To use the Web, you need a client called a Web browser. One of the first Web browsers was Mosaic, created by Netscape founder Marc Andreessen. An early version of Mosaic was released by NCSA (National Center for Supercomputing Applications) in 1993. In 1994, Andreessen left NCSA to start Mosaic Communications, now called Netscape. Since then, Netscape Navigator has gone through four major revisions and now competes with an equally solid Microsoft browser--Internet Explorer.

## **Understanding How the Web Works**

In theory, hypertext describes a method for organizing information. In practice, however, hypertext describes the ability to link related

documents together using words and phrases. You're quite familiar with how hypertext works, as you've used online help for some time now. Take, for example, a document that contains an overview of several different travel destinations. The name of each destination might be linked to another document that contains more detail. When a user clicks a link in the first document, the browser opens the second, more detailed document.

Hypertext has two interesting characteristics:

- It imposes no order or hierarchy on information--only random relationships. Whereas many methods focus on ordering information or putting information in a hierarchy, hypertext focuses on creating relationships between information. In this way, hypertext supports the organization of information in the "brain-like" manner that Berners-Lee imagined.
- It allows information to have different relationships to other information. When creating an ordered list or outline, you put a single piece of information in a single place within the organization. However, hypertext easily allows each piece of information to occupy many different locations within an organization. You can relate a document about Scotland to other documents about travel, golf, Celtic history, and so on.

The term hypermedia describes what you find on the Web. Hypermedia is a natural extension of hypertext in that the contents of each document include much more than text. They include multimedia, like images, videos, and sounds. Many types of media that you find in a hypermedia document can be linked to other hypermedia documents. In a Web page, for example, you can link images to documents so that when a user clicks the image, the browser opens the document to which it refers.

## **Web Pages (HTML Documents)**

A Web page is a hypermedia document. It is a specific implementation, though, that has its own characteristics.

HTML (Hypertext Markup Language) describes a Web page, which is stored in a plain text file with the HTM or HTML file extension. The primary purpose of HTML is to describe how to format the text content of a Web page. For example, HTML tags specify when text is displayed with

emphasis or when a block of text is formatted as a heading. Some HTML tags also cause the Web browser to display an image or to format content as a table. Most importantly, HTML tags can link a word or phrase to other documents on the Internet. In short, a Web page is a combination of text and the HTML tags that format that text.

HTML is a standard defined by the World Wide Web Consortium (W3C). The latest HTML recommendation (W3C doesn't use the term standard) is 4. (Many vendors such as Microsoft and Netscape define their own extensions to HTML, however). This organization also takes responsibility for defining other standards like HTTP (Hypertext Transport Protocol) and PICS (Platform for Internet Content Selection). Tim Berners-Lee, the inventor of the Web, directs the activities of W3C. You can learn more about this organization by opening <http://www.w3.org> in your Web browser.

**NOTE:** The term Web page refers to documents on the Web that use HTML to describe how they're formatted. Another name for a Web page is HTML document, which is more accurate because HTML is important to more than just the Web. HTML is also used to build online help, format Internet mail messages, and, as is the case with Internet Explorer 4.0, display the contents of a folder.

## **Uniform Resource Locators (URLs)**

You open a resource using its address, or URL. An URL is the standard convention for specifying the location of every resource on the Internet, whether that resource is a document or service. An URL has three parts:

- **Scheme** The scheme describes the protocol that the client should use to access the resource. The protocol is usually followed by `://` except in the case of `file`, which looks like `file:///`, `mailto`, which looks like `mailto:`, and `news`, which looks like `news:`. The table below shows you how to form an URL for each type of protocol.
- **Host** The host is the domain name of the Internet host on which the resource resides. The third-level domain for the host name usually tells you what kind of information you find at the host. For example, you probably find Web pages at a host called `www.host.com` and you probably find an FTP site at a host called `ftp.host.com`.
- **Path** The path is the full path and possibly the file name of a document. Although the protocol and host are always required, the path

is not. If you're opening an http URL, leaving off the path causes the Web server to return the default home page for the site.

Format	Example
<code>gopher://</code>	<code>gopher://gopher.umc.edu</code>
<code>ftp://</code>	<code>ftp://ftp.mcp.com/que</code>
<code>file:///</code>	<code>file:///c win95/readme.bat</code>
<code>http://</code>	<code>http://www.microsoft.com/ie/ie40</code>
<code>mailto:</code>	<code>mailto:john@server.com</code>
<code>wais://</code>	<code>wais://wais.tihs.com</code>
<code>news:</code>	<code>news:alt.fan.enya</code>
<code>telnet://</code>	<code>telnet://onramp.net</code>

**NOTE:** The term *Web page* fits equally well for every HTML document on the Internet. The term *home page*, however, has two different definitions. First, a home page is typically the top-level page on a large Web site. This is the Web page that the browser opens if the user opens a URL without specifying a path or file name. In such a case, the Web server opens a file named **index.htm** or **default.htm** and displays the home page. Second, a home page is also a term used to describe the personal Web pages that individuals put on the Internet.

## HTTP (HyperText Transport Protocol)

Since 1990, the Web has used HTTP, the protocol that defines how clients form requests and how Web servers respond to those requests. In short, HTTP describes how the Web transports Web pages.

HTTP is a plain text protocol, which means that you could read and understand each command sent to the server (try Telnetting to a Web server on port 80). A browser sends a request to the Web server, which replies in turn with a response. The first line of a request is a verb, usually GET or POST, which contains the URL of the resource (file, script, program, and so on. and the version of HTTP that the browser is using. A handful of headers follow the verb and provide additional information required by the Web server. Each header has the same format: name: value. The user-agent header describes the user's browser to the server, for example. HTTP defines a large variety of other headers, too. A complete request might look something like this:

```
GET example.html HTTP/1.0 accept: image/gif, image/jpeg,
image/png, */* user-agent: Mozilla/4.0 (compatible; MSIE
4.0; Windows 95) connection: Keep-Alive
```

The browser replies with a response that begins with the server's HTTP version and a response code. The response codes, which you might see from time to time when a server returns an error. The response then follows with a handful of headers that describe the server, the length of the content, and the content type (MIME type). The actual content follows these headers. A complete HTTP response could look like this:

```
HTTP/1.0 200 OK Server: Netscape-Communications/1.12 Last-
modified: Wednesday, 01-Oct-97 02:55:33 GMT Content-length:
20 Content-type: text/plain This is the content!
```

## Using a Web Browser

The client program that you use to access the Web is called a Web browser. The two most popular browsers are Netscape Navigator and Internet Explorer.

Internally, Web browsers are incredibly complex programs. They support a plethora of standards that govern everything from how to transfer a Web page using HTTP to how to display a document with HTML to how to secure a transaction using SSL. Externally, however, Web browsers are extremely easy to use. To get you up and running, take a look at the capabilities you find in most every Web browser:

- **Browsing URLs** Each Web browser provides the same methods for opening a Web page. You can click a link, type the URL of the Web page in the Address Bar, use the menu, pick a site from your list of favorites, and so on.
- **Toolbar buttons** Most Web browsers provide the same buttons in their toolbars.
- **History Lists** Your browser keeps three different history lists. First, it keeps a list of all the Web pages you've visited during the current session; you can move forward and backward through this list using the Forward and Back buttons. Second, it keeps a history of all the Web pages you've opened by typing their URLs in the Address Bar. You see

this list by clicking the arrow next to the Address Bar. Third, it keeps a separate list of every Web page you've visited.

- **Printing, Saving, and Sending** Your Web browser provides a number of ways to keep a copy of a Web page. You can print it. You can save the page to your disk. You can also send a page or a link to a page to someone else via Internet mail.

- **Bookmarking Your Favorites** The Web contains millions upon millions of Web pages. Your browser makes it easier to go back to your favorite pages by allowing you to put links to those pages in a bookmarks or favorites list.

- **Offline Browsing** A recent innovation is offline browsing. You specify parameters for a Web search and download the content to your computer while you're not using it. Then, you can browse that content while you're not connected to the Internet. Because you don't have to wait for pages to download before your eyes, offline browsing makes the whole Web searching experience seem faster.

- **Channel Support** The latest browsers support channel casting. With channel casting, you subscribe to a channel, and the publisher determines the amount of content and the schedule on which the content is updated on your computer. Then, you can view that content offline. Channel content is typically much better than a normal Web site because most users' expectations are much higher for channel content.

- **Customization** Early browsers allowed you very little customization. These days you can customize everything, though. You can change your start page. You can customize the location, size, and contents of your toolbars. Most browsers also provide an abundance of advanced options you can set.

### ***Typical toolbar buttons found in a browser***

<b>Button</b>	<b>Description</b>
Forward	Moves forward in a list of recently visited pages. Might also drop-down a list of pages.
Backward	Moves backward in a list of recently visited pages. Might also drop-down a list of pages.
Stop	Forces the browser to abort the current action. If you're opening a Web page, the browser stops loading it when you click Stop.
Refresh	Forces the browser to reload the Web page.

Home	Opens the browser's start page, which is the Web page that the browser opens when it first starts. You can configure the start page.
Print	Prints the open Web page. In many cases, the browser lets you choose how to print framed sites.

**TIP:** Caching Improves Performance

Your Web browser supports caching (pronounced "cashing"). The first time that your browser opens a file, it loads the file from the Internet and stores it in a cache. The next time your browser opens the file, it checks to see if the file is out of date. If the file is not out of date, it loads the file from the cache instead of from the Internet. This produces significant time savings.

Caching allows the browser to open some files much faster because it doesn't have to load the files from the Internet every time it needs them.

Your disk would fill up quickly if the browser left every file it ever opened in the cache. Thus, your browser gets rid of the older files, keeping the most recently used files (thus the term *MRU caching*). Your browser gives you a fair amount of control over how much space is allocated to the cache.

**Make Sure You Have the Right Hardware**

Many Web pages take advantage of hardware that you might not have installed in your computer. If you want to get full advantage of multimedia on the Web, consider the following:

- Install additional memory in your computer. The more memory your computer has the better your browser will perform. Windows 95 users should consider 32M the minimum amount of memory to be satisfied.
- Make sure your drive is big enough to have at least 20M available at all times. You can give more space to your browser's cache, which will improve its performance.
- Purchase the fastest connection possible. Don't settle for anything less than a 56K modem, and if you have access to and can afford it, consider using ISDN, cable modems, or digital satellite service.
- Install a sound card and speakers. The brand name isn't important, as long as you can play .WAV and .MID files on the speakers. You can also take advantage of audio-on-demand.

- Install a high-performance video adapter because video is becoming increasingly important in some sites. More and more Web sites are using animation. Still other sites provide video-on-demand. The better your video adapter, the better quality you'll get.

## Searching for Content on the Web

The Web is a complex jumble of documents in which you'll never find the information for which you're looking without help. A variety of search tools scour the Web, cataloging the sites that they find and the keywords in those sites. You then can search these huge databases using keywords that describe what you're looking for, and the search tool returns a list of matching URLs. There are two different types of search tools, and some search tools are actually a combination of both types:

- **Directories** Directories search the Web for content and store the results in a huge database that you search.
- **Indexes** Indexes catalog submissions from users, but they generally don't scour the Internet for content themselves. Most indexes are organized hierarchically.

### *Search Tools*

<b>Name</b>	<b>Type</b>	<b>URL</b>
AltaVista	Directory	<a href="http://www.altavista.digital.com">http://www.altavista.digital.com</a>
Excite	Both	<a href="http://www.excite.com">http://www.excite.com</a>
HotBot	Both	<a href="http://www.hotbot.com">http://www.hotbot.com</a>
Infoseek	Both	<a href="http://www.infoseek.com">http://www.infoseek.com</a>
Lycos	Both	<a href="http://www.lycos.com">http://www.lycos.com</a>
Yahoo!	Index	<a href="http://www.yahoo.com">http://www.yahoo.com</a>
WebCrawler	Both	<a href="http://www.webcrawler.com">http://www.webcrawler.com</a>