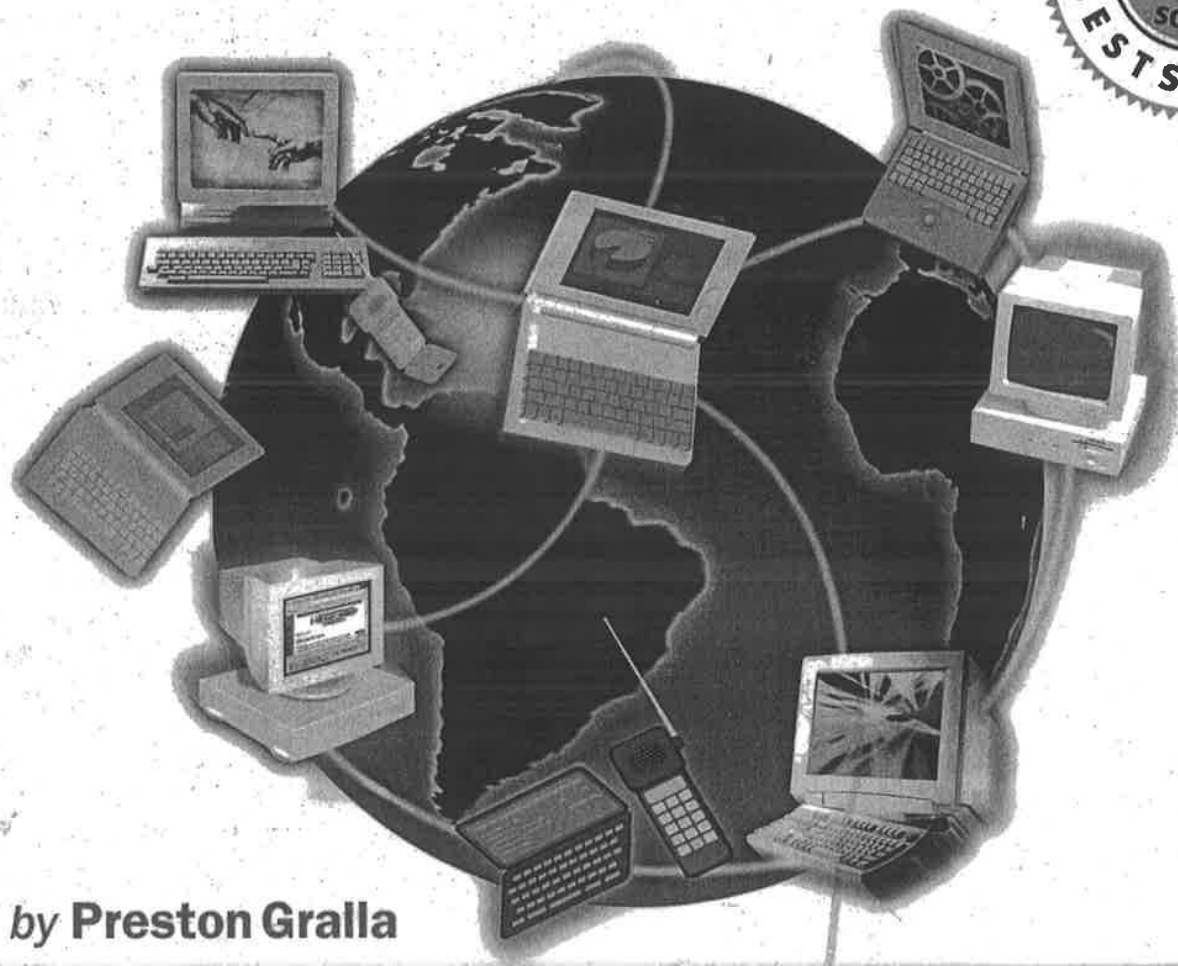


MILLENNIUM EDITION

HOW THE INTERNET WORKS

NATIONAL
BESTSELLER
MORE
THAN
2,000,000
How It Works
SOLD



by Preston Gralla

que

Illustrated by SARAH ISHIDA,
MINA REIMER, & STEPHEN ADAMS

How the Internet Works

Millennium Edition

Preston Gralla

que[®]

A division of Macmillan Computer Publishing, USA
201 W. 103rd Street
Indianapolis, IN 46290

How the Internet Works, Millennium Edition **Copyright® 1999 by Que®**

Executive Editor	Greg Wiegand
Acquisitions Editor	Stephanie J. McComb
Development Editor	Nicholas Goetz
Managing Editor	Thomas F. Hayes
Project Editor	Karen S. Shields
Copy Editor	Kay Hoskin
Indexer	Christine Nelsen
Proofreader	Maribeth Echard
Technical Editor	Bill Bruns
Illustrators	Sarah Ishida, Mina Reimer, Stephen Adams, and Shelley Norris
Book Designers	Carrie English and Bruce Lundquist
Copy Writer	Eric Borgert
Layout Technicians	Lisa England, Cyndi Davis-Hubler

All rights reserved. No part of this book shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher. No patent liability is assumed with respect to the use of the information contained herein. Although every precaution has been taken in the preparation of this book, the publisher and author assume no responsibility for errors or omissions. Neither is any liability assumed for damages resulting from the use of the information contained herein.

International Standard Book Number: 0-7897-2132-5

Library of Congress Catalog Card Number: 99-63011

Printed in the United States of America

First Printing: August 1999

01 00 5

This book was produced digitally by Macmillan Computer Publishing and manufactured using computer-to-plate technology (a filmless process) by GAC, Indianapolis, Indiana.

Trademarks

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Que cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

Warning and Disclaimer

Every effort has been made to make this book as complete and as accurate as possible, but no warranty or fitness is implied. The information provided is on an "as is" basis. The authors and the publisher shall have neither liability nor responsibility to any person or entity with respect to any loss or damages arising from the information contained in this book.

Introduction

Part 1: What Is the Internet?**I**

Chapter 1	
The Wired World of the Internet	4
Chapter 2	
How Information Travels Across the Internet	8
Chapter 3	
How TCP/IP Works	12
Chapter 4	
Understanding Internet Addresses and Domains	16
Chapter 5	
Anatomy of a Web Connection	20
Chapter 6	
Internet File Types	24

Part 2: The Internet's Underlying Architecture **28**

Chapter 7	
How the Domain Name System Works	32
Chapter 8	
How Routers Work	36
Chapter 9	
The Internet's Client/Server Architecture	40

Part 3: Connecting to the Internet **44**

Chapter 10	
Connecting Your Computer	48
Chapter 11	
Connecting to the Internet from Online Services	52
Chapter 12	
How ISDN Works	56

Chapter 13	
The Internet/Television Connection	60

Chapter 14	
How Network Computers (NCs) Work	66

Chapter 15	
How the Digital Subscriber Line (DSL) Works	70

Chapter 16	
How Satellite and Palmtop Connections Work	74

Part 4: Communicating on the Internet **80**

Chapter 17	
How Email Works	84

Chapter 18	
Email Spam and Blocking	98

Chapter 19	
How Usenet Newsgroups Work	104

Chapter 20	
How Internet Chat and Instant Messaging Work	110

Chapter 21	
Making Phone Calls on the Internet	116

Part 5: How the World Wide Web Works **122**

Chapter 22	
How Web Pages Work	126

Chapter 23	
How Web Browsers Work	134

Chapter 24	
How Markup Languages Work	144

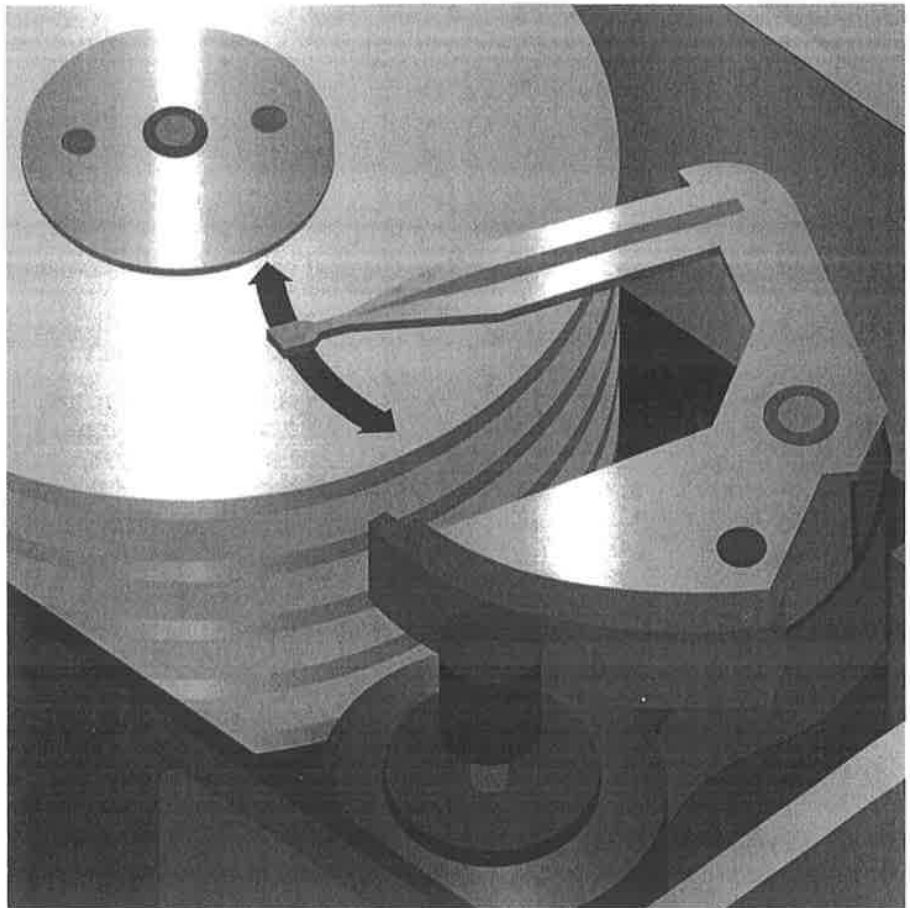
Chapter 25	
How Hypertext Works	150

Chapter 26			
How URLs Work	154		
Chapter 27			
How Imagemaps and Interactive Forms Work	158		
Chapter 28			
How Web Host Servers Work	162		
Chapter 29			
How Web Sites Work with Databases	166		
Part 6: Common Internet Tools	170		
Chapter 30			
How Telnet Works	174		
Chapter 31			
FTP and Downloading Files	178		
Chapter 32			
Searching the Internet	184		
Chapter 33			
How Push Technology Works	190		
Chapter 34			
How Java, JavaScript, and ActiveX Work	198		
Chapter 35			
How Agents Work	206		
Chapter 36			
How CGI Scripting Works	210		
Part 7: Multimedia on the Internet	214		
Chapter 37			
Audio on the Internet	218		
Chapter 38			
Video on the Internet	226		
Chapter 39			
How Multicast IP and the Mbone Work	234		
Chapter 40			
Virtual Reality on the Internet	238		
Chapter 41			
Animation on the Web	242		
Part 8: Intranets and Shopping on the Internet	248		
Chapter 42			
How Intranets Work	252		
Chapter 43			
How Workgroup Software Works	256		
Chapter 44			
Shopping on the Internet	260		
Part 9: Safeguarding the Internet	270		
Chapter 45			
How Firewalls Work	274		
Chapter 46			
How "Smurf Attacks" Can Cripple Internet Service Providers	280		
Chapter 47			
How Viruses Work	284		
Chapter 48			
How Cookies, Passports, and Web Tracking Work	290		
Chapter 49			
Cryptography, Privacy, and Digital Certificates	298		
Chapter 50			
Parental Controls on the Internet	304		
Index	308		

CHAPTER

9

The Internet's Client/Server Architecture



THE Internet works on the client/server model of information delivery. In this model, a client computer connects to a server computer on which information resides; the client depends on the server to deliver information. In effect, the client requests the services of the larger computer. These services may involve searching for information and sending it back to the client, such as when a database on the Web is queried. Other examples of these services are delivering Web pages, and handling incoming and outgoing email. Whenever you use the Internet, you're connected to a server computer and requesting the use of that server's resources.

Typically, the client is a local personal computer and the server (also known as the *host*) is usually a more powerful computer that houses the data. These computers can be of many different kinds: powerful Windows-based PCs, Macintoshes, and a wide variety of hardware that runs the UNIX operating system, such as computers made by the Digital Equipment Corporation.

The connection to the server is made via a LAN (local area network), a phone line, or a TCP/IP-based WAN (wide area network) on the Internet. A primary reason to set up a client/server network is to allow many clients to access the same applications and files that are stored on a server.

In the case of the Internet's World Wide Web, the client is actually the browser on your PC and the server is a host computer located somewhere on the Internet. Typically, the browser sends the server a request for a specific Web page. The server processes that request and sends an answer back to the browser (again, most often in the form of a Web page).

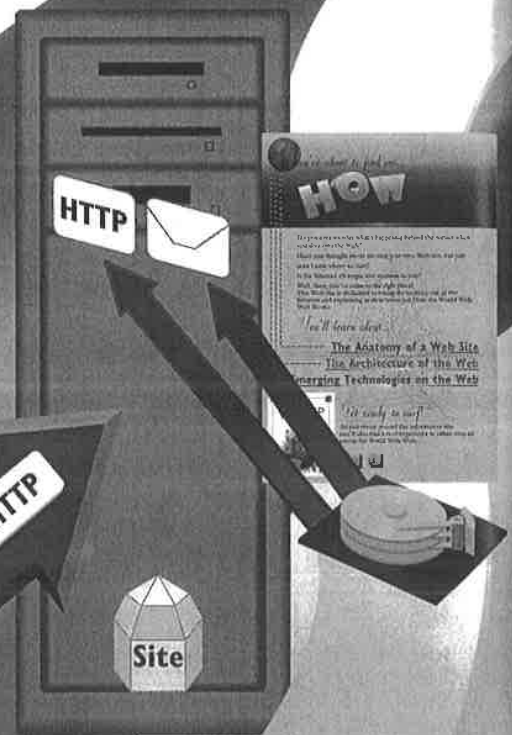
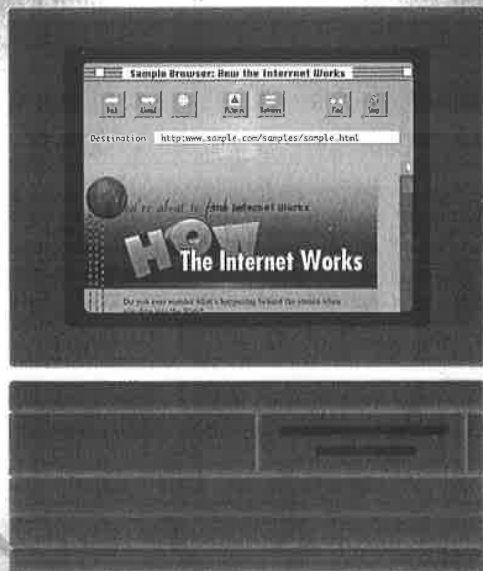
The connection between the client and server is maintained only during the actual exchange of information. Thus, after a Web page is transferred from the host (or server) computer, the HTTP connection between that computer and the client is broken. (HTTP stands for Hypertext Transfer Protocol; it's the protocol used by the World Wide Web.) Even though the HTTP connection is closed, the ISP maintains the TCP/IP connection to the Internet.

The client/server model enables the desktop PC to run the browser software to search the Web, yet still access host servers around the Internet to execute search and retrieval functions. In essence, this architecture enables the Web to be conceived of as a limitless file storage medium and database, distributed among thousands of host computers, all accessible by any individual PC.

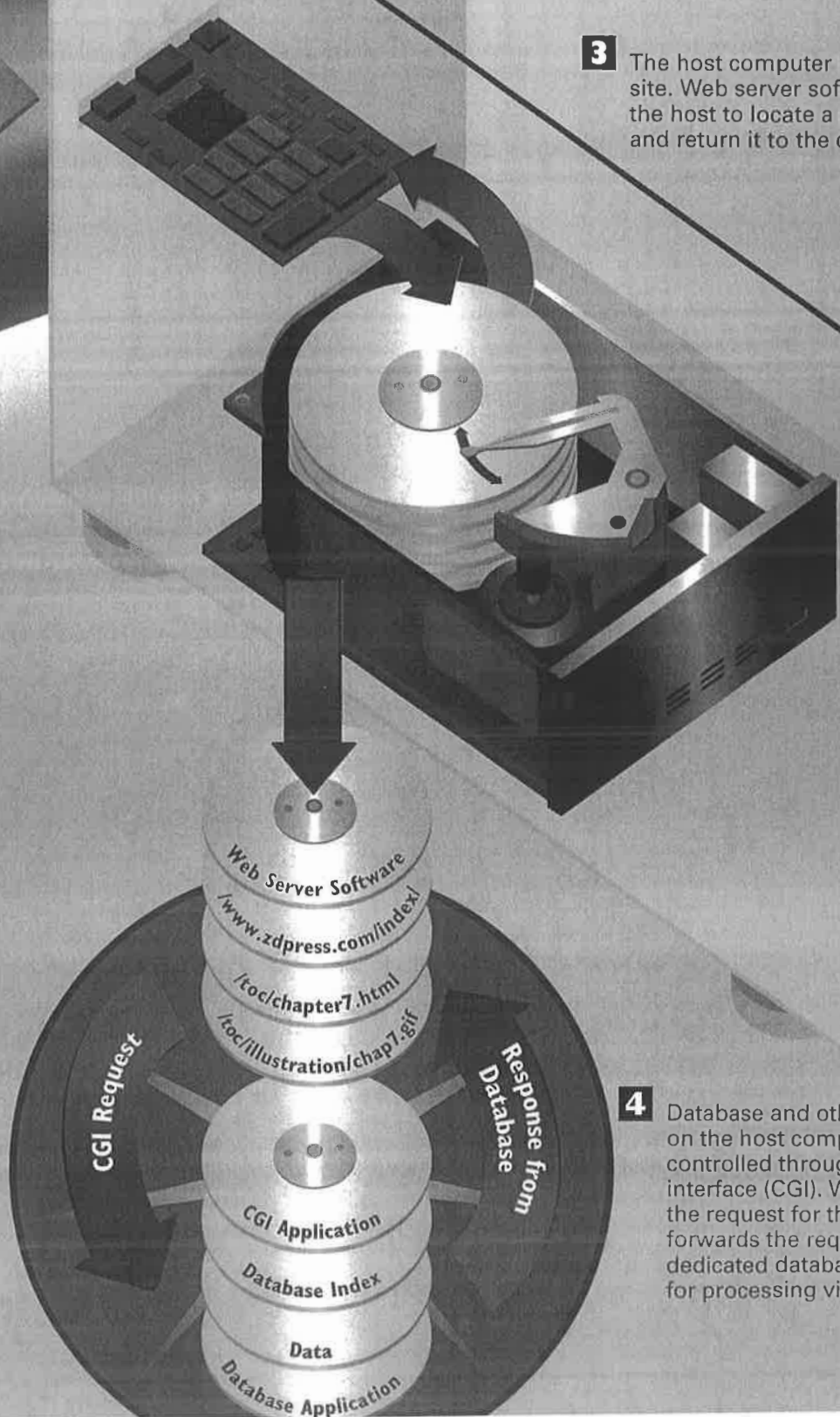
The following illustration shows how the Web runs on a client/server architecture. Keep in mind that all other resources on the Internet run on the client/server model as well. For example, in email transactions the client would be the email software on your computer while the server would be the email server into which you connect.

How Client/Server Architecture Works

1 The PC-based browser software controls the client end of the Web application. Using TCP/IP, the browser issues HTTP requests to the host server. The browser can request a specific Web page or it can ask the host server to perform a database query. In either instance, the request is broken into HTTP packets that are sent across the Internet's TCP/IP communications infrastructure to the host computer.



2 The host computer runs the server software that enables the host to separate the actual request from the packets and perform the asked-for services. This will either involve retrieving and sending back to the client PC the specified Web page, or executing a database search and sending back to the client the result in the form of a Web page.



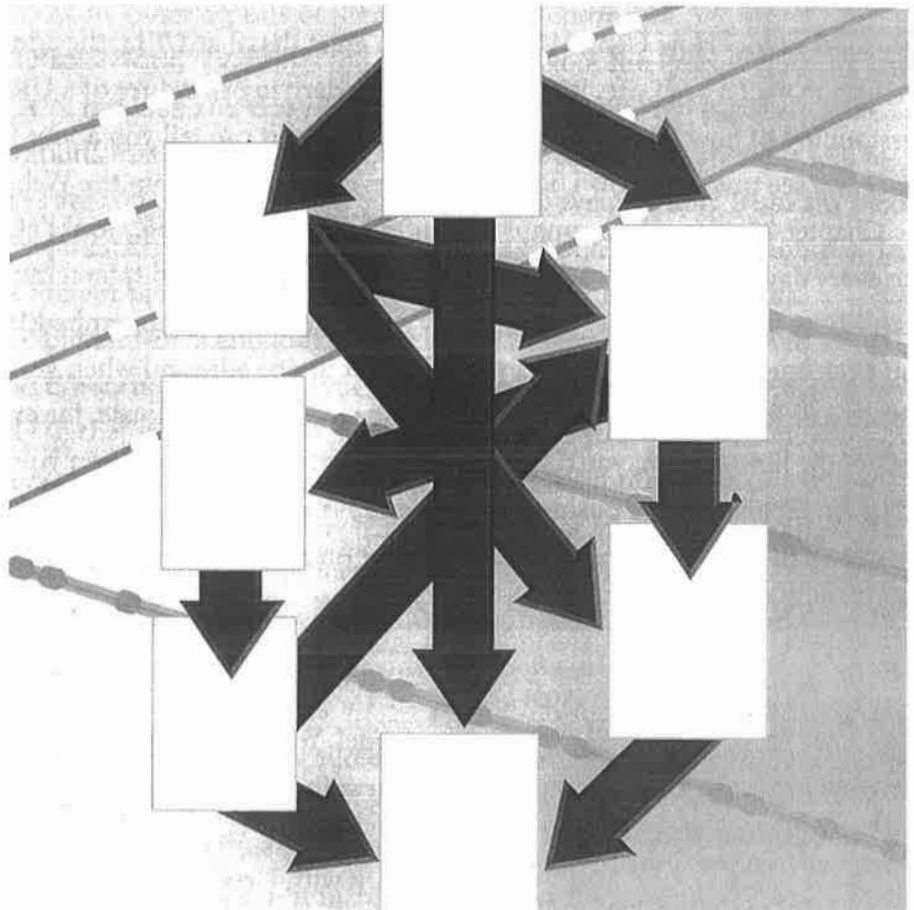
3 The host computer houses the Web site. Web server software enables the host to locate a requested page and return it to the client.

4 Database and other similar applications on the host computer are accessed and controlled through a common gateway interface (CGI). When the server receives the request for the database query, it forwards the request to a separate, dedicated database server or application for processing via the CGI.

CHAPTER

22

How Web Pages Work



THE World Wide Web is the fastest growing, and in many ways, the most exciting and intriguing part of the Internet. When people refer to “surfing the Net,” more often than not they’re talking about using the World Wide Web.

As its name implies, the World Wide Web is a globally connected network. The Web contains many things, but what makes it so fascinating to so many are the Web “pages” that incorporate text, graphics, sound, animation, and other multimedia elements. In essence, each page is an interactive multimedia publication that can include videos and music as well as graphics and text.

Pages are connected to one another using *hypertext* that allows you to move from any page to any other page, and to graphics, binary files, multimedia files, as well as any Internet resource. To jump from one page to another, click on a hypertext link.

The Web operates on a client/server model. You run Web client browser software such as Netscape Navigator or Microsoft’s Internet Explorer on your computer. That client contacts a Web server and requests information or resources. The Web server locates and then sends the information to the Web browser, which displays the results.

Pages on the Web are built using a markup language called HTML (Hypertext Markup Language). The language contains commands that tell your browser how to display text, graphics, and multimedia files. It also contains commands for linking the page to other pages, and to other Internet resources.

The term “home page” is often used to refer to the first, or top, page in a collection of pages that make up a Web site. This is to distinguish it from the many pages collected together as a single “package” of sorts that often make up Web sites. The home page is like a magazine cover or the front page of a newspaper. Usually, it acts as an introduction to the site, explaining its purpose and describing the information found on other pages throughout the site. In this way, the home page often acts as the table of contents for the rest of the site.

In general, Web sites use three kinds of organizational structures to organize their pages. In a *tree structure*, a pyramid or outline format makes it easy for users to navigate through the site and find the information they want. In a *linear structure*, one page leads to the next, which then leads to the next, and so on, in a straight line. Finally, in a *random structure*, pages are connected to one another, seemingly at random.

The last illustration in this chapter shows how someone might build his or her own pages using an HTML editor. After you create your pages, you use FTP software to post them onto a small portion of a Web server. You can either rent the server space from a local ISP or set up your own Web server.

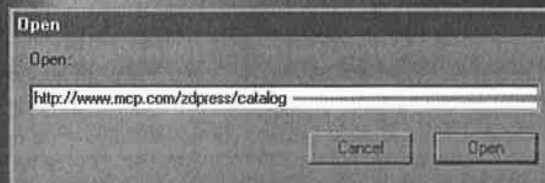
How the World Wide Web Works

1 The World Wide Web is the fastest growing and most innovative part of the Internet. When you browse the Web, you view multimedia pages composed of text, graphics, sound, and video. The Web uses hypertext links that allow you to jump from one place to another on the Web. The language that allows you to use hypertext links and to view Web pages is called Hypertext Markup Language, more commonly known as HTML.

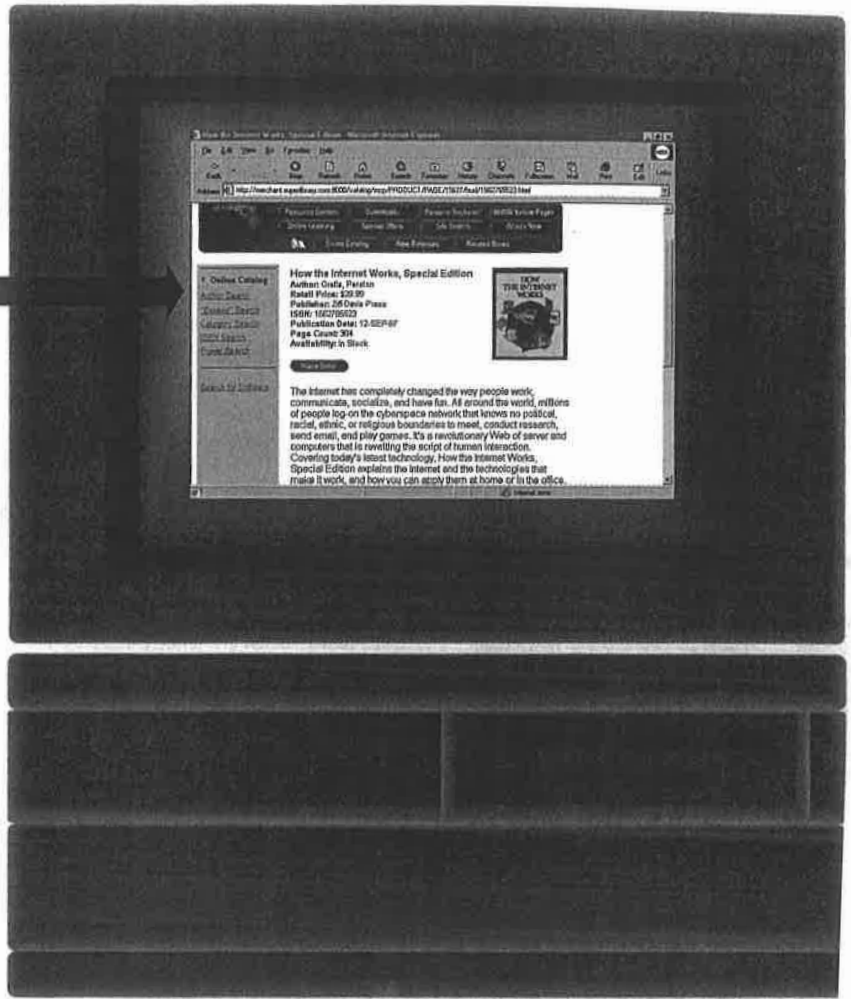
2 The Web works on a client/server model in which client software—known as a Web browser—runs on a local computer. The server software runs on a Web host. To use the Web, you first make an Internet connection, and then launch your Web browser.



3 In a Web browser, you type the URL for a location you want to visit or click on a link that will send you to the desired location. The names for Web locations are URLs (uniform resource locators). Your Web browser sends the URL request using HTTP (Hypertext Transfer Protocol), which defines the way the Web browser and the Web server communicate with each other.



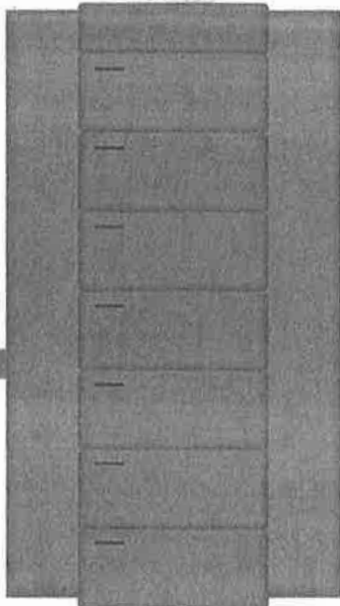
7 When the server finds the requested home page, document, or object, it sends that home page, document, or object back to the Web browser client. The information is then displayed on the computer screen in the Web browser. When the page is sent from the server, the HTTP connection is closed and can be reopened.



6 The Web server receives the request using the HTTP protocol. It is told which specific document is being requested.

<http://www.mcp.com/zdpress/catalog>

<http://www.mcp.com/zdpress/catalog>



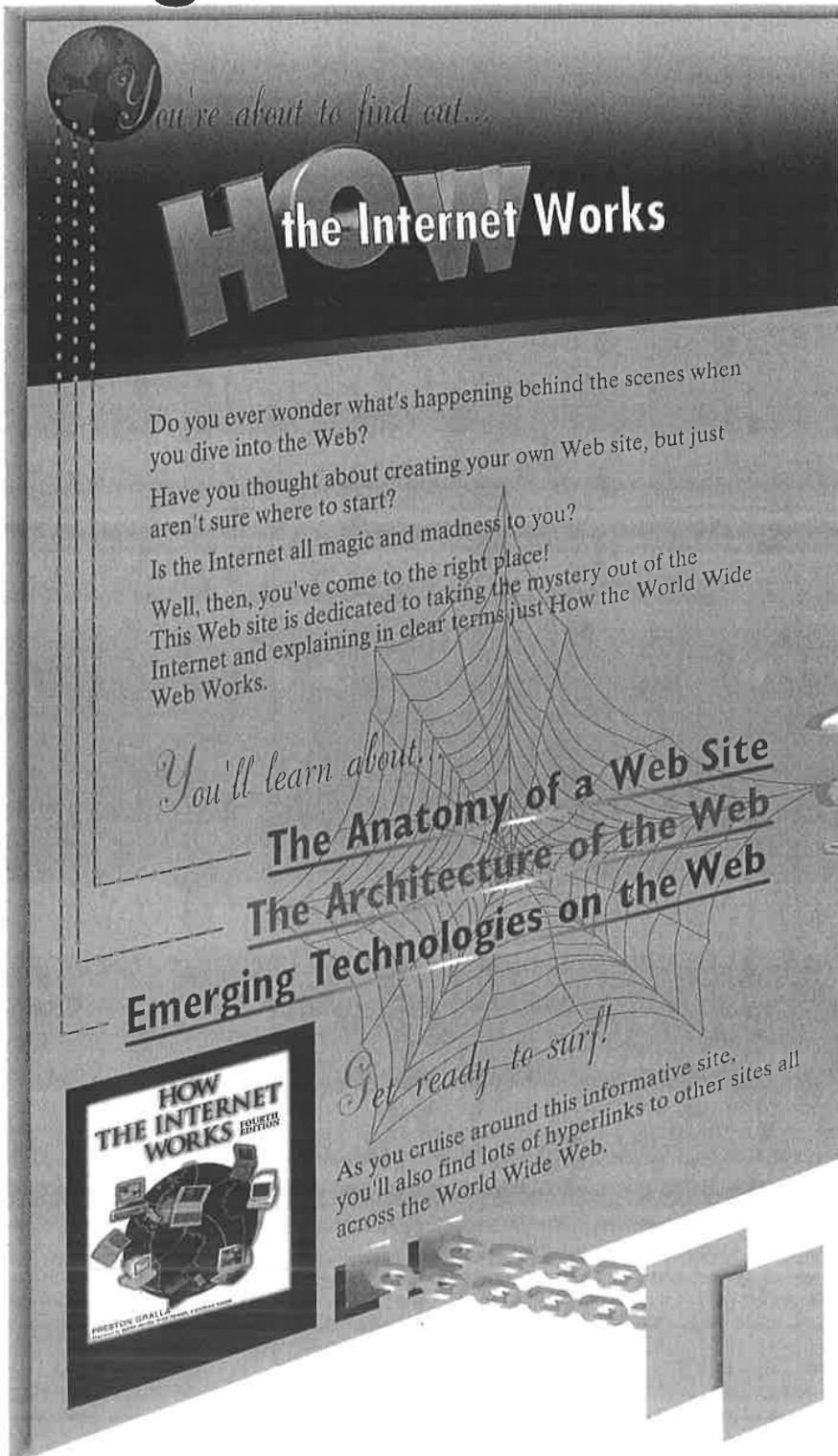
<http://www.mcp.com/zdpress>

5 The request is sent to the Internet. Internet routers examine the request to determine which server to send the request to. The information just to the right of the http:// in the URL tells the Internet on which Web server the requested information can be found. Routers send the request to that Web server.

<http://www.mcp.com/zdpress/catalog>

4 URLs contain several parts. The first part—the http://—details which Internet protocol to use. The second part—the part that usually has a www in it—sometimes tells what kind of Internet resource is being contacted. The third part—such as zdnet.com—can vary in length and identifies the Web server to be contacted. The final part identifies a specific directory on the server and a home page, document, or other Internet object.

How Web Pages Are Organized on a Web Site



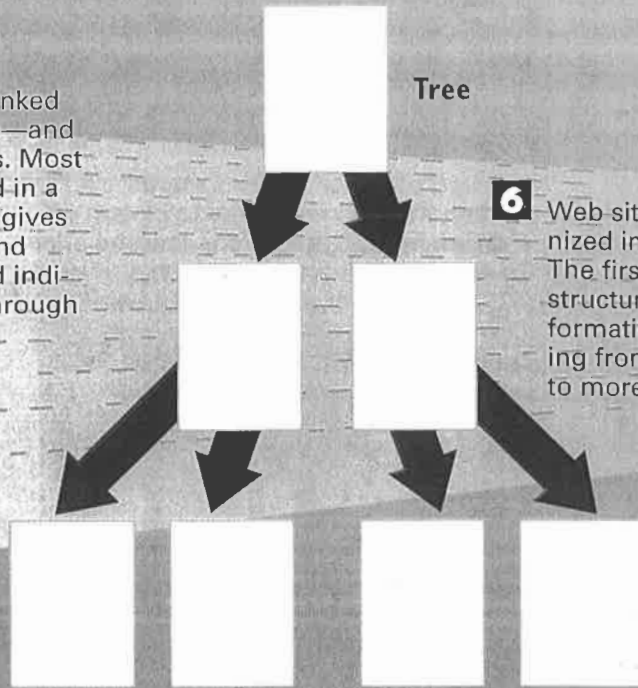
1 The home page is the first or top page of any Web site. A site can be just one page, or can comprise dozens or even hundreds of pages. In the latter case, the home page acts as a table of contents to organize the site and help users find information available on the site.

2 Underlined or highlighted hyperlink text is often embedded in the home page. The hyperlinks serve to connect the top page with other pages throughout the site.

3 Related documents residing together on a Web host computer make up a Web site. However, a single server can host multiple Web sites, each contained in a separate area or directory, much like a hard drive can accommodate multiple directories. Some Web sites are so large and heavily trafficked that they cannot fit on a single server and require multiple servers.

4 Good Web design principles suggest that pages throughout a site link back to the home page. This approach allows users to always find their way back to the top of a site to navigate in other directions.

5 Documents within a site can be linked to any other document in the site—and even to documents on other sites. Most Web sites, however, are designed in a pyramid or outline structure that gives users a visual model to understand how information is arranged, and indicates how to find and navigate through the site's documents.



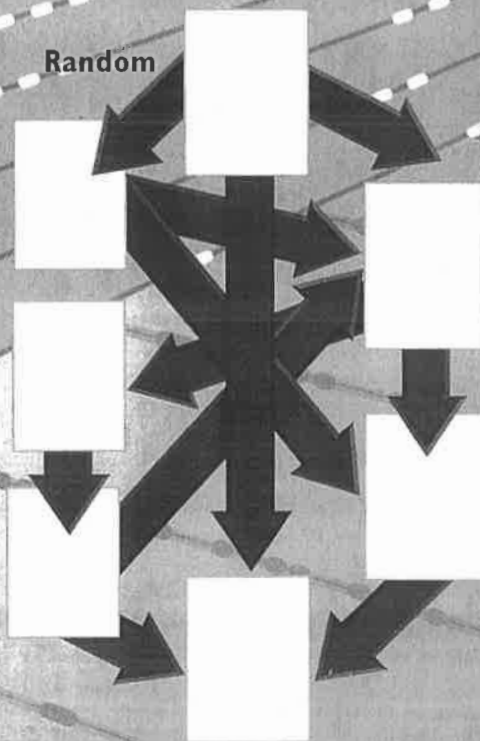
6 Web sites are typically organized in one of three ways. The first is an outline or tree structure, which arranges information hierarchically, moving from general information to more specific data.

Linear



7 The second organizational method is linear, in which one page leads to the next, which leads to the next one, and so on.

Random



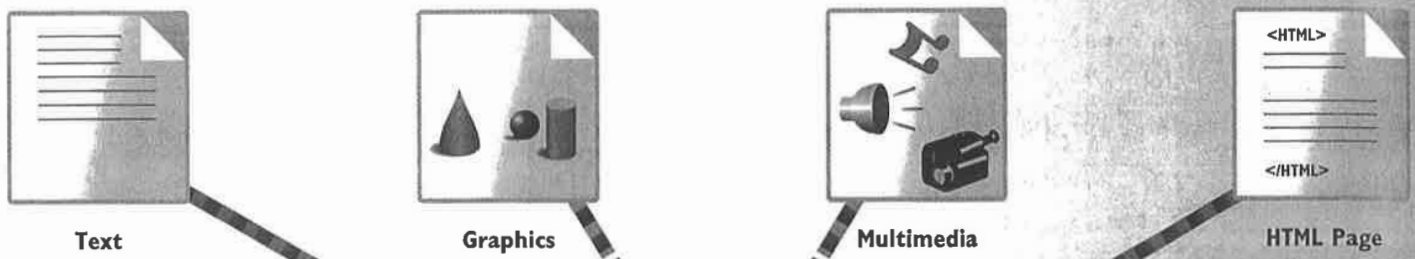
8 The third organizational structure is really a lack of structure, in which pages are connected to one another seemingly at random. (It's this last structure, though, that makes it clear why the Web is called the Web.)

The Anatomy of a Web Site

The Architecture of the Web

Emerging Technologies on the Web

How Web Sites Work



1 First, the raw materials are collected for the Web page. Content may be any number of things—family photos, poems, personal opinions, text of any kind, even sound clips and movies. Text can be easily compiled with any word processor, such as Microsoft Word.

3 Software is available that facilitates the arrangement of text, images, and other media without requiring a single line of HTML to be typed; one such program is Adobe's PageMill. However, if the pages need to be tweaked, there is no substitute for knowing a little HTML. Before posting new pages, they should be previewed from the hard disk using a Web browser. If any changes are necessary, it is easier to make them at this stage.

Text Editor

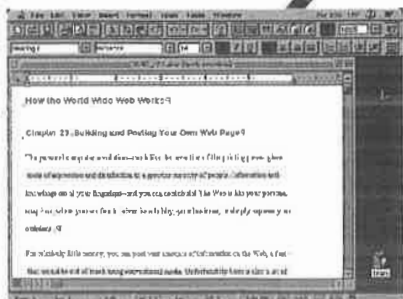
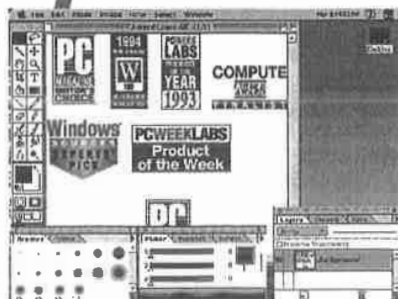


Image Editor



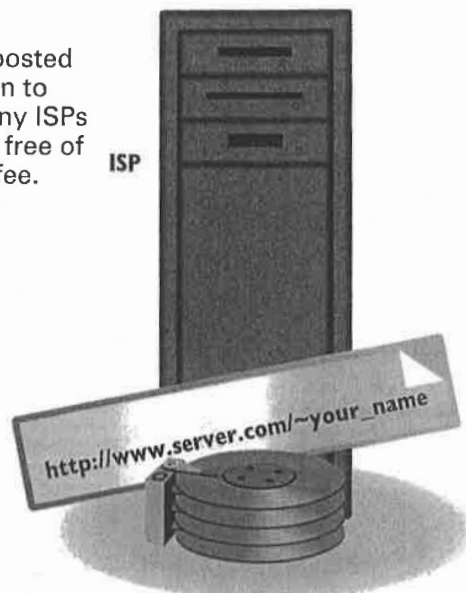
Media Editors



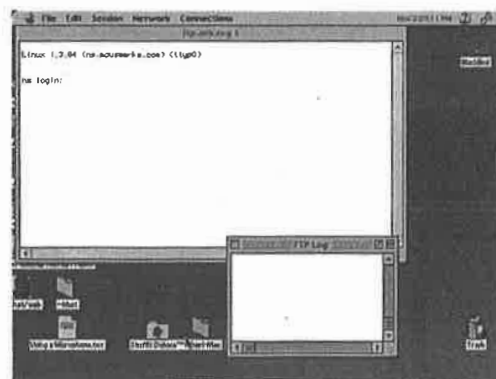
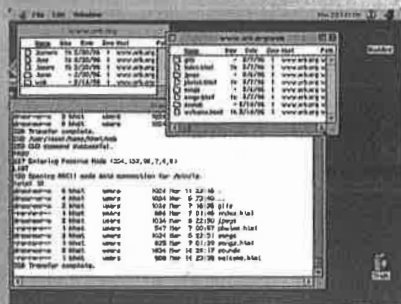


5 The Web page can typically be posted by means of a dial-up connection to the Internet (via a local ISP). Many ISPs provide the space for doing this free of charge as part of their monthly fee.

ISP



6 An FTP program is used to log on to the ISP's Web server; many good ones are available as shareware on the Internet. The ISP issues a username and password at sign-up time, which are used to log on to the directory where the HTML page, graphics, and other media will be placed. After the various elements are put in the right folder, the "live" pages can be checked on the Web using the URL provided by the ISP.



2 A page can come to life with a few well-chosen images. In designing a family home page, for example, a scanned photo might work well on the Web page. Icons or colored bullets spruce up a text list, and navigational icons such as arrows or pointers let a user move logically around the site. A whole range of shareware image editors is available on the Internet. The most full-featured commercial image editor is Photoshop by Adobe. Photoshop is a professional-grade program that allows you to modify, create, or resize virtually any digital image.

HTML Editor

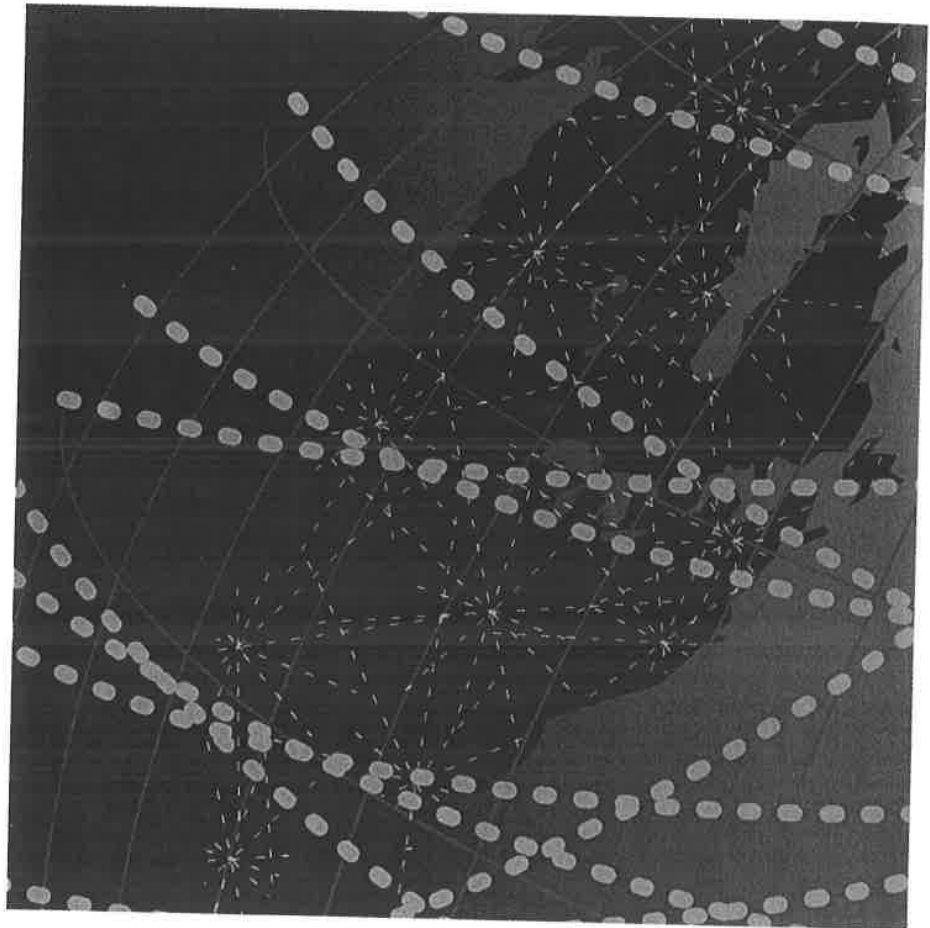


4 If you want to add other media to your pages—such as digitized sound, music, or video—you will need access to the digital files, or you can digitize them yourself with additional computer hardware. When you are dealing with media with large file sizes—such as a 4MB movie—it will take users a very long time to download the file. During editing, try to trim all the fat out of video and audio clips. Also try to scale back the length and resolution of audio and video to decrease file size.

CHAPTER

23

How Web Browsers Work



LIKE much of the Internet, the World Wide Web operates on a client/server model. You run a Web client on your computer—called a Web browser—such as Netscape Communicator, or Microsoft's Internet Explorer. That client contacts a Web server and requests information or resources. The Web server locates and then sends the information to the Web browser, which displays the results.

When Web browsers contact servers, they're asking to view pages built with Hypertext Markup Language (HTML). They interpret those pages and display them on your computer. They also can display applications, programs, animations, and similar material created with programming languages such as Java and ActiveX, and scripting languages such as JavaScript.

Sometimes, home pages contain links to files that the Web browser can't play or display, such as sound or animation files. In that case, you'll need a plug-in or a helper application. You configure your Web browser to use the helper application or plug-in whenever it encounters a sound or animation file that the browser can't run or play.

Over the years, Web browsers have become increasingly sophisticated. Browsers have now become full-blown software suites that can do everything from videoconferencing to letting you create and publish HTML pages. They have also begun to blur the line between your local computer and the Internet—in essence, they can make your computer and the Internet function as a single computer system.

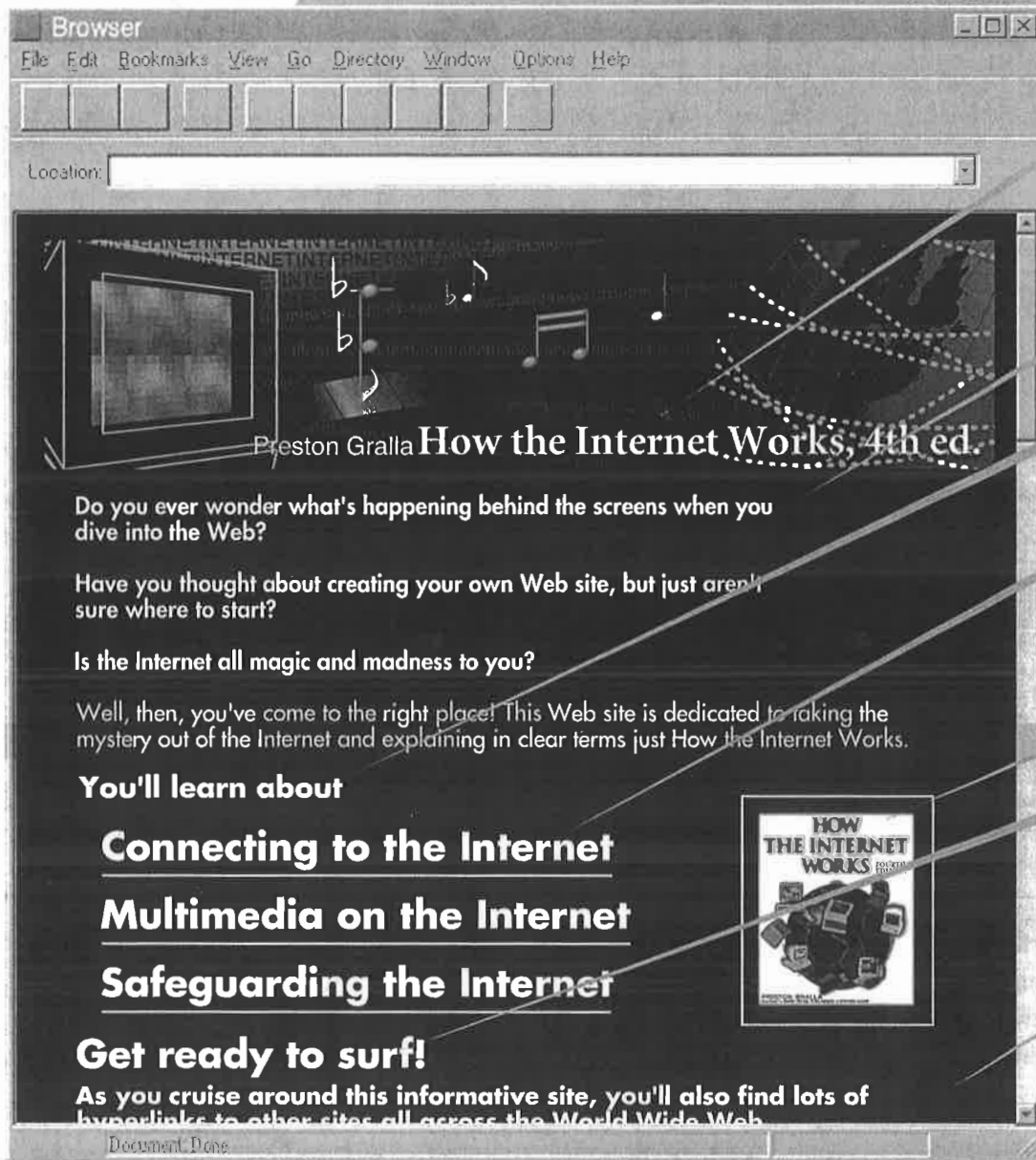
Toward this end of bringing the Internet more directly into computers, Microsoft has integrated Web browsing and the Internet directly into the operating system. For example, with Internet Explorer 4.0 and later, and with Windows 98, the Windows desktop can be HTML-based. This means Web links can be directly embedded into the desktop. So, for example, you can have links to your favorite Web pages right on the desktop. And even applications such as word processors now have Web capabilities built into them, such as being able to browse the Web, or build home pages. Even more significantly, using technology that Microsoft calls Active Desktop, Internet-based *desktop components* can live on the desktop. These components can be things such as stock tickers, which deliver live Web content directly to the desktop. You don't need to go fire up your Web browser to get the information; it's delivered straight to your Windows desktop without your having to do anything.

Both Microsoft and Netscape have also built entire suites of software around their browsers. Netscape, for example, calls its suite Netscape Communicator. Communicator includes modules for reading newsgroups; for reading, sending and managing Internet mail; for audio conferencing; for collaborative work on *whiteboard applications* in which people can view and mark up the same documents simultaneously; and more. These enhancements will help usher in an era of collaborative computing. Not only will the Internet be used to transmit and receive information, but it will also alter the way we do business and help us communicate more effectively.

How a Web Browser Works

1 Web browsers consist of client software that runs on your computer and displays home pages on the Web. There are clients for PC, Macintosh, and UNIX computers.

2 A Web browser displays information on your computer by interpreting the Hypertext Markup Language (HTML) that is used to build home pages on the Web. Home pages usually display graphics, sound, and multimedia files, as well as links to other pages, files that can be downloaded, and other Internet resources.



<HTML>
 <HEAD>
 <TITLE>
 </HEAD>

3 The coding in the HTML files tells your browser how to display the text, graphics, links, and multimedia files on the home page. The HTML file that your browser loads to display the home page doesn't actually have the graphics, sound, multimedia files, and other resources on it. Instead, it contains HTML references to those graphics and files. Your browser uses those references to find the files on the server and then display them on the home page.

<BODY>

<P> Do you ever wonder what's happening behind the screens when you dive into the Web?

<H3> You'll learn about:

 Go TO URL

4 The Web browser also interprets HTML tags as links to other Web sites, or to other Web resources, such as graphics, multimedia files, newsgroups, or files to download. Depending on the link, it will perform different actions. For example, if the HTML code specifies the link as another home page, the browser will retrieve the URL specified in the HTML file when the user clicks on the underlined link on the page. If the HTML code specifies a file to be downloaded, the browser will download the file to your computer.

<P> <IMG SRC="BOOK.GIF"

<H2> Get ready to surf!

<P> As you cruise around this informative site, you'll also find lots of hyperlinks to other sites all across the World Wide Web.

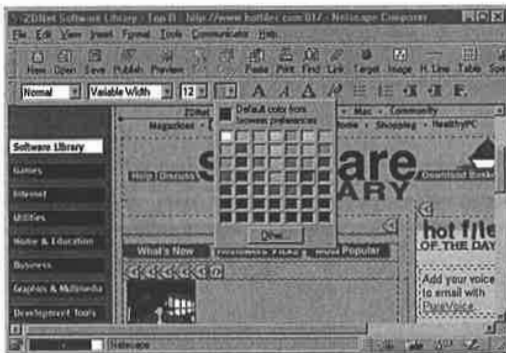
</BODY>
 <HTML>

NOTE There are many kinds of files on the Internet that Web browsers cannot display. In particular, this pertains to many kinds of multimedia files such as sound, video, and animation files. Still, there are often references to these kinds of files on Web pages. To view or play these files, you'll need what are called "helper applications" and plug-ins. You must configure your Web browser to launch these helper applications and plug-ins whenever you click on an object that needs them in order to be viewed. Helper applications and plug-ins can also be used for displaying virtual reality pages, for chatting on the Internet, and for doing other Internet tasks.

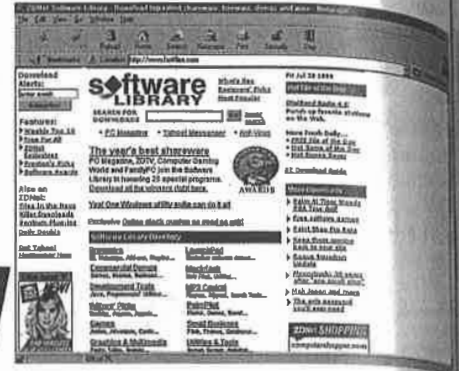
NOTE The meanings of tags are easily decipherable. Every HTML tag, or instruction, is surrounded by a less-than and a greater-than sign—<P>. Often tags appear in pairs, the beginning tag and the ending tag. They are identical except for a simple slash in the end tag. So a paragraph of text will frequently be surrounded by tags like this: <P> Paragraph of text.</P>. Also, tags are not case sensitive. <P> equals <p>.

How Netscape Communicator Works

1 Netscape Communicator is a complete suite of Internet applications, not merely a Web browser. Any individual component can be run alone, or it can be run in concert with other components.



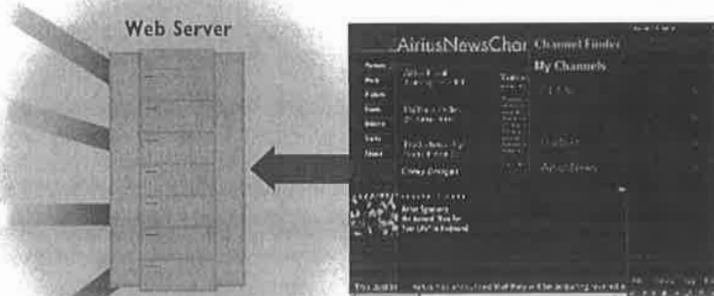
2 The centerpiece of Communicator is Netscape Navigator, the Web browser component.



7 Communicator also includes an HTML editor that allows you to edit, create, and post HTML to the Web.



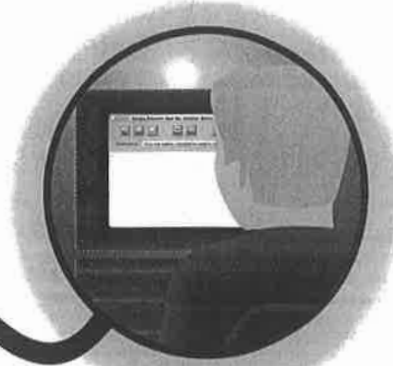
6 The Netcaster component of Communicator allows for push technology to send information to you over the Internet, and for "subscriptions" to Web sites. This saves time and allows information to be read offline while you're not connected to the Internet.





3 Netscape Conference is the workgroup and multimedia communications component of Communicator, used primarily on intranets. It allows people to hold group conference calls from their computers, as well as view and use shared files and documents together. This is called a *whiteboard* application.

Conference

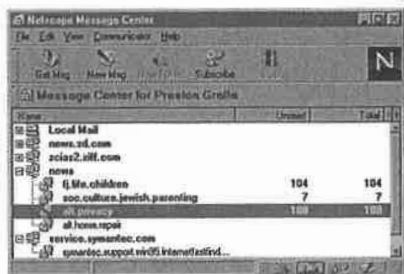


4 Communicator includes an email component called Netscape Messenger. Unlike some other email packages, Messenger can read HTML documents so you can receive entire Web pages as well as messages that include graphics and multimedia content.

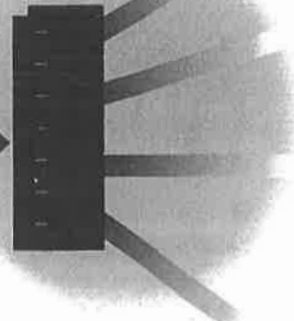
Messenger



5 Two other Communicator components allow for reading newsgroups and online discussions on the Internet. A Usenet newsgroup reader gives access to Usenet, whereas software called Collabra allows for intranet-based discussions.

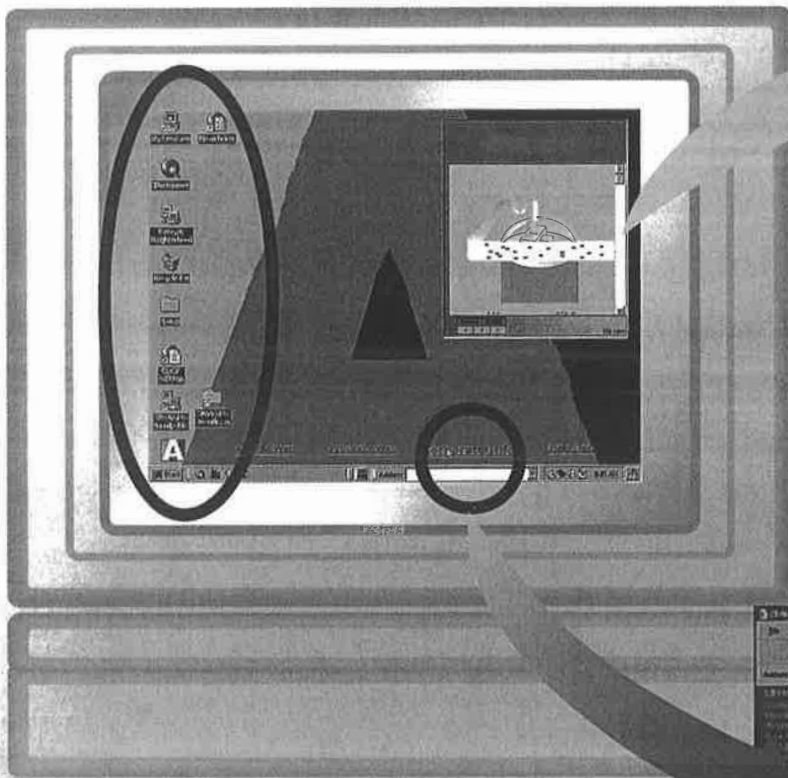


Usenet



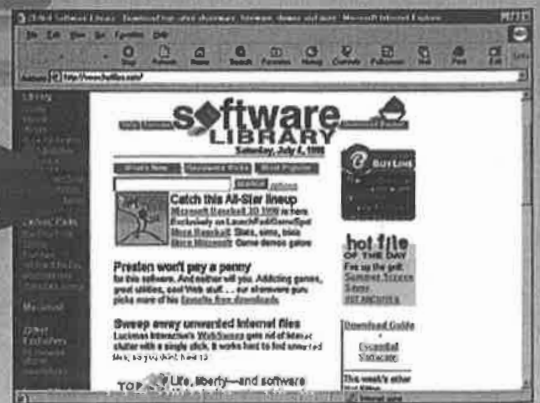
How Internet Explorer Integrates into Windows

1 Internet Explorer 4.0 and later—and Windows 98—blur the distinction between your computer and the Internet. This change has a great impact on Web browsing. The Active Desktop is primarily responsible for this enhancement.

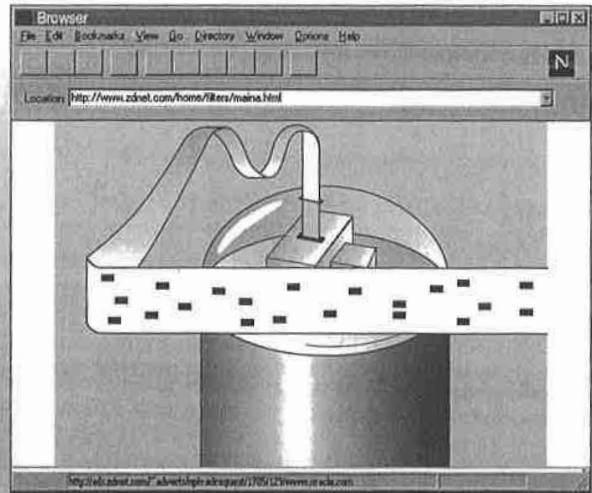


2 The Active Desktop is built with several layers. One layer, the background, allows your normal Windows 95 background to function as an HTML page instead of wallpaper or a color. So, for example, Web links could be placed on the background of your desktop. When you click on the link, it launches Internet Explorer and sends you to the Web site.

3 On top of the background HTML layer is the icon layer. The icons are the same shortcuts that exist in Windows. When icons are clicked upon, they launch a program or a file, open a folder, or do anything else that icons normally do in Windows.

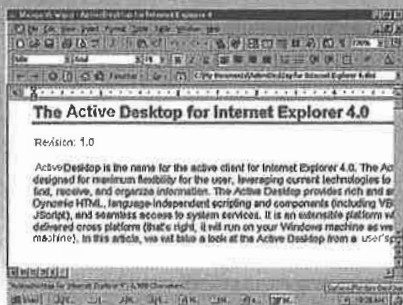
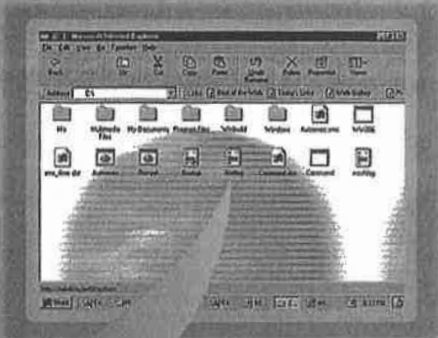


3 Desktop components can be placed on this background HTML layer. Desktop components are HTML frames that contain HTML-based content, sites, and applications. For example, a desktop component might be a stock ticker that displays constantly changing stock prices.



4 A desktop component, such as a stock ticker, can actually be a Web page or component that exists on the Internet. The Active Desktop contacts the Web server hosting the page, downloads it to your local PC, and then displays it within the desktop component frame, resizing the component or page to fit into the frame.

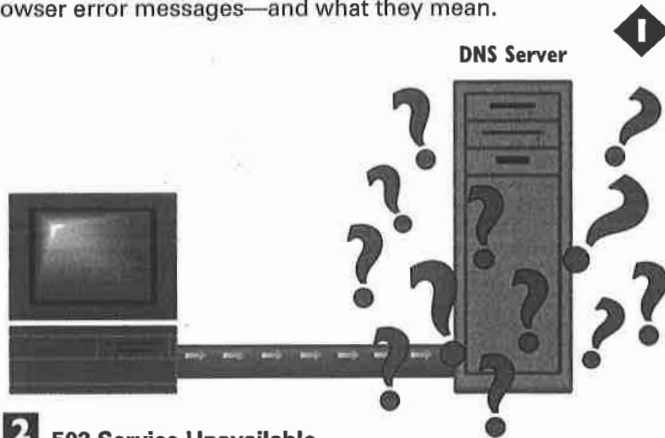
6 Another way Internet Explorer 4.0 and later and Windows 98 integrate your local PC with the Internet is by allowing you to browse the contents of your computer while you're using your browser on the Internet. When you type the drive letter of your computer (such as C:), you will be able to see the folder structure of your computer from within Internet Explorer.



7 You can then launch any document by double-clicking it, just as you would normally do in Windows. You'll read the document in your normal application, not within Internet Explorer.

What Common Browser Error Messages Mean

One of the most frustrating experiences in browsing the Web is the error messages that browsers display when they're having trouble contacting a Web site. Depending on what browser you use, and what version of the browser you're using, those messages might be different. Version 5.0 and later of Internet Explorer, for example, sometimes displays error messages in plain English. But most browsers don't. Here's a list of the most common browser error messages—and what they mean.

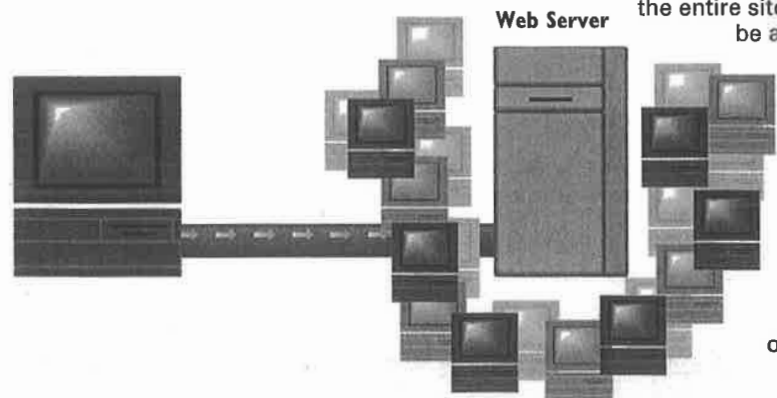


1 Server Does Not Have a DNS Entry
 When you type a URL in your browser to visit a site, your computer contacts a server called a DNS (Domain Name System) server. The DNS server translates the URL into the IP number that computers can understand—and after it does that, your browser can go to the site. (For more information about DNS servers, see Chapter 7, "How the Domain Name System Works.") If you get an error message telling you the server doesn't have a DNS entry, it means that the server doesn't have a listing for the URL you typed. This usually either means that you typed the URL incorrectly, or there is something wrong with the DNS server. Check the URL and retype it.

2 503 Service Unavailable
 This is a catch-all error message for a variety of problems, but all of them mean that the Web site hasn't been able to be contacted by your browser. The problem might be that the site's server has crashed because of too much traffic, or that there's network congestion.



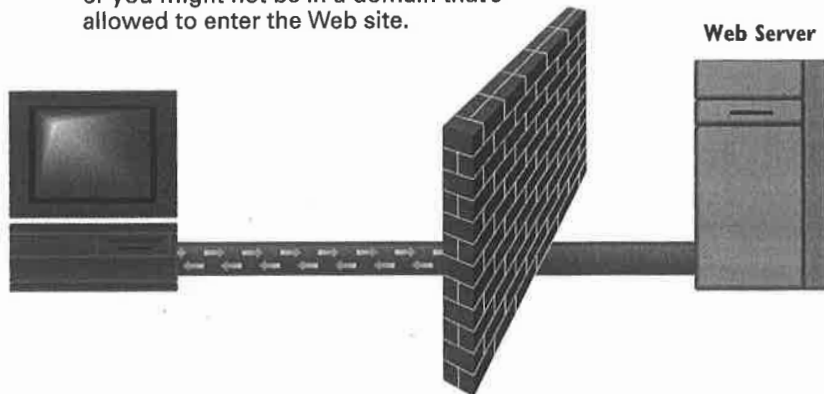
3 403.9 Access Forbidden. Too Many Users Are Connected.
 Some Web sites recognize that if they get too much traffic at once, the entire site can be brought down, and no one will be able to visit. Those sites put a limit on the number of people who can come to the site at once—in that way, the site will always be available, even if everyone who wants to visit can't get in. If you get this "Too Many Users Are Connected" message, it usually means that the Web site is up and running, but you can't get in because the maximum number of people are already on the site. Keep trying—when one person leaves, another can come in, and it might be you.



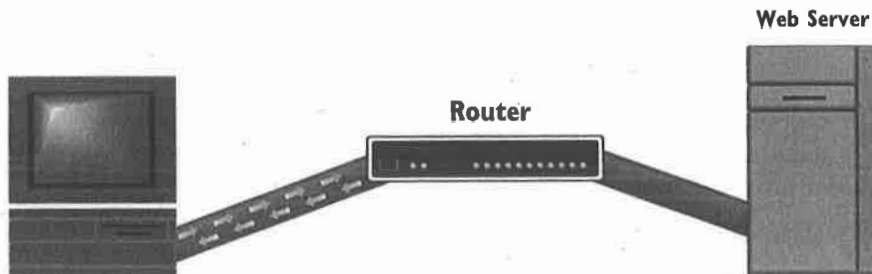


4 404 Not Found
 When you get this message, it means that you've arrived at the right Web site, but the specific page you're looking for can't be found. That specific page might have been deleted from the site or moved—or you might have typed the location incorrectly.

5 401 Unauthorized and 403 Forbidden
 If you get either of these error messages, it means that you're trying to enter a Web site that allows only certain people in—and it's not allowing you to enter. Typically, these kinds of Web sites are password protected, and might also allow people to visit who are only from certain domains, such as zd.com. If you've entered a password, you might have entered it improperly, or you might not be in a domain that's allowed to enter the Web site.



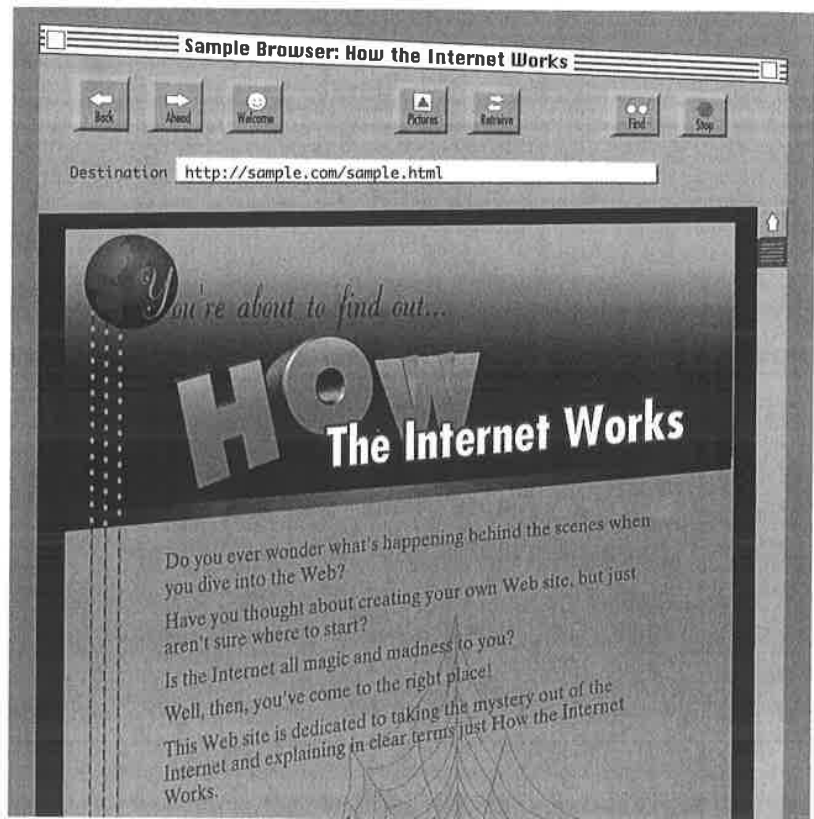
6 Spinning Hourglass
 This isn't an error message that your browser displays—instead, your Windows cursor turns into a spinning hourglass. The spinning hourglass tells you that your browser is trying to make a connection to a Web site. If it keeps spinning and a connection is never made, it can mean that an Internet router someplace between where you are and where you're trying to visit has crashed and you can't make the connection.



CHAPTER

24

How Markup Languages Work



MARKUP languages are the road signs of a Web page. They are sets of directions that tell the browser software how to display and manage a Web document, much like written music scores are instructions that tell a musician how to play a particular song. These instructions (called *tags* or *markups*) are embedded in the source document that creates the Web page.

Tags reference graphic images located in separate files, and they instruct the browser to retrieve and display these images within the page. Tags can also tell a browser to connect a user to another file or URL when he or she clicks an active hyperlink. So each Web page has everything it needs to be displayed on any computer with a browser that can interpret the markup language.

Your original text will probably have headings, multiple paragraphs, and some simple formatting. A Web browser will not understand all these layout instructions because the original text isn't formatted with HTML, the language of the Web (discussed later in this chapter). Paragraphs, carriage returns, indents, and multiple spaces will be shown instead as a single space if no HTML markup is added.

Markup languages should not be confused with programming languages, such as C+ or Pascal. Programming languages are used to write complex applications, such as word processors or spreadsheets. Markup languages, in contrast, are much simpler and describe the way that information should be displayed—for example, by defining when text should be boldface. In markup languages, tags are embedded within documents to describe how the documents should be formatted and displayed.

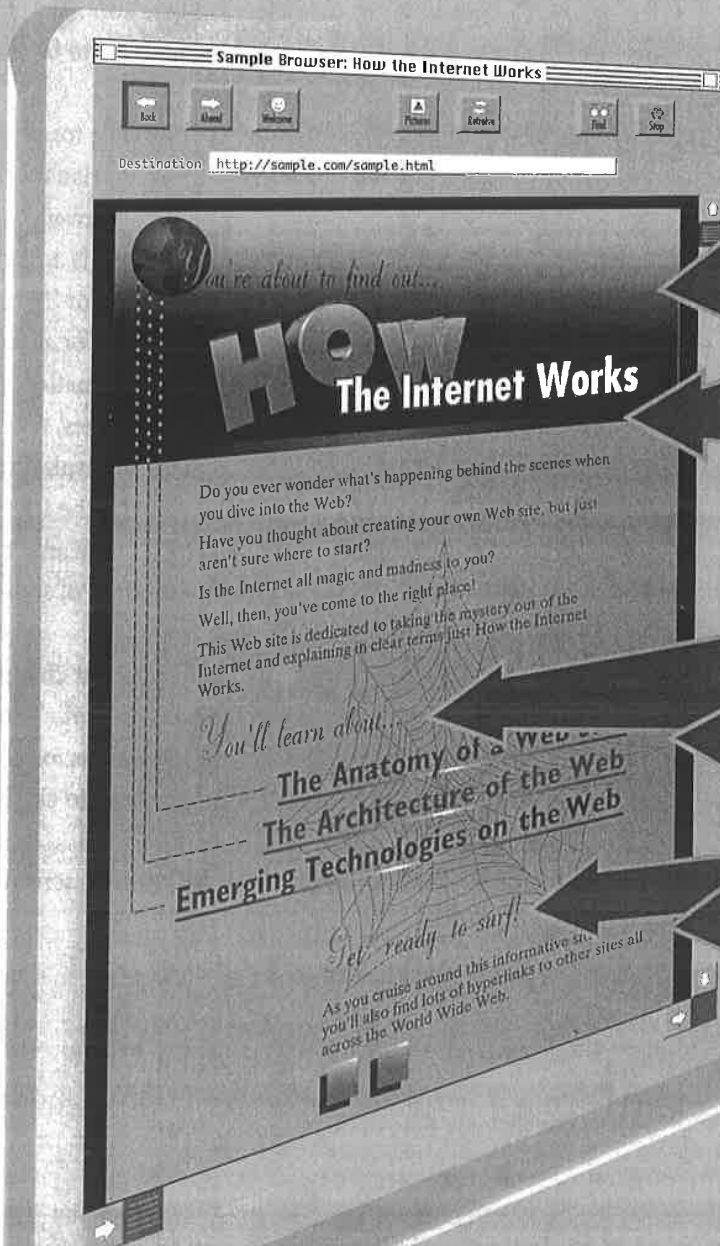
Hypertext Markup Language (HTML) is the markup language of the Web. It defines the format of a Web document and enables hypertext links to be embedded in the document. You can use any text editor or word processor to add HTML tags to an ASCII text document, although a number of shareware and commercially available HTML editors can assist Web page authors as well.

The Web evolves daily, and HTML also expands and changes along with it. The newest changes to HTML are a group of technologies, which together are termed *Dynamic HTML* or DHTML. These technologies allow HTML to be more than a static language, and they enable HTML to perform animations and become more interactive and flexible. It might take a while for this technology to catch on, but when it does, it should change the way everyone uses the Web.

Perhaps the best way to get a feel for HTML is to look at text as it is displayed on your screen and text that is "marked up" using HTML. Turn the page for an example of this.

How HTML Works

I To display Web pages in any browser, you need to add HTML tags to your original text. This process is called *tagging*.



2 Use HTML to give your text structure. All HTML files begin and end with the HTML tags. Headings are marked as such, as are paragraphs, line breaks, block quotes, and special character emphasis. Any carriage returns or indentations within the source text do not affect the browser's display of the page. HTML tags need to be put in if they are to be displayed in a browser.

3 The finished HTML document will be the source page for any browser on any computer. This simplicity of HTML makes cross-platform compatibility easy and reliable. The more complex and specialized the HTML tagging, the longer it will take to download and display the document.

Display text "How the Internet Works"

Display text "Do you ever wonder..."

Display text "You'll learn about. . ."

Create a link to this site

Display text "Get ready to surf!"

Display text "As you cruise..."

```

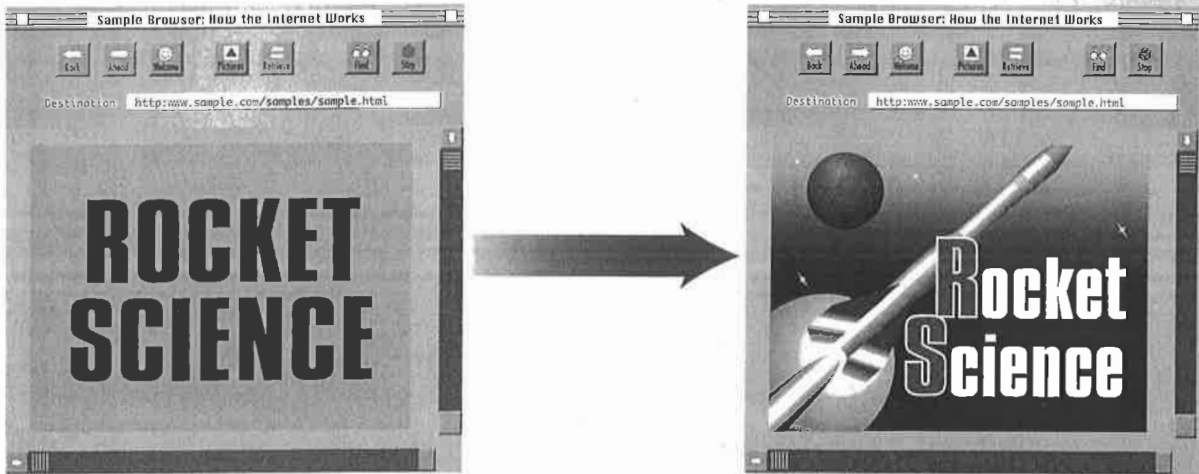
<HTML>
<HEAD>
<TITLE>Sample Browser: How The Internet Works</TITLE>
</HEAD>
-----
<BODY background= "Spider.GIF">
<IMG SRC = "top.gif">
<P>
<BLOCKQUOTE>
Do you ever wonder what's happening behind the scenes when you dive into the Web?
<P>
Have you thought about creating your own Web site, but just aren't sure where to
start?
<P>
Is the Internet all magic and madness to you?
<P>
Well, then, you've come to the right place!
<P>
This web site is dedicated to taking the mystery out of the Internet and explaining in
clear terms just How the Internet Works.
-----
<H1> You'll learn about...</H1>
<BLOCKQUOTE>
<A HREF = "anatomy.html">The Anatomy of a Web Site</A>
<P>
<A HREF = "architecture.html">The Architecture of the Web</A>
<P>
<A HREF = "emerging.html">Emerging Technologies on the Web</A>
</BLOCKQUOTE>
</BLOCKQUOTE>
-----
<IMG SRC = "footer.gif" ALIGN=LEFT>
-----
<BLOCKQUOTE>
<H2> Get ready to surf!</H2>
-----
As you cruise around this informative site, you'll also find lots of hyperlinks to other
sites all across the World Wide Web.
</BLOCKQUOTE>
-----
</BODY>
</HTML>

```

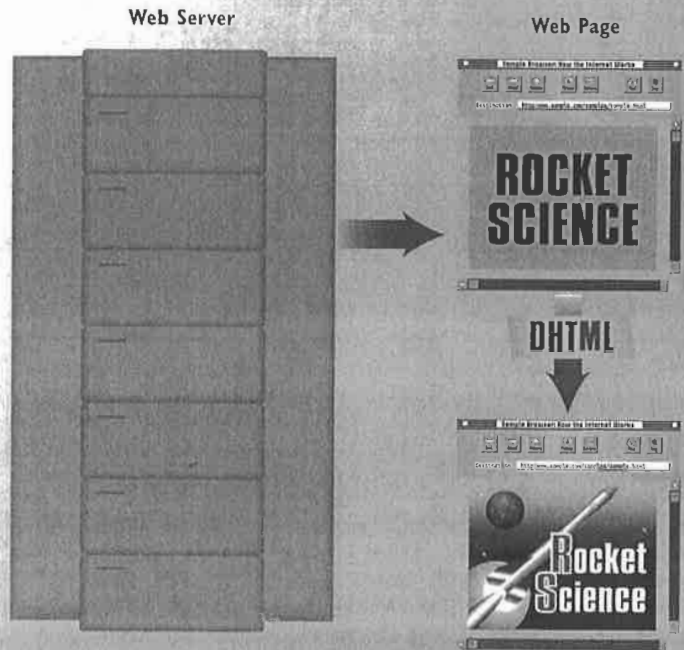
4 Most Web browsers will enable your document to retain its structural integrity when you display, or *parse*, it. Headings will appear in a larger font size than text within paragraphs, for example, and block quotes will be uniformly indented. However, the look might vary from browser to browser. Note that browsers determine the exact font, size, and color. Also be aware that the relative importance of the elements is always kept intact.

How Dynamic HTML Works

1 Dynamic HTML (DHTML) differs from traditional HTML in that it enables Web pages to be changed on-the-fly, after they've been downloaded. In plain HTML, after a page is downloaded, it is static and can be changed only when a user takes an action of some kind. But DHTML, for example, could cause an animation of a rocket to fly across your browser window several seconds after the page has been downloaded—without your doing anything.



2 DHTML does its work without having to contact the server after the page downloads, which means that it can perform some interactive functions more quickly than other technologies that have to contact the server. The instructions for performing the commands are in the HTML commands that are in the page itself.



3 Although DHTML is often referred to as if it were a single technology, it is, in fact, a general term used for a group of technologies that can work together or by themselves to change a Web page after the page has been downloaded to your computer. These technologies are the Document Object Model (DOM), Cascading Style Sheets (CSS), and client-side scripting languages, such as JavaScript.

Elements of DHTML



4 The *DOM* defines every object and element on a Web page and enables those objects to be manipulated or accessed. This includes fonts, graphics, tables, and visual elements, as well as elements you can't necessarily see, such as the browser's version number and the current date and time. Without *DOM*, all the elements on a page are static. So on the simplest level, DHTML could use the *DOM* to change the font of every letter, individually, on a Web page.



5 *Cascading Style Sheets* are, in essence, templates that apply formatting and style information to the elements of a Web page. They're called cascading because any single page can have more than one style sheet associated with it. Additionally, *Cascading Style Sheets* enable images to overlap one another. This enables animations to be created easily on a page.

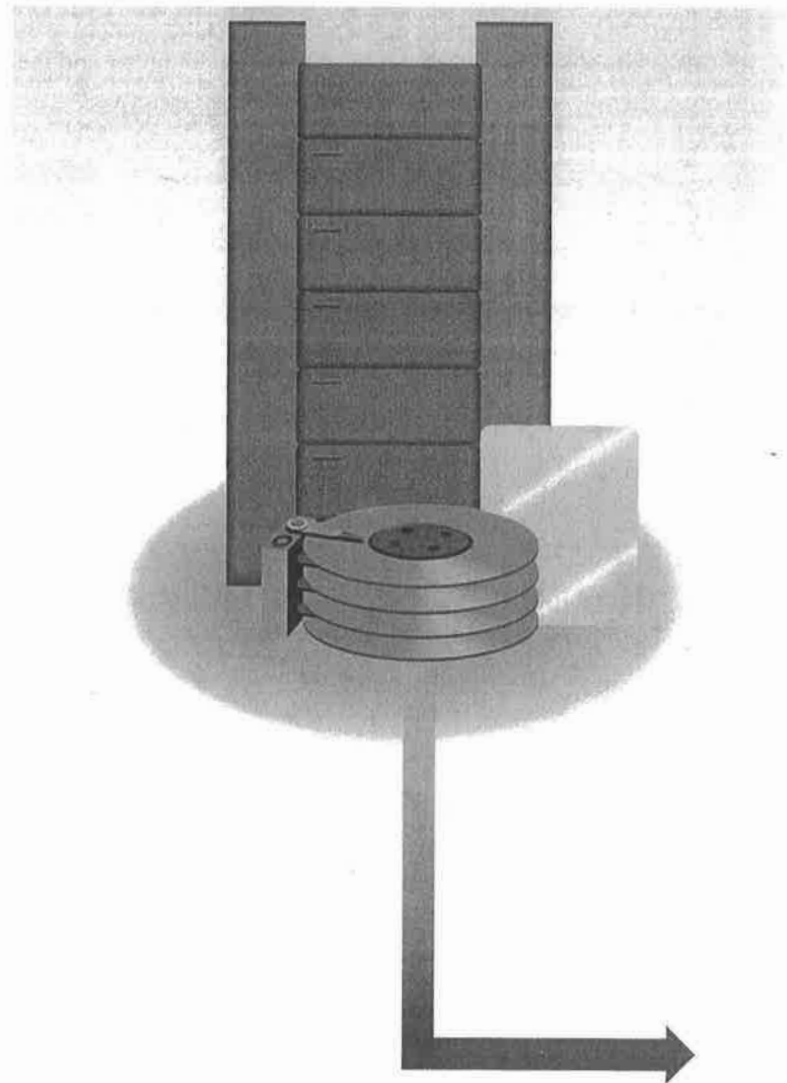


6 *Client-side scripting languages* perform much of the work of DHTML. These languages access the *DOM* and manipulate its elements, and they do the same to *Cascading Style Sheets*. They perform the actions of DHTML. So a script, for example, could turn a word a different color when a mouse moves across it, or it could create easy-to-use collapsible navigation on every page on a Web site.

CHAPTER

25

How Hypertext Works



IN the late 1960s, a computer scientist named Ted Nelson introduced hypertext, a concept that lays the foundation for the World Wide Web and its connections between documents, or pages.

Nelson wanted to create a new way of exploring information. He wanted to provide the reader with a spontaneous means of accessing more and more in-depth information about something that sparked the reader's interest when reading text on the page. Rather than reading a document from beginning to end, digesting the material in a sequential order, the reader could highlight a word and receive more information on the meaning of that word, for example.

Nelson imagined that you could read the United States Constitution, come across the term "electoral college," and then open another document that explained how the electoral college works. From that document, you might open another document that listed the votes of the electoral college from its founding to the present. From there, you might choose to open a document about William Jefferson Clinton, then another about the First Cat, Socks, then another about the care and feeding of felines, and so forth. Ultimately, you could find a subject that wasn't even remotely connected to the Constitution, but which you would find interesting or entertaining.

This hypertext concept was obviously on Tim Berners-Lee's mind when he began thinking about how researchers could share their work across the Internet. He envisioned a system where a document could be linked to other documents, enabling researchers to easily find more and related information simply by following a link from one document on the network to another.

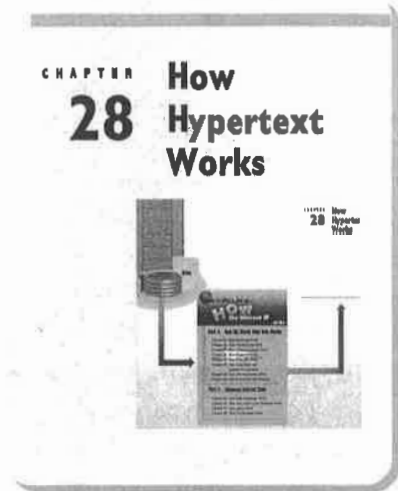
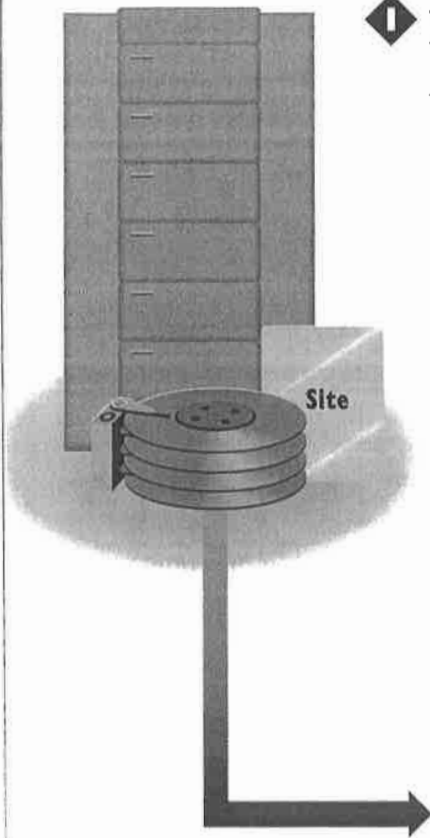
Typically, hypertext consists of a hyperlink that appears onscreen as a highlighted word, icon, or graphic. By moving a mouse cursor over the item, or object, and clicking on it, you easily navigate to additional information. On the Web, that information can be located at any other place on the Web, be it on the same host server or one across the globe. A linked object can be various media, such as text (linking from one character to a whole document, for example), a graphical button (such as direction arrows that move from page to page), or still images (photos, icons, or a comic strip), for example. The documents and objects that are being linked to can be on the same site as the original document, or in an entirely different document.

Hypertext links are embedded into a Web document using HTML, Hypertext Markup Language. A text link usually appears on the screen as an underlined word or phrase and is sometimes rendered in a different color from other text, depending on how your Web browser interprets the HTML codes. When you place the mouse cursor on this underlined text and click the mouse button, you initiate a request by the browser for a new Web page or—if the text references an internal link to information in the same document—direct your browser to scroll to another, specific point within the same document.

Images or icons can also act as hyperlinks. When you move the mouse cursor over the icon or graphic and click the mouse button, you launch the request to retrieve the linked information.

How Hyperlinks Work

I The "hyperlinking" begins when you first retrieve a Web page from a remote Web server. Target links within the page move you quickly from one part of the page to the next.



You're about to find out...
HOW
the Internet Works

Part 6: How the World Wide Web Works

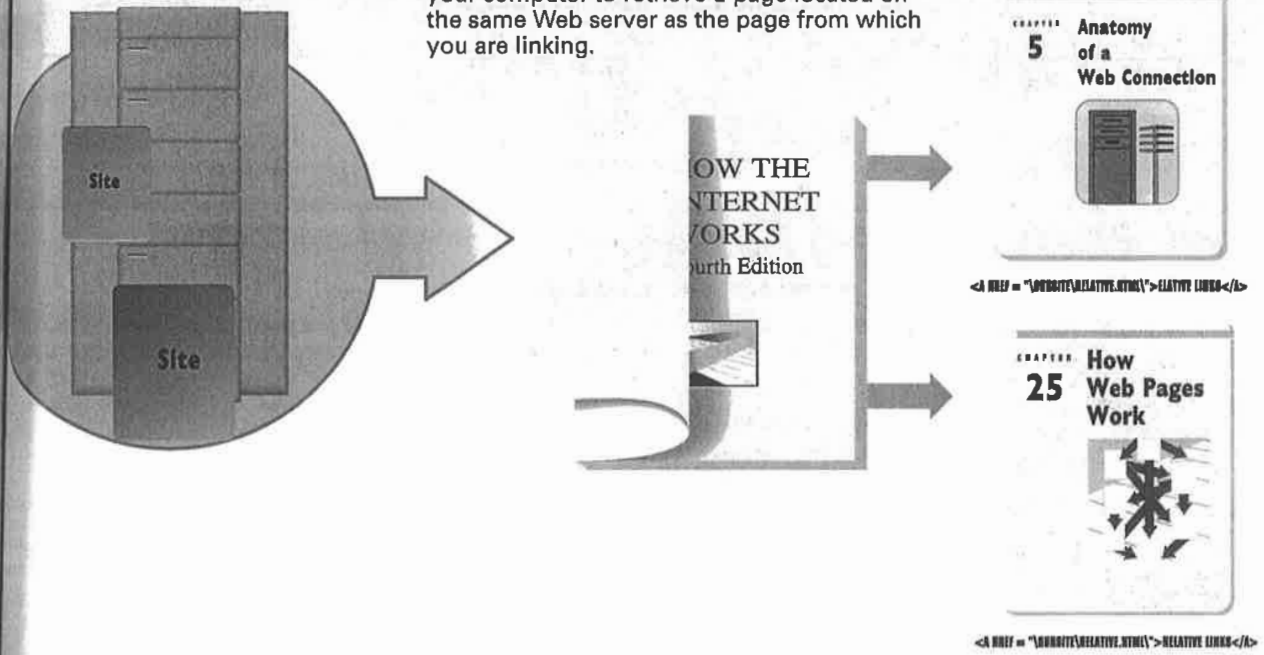
- Chapter 25: How Web Pages Work
- Chapter 26: How Web Browsers Work
- Chapter 27: How Markup Languages Work
- Chapter 28: How Hypertext Works**
- Chapter 29: How Web URLs Work
- Chapter 30: How Image Maps and Interactive Forms Work
- Chapter 31: How Web Host Servers Work
- Chapter 32: How Web Sites Work with Databases

Part 7: Advanced Internet Tools

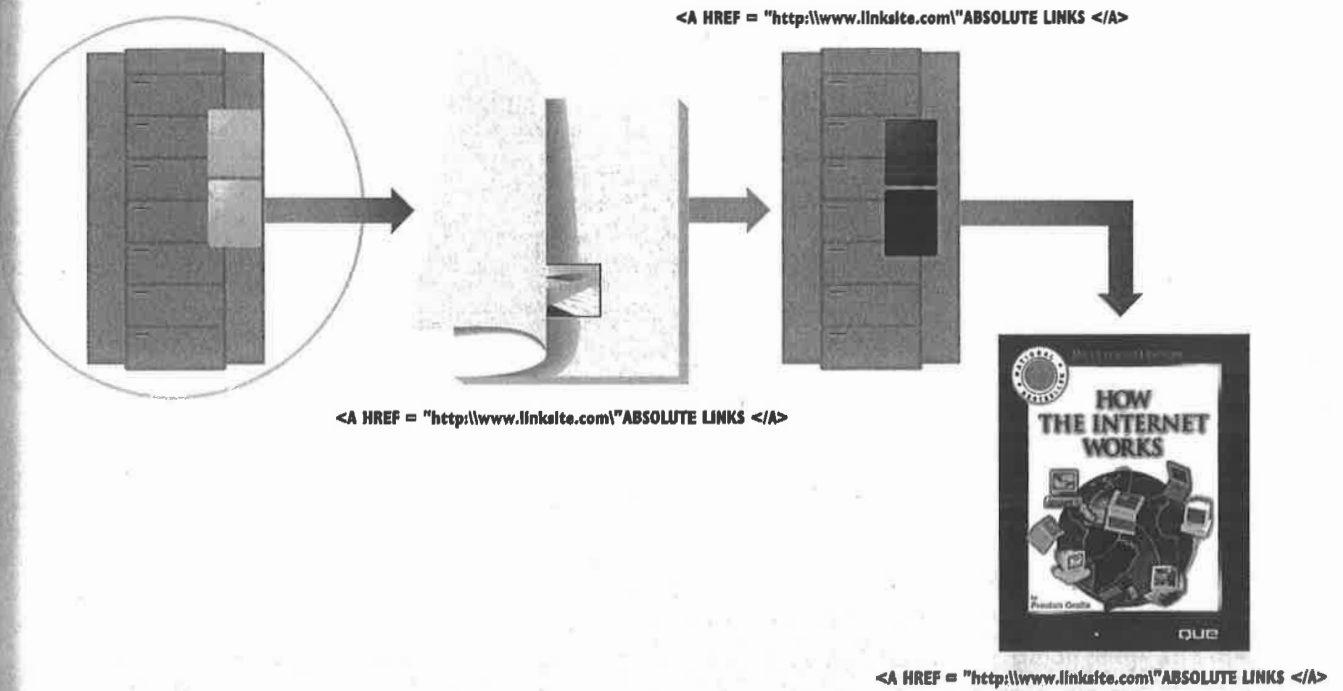
- Chapter 33: How Push Technology Works
- Chapter 34: How Java, ActiveX, and JavaScript Work
- Chapter 35: How Agents Work
- Chapter 36: How CGI Scripting Works

```
<A HREF = "#TARGET">TARGET LINKS</A>
```

2 A relative link is used to initiate a request by your computer to retrieve a page located on the same Web server as the page from which you are linking.



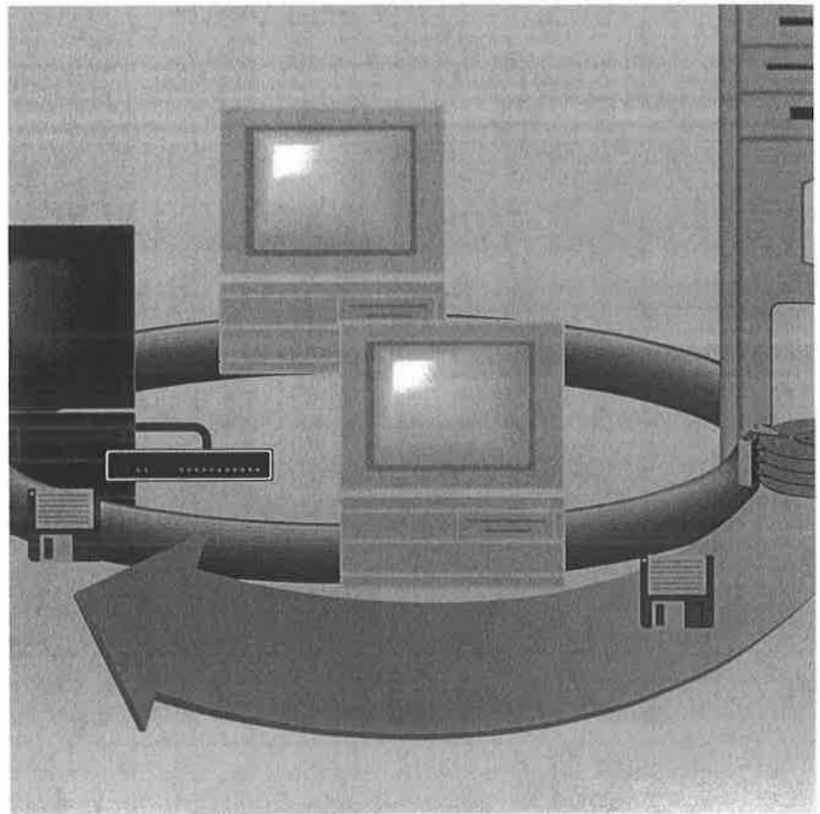
3 A hyperlink that leads to a completely different Web server uses an absolute link.



CHAPTER

26

How URLs Work



THE Web pages and the hosts that make up the World Wide Web must have unique locations so that your computer can locate and retrieve the pages. The unique identifier for a host is called the IP (Internet Protocol) address and the unique identifier for a page is called the URL, or uniform resource locator. A URL functions much like a postal or email address. Just as postal and email addresses list a name and specific location, a URL, or Web address, indicates where the host computer is located, the location of the Web site on the host, and the name of the Web page and the file type of each document, among other information.

A typical URL looks like this:

```
http://www.zdpress.com/internetworks/index.html/
```

If you were to interpret the instructions in this URL from left to right, it would translate to: “Go to the host computer called zdpress (a commercial business), in a directory called internetworks, and retrieve a hypertext document with the filename index.html.” The URL, or address, tells the browser what document to fetch and exactly where to find it on a specific remote host computer somewhere on the Internet.

The first part of the URL indicates what type of transfer protocol will be used to retrieve the specified document. The most common request is for a hypertext document that uses HTTP (Hypertext Transfer Protocol).

The second portion of the URL refers to the specific host computer on which the document resides, which is to be contacted by the browser software. This part of the address is also called the *domain name*. See Chapter 4, “Understanding Internet Addresses and Domains,” and Chapter 6, “Internet File Types,” for more information about domains.

The third part of the URL is the directory on the host computer that contains a specific Web site or multiple Web sites. This is always located after the first single slash in the URL and is essentially the subdirectory on the hard disk that houses the Web site. Subdirectories might also be indicated in this part of the address. For example, if the above URL were changed to

```
http://www.zdpress.com/internetworks/partone/chapters/chapte.html
```

there would be two subdirectories—part one and chapters.

In the preceding example, the filename is chapte.html. This is always the last portion of the URL. If you see an address without a filename, it is assumed that the filename index.html contains the requested Web page. The default document a Web server will deliver to the client when no other filename is listed is index.html.

The illustration in this chapter shows the process necessary to request and retrieve a Web document. When a request for a document occurs for the first time in a Web-browsing session, the host computer must first be located in order to find the file. After that, the specific subdirectory and document are retrieved.

The Structure of a Web URL

1 The first part of the URL indicates what type of transfer protocol will be used to retrieve the specified document. The most common request is for a hypertext document that uses the HTTP protocol.

3 The third part of the URL is the directory on the host computer that contains a specific Web site. A host computer can house multiple Web sites. This third segment of the address is essentially the root directory that houses the site. Subdirectories might also be indicated in this part of the address.

Destination

`http://www.sample.com/samples/sample.html`

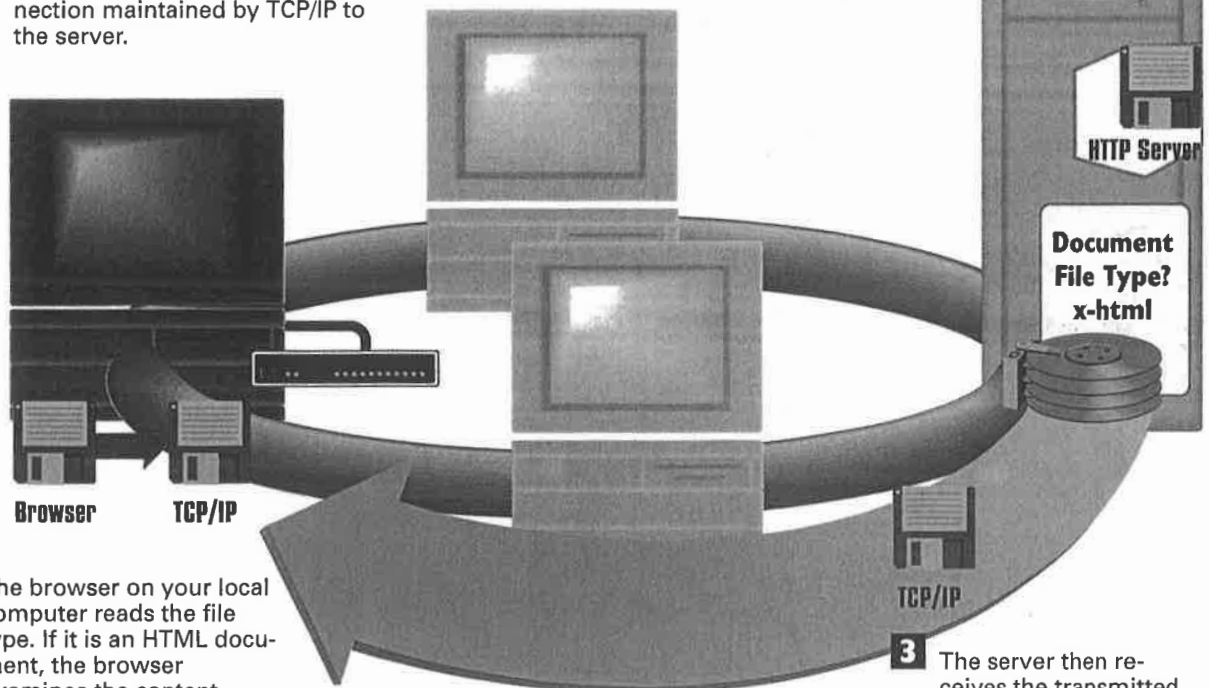
2 The second portion of the URL is the specific host computer on which the document resides, which is to be contacted by the browser software. This part of the address is also called the domain. Domain names end in a suffix that indicates what kind of organization the domain is. For example, .com indicates a commercial business, .edu indicates a college or university, .gov indicates a government office, .mil a military facility, and .org a not-for-profit organization. The suffix can also indicate the country in which the host computer is located. For example, .ca is in Canada and .au is in Australia.

4 The last segment of the URL is the filename of the specific Web page you are requesting. If no filename is indicated, the browser will assume a default page, usually called index.html.

How URLs Help Retrieve Web Documents

1 The Web browser installed on your local computer sends your TCP/IP software a signal that it is ready to request a document. TCP/IP makes a connection with the host TCP/IP software. After the connection is established, your browser makes a request for a document by sending its URL through the two-way connection maintained by TCP/IP to the server.

2 The HTTP server is the portion of the host computer that runs HTTP server software. TCP/IP makes and maintains the connection this way. The browser can use HTTP to send requests and receive pages through the host's Web server software. This software allows the host to communicate with the client browser, in HTTP, over TCP/IP.



5 The browser on your local computer reads the file type. If it is an HTML document, the browser examines the content, breaking it down into meaningful parts. Two general parts include text, which is displayed by the browser word for word; the other part consists of HTML markup information called *tags* that are not displayed, but display formatting information, such as normal text, bold headers, or colored hypertext. The results are displayed on your monitor.

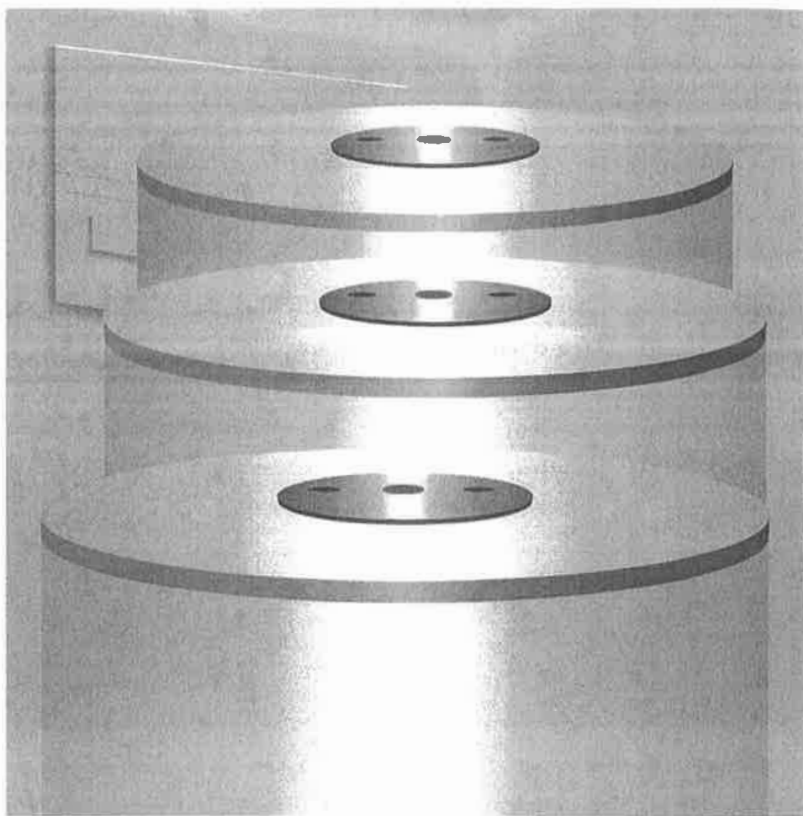
4 If the document is found, the host checks its file type (usually either x-html or x-text) and sends this information to the client with the requested page. When the client receives the page, it first checks the file type. If the type is one it can display, it does so; otherwise, it prompts the user to see whether he or she would like to save it to disk or open it using a helper application. The x-html file type is by far the most common one used when transmitting Web pages.

3 The server then receives the transmitted URL and responds in one of three ways. It follows the directory path given in the URL; the server finds the file on its local hard disk and opens it; the server runs a CGI script or it detects an error (such as "file not found") and generates an error document to be sent back to the client.

CHAPTER

28

How Web Host Servers Work



TO serve up pages, Web sites need a host—a computer—and server software that runs on the host. The host manages the communications protocols and houses the pages and the related software required to create a Web site on the Internet. The host machine often uses the UNIX, Windows NT, or Macintosh operating systems, which have the TCP/IP protocols built in.

The server software resides on the host and serves up the pages and otherwise acts on the requests sent from the client browser software. The server is not responsible for TCP/IP communications—the host operating system does that—but instead the server handles the HTTP requests and communications with the host operating system.

There are different types of server software (database servers or network servers, for example) that perform different types of services for different types of clients. Specifically, a Web server is an HTTP server and its function is to send information to the client software (typically a browser) using the Hypertext Transfer Protocol.

Usually, the client browser requests that the server return an HTML document. The server receives this request and sends back a response. The top portion of the response includes transmission information and the rest of the response is the HTML file.

A Web server does more than send pages to the browser, however. It passes requests to run CGI (Common Gateway Interface) scripts to the CGI applications. These scripts run external mini-programs, such as a database lookup or interactive forms processing. The server sends the script to the application via CGI and communicates the results of the script back to the browser, if appropriate. Moreover, the server software includes configuration files and utilities to secure and manage the Web site in a variety of ways.

How Web Server Software Works

1 Client (browser) software sends its request for data to the host, where the Web server software processes the request.

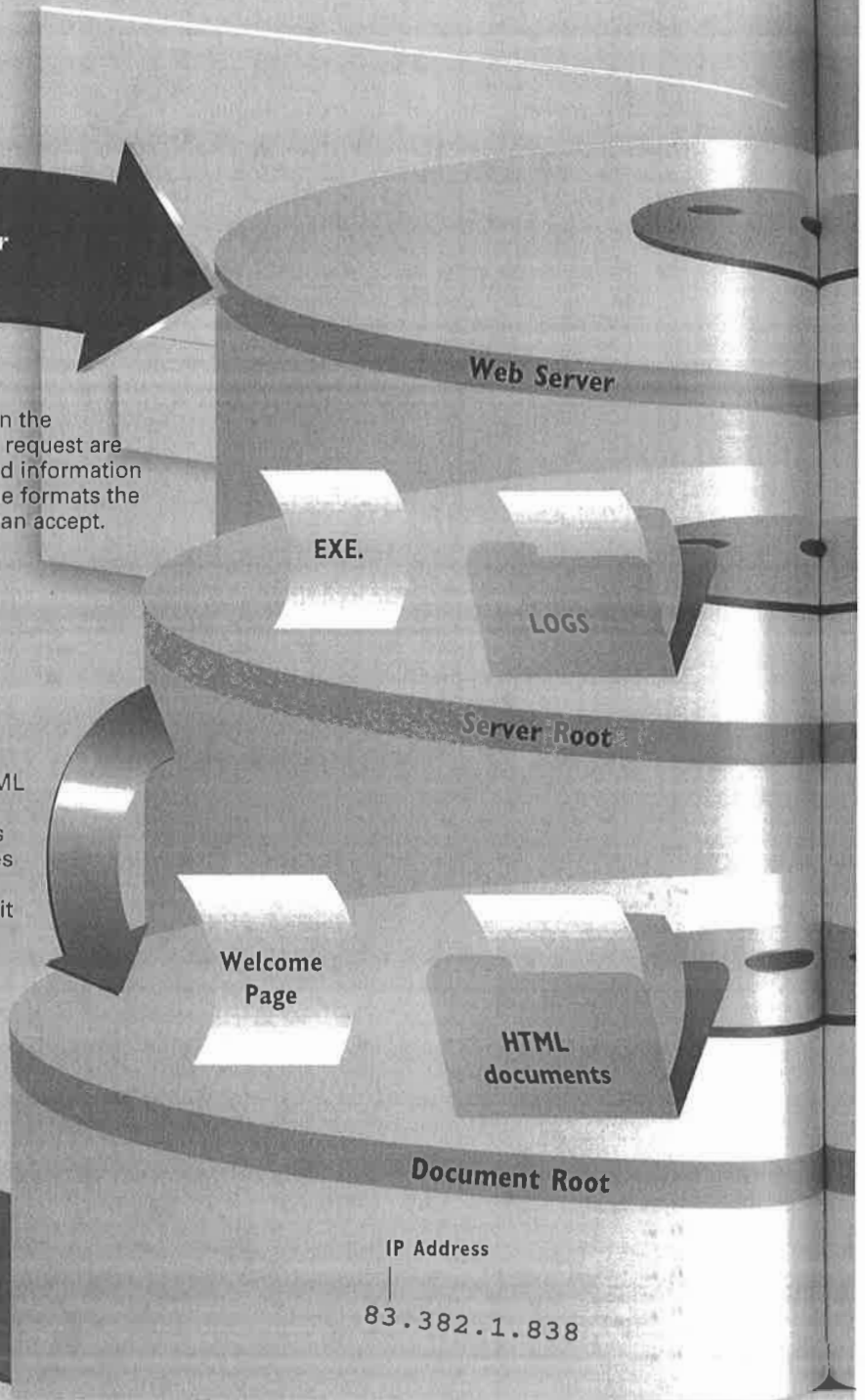


Request from Browser

2 Included in the browser's request are the desired information and the file formats the browser can accept.

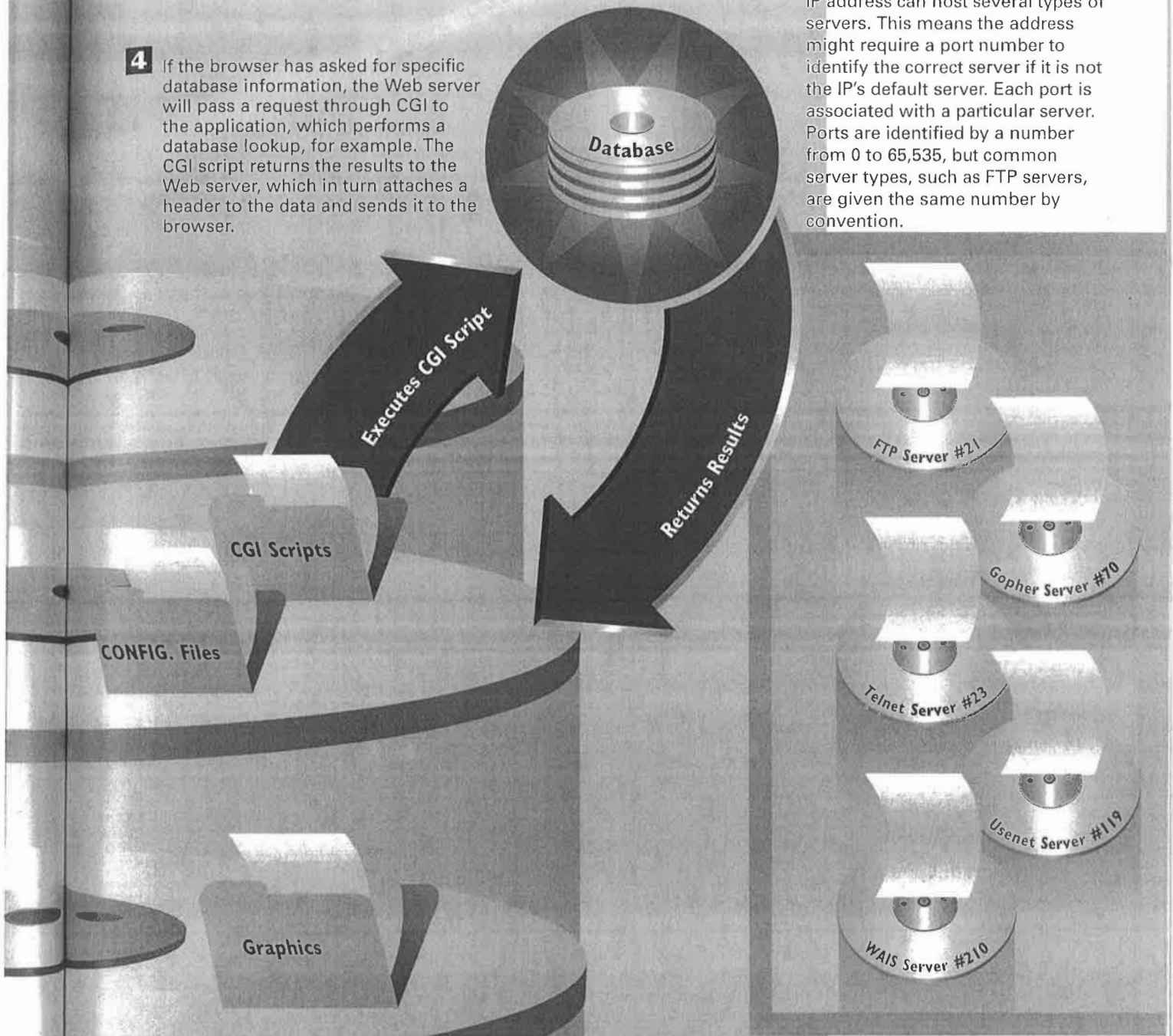
3 If the browser asks for an HTML file, the Web server retrieves the file, attaches a header to the file, and sends it to the browser.

Returns HTML page to Client



4 If the browser has asked for specific database information, the Web server will pass a request through CGI to the application, which performs a database lookup, for example. The CGI script returns the results to the Web server, which in turn attaches a header to the data and sends it to the browser.

NOTE A computer with a single IP address can host several types of servers. This means the address might require a port number to identify the correct server if it is not the IP's default server. Each port is associated with a particular server. Ports are identified by a number from 0 to 65,535, but common server types, such as FTP servers, are given the same number by convention.



Domain Name

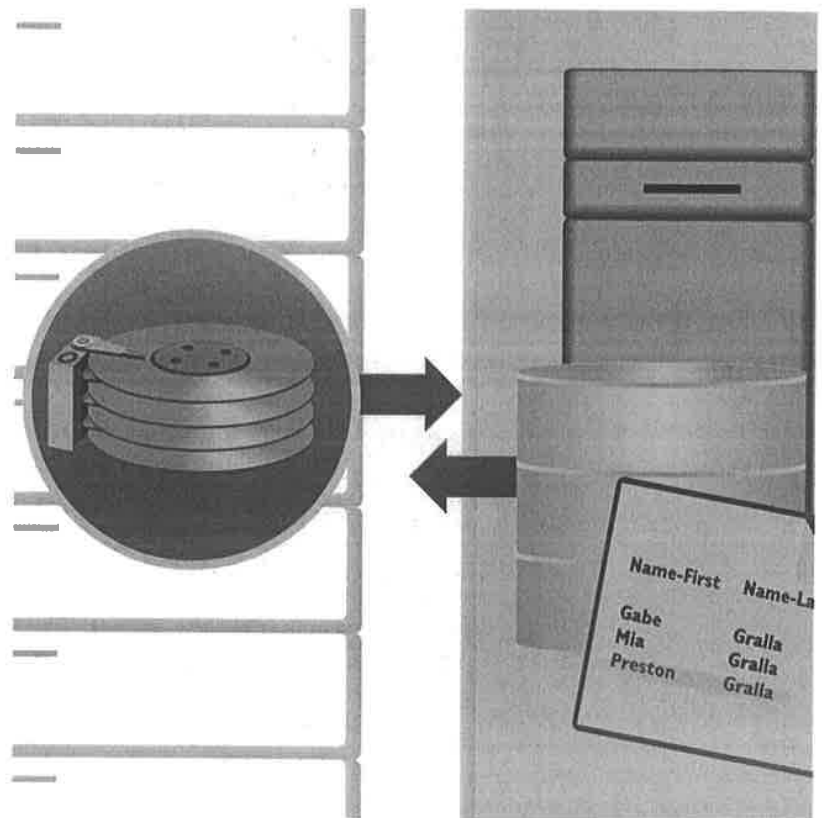
books.zdp.com

NOTE The host computer must have a unique IP address to send and receive information across the Internet. Because raw IP addresses are very intimidating, they are assigned a unique domain name, which is less daunting. The domain name is part of a hierarchical lookup system called the Domain Name System (DNS).

CHAPTER

29

How Web Sites Work with Databases



ONE of the most useful applications of the Web is its capability to link a Web site with a database so Web surfers can search for information. In essence, the Web page becomes the front-end for database applications, enabling you to select search criteria and execute even complex searches of a database that resides on the host computer.

A well-known and widely used example of this type of linking between Web sites and databases is the popular Yahoo! Web site. The Yahoo! site serves as a front-end to an extensive database of Web site descriptions, which can be searched according to keywords. The Welcome page includes a search dialog box in which you enter a keyword that represents the subject matter you are looking for. Selecting "Search" from the page sends a request from the browser to the Web server to bring back a list of all Web sites that contain your keyword.

Furthermore, not only can the Web serve up data, it can also collect it. For example, many Web sites ask users to "register" their names, addresses, and other demographic information that is captured and stored in a database.

But how does this all work? You don't have to be a corporate giant—or for that matter even an able programmer—to link your Web site to a database. In fact, linking a Web site to a database can be relatively simple. The database can take just about any form and can be as simple as a FileMaker Pro database, or as complex as an Oracle SQL database. The bridge that brings together Web sites and databases is the Common Gateway Interface (CGI).

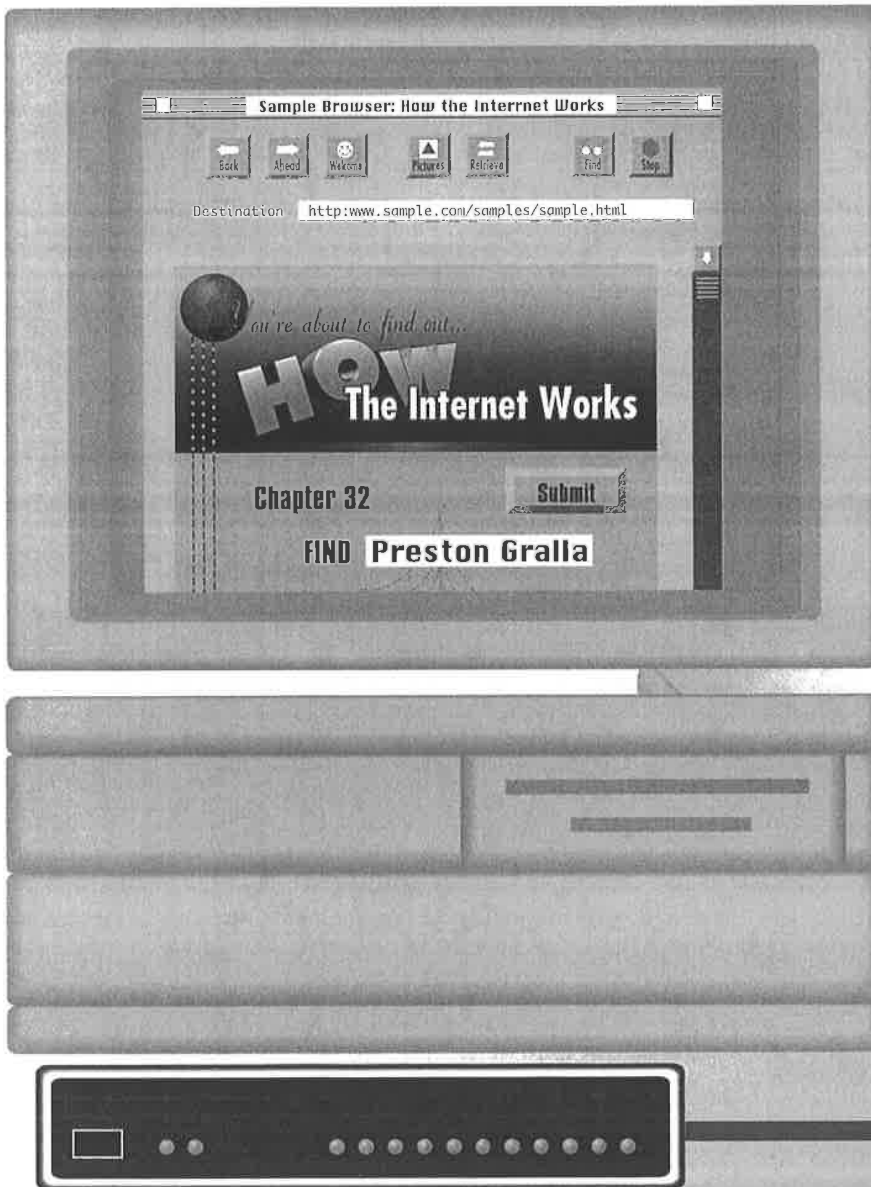
On the client side of the database, you see a Web page that includes a form in which you enter your search terms. By executing the search, you launch a CGI script that sends a search command to the Web server in the form of a link to the CGI bin on the Web server. So a search on the Yahoo! site for public relations firms looks like this:

```
http://search.yahoo.com/bin/search?p=public+relations
```

When the Web server receives this URL, it identifies the URL as a trigger for a CGI script (called "script" in this example) and passes it along with the search criteria ("public relations," in this example) to the mini-program using CGI. The CGI script then sends the search to the database, receives the results of the query along with the HTML page created by the database to contain the result, and passes it on to the Web server to be sent back to the client. That's a lot of handing off of requests and data, but typically even a search of a large database is very fast because the majority of UNIX and Windows NT databases—the types most often used—can perform these tasks simultaneously. All of this happens behind the scenes, of course—you won't need to do any kind of database work or scripting yourself. Instead, the Web sites you visit have easy-to-use interfaces that take care of interacting with databases; you'll have to type only what you're looking for.

How the Web Works with Databases

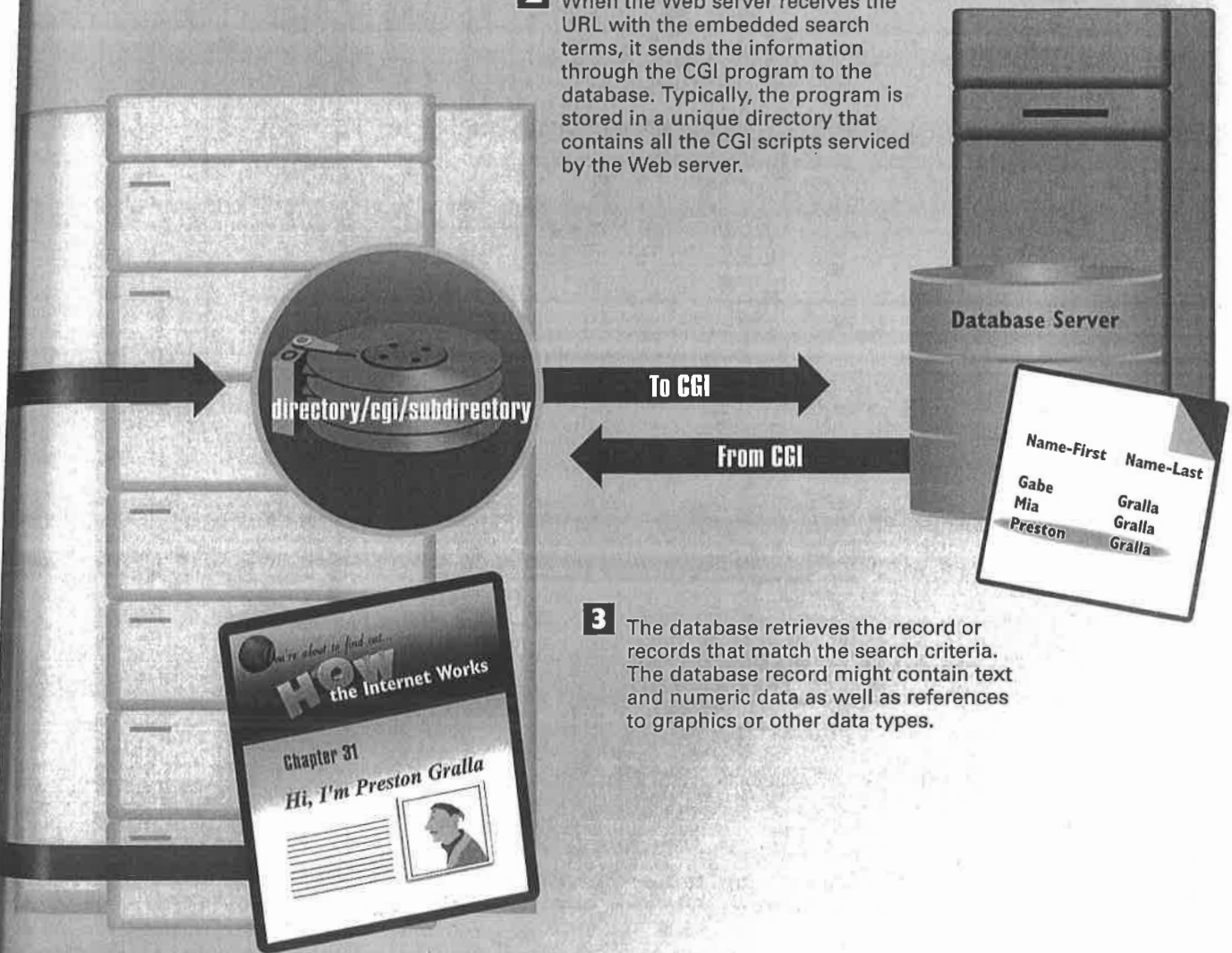
❏ The search begins on a Web page that includes a form field to accept search terms and HTML codes to execute a CGI script. The browser may pass the data to the Web server in a query string. The *query string* contains the name of the CGI script in a directory called *cgi-bin*. This directory is followed by a subdirectory that includes the search terms, often separated by a question mark or slashes. The HTML code for an extra path might look like this: ``.



```
<a href="cgi-bin/search?preston+gralla/">
```

HTTP Response

2 When the Web server receives the URL with the embedded search terms, it sends the information through the CGI program to the database. Typically, the program is stored in a unique directory that contains all the CGI scripts serviced by the Web server.



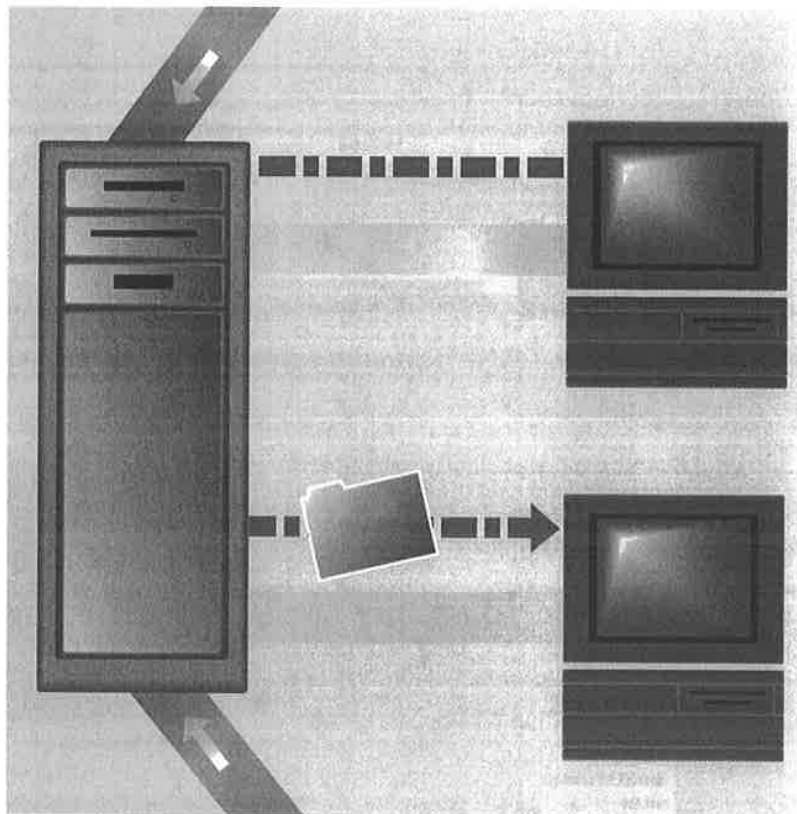
3 The database retrieves the record or records that match the search criteria. The database record might contain text and numeric data as well as references to graphics or other data types.

4 The database returns the data to the Web server via CGI in the form of a new HTML page. The server then sends the page back to the client browser as a new HTML page.

CHAPTER

3

FTP and Downloading Files



ONE of the most popular uses of the Internet is to download files—that is, transfer files from a computer on the Internet to your computer. These files can be of many types: programs that you can run on your own computer; graphics you can view; sounds and music you can listen to; or text files that you can read. Many tens of thousands of files are downloaded every day on the Internet. Most of those files are downloaded using the Internet's File Transfer Protocol, commonly referred to as FTP. FTP can also be used to upload files from your computer to another computer on the Internet.

FTP, like many Internet resources, works on a client/server model. You run FTP client software on your computer to connect to an FTP server on the Internet. On the FTP server, a program called an FTP daemon (pronounced "demon") allows you to download and upload files.

To log on to an FTP site and download files, an account number (or username) and a password must be typed in before the daemon will allow you to enter. Some sites allow anyone to enter and download files—but an account number (or username) and password must still be entered. Often, to get in, you use "anonymous" as your username and your email address as your password. Because of this, these sites are often referred to as anonymous FTP sites. Some FTP sites are private, allowing only certain people with the proper account number and password to enter.

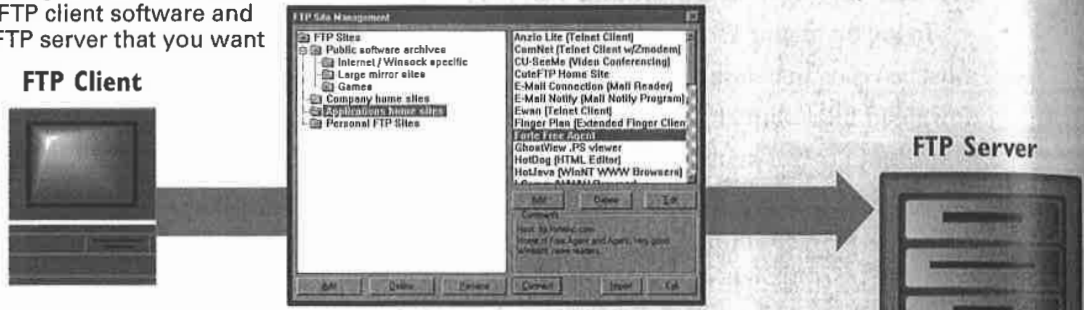
FTP is fairly simple to use. When you log on to an FTP site, you can browse through the available files by changing directories and you can see a listing of all the files available in each directory. When you see a file you want to download, you use your client software to instruct the FTP server to send you the file.

As the World Wide Web gains popularity, downloading software is becoming even easier. You can use your Web browser and click on links to files; behind the scenes, FTP is often still downloading the files. FTP remains the most popular way to download files from the Web and the Internet. The HTTP protocol of the Web can be used for downloading files from the Web, but it's not as efficient as FTP, so it isn't used as frequently.

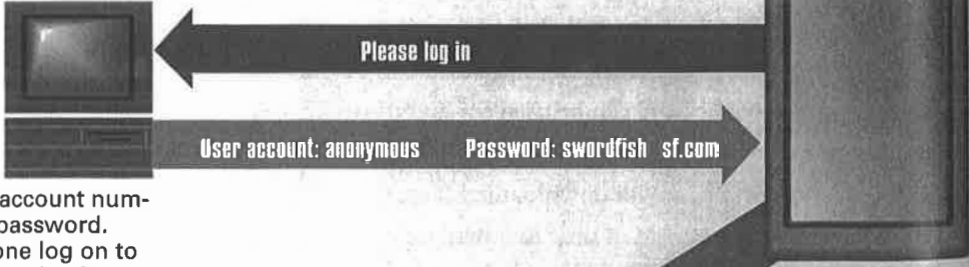
One problem with downloading files over the Internet is that some files are so large it can take a tremendous amount of time to download them, especially if the connection is made via modem. Even at 28,800bps, downloading files can be slow. As a way to speed up file transfers and save space on the FTP server, files are commonly compressed—shrunk in size using special compression software. Many different methods are used to compress files. Depending on the file type, files are usually compressed from 10 to 50%. After the files have been downloaded, you'll need to run the compression software on your own computer to decompress the files so you can use them.

How an FTP Session Works

1 FTP, like many other Internet resources, runs on a client/server model. So, in order to use it you'll need client software to run on your computer. To begin an FTP session, you run the FTP client software and contact the FTP server that you want to download files from. You can get FTP client software in hundreds of places on the Internet, including the ZDNet Software Library at www.hotfiles.com.



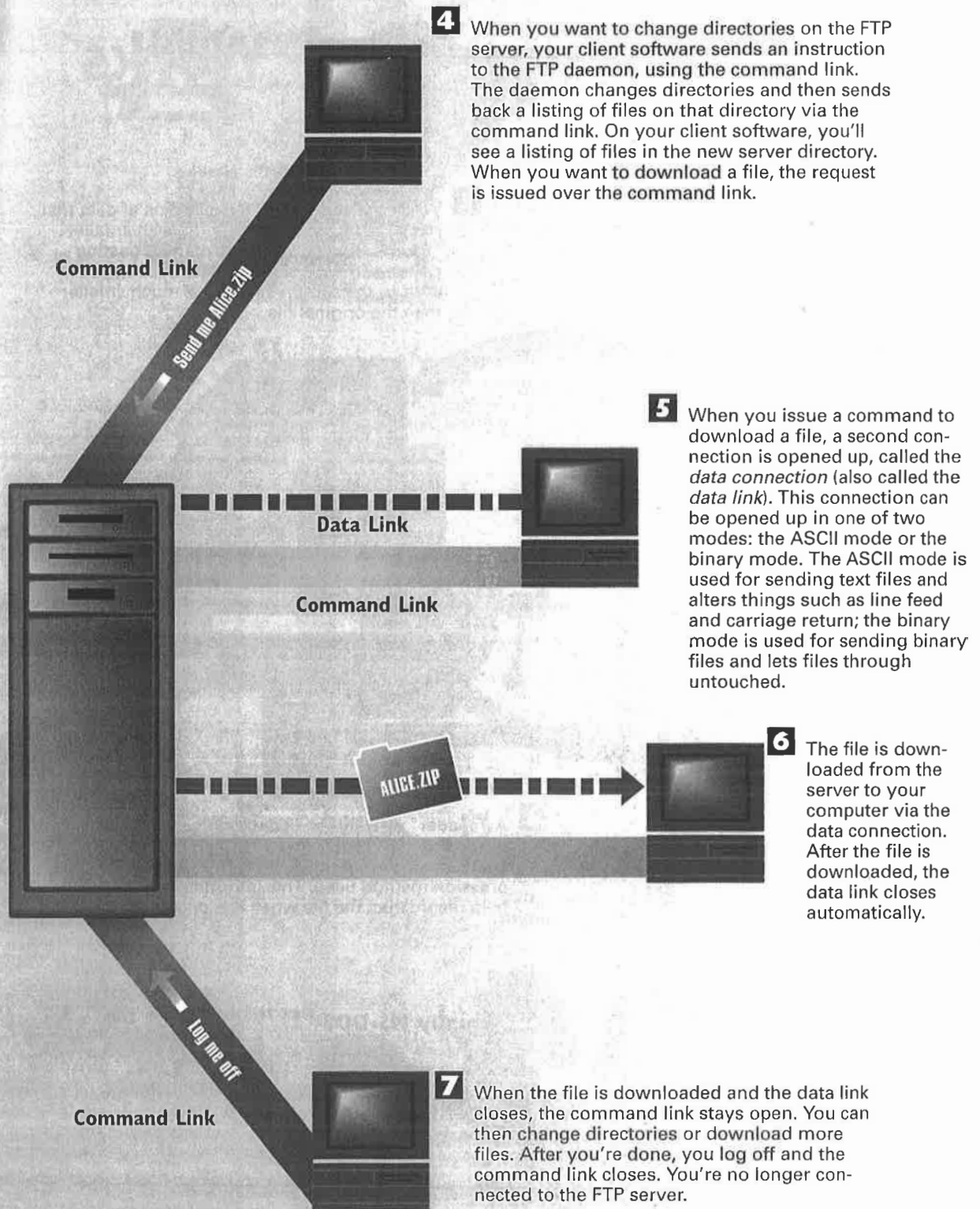
2 The FTP daemon runs on the FTP server. This daemon handles all FTP transactions. When an FTP client contacts a server, the daemon will ask for an account number (or username) and password. Many FTP sites let anyone log on to them to download files and software. This is called *anonymous FTP*. With anonymous FTP, you often use "anonymous" for your account number and your email address for your password. Note that some FTP clients will automatically log on to the FTP server for you when you connect so you won't be asked to log on—your software will do it for you automatically.



Command Link

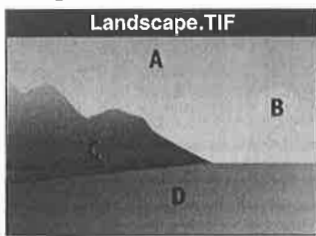
3 When you log on to the FTP server, a connection called a *command link* is opened up between your computer and the server. This link is used for sending commands to the server from your computer, and sending messages and information back from the server to your computer.





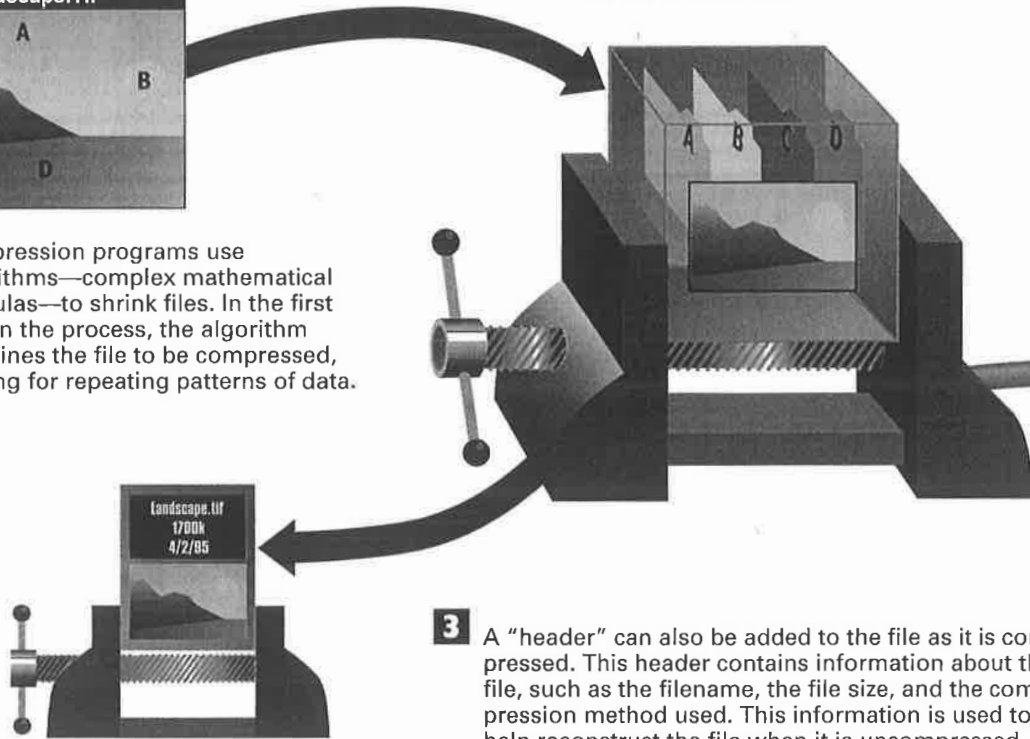
How File Compression Works

$$(a_2b^3 \sim q_4) \div C^4 \sim \text{cosine}b^{-1^4}$$



1 Compression programs use algorithms—complex mathematical formulas—to shrink files. In the first step in the process, the algorithm examines the file to be compressed, looking for repeating patterns of data.

2 When the algorithm finds patterns of data that repeat, it replaces the patterns with smaller “tokens.” In a file that has many repeating patterns, many tokens are used to replace data so the compressed file is much smaller than the original file.

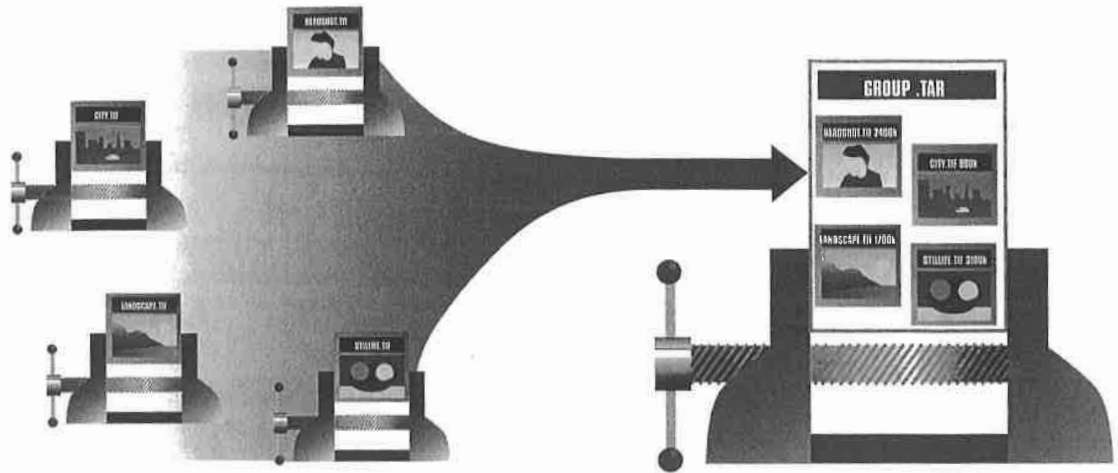


3 A “header” can also be added to the file as it is compressed. This header contains information about the file, such as the filename, the file size, and the compression method used. This information is used to help reconstruct the file when it is uncompressed.

7 File extensions, the letters that appear after the period at the end of a filename, can tell you whether and how a file is compressed.

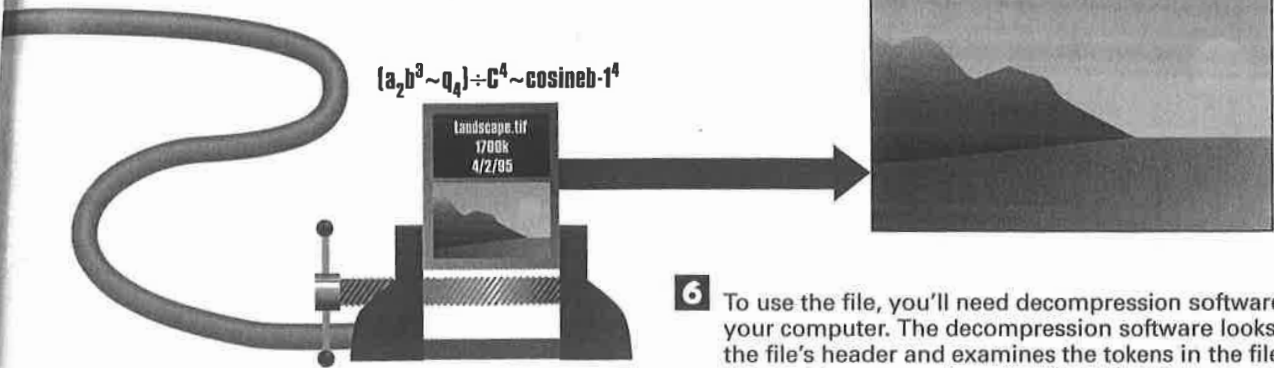


Types of File-Compression Schemes You'll Find on the Internet

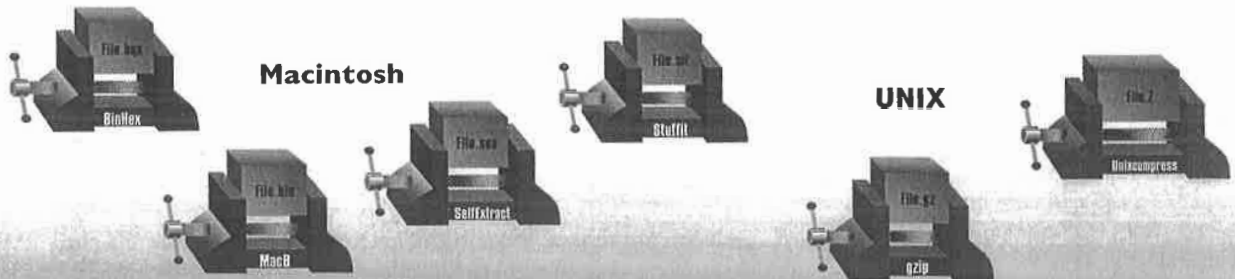


4 Some compression software, such as PKZIP for the PC, can also archive files, combining several compressed files. The UNIX command TAR can also combine many files into a single archive.

5 When you want to use a compressed file you find on the Internet, transfer it over the Internet to your computer.



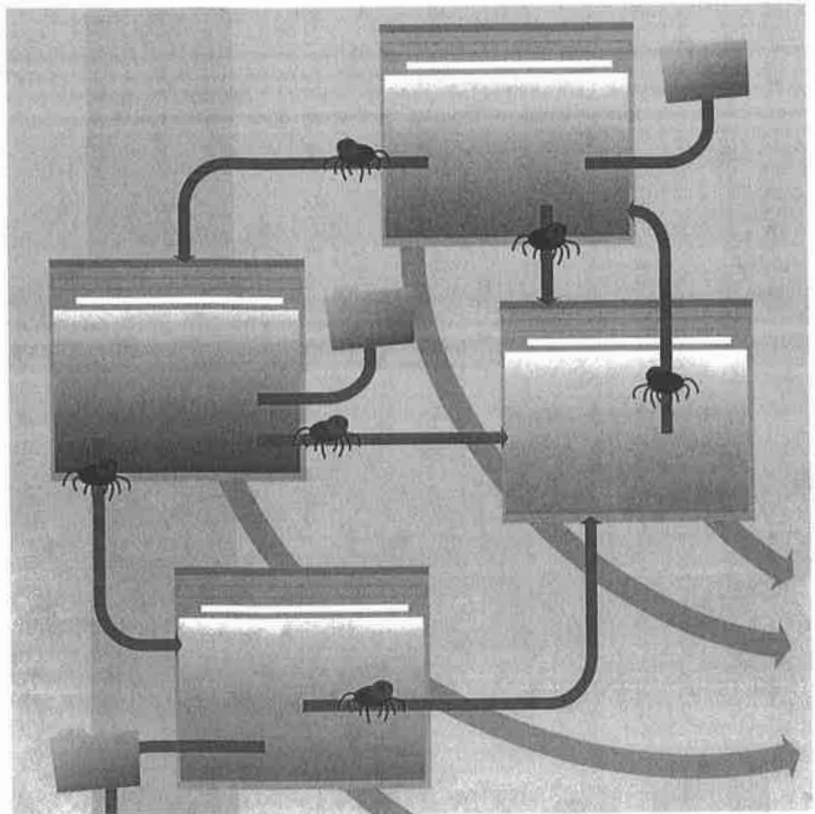
6 To use the file, you'll need decompression software on your computer. The decompression software looks into the file's header and examines the tokens in the file. Using a decompression algorithm, it reconstructs the original file, which you can then use on your computer.



CHAPTER

32

Searching the Internet



SO much information is available on the Internet, but there is so little organization to the Internet that it can seem impossible to find the information or documents you want. A number of solutions have sprung up to solve the problem. The two most popular ones are indexes and search engines.

Indexes present a highly structured way to find information. They enable you to browse through information by categories, such as arts, computers, entertainment, sports, and so on. In a Web browser, you click a category, and you are then presented with a series of subcategories. Under sports, for example, you'll find baseball, basketball, football, hockey, and soccer. Depending on the size of the index, several layers of subcategories may be available. When you get to the subcategory you're interested in, you'll be presented with a list of relevant documents. To get to those documents, you click the links to them. Yahoo! (<http://www.yahoo.com/>) is the largest and most popular index on the Internet. Yahoo! and other indexes also enable you to search by typing words that describe the information you're looking for. You then get a set of search results—links to documents that match your search. To get the information, you click a link.

Another popular way of finding information is to use *search engines*, also called *search tools* and sometimes called *Web crawlers* or *spiders*. Search engines operate differently from indexes. They are essentially massive databases that cover wide swaths of the Internet. Search engines don't present information in a hierarchical fashion. Instead, you search through them as you would a database, by typing keywords that describe the information you want.

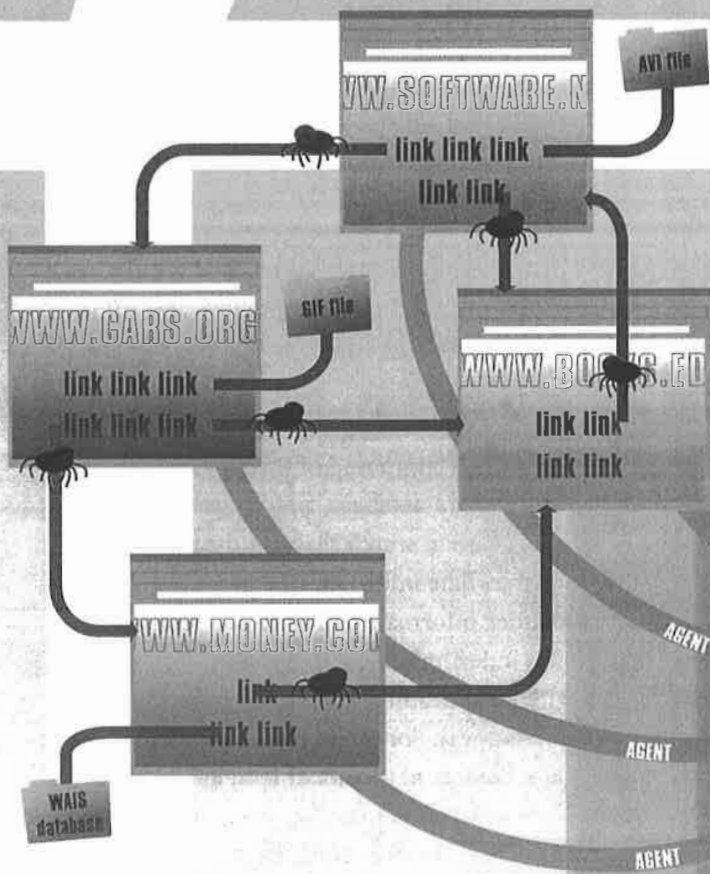
Many popular Internet search engines exist, including Lycos, Excite, and AltaVista. Although the specifics of how they operate differ somewhat, generally they are all composed of three parts: at least one spider, which crawls across the Internet gathering information; a database, which contains all the information the spiders gather; and a search tool, which people use to search through the database. Search engines are constantly updated to present the most up-to-date information, and they hold enormous amounts of information. Search engines extract and index information differently. Some index every word they find in a document, for example, and others index only the key 100 words in each document. Some index the size of the document; some index the title, headings, subheadings, and so on.

Additionally, each search engine returns results in a different way. Some weigh the results to show the relevance of the documents; some show the first several sentences of the document; and some show the title of the document as well as the URL.

Many search engines and indexes are on the Internet, each with its own strengths and weaknesses. To cast the widest possible net when looking for information, you'd like to search as many of them as you can. The problem is that doing so is too time-consuming. So a type of software called meta-search software has been developed. With this software, you type a search on your own computer. The software then submits the search to many Internet search engines and indexes simultaneously, compiles the results for you, and then delivers the results to your computer. To visit any resulting site, just click the link, the same as if you were on an index or search engine site.

How Internet Search Engines Work

1 Each search engine uses a crawler or spider with its own set of rules guiding how documents are gathered. Some follow every link on every home page that they find and then, in turn, examine every link on each of those new home pages, and so on. Some spiders ignore links that lead to graphics files, sound files, and animation files. Some ignore links to certain Internet resources such as WAIS databases, and some are instructed to look primarily for the most popular home pages.



2 As the spider discovers documents and URLs, software agents are instructed to get the URLs and documents and send information about them to indexing software.

INDEXING SOFTWARE

And/Or/Not/End/Cancel [End]
 Searching for #3 AND #4
 Press any key to interrupt processing

Set number 6 resulted in 569 references found in 88 citations

Command/HELP/BYE [View 6]: c

First Set number (1-6) to Combine/List/BYE [CANCEL]: 5

And/Or/Not/End/Cancel [And]: or

Next Set number (1-6) to Combine/List/BYE [CANCEL]: 6

And/Or/Not/End/Cancel [End]:

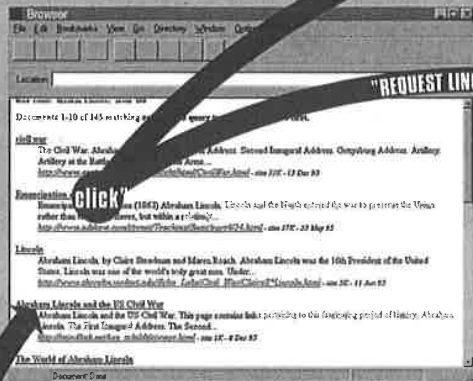
Searching for #5 AND #6
 Press any key to interrupt processing

Set number 7 resulted in 748 references found in 118 citations

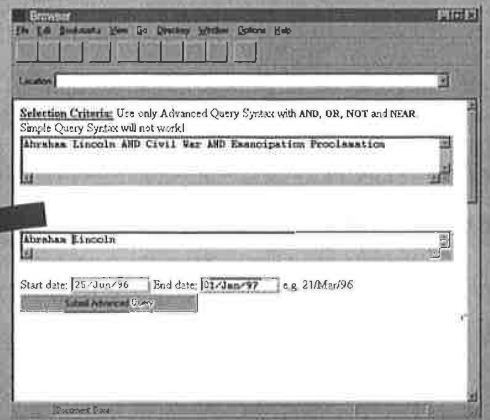
Command/HELP/BYE [View 7]

SET NUMBER	CITATIONS	SEARCH HITS	SET CRITERIA
1	911	2269	Subject - Arts
2	7150	14448	Abstract - Computers
3	1031	4421	Abstract - Entertainment
4	49466	109752	Subject - Sports
5	31	179	#1 AND #2
6	88	569	#3 AND #4
7	118	748	#5 OR #6

6 When you click a link to one of the documents that interest you, you're sent straight to that document. The document itself is not in the database or on the search engine site.



5 The database is searched, based on the criteria you've set. Results are returned in HTML pages. Each search engine returns results in a different way. Some weigh the results to show how relevant the document is to your search; some show the URL, as well as the first several sentences of the document; and some show the title of the document and the URL.



4 When you visit a search engine and want to search the Internet for information, you'll type words on a Web page that describe the information you want to find. Depending on the search engine, more than just keywords can be used. For example, you can search by date and other criteria with some search engines.

3 The indexing software receives the documents and URLs from the agent. The software extracts information from the documents and indexes it by putting the information into a database. Each search engine extracts and indexes different kinds of information. Some index every word in each document, for example, but others index only the key 100 words in each; some index the size of the document and the number of words in it; some index the title, headings and sub-headings, and so on. The kind of index built will determine what kind of searching can be done with the search engine and how the information will be displayed.

Database

How Meta-Search Software Works

Meta-search software is software that sits on your computer and enables you to search through many Internet search engines simultaneously and to view and use the results. You can browse through the results on your computer and click those pages that you want to visit. This illustration shows how one such piece of meta-search software, called Copernic, works.

2 The meta-search software sends many "agents" simultaneously—depending on the speed of your connection, usually from 4 to 8, but it can be as many as 32 different agents. Each agent contacts one or more search engines or indexes, such as Yahoo!, Lycos, or Excite.

1 When you want to search for something on the Internet, you type descriptive words or a search term into the meta-search software.



5 The agent sends the results back to the meta-search software. After the agent sends its report back to the meta-search software, it goes to another search engine and submits a search in that engine's proper syntax, and then again sends the results back to the meta-search software.

6 The meta-search software takes all the results from all the search engines and examines them for duplicate results. If it finds duplicate results, it deletes them. It then displays the results of the search, ranking each "hit" by the likelihood that it contains the information you requested. It figures out the ranking by examining the title of the site found, the header information in the site, and the words on the site.

Preston's Picks

<http://www.hotfiles.com/home.html>
<http://www.hotfiles.com/index.html>
<http://www.hotfiles.com/prespick/presmain.html>

Preston's Picks

<http://www.hotfiles.com/index.html>
<http://www.hotfiles.com/prespick/presmain.html>

Preston's Picks

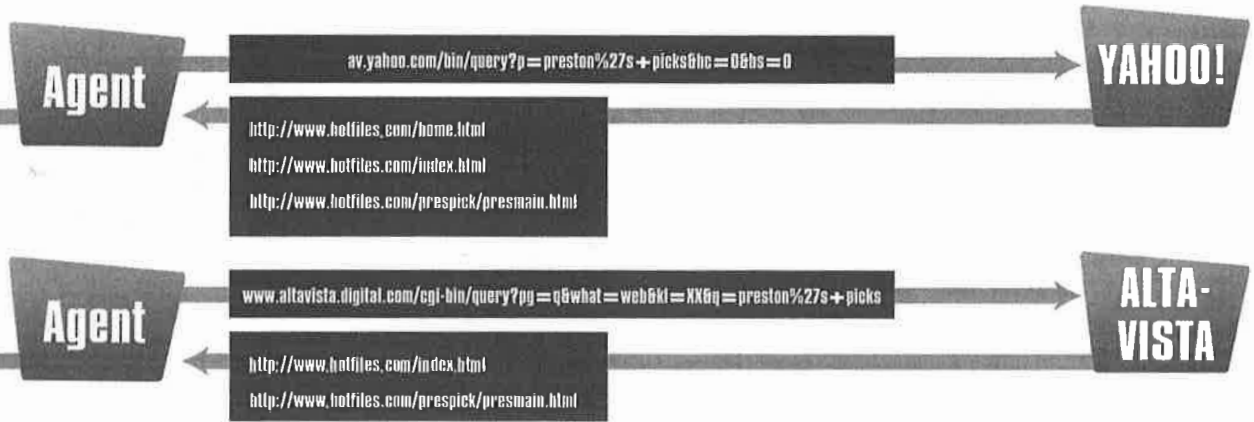
<http://www.hotfiles.com/home.html>
<http://www.hotfiles.com/index.html>
<http://www.hotfiles.com/prespick/0498/pc.html>

Preston's Picks

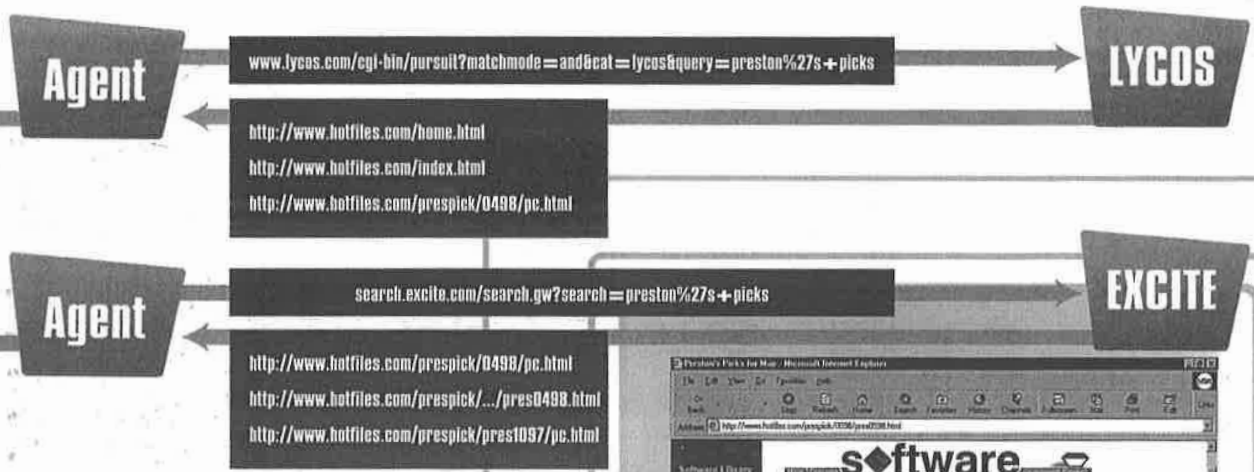
<http://www.hotfiles.com/prespick/0498/pc.html>
<http://www.hotfiles.com/prespick/.../pres0498.html>
<http://www.hotfiles.com/prespick/pres1097/pc.html>

Icon	Address	Rank	File Count	Date Found
1	ZDNet Software Library - Top Rated Home &... http://www.hotfiles.com/home.html	0	2	5/8/98 4:16:10
2	ZDNet Software Library - Top Rated Servers http://www.hotfiles.com/index.html	2	3	5/8/98 4:16:10
3	Preston's Picks for April http://www.hotfiles.com/prespick/0498/presmain.html	4	3	5/8/98 4:16:10
4	Preston's Picks for April http://www.hotfiles.com/prespick/0498/pc.html	5	1	5/8/98 4:16:10
5	Preston's Picks for April http://www.hotfiles.com/prespick/.../pres0498.html	6	2	5/8/98 4:16:10
6	ZDNet Software Library - Preston's Picks for... http://www.hotfiles.com/prespick/pres1097.html	6	1	5/8/98 4:16:10
7	ZDNet Software Library - Preston's Picks http://www.hotfiles.com/prespick/presmain.html	1	5	5/8/98 4:16:10

3 The agents are intelligent enough to know how each search engine functions—for example, whether a particular engine allows for Boolean searches (searching by using AND, OR, and other variables). The agents also know the exact syntax that each engine requires. The agents put the search terms in the proper syntax required at each specific search engine and submit the search—they don't have to fill out forms, as users normally do at search engines.



4 The search engines report the results of the search to each agent. The results typically include the URL of each site that matches the search, and often a summary of information found on the site, the date the site was last updated, and other data.



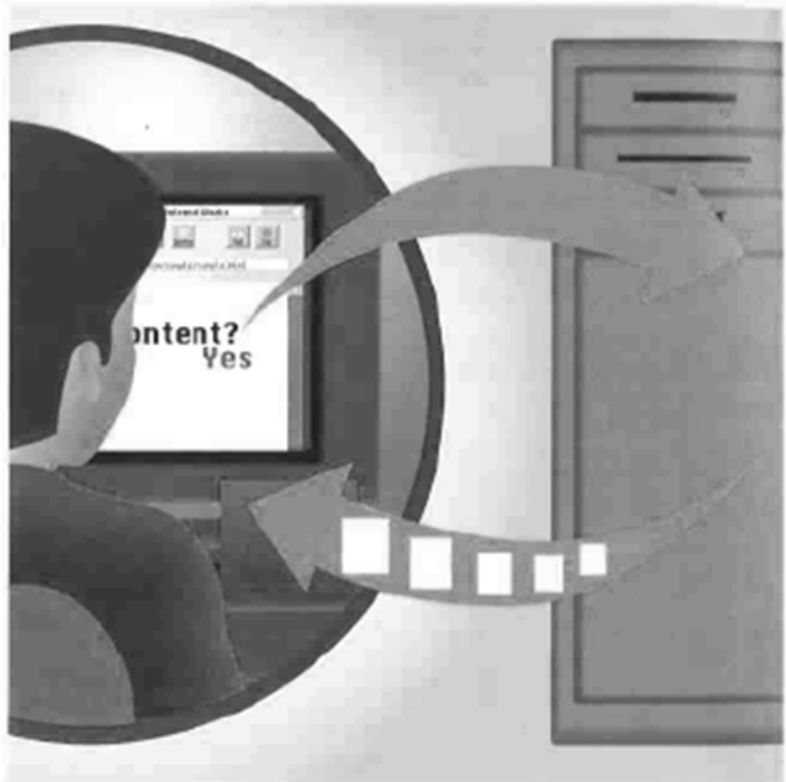
7 You browse through the results in the meta-search software. When you see a page you're interested in, you double-click it. You'll then be sent to that site.



CHAPTER

33

How Push Technology Works



INCREASINGLY, the Internet is becoming a broadcast medium. It relies less each day on people taking it upon themselves to visit certain sites. Instead, information, entire Web sites, and applications can be sent via the Internet straight to people's computers with a variety of technologies collectively known as *push*.

In push technology, you "subscribe" to Web sites, often called channels. A *channel* generally refers to an area of interest that a publisher builds, which can include HTML pages, Java applets, ActiveX components, multimedia objects, and other information packaged together to deliver customized information to people via push technology. Those sites you subscribe to and the information they contain are sent to your computer automatically at intervals that you specify. You don't pay for these subscriptions; you merely ask to receive the desired information on a regular schedule.

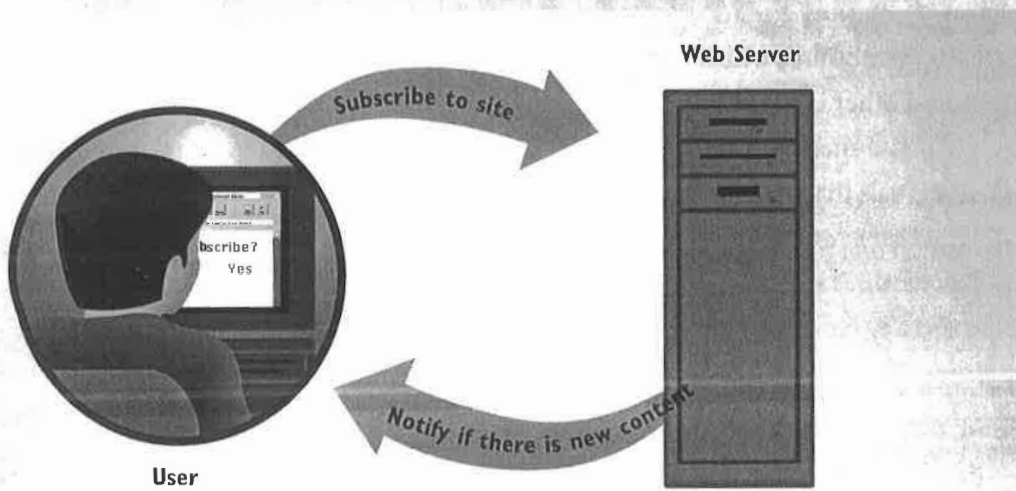
Receiving these channels requires either special client software or a push-enabled browser on your desktop. This software is generally free. Push technology is built into browsers such as Netscape Communicator and Microsoft Internet Explorer. Push client software is available from companies such as PointCast and BackWeb. Most push technologies also allow you to customize the kind of information that you receive.

This kind of customization is one benefit that push technologies offer. Another benefit is the time savings—instead of you having to go out and gather the information, it's delivered right to you with no effort on your part, and it's precisely the kind of information that interests you. There's one more benefit as well: When the information is sent to you, it resides locally on your computer. So you can read it whenever you want—you don't have to be connected to the Internet. You'll also be able to view the pages and information more quickly than if they were on the Internet instead of your local computer because you're reading it at high speed from your hard disk instead of across the slow wires of the Internet.

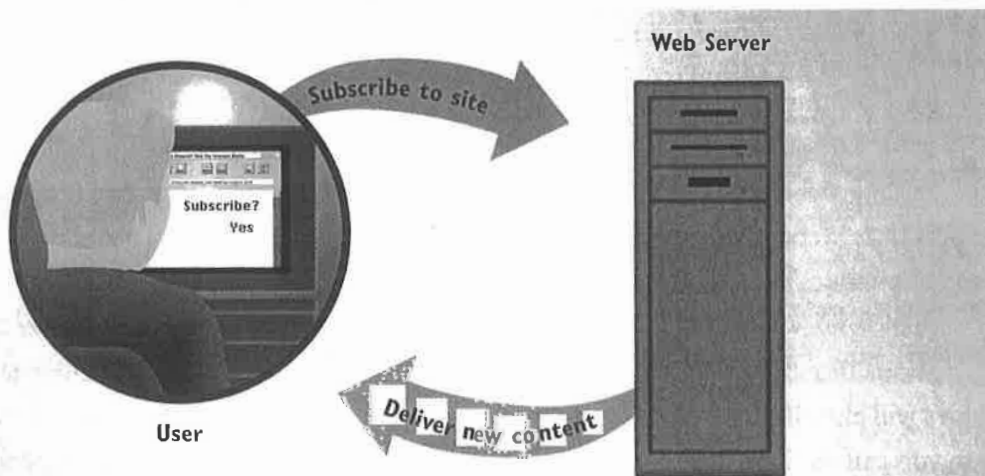
There is no single standard for push technology. Netscape has established its own standard and software, called Netcaster, which is part of its Communicator package. Netscape refers to its version of push technology as Netcasting. Microsoft has its own version of push technology built into Internet Explorer 4.0 and later, which it refers to as Webcasting. Aside from Netscape and Microsoft there are other push technologies as well, from companies such as PointCast, Marimba, and BackWeb.

Most push technologies have a few basics in common. For example, most allow you to customize how you'll receive your information. To receive that information, you'll have to be connected to the Internet, so many push technologies will also dial into the Internet for you to get updates. They deliver the information to your computer so you can read the information locally at your leisure. Also, most use similar terminologies, such as the word "channel."

How Internet Explorer's Webcasting Works



1 Internet Explorer 4.0 offers a variety of ways to use push technology and offline reading—what Microsoft calls Webcasting. In the simplest form of Webcasting, you “subscribe” to a Web site and ask to be told when the content on the site changes. Your browser will regularly visit that site and notify you if new content is available. It won’t, however, automatically download any of that content.



2 In a slightly more sophisticated version of Webcasting, you subscribe to a Web site and ask that whenever there is new content, it be automatically delivered to you. In this instance, your browser regularly contacts the Web site, “crawls” the pages by looking at every page, and downloads to your computer only those pages that have changed since your last visit. You’ll now be able to view those pages offline, when you’re not connected to the Internet.

3 A much more sophisticated form of Webcasting uses channels. A channel is designed from the ground up to be delivered via push technology, and incorporates multimedia, ActiveX, and other interactive kinds of content. To build a channel, developers create HTML files, ActiveX controls, and similar content, and post them on their Web pages. They then create a Channel Definition Format (CDF) file that describes the channel, how it functions, and where its content is on a Web server. This CDF file is linked to a button on a Web page.

```

<?XML version="1.0"?>
<!DOCTYPE Channel SYSTEM "http://www.w3c.org/Channel.dtd">
<CHANNEL HREF="http://www.hotfiles.com">
<SELF HREF="http://www.hotfiles.com/hf.cdf" />
<TITLE> ZNet Software Library </TITLE>
<LOGO HREF="http://www.hotfiles.com/hf.ico" STYLE="ICON"/>
<LOGO HREF="http://www.hotfiles.com/hf.gif" STYLE="IMAGE"/>
<SCHEDULE>
<INTERVAL TIME HOUR="2" />
<LATEST TIME MIN="30" />
</SCHEDULE>
<CHANNEL>
<TITLE> Hottest New Files</TITLE>
<ABSTRACT> Best new files.</ABSTRACT>
<ITEM HREF="http://www.hotfiles.com/latest/new.htm">
    
```

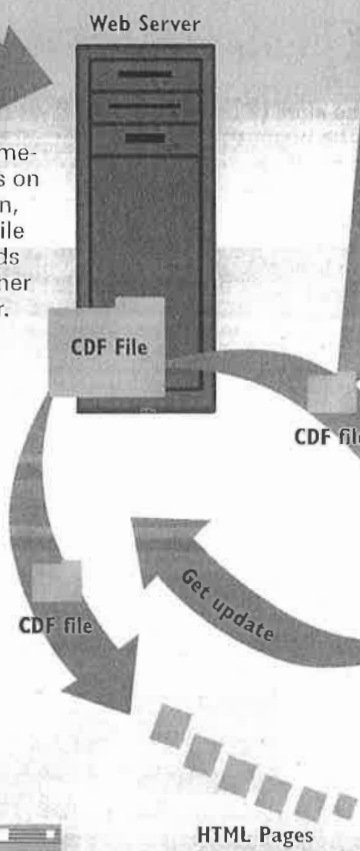


Developer/Editor

4 When someone clicks on the button, the CDF file downloads to his or her computer.

5 The CDF file determines how the channel functions. It contains data such as how often to schedule updates, information about the structure of the channel (such as whether it has subchannels), and links to specific Web pages and contents, such as ActiveX controls and multimedia content.

6 The browser visits the Web site on a schedule determined in the CDF. Then it downloads the CDF file in case the file has changed, and also downloads the associated content referenced in the CDF file such as Web pages, ActiveX controls, and multimedia content.



User

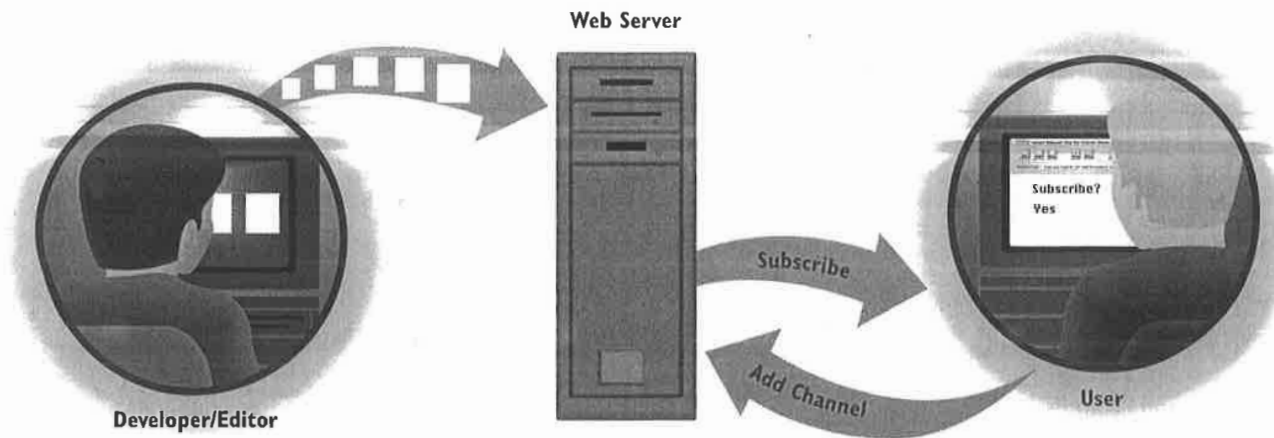
7 The channel pages and associated files are now on the local computer. Because the information and pages are local instead of residing on the Internet, they can be accessed while the user is not connected to the Internet. The pages and information will be accessed quickly because they are being read from a hard disk instead of across the Internet.



How Netscape's Netcasting Works

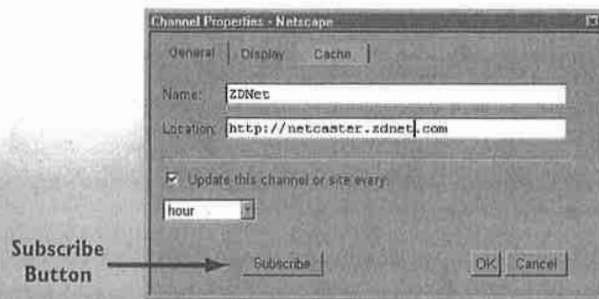
1 Netcaster is the part of Netscape Communicator that allows for information and sites to be pushed to someone's desktop, and for those sites to be browsed offline. The sites that can be pushed are called channels. Channels are built by developers using HTML and JavaScript and can also incorporate Java applets and multimedia content. Before a channel can be pushed, the developer posts all the content and channel information on a Web site, and sets default properties for that channel, such as how often Netcaster should automatically update it.

3 When the button is clicked on, an "addChannel" JavaScript call is made. A dialog box pops up, listing the properties of the channel (such as how often to update it). The user can change the properties or keep the defaults. After the user clicks on OK, the channel is added to Netcaster. The subscription to that channel is now in effect.

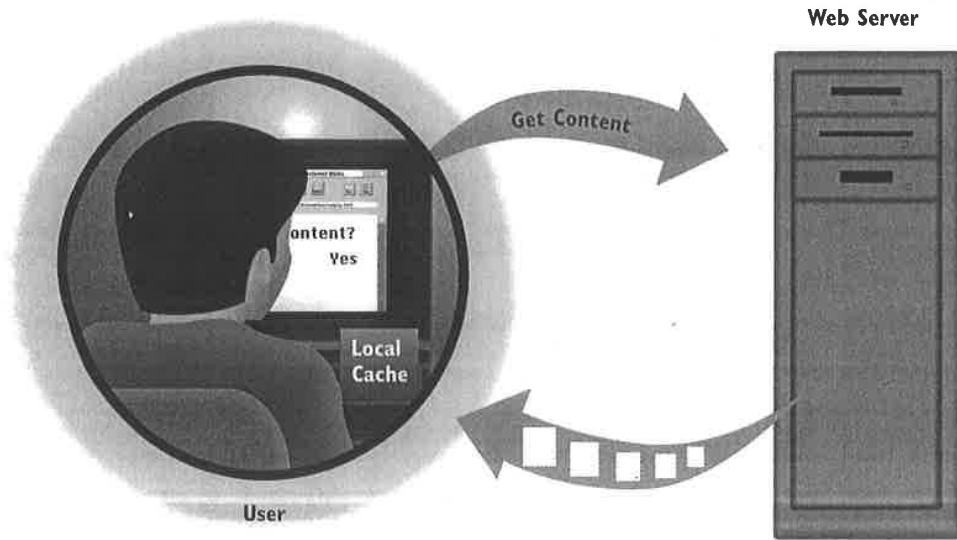


2 After the channel is created, the developer creates a button on a Web site that, when clicked, will "subscribe" people to the channel.

4 Through the Channel Properties box, the user can change many aspects of how to interact with the channel, such as how often to check the channel for new content, and how many levels of Web pages should be checked.



- 5 Netscaster connects to the channel at intervals the user specifies, or at intervals in the default information about the channel.



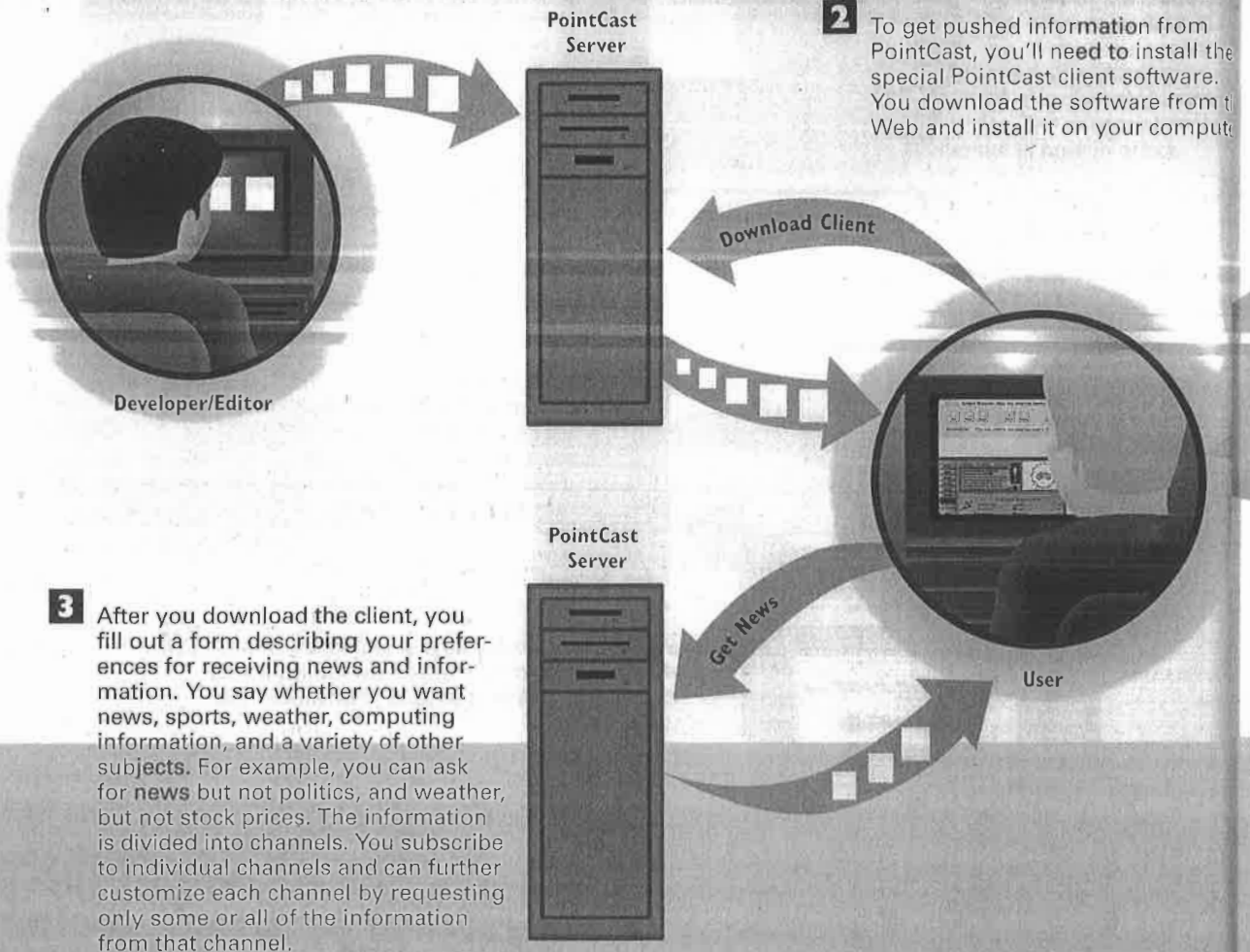
- 6 After Netscaster connects, it "crawls" the channel—that is, it downloads the information, data, and pages in the channel. First it downloads the HTML pages, then the associated graphics, then embedded objects such as Java applets or audio and video files. The pages, graphics, and embedded objects are all put into a cache on the local machine.



- 7 The content of the channel can be read from the cache. Because the information and pages are local instead of on the Internet they can be accessed when the user is not connected. They can also be accessed quickly because they're being read from a hard disk instead of across the Internet.

How PointCast Works

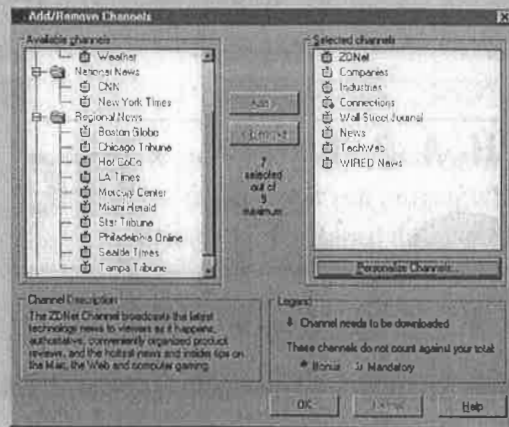
- 1** PointCast, the first push technology to gain widespread acceptance, delivers customized news and information to people's desktops. Publishers who want to deliver information via PointCast have to create special files and automated feeds that are placed on PointCast servers.



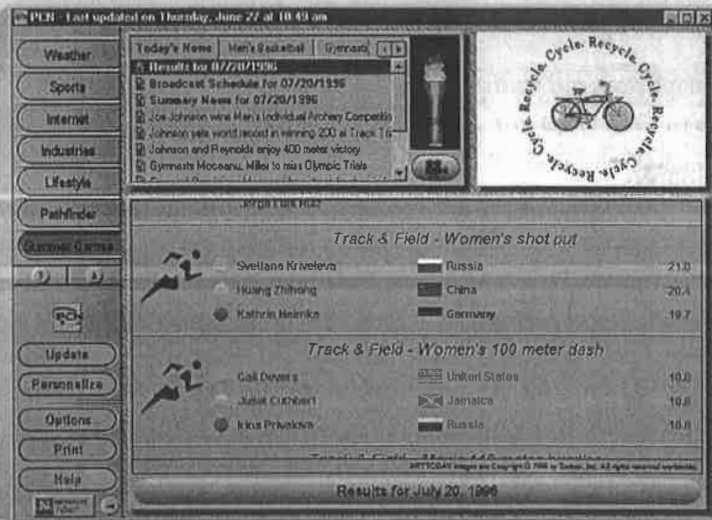
- 2** To get pushed information from PointCast, you'll need to install the special PointCast client software. You download the software from the Web and install it on your computer.

- 3** After you download the client, you fill out a form describing your preferences for receiving news and information. You say whether you want news, sports, weather, computing information, and a variety of other subjects. For example, you can ask for news but not politics, and weather, but not stock prices. The information is divided into channels. You subscribe to individual channels and can further customize each channel by requesting only some or all of the information from that channel.

4 At set times that you determine in your preferences, the PointCast client software automatically connects to the PointCast server over the Internet. The software downloads the customized news that you've requested. You can also manually tell the client software to download the news at any time you want.



5 Using the PointCast client, you can now read the news on your own computer. The software can also be set up as a screen saver so that when the computer is idle for a certain amount of time, the news will flash across your screen.

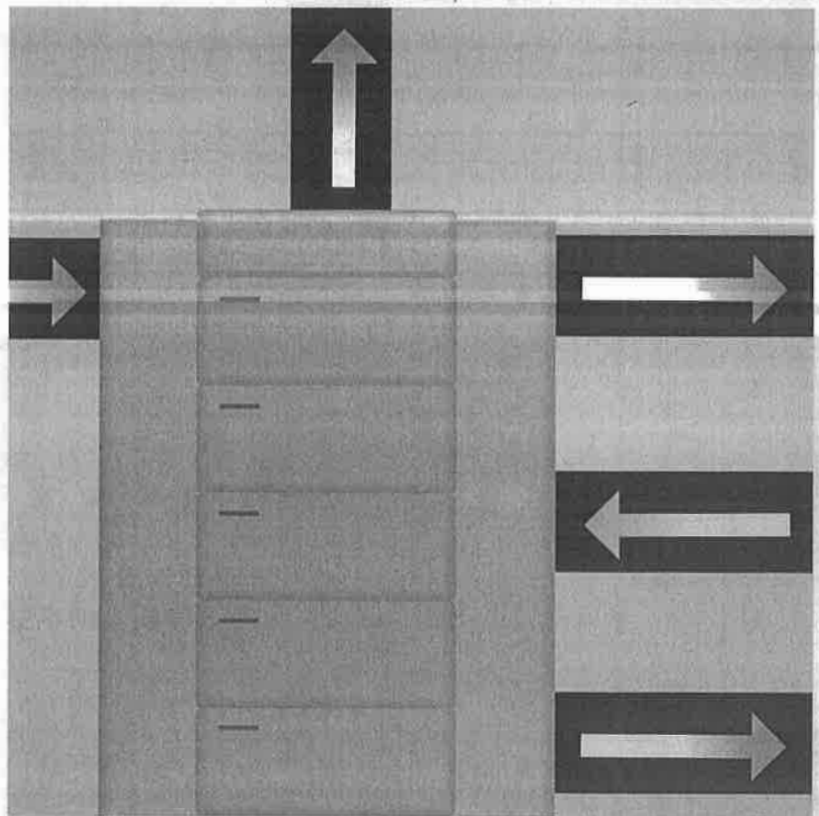


6 Often, in news stories, there are URLs to sites that have additional information. When that link is clicked on, an Internet browser is launched, and the specified site is contacted. The person can now read more in-depth news. For example, there may be a brief news story in PointCast about a company's quarterly earnings. That story could contain links to stories on the Web with more in-depth financial information, as well as other articles about that company.

CHAPTER

34

How Java, JavaScript, and ActiveX Work



THE Internet is no longer a place that you visit with your computer and merely look at documents or gather information—increasingly, it is an extension of your computer. You can now run programs that reside on the Internet rather than on your own computer, and tools have been developed that enable your computer and the Internet to interact as if they were one large computer system. This allows for all kinds of things never before possible: news tickers that flash breaking news; interactive games; multimedia presentations combining animation, sound, and music; and much more.

The three most important programming tools associated with Web technology are Java, ActiveX, and JavaScript. *Java*, a computer language developed by Sun Microsystems, enables applications to be run from the Internet, the same as word processing and spreadsheet programs are run on your computer. It is similar to the C++ computer programming language, and it is object-oriented, which means programs can be created by using many preexisting components instead of by a programmer writing the entire program from scratch.

Java programs run inside your Web browser if you have a Java-enabled browser, such as Netscape. When Java programs are run inside a browser, they are called *applets*. You don't need to do anything to run a Java applet. When you visit a Web site that has a Java applet on it, the applet is downloaded automatically from a Web server, and then run automatically in your browser. Java applets can be run on any computer, such as a PC, a Macintosh, or a UNIX workstation.

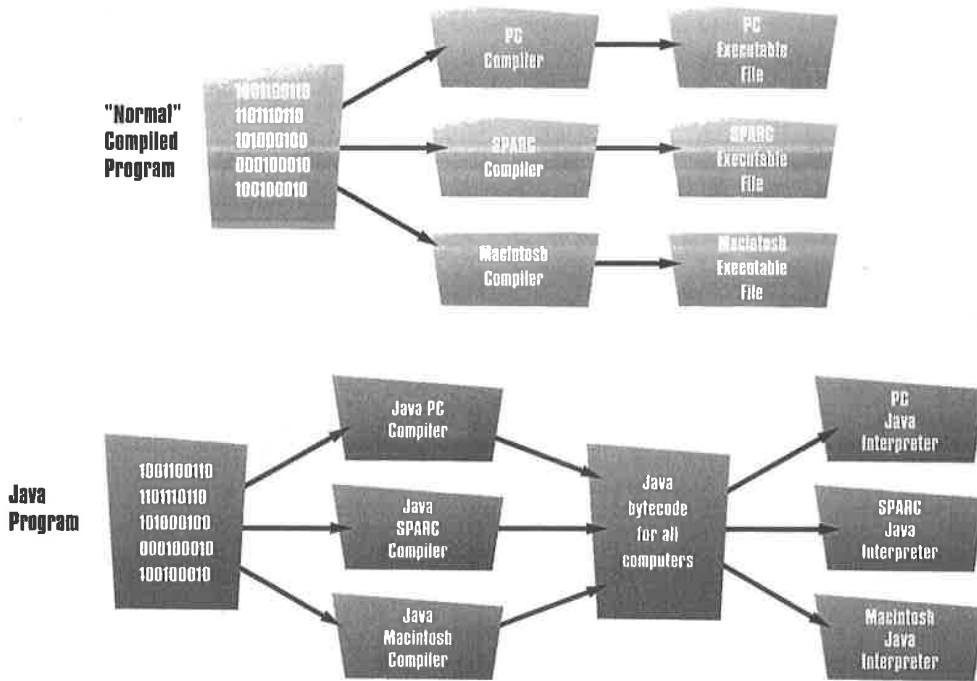
ActiveX, a technology from Microsoft, allows Internet programmers to create programs—commonly referred to as ActiveX *controls* or *components*—that can essentially turn the Internet into an extension of your computer. Like Java applets, these controls are downloaded to your computer and run there. They can do anything that a normal application can do, in addition to interacting with the Web, the Internet, and other computers connected to the Internet. To run these controls, a browser that supports ActiveX, such as Internet Explorer, is needed.

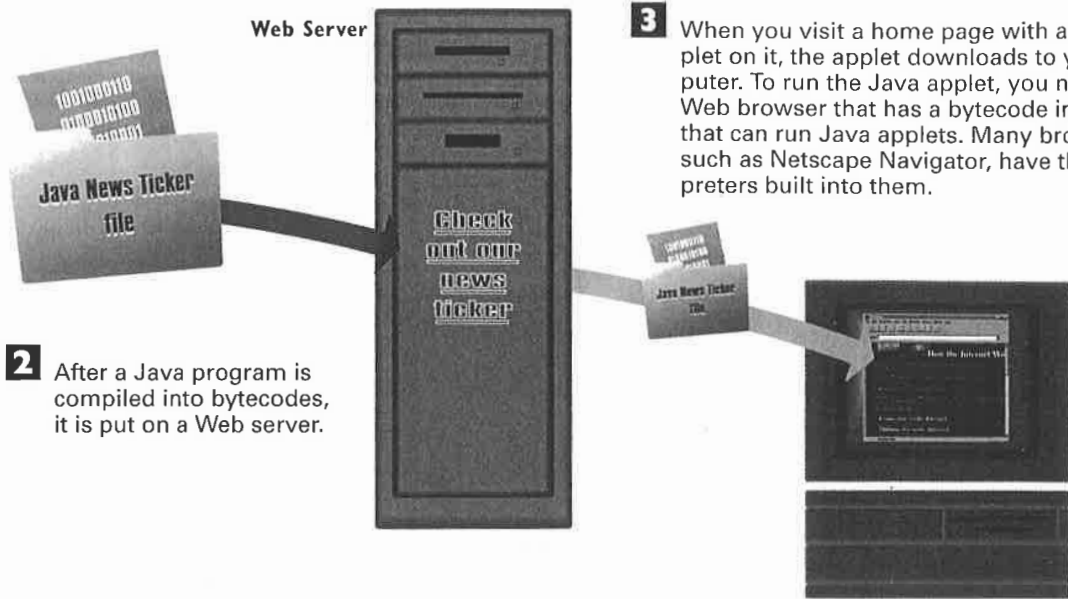
One benefit of ActiveX controls is that they are written as components, which means they can be put together, much like building blocks, to build larger and more complex applications. Another benefit is that because you have already downloaded a component, you won't ever need to download it again. This means when you visit a page with a complex ActiveX application on it, you may need to download only a small portion of it because you may already have the other components on your computer.

JavaScript is a scripting language that is less complex and, therefore, much easier to learn than Java and ActiveX. People without substantial programming experience can write scripts with JavaScript. It's also an interpreted language, which means that its commands are executed by the browser in the order in which the browser reads them. It's commonly used for things such as creating drop-down boxes, navigational aids, and interactive forms, although it can be used for creating more complex applications, as well.

How Java Works

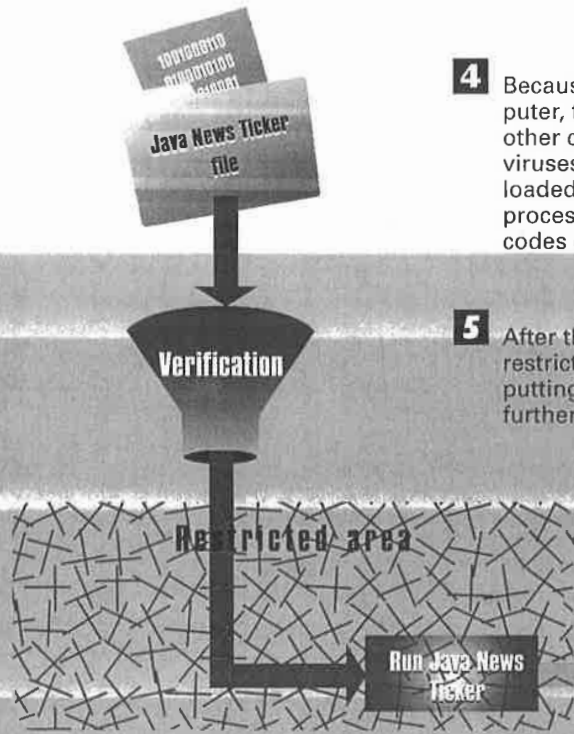
I Java is a *compiled language*, which means that after a Java program is written, the program must be run through a compiler to turn the program into a language that a computer can read. Java differs from other compiled languages, however. In other compiled languages, computer-specific compilers create distinct executable binary code for all the different computers on which the program can run. In Java, by contrast, a single compiled version of the program—called *Java bytecode*—is created by a compiler. Interpreters on different computers, such as a PC, Macintosh, or SPARC workstation, understand the Java bytecode and run the program. In this way, a Java program can be created once and then used on many kinds of computers. Java programs designed to run inside a Web browser on the World Wide Web are called *applets*. Java-enabled browsers contain Java bytecode interpreters.





2 After a Java program is compiled into bytecodes, it is put on a Web server.

3 When you visit a home page with a Java applet on it, the applet downloads to your computer. To run the Java applet, you need a Web browser that has a bytecode interpreter that can run Java applets. Many browsers, such as Netscape Navigator, have these interpreters built into them.



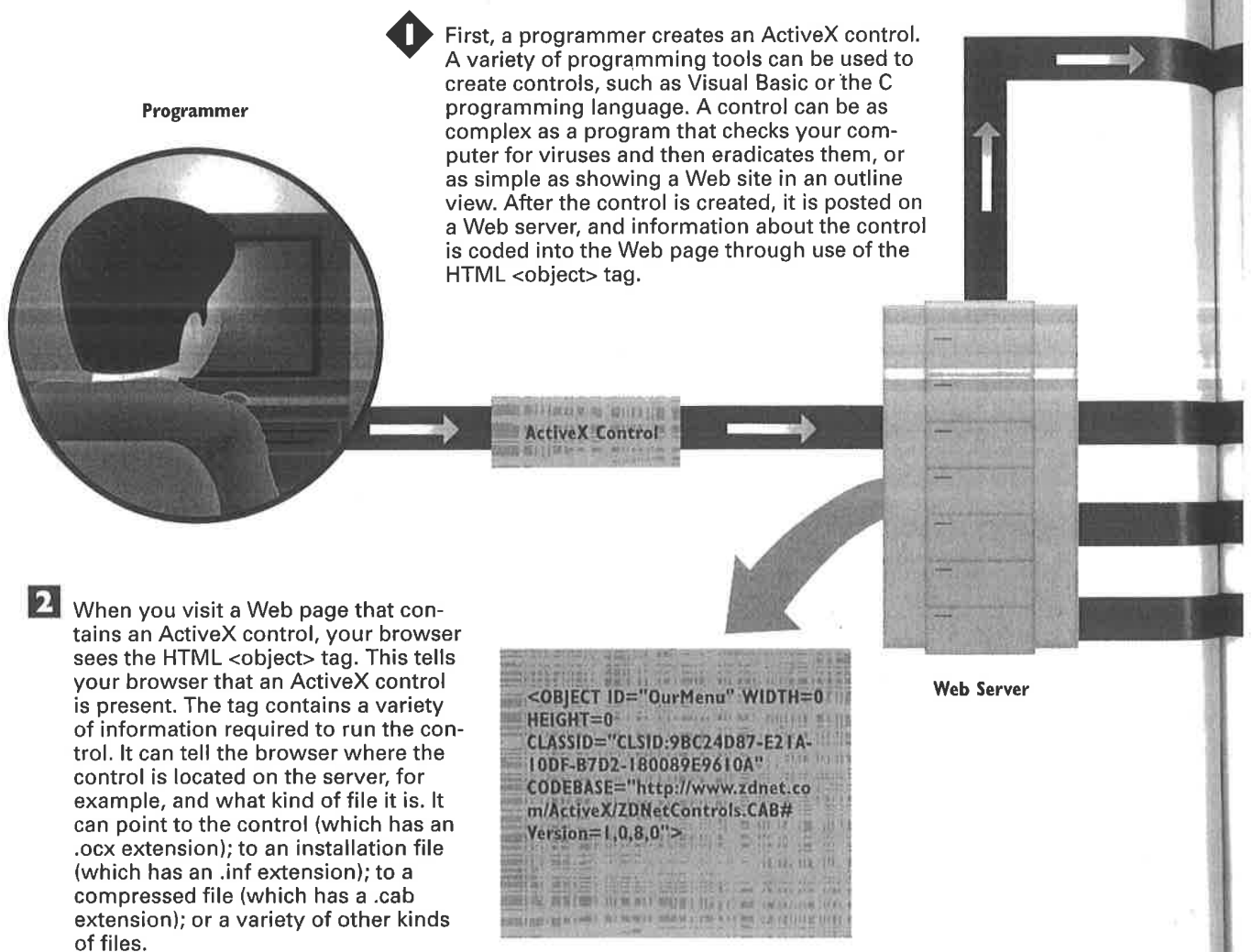
4 Because Java applets are programs that run on your computer, theoretically they could carry a virus just like any other computer program could. To help ensure that no viruses infect your computer, when a Java applet is downloaded to your computer, the applet first goes through a process of *verification*—a process that makes sure the bytecodes can be run safely.

5 After the bytecodes have been verified, they are put into a restricted area in your computer's memory and run. By putting the applet into this special area of your computer, further care is taken that no virus can harm your computer.

6 The Java applet is run; because it allows for animations and interactivity, it opens many other kinds of multimedia possibilities.



How ActiveX Works



3 Some ActiveX controls need more than a single control to work. In that event, the HTML page will contain multiple references to ActiveX controls needed to run the control on that page. The controls can be located on the same server, on a different server on the same Web site, or on another site and server on the Internet.

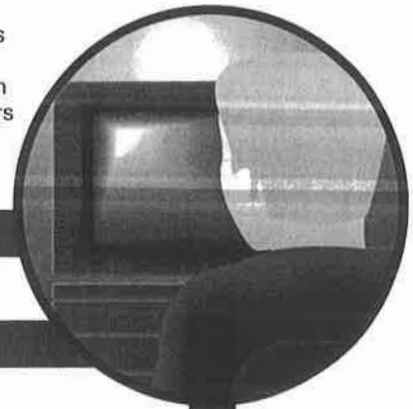
Web Server



→ → Get ActiveX Control

4 Using the information in the <object> tag, the browser starts to download the ActiveX controls. If you have set your browser to a certain level of security, you will get a message asking whether you want to download the control. To ensure even more security, ActiveX controls can be digitally "signed" by a digital certificate authority such as VeriSign. This signing will assure you that the control you want to run was written by the person to whom it is attributed. If a problem occurs with the control, you will be able to contact that person.

Web Browser



→ Do you want to Download Control? →

← Yes ←

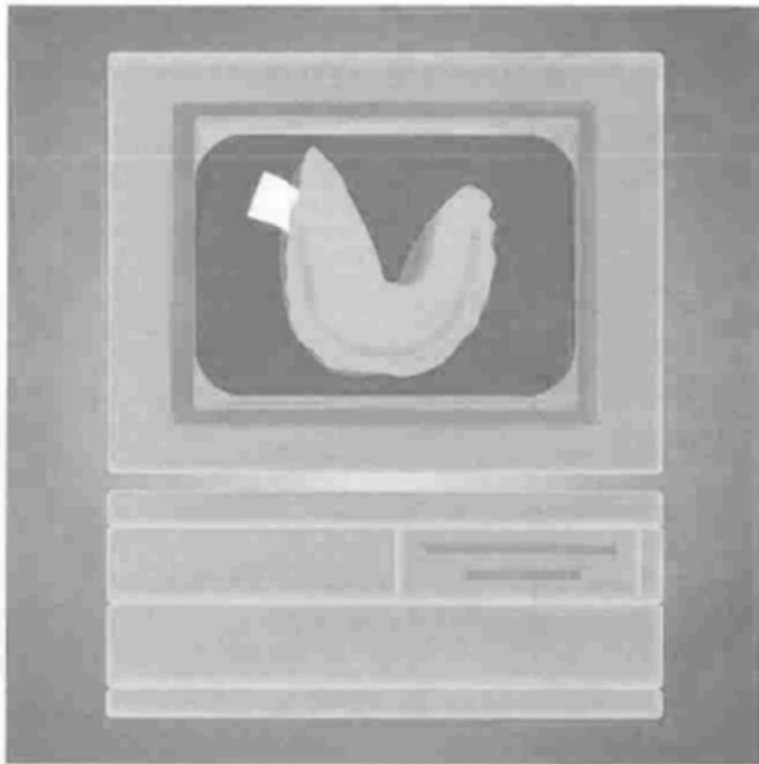
→ ActiveX Control →

5 If you have low security set on your browser, or if you give the okay to download the control, the control and its related ActiveX controls are downloaded to your computer. Some of the controls may already be on your system, so you won't need to download them. After the control is downloaded, the file is decompressed (if it was compressed), information about it is put into the Windows Registry, and it is installed on your computer. The control will then run. An ActiveX control can do anything that any other program can do. It can interact with your computer and with any Internet resource, such as the Web, FTP, Telnet, or virtually any other Internet resource. It can also directly use the Internet's TCP/IP protocols so that it need not ride on top of another Internet resource.

CHAPTER

48

How Cookies, Passports, and Web Tracking Work



PRIVACY issues are a big concern on the Net. Much information can be gathered about people when they use the Net, and it's not always clear who will use that information or how it will be used. In particular, two technologies that concern people are cookies and Web tracking. Both serve useful purposes, but many people worry that there is a "Big Brother" aspect to them. One technology, Internet "passports," might enable people to ensure that their privacy isn't invaded while still enabling Web sites to gather information that can be used to deliver specialized services to Web surfers.

Cookies are bits of data put on a hard disk when someone visits certain Web sites. The most common use of this data is to make it easier for people to use Web sites that require a username and password. The cookie on the hard disk has the username and password in it, so people don't have to log in to every page that requires that information. Instead, the cookie sends the information to the server and the person can visit the page freely.

Cookies can contain virtually any kind of information, such as the last time a person visited the site, the person's favorite sites, and similar, customizable information. They can be used to track people as they go through a Web site and to help gain statistics about what kinds of pages people like to visit. Although some people view them as invaders of privacy, they can also make the Web a much better place to visit by doing things such as making it easier to conduct electronic commerce.

Web sites can't read cookies that other sites have placed on a hard disk, so if you visit a certain news site, for example, no other site can get the information out of the cookie. Users always have the option of not allowing cookies to be placed on their hard disks—browsers give people this option.

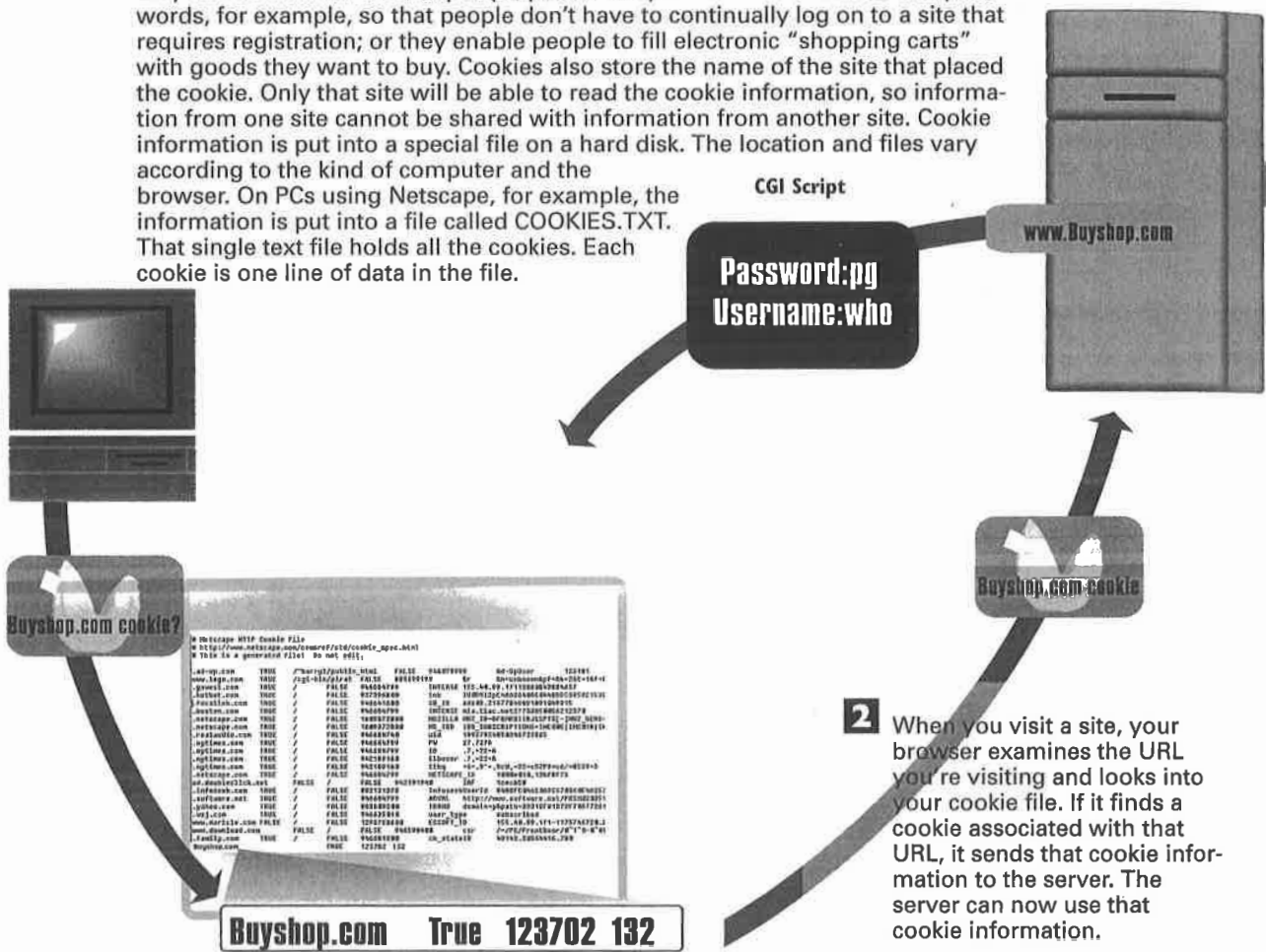
Although cookies can be used to track how people use a Web site, many other methods can be used, as well. In one method, Web server logs are examined in detail. This would make it possible, for example, to identify the most popular pages on the site, the sites people have just visited, how many pages people read in a typical visit, and similar information. Other methods include using software "sniffers" that examine every packet coming into or going out of a Web site. Webmasters can use this tracking information to help create better sites—but they can also use it to assemble demographic information to sell to advertisers.

In fact, Webmasters need to know a good deal of information about how their site is being used, such as the total number of visitors per day, the total number of pages viewed, how people travel through the site, where people come to the site from, and where they go when they leave. The third illustration in this chapter shows the functionality of Web tracking software from a company called Accrue.

To allay people's privacy concerns, a variety of technologies and standards are being developed. They include the Platform for Privacy Preferences (P3P), the Internet Content and Exchange standard (ICE), and the Open Profiling Standard (OPS). Generically, we'll call them Internet passports. These Internet passports let people control which information about themselves they'll allow to be released to Web sites—and how that information can be used. And they let people control what kind of information can be gathered about their surfing and how that information can be used, as well. In general, the more information that people allow to be gathered about them, the more specialized services they'll gain on the Web, such as customized news feeds. We'll look at Internet passports in the second illustration in this chapter.

How Cookies Work

1 Cookies are pieces of data placed on a computer's hard drive by a Web server; they can be used for a variety of purposes. They can store usernames and passwords, for example, so that people don't have to continually log on to a site that requires registration; or they enable people to fill electronic "shopping carts" with goods they want to buy. Cookies also store the name of the site that placed the cookie. Only that site will be able to read the cookie information, so information from one site cannot be shared with information from another site. Cookie information is put into a special file on a hard disk. The location and files vary according to the kind of computer and the browser. On PCs using Netscape, for example, the information is put into a file called COOKIES.TXT. That single text file holds all the cookies. Each cookie is one line of data in the file.



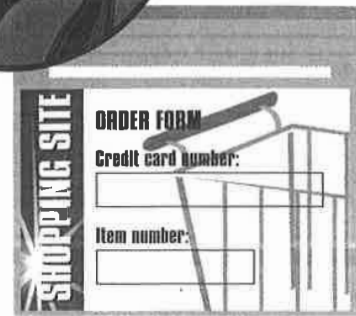
2 When you visit a site, your browser examines the URL you're visiting and looks into your cookie file. If it finds a cookie associated with that URL, it sends that cookie information to the server. The server can now use that cookie information.

3 If no cookie is associated with the URL, the server will place a cookie inside the cookie file. Some sites may first ask a series of questions, such as name and password, and will then place a cookie on the hard disk with that information in it. This is typical of sites that require registration. Commonly, a CGI script on the server takes the information that the user has entered and then writes the cookie onto the hard disk.

4 As you travel through a Web site, more information might need to be put into your cookie. On a site where you can purchase goods online, for example, you might put goods into an electronic shopping cart. Every time you did this, new cookie information would be added, detailing the goods you wanted to buy. When new cookie information is put in, a CGI script deletes the old cookie information and puts in a new cookie. When you leave a site, your cookie information remains on your hard disk so that the site can recognize you the next time you decide to visit—unless the cookie has specifically been written to expire when you leave the site.



5 The server takes actions based on your cookies—for example, displaying your electronic shopping cart. If the site enables you to buy online, it might have asked for your credit card number. For security reasons, that number will not be stored in your cookie. Instead, it will be stored on a secure server. When you decide to buy something, you enter a secure area with your browser. Your cookie then sends an ID to the server that identifies you, and the server then displays your credit card information, enabling you to buy online.

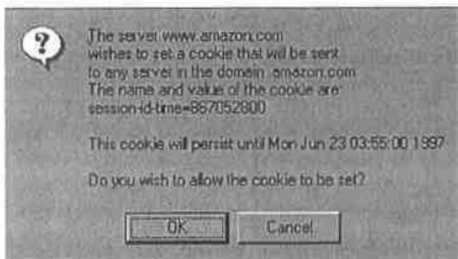


Delete old cookies; put on new cookie

6 After you order something from your electronic shopping cart—or after you decide to delete something out of the shopping cart—a new cookie is put on your hard disk; this one does not include the goods you bought or decided to take out of your shopping cart.



Web Server



7 Because some people don't like cookies to be placed on their hard disk, browsers give people control over whether to accept cookies, to not accept cookies, or to ask each time a cookie is being placed on the hard disk. Pictured is the message you get if you've asked to be told each time a cookie is placed on your hard disk.