CHAPTER 2

TECHNOLOGY INFRASTRUCTURE: THE INTERNET AND THE WORLD WIDE WEB

LEARNING OBJECTIVES

In this chapter, you will learn about:

- The origin, growth, and current structure of the Internet
- How packet-switched networks are combined to form the Internet
- How Internet protocols and Internet addressing work
- The history and use of markup languages on the Web, including SGML, HTML, and XML
- How HTML tags and links work on the World Wide Web
- The differences among internets, intranets, and extranets
- Options for connecting to the Internet, including cost and bandwidth factors
- Internet2 and the Semantic Web

INTRODUCTION

Most people who use the Internet and the Web today do so using a computer. As you will learn in this chapter, the ability to access the Internet generally costs money. The cost of Internet access combined with the cost of owning a computer puts the Web beyond the reach of many people around the world. In 2009, about 70 percent of the U.S. population had regular access to the Internet, but in China (with 1.4 billion residents, it is the most populous country on Earth), fewer than 25 percent of the population did. In the United States, most people access the Internet through a computer. In China, fully half of all Internet access is now through mobile phones or smart phones (nearly 200 million of them), and the proportion is increasing rapidly. Another 500 million people in China use mobile phones without Internet access, indicating great growth potential for China both in total mobile phone use and Internet access through mobile and smart phones.

In India (with a population of 1.2 billion), fewer than 250 million people have mobile phones and fewer than 1 percent of those have reliable Internet access through their phones. Fewer than 5 percent of the Indian population has any Internet access at all. But in 2009, India's telecom companies began a rapid expansion of the infrastructure that will allow them to offer better Internet access to their phone customers. Industry analysts expect that India mobile and smart phone use will soon be growing at annual rates of 15 percent to 20 percent, similar to the recent growth observed in China.

Although the first Internet-capable mobile phones were developed in the late 1990s, a number of technological issues prevented them from being very useful as a way to browse the Internet. Their screens were small and lacked color, they did not have alphanumeric keyboards, their ability to store information was limited, and the networks through which they connected to the Internet were slow and unreliable.

In 2001, Handspring introduced its Treo phones and Research in Motion (RIM) introduced its BlackBerry phones. These mobile phones included small alphanumeric keyboards, significantly larger memory capacities than other phones of the time, and were designed for quick access to e-mail. Nokia was quick to follow with smart phones that had similar features. By 2009, every major phone manufacturer offered a range of smart phones and Internet-capable mobile phones. Although many of these offerings were too expensive for markets in developing countries, some were not. Nokia has been especially effective in developing lower-cost phones specifically for these markets.

Although some Web sites have created pages for their mobile users that are designed to be used without a mouse and that are readable on the relatively small screens of phones, most have not. This can limit the usefulness of mobile phones as tools of electronic commerce. As more online businesses realize that mobile phone users are potential customers, more Web sites will be redesigned to give mobile users a better experience.

In the developed industrial countries, Internet-capable phones are tools of convenience; they provide continual access to e-mail and the Web for busy people who work from multiple locations. In the rest of the world, they are often the only affordable way to access the Internet. The rapid growth expected in the use of Internet-capable phones in parts of the world that have never had reliable access to the Internet and the Web offers the potential for vast increases in international electronic commerce.

THE INTERNET AND THE WORLD WIDE WEB

A **computer network** is any technology that allows people to connect computers to each other. An **internet** (small "i") is a group of computer networks that have been interconnected. In fact, "internet" is short for "interconnected network." One particular internet, which uses a specific set of rules and connects networks all over the world to each other, is called the **Internet** (capital "i"). Networks of computers and the Internet that connects them to each other form the basic technological structure that underlies virtually all electronic commerce. This chapter introduces you to many of the hardware and software technologies that make electronic commerce possible. First, you will learn how the Internet and the World Wide Web work. Then, you will learn about other technologies that support the Internet, the Web, and electronic commerce. In this chapter, you will be introduced to several complex networking technologies. If you are interested in learning more about how computer networks operate, you can consult one of the computer networking books cited in the For Further Study and Research section at the end of this chapter, or you can take courses in data communications and networking.

The part of the Internet known as the **World Wide Web**, or, more simply, the **Web**, is a subset of the computers on the Internet that are connected to one another in a specific way that makes them and their contents easily accessible to each other. The most important thing about the Web is that it includes an easy-to-use standard interface. This interface makes it possible for people who are not computer experts to use the Web to access a variety of Internet resources.

Origins of the Internet

In the early 1960s, the U.S. Department of Defense became concerned about the possible effects of nuclear attack on its computing facilities. The Defense Department realized that the weapons of the future would require powerful computers for coordination and control. The powerful computers of that time were all large mainframe computers.

Defense Department began examining ways to connect these computers to each other and also to connect them to weapons installations distributed all over the world. Employing many of the best communications technology researchers, the Defense Department funded research at leading universities and institutes. The goal of this research was to design a worldwide network that could remain operational, even if parts of the network were destroyed by enemy military action or sabotage. These researchers determined that the best path to accomplishing their goals was to create networks that did not require a central computer to control network operations.

The computer networks that existed at that time used leased telephone company lines for their connections. These telephone company systems established a single connection between sender and receiver for each telephone call, then that connection carried all data along a single path. When a company wanted to connect computers it owned at two different locations, the company placed a telephone call to establish the connection, and then connected one computer to each end of that single connection.

The Defense Department was concerned about the inherent risk of this single-channel method for connecting computers, and its researchers developed a different method of sending information through multiple channels. In this method, files and messages are broken into packets that are labeled electronically with codes for their origins, sequences, and destinations. You will learn more about how packet networks operate later in this chapter.

In 1969, Defense Department researchers in the Advanced Research Projects Agency (ARPA) used this direct connnection network model to connect four computers—one each at the University of California at Los Angeles, SRI International, the University of California at Santa Barbara, and the University of Utah—into a network called the ARPANET. The ARPANET was the earliest of the networks that eventually combined to become what we now call the Internet. Throughout the 1970s and 1980s, many researchers in the academic community connected to the ARPANET and contributed to the technological developments that increased its speed and efficiency. At the same time, researchers at other universities were creating their own networks using similar technologies.

New Uses for the Internet

Although the goals of the Defense Department network were to control weapons systems and transfer research files, other uses for this vast network began to appear in the early 1970s. E-mail was born in 1972 when Ray Tomlinson, a researcher who used the network, wrote a program that could send and receive messages over the network. This new method of communicating became widely used very quickly. The number of network users in the military and education research communities continued to grow. Many of these new participants used the networking technology to transfer files and access computers remotely.

The first e-mail mailing lists also appeared on these military and education research networks. A **mailing list** is an e-mail address that forwards any message it receives to any user who has subscribed to the list. In 1979, a group of students and programmers at Duke

University and the University of North Carolina started **Usenet**, an abbreviation for **User's News Network**. Usenet allows anyone who connects to the network to read and post articles on a variety of subjects. Usenet survives on the Internet today, with more than 1000 different topic areas that are called **newsgroups**. Other researchers even created game-playing software for use on these interconnected networks.

Although the people using these networks were developing many creative applications, use of the networks was limited to those members of the research and academic communities who could access them. Between 1979 and 1989, these network applications were improved and tested by an increasing number of users. The Defense Department's networking software became more widely used in academic and research institutions as these organizations recognized the benefits of having a common communications network. As the number of people in different organizations using these networks increased, security problems were recognized. These problems have continued to become more important, and you will learn more about network security issues in Chapter 10. The explosion of personal computer use during the 1980s also helped more people become comfortable with computers. During the 1980s, other independent networks were developed by academics worldwide (such as Bitnet) and researchers in specific countries other than the United States (such as the United Kingdom's academic research network, Janet). In the late 1980s, these independent academic and research networks from all over the world merged into what we now call the Internet.

Commercial Use of the Internet

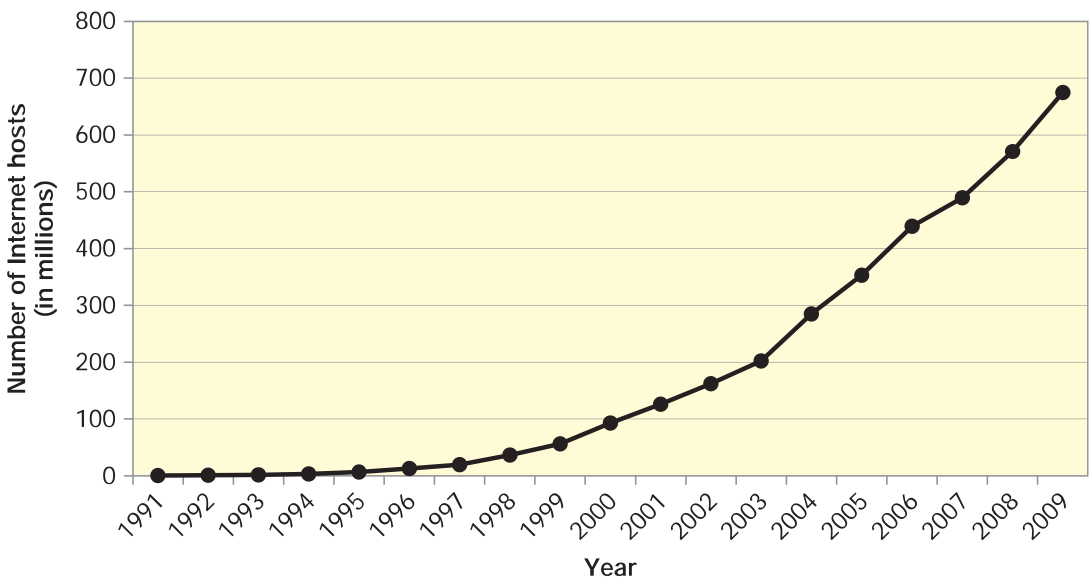
As personal computers became more powerful, affordable, and available during the 1980s, companies increasingly used them to construct their own internal networks. Although these networks included e-mail software that employees could use to send messages to each other, businesses wanted their employees to be able to communicate with people outside their corporate networks. The Defense Department network and most of the academic networks that had teamed up with it were receiving funding from the **National Science Foundation (NSF)**. The NSF prohibited commercial network traffic on its networks, so businesses turned to commercial e-mail service providers to handle their e-mail needs. Larger firms built their own networks that used leased telephone lines to connect field offices to corporate headquarters.

In 1989, the NSF permitted two commercial e-mail services, MCI Mail and CompuServe, to establish limited connections to the Internet for the sole purpose of exchanging e-mail transmissions with users of the Internet. These connections allowed commercial enterprises to send e-mail directly to Internet addresses, and allowed members of the research and education communities on the Internet to send e-mail directly to MCI Mail and CompuServe addresses. The NSF justified this limited commercial use of the Internet as a service that would primarily benefit the Internet's noncommercial users. As the 1990s began, people from all walks of life—not just scientists or academic researchers—started thinking of these networks as the global resource that we now know as the Internet. Although this network of networks had grown from four Defense Department computers in 1969 to more than 300,000 computers on many interconnected networks by 1990, the greatest growth of the Internet was yet to come.

Growth of the Internet

In 1991, the NSF further eased its restrictions on commercial Internet activity and began implementing plans to privatize the Internet. The privatization of the Internet was substantially completed in 1995, when the NSF turned over the operation of the main Internet connections to a group of privately owned companies. The new structure of the Internet was based on four **network access points (NAPs)** located in San Francisco, New York, Chicago, and Washington, D.C., each operated by a separate telecommunications company. As the Internet grew, more companies opened more NAPs in more locations. These companies, known as **network access providers**, sell Internet access rights directly to larger customers and indirectly to smaller firms and individuals through other companies, called **Internet service providers (ISPs)**.

The Internet was a phenomenon that had truly sneaked up on an unsuspecting world. The researchers who had been so involved in the creation and growth of the Internet just accepted it as part of their working environment. However, people outside the research community were largely unaware of the potential offered by a large interconnected set of computer networks. Figure 2-1 shows the consistent and dramatic growth in the number of **Internet hosts**, which are computers directly connected to the Internet.



Source: Internet Software Consortium (http://www.isc.org/) and author's estimates

FIGURE 2-1 Growth of the Internet

In 40 years, the Internet has grown to become one of the most amazing technological and social accomplishments of the last millennium. Millions of people, from elementary schoolers to research scientists, now use this complex, interconnected network of computers. These computers run thousands of different software packages. The computers are located in almost every country of the world. Every year, billions of dollars change hands over the Internet in exchange for all kinds of products and services. All of this activity occurs with no central coordination point or control, which is especially ironic given that the Internet began as a way for the military to maintain control while under attack.

The opening of the Internet to business activity helped to dramatically increase its growth; however, there was another development that worked hand in hand with the commercialization of the Internet to spur its growth. That development was the World Wide Web.

Emergence of the World Wide Web

At a technological level, the Web is nothing more than software that runs on computers that are connected to the Internet. The network traffic generated by Web software is the largest single category of traffic on the Internet today, outpacing e-mail, file transfers, and other data transmission traffic. But the ideas behind the Web developed from innovative ways of thinking about and organizing information storage and retrieval. These ideas go back many years. Two important ideas that became key technological elements of the Web are hypertext and graphical user interfaces.

The Development of Hypertext

In 1945, **Vannevar Bush**, who was director of the U.S. Office of Scientific Research and Development, wrote an article in *The Atlantic Monthly* about ways that scientists could apply the skills they learned during World War II to peacetime activities. The article included a number of visionary ideas about future uses of technology to organize and facilitate efficient access to information. Bush speculated that engineers would eventually build a machine that he called the Memex, a memory extension device that would store all of a person's books, records, letters, and research results on microfilm. Bush's Memex would include mechanical aids, such as microfilm readers and indexes, that would help users quickly and flexibly consult their collected knowledge.

In the 1960s, Ted Nelson described a similar system in which text on one page links to text on other pages. Nelson called his page-linking system **hypertext**. Douglas Engelbart, who also invented the computer mouse, created the first experimental hypertext system on one of the large computers of the 1960s. In 1987, Nelson published *Literary Machines*, a book in which he outlined project Xanadu, a global system for online hypertext publishing and commerce. Nelson used the term *hypertext* to describe a page-linking system that would interconnect related pages of information, regardless of where in the world they were stored.

In 1989, Tim Berners-Lee was trying to improve the laboratory research document-handling procedures for his employer, CERN: European Laboratory for Particle Physics. CERN had been connected to the Internet for two years, but its scientists wanted to find better ways to circulate their scientific papers and data among the high-energy physics research community throughout the world. Berners-Lee proposed a hypertext development project intended to provide this data-sharing functionality.

Over the next two years, Berners-Lee developed the code for a hypertext server program and made it available on the Internet. A **hypertext server** is a computer that stores files written in **Hypertext Markup Language (HTML)**, the language used for the creation of Web pages. The hypertext server is connected through the Internet to other computers that can connect to the hypertext server and read those HTML files. Hypertext servers used on the Web today are usually called **Web servers**. HTML, which Berners-Lee developed from his original hypertext server program, is a language that includes a set of codes (or tags) attached to text. These codes describe the relationships among text elements. For example, HTML includes tags that indicate which text is part of a header element, which text is part of a paragraph element, and which text is part of a numbered list element. One important type of tag is the hypertext link tag. A **hypertext link**, or **hyperlink**, points to another location in the same or another HTML document. The details of HTML and other markup languages are covered later in this chapter.

Graphical Interfaces for Hypertext

Several different types of software are available to read HTML documents, but most people use a Web browser such as Mozilla Firefox or Microsoft Internet Explorer. A **Web browser** is a software interface that lets users read (or browse) HTML documents and move from one HTML document to another through text formatted with hypertext link tags in each file. If the HTML documents are on computers connected to the Internet, you can use a Web browser to move from an HTML document on one computer to an HTML document on any other computer on the Internet.

An HTML document differs from a word-processing document in that it does not specify how a particular text element will appear. For example, you might use word-processing software to create a document heading by setting the heading text font to Arial, its font size to 14 points, and its position to centered. The document displays and prints these exact settings whenever you open the document in that word processor. In contrast, an HTML document simply includes a heading tag with the heading text. Many different browser programs can read an HTML document. Each program recognizes the heading tag and displays the text in whatever manner each program normally displays headings. Different Web browser programs might each display the text differently, but all of them display the text with the characteristics of a heading.

A Web browser presents an HTML document in an easy-to-read format in the browser's graphical user interface. A **graphical user interface (GUI)** is a way of presenting program control functions and program output to users and accepting their input. It uses pictures, icons, and other graphical elements instead of displaying just text. Almost all personal computers today use a GUI such as Microsoft Windows or the Macintosh user interface.

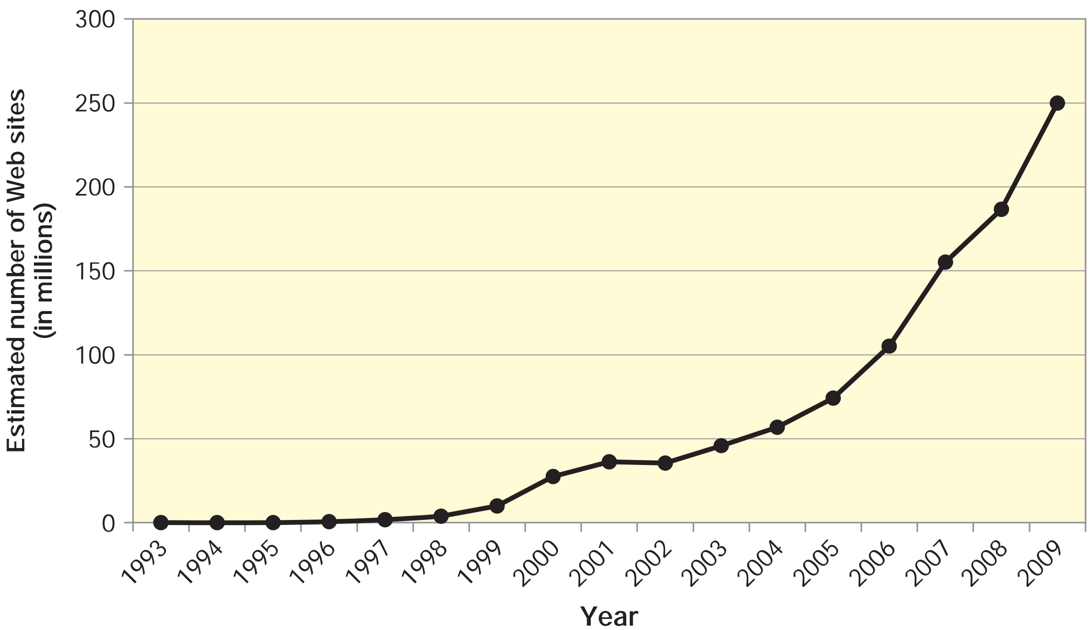
The World Wide Web

Berners-Lee called his system of hyperlinked HTML documents the World Wide Web. The Web caught on quickly in the scientific research community, but few people outside that community had software that could read the HTML documents. In 1993, a group of students led by Marc Andreessen at the University of Illinois wrote Mosaic, the first GUI program that could read HTML and use HTML hyperlinks to navigate from page to page on computers anywhere on the Internet. Mosaic was the first Web browser that became widely available for personal computers, and some Web surfers still use it today.

Programmers quickly realized that a system of pages connected by hypertext links would provide many new Internet users with an easy way to access information on the Internet. Businesses recognized the profit-making potential offered by a worldwide network of easy-to-use computers. In 1994, Andreessen and other members of the University of Illinois Mosaic team joined with James Clark of Silicon Graphics to found Netscape Communications (which is now owned by Time Warner). Its first product, the Netscape Navigator Web browser program based on Mosaic, was an instant success. Netscape became one of the fastest-growing software companies ever. Microsoft created its

Internet Explorer Web browser and entered the market soon after Netscape's success became apparent. Today, Internet Explorer is the most widely used Web browser in the world. Its main competitor, Mozilla Firefox, is a descendant of Netscape Navigator.

The number of Web sites has grown even more rapidly than the Internet itself. The number of Web sites is currently estimated at more than 250 million, and individual Web pages number more than 50 billion because each Web site might include hundreds or even thousands of individual Web pages. Therefore, nobody really knows how many Web pages exist. Figure 2-2 shows the overall rapid growth rate of the Web. Other than a brief consolidation period during the 2001–2002 economic downturn, the Web has grown at a consistently rapid rate.



Adapted from Netcraft Computer Surveys (http://www.netcraft.com) and author's estimates

FIGURE 2-2 Growth of the World Wide Web

In addition to Web pages that are specifically programmed to exist in a permanent form, the Web provides access to customized Web pages that are created in response to a particular user's query. Such Web pages pull their content from databases. For example, if you visit Amazon.com and search for a book about "online business," computers at Amazon.com query their databases of information about books and create a Web page that is a customized response to your search. The Web page that lists your search results never existed before your visit. This store of information that is available though the Web is called the **deep Web**. Researchers, such as those at **Bright Planet**, estimate the number of possible pages in the deep Web to be in the trillions.

As more people gain access to the Web, commercial interest in using the Web to conduct business will continue to increase, and the variety of nonbusiness uses will become even greater. In the rest of this chapter, you will learn how Internet and Web technologies work to enable electronic commerce.

PACKET-SWITCHED NETWORKS

A network of computers that are located close together—for example, in the same building—is called a **local area network (LAN)**. Networks of computers that are connected over greater distances are called **wide area networks (WANs)**.

The early models (dating back to the 1950s) for WANs were the circuits of the local and long-distance telephone companies of the time, because the first early WANs used leased telephone company lines for their connections. A telephone call establishes a single connection path between the caller and receiver. Once that connection is established, data travels along that single path. Telephone company equipment (originally mechanical, now electronic) selects specific telephone lines to connect to one another by closing switches. These switches work like the switches you use to turn lights on and off in your home, except that they open and close much faster, and are controlled by mechanical or electronic devices instead of human hands.

The combination of telephone lines and the closed switches that connect them to each other is called a **circuit**. This circuit forms a single electrical path between caller and receiver. This single path of connected circuits switched into each other is maintained for the entire length of the call. This type of centrally controlled, single-connection model is known as **circuit switching**.

Although circuit switching works well for telephone calls, it does not work as well for sending data across a large WAN or an interconnected network like the Internet. The Internet was designed to be resistant to failure. In a circuit-switched network, a failure in any one of the connected circuits causes the connection to be interrupted and data to be lost. Instead, the Internet uses packet switching to move data between two points. On a **packet-switched** network, files and e-mail messages are broken down into small pieces, called **packets**, that are labeled electronically with their origins, sequences, and destination addresses. Packets travel from computer to computer along the interconnected networks until they reach their destinations. Each packet can take a different path through the interconnected networks, and the packets may arrive out of order. The destination computer collects the packets and reassembles the original file or e-mail message from the pieces in each packet.

Routing Packets

As an individual packet travels from one network to another, the computers through which the packet travels determine the most efficient route for getting the packet to its destination. The most efficient route changes from second to second, depending on how much traffic each computer on the Internet is handling at each moment. The computers that decide how best to forward each packet are called **routing computers**, **router computers**, **routers**, **gateway computers** (because they act as the gateway from a LAN or WAN to the Internet), or **border routers** (because they are located at the border between the organization and the Internet). The programs on router computers that determine the best path on which to send each packet contain rules called **routing algorithms**. The programs apply their routing algorithms to information they have stored in **routing tables** or **configuration tables**. This information includes lists of connections that lead to particular groups of other routers, rules that specify which connections to use first, and rules for handling instances of heavy packet traffic and network congestion.

Individual LANs and WANs can use a variety of different rules and standards for creating packets within their networks. The network devices that move packets from one part of a network to another are called hubs, switches, and bridges. Routers are used to connect networks to other networks. You can take a data communications and networking class to learn more about these network devices and how they work.

When packets leave a network to travel on the Internet, they must be translated into a standard format. Routers usually perform this translation function. As you can see, routers are an important part of the infrastructure of the Internet. When a company or organization becomes part of the Internet, it must connect at least one router to the other routers (owned by other companies or organizations) that make up the Internet. Figure 2-3 is a diagram of a small portion of the Internet that shows its router-based architecture. The figure shows only the routers that connect each organization's WANs and LANs to the Internet, not the other routers that are inside the WANs and LANs or that connect them to each other within the organization.

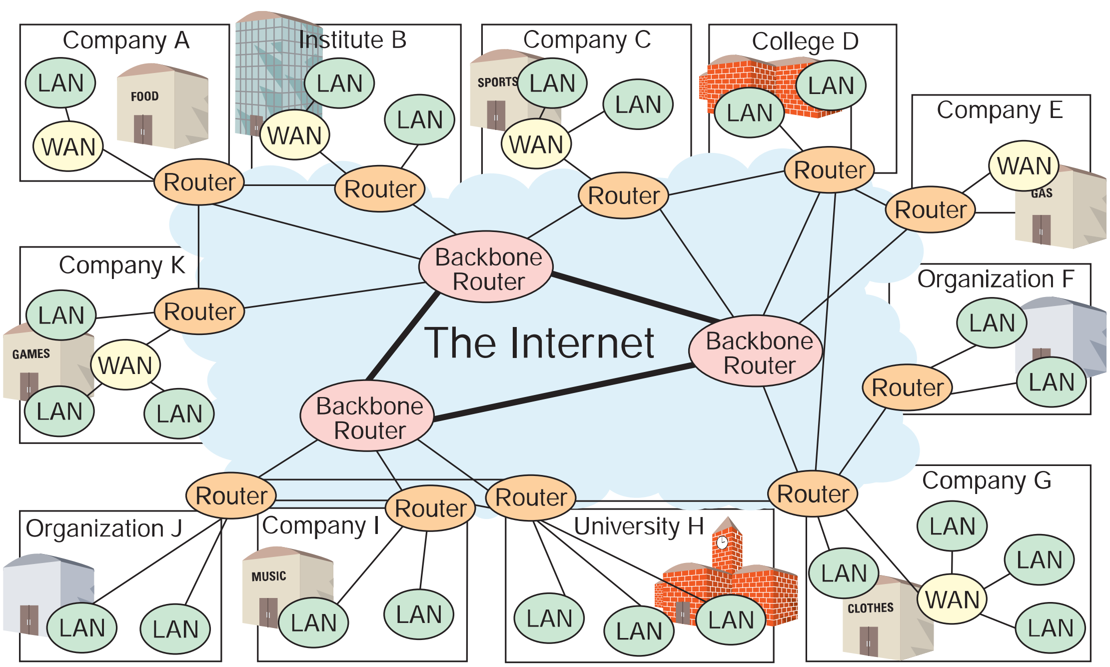


FIGURE 2-3 Router-based architecture of the Internet

The Internet also has routers that handle packet traffic along the Internet's main connecting points. These routers and the telecommunications lines connecting them are collectively referred to as the **Internet backbone**. These routers, sometimes called **backbone routers**, are very large computers that can each handle more than 3 billion packets per second! You can see in Figure 2-3 that a router connected to the Internet always has more than one path to which it can direct a packet. By building in multiple packet paths, the designers of the Internet created a degree of redundancy in the system that allows it to keep moving packets, even if one or more of the routers or connecting lines fails.

INTERNET PROTOCOLS

A **protocol** is a collection of rules for formatting, ordering, and error checking data sent across a network. For example, protocols determine how the sending device indicates that it has finished sending a message and how the receiving device indicates that it has received (or not received) the message. A protocol also includes rules about what is allowed in a transmission and how it is formatted. Computers that communicate with each other must use the same protocol for data transmission. As you learned earlier in this chapter, the first packet-switched network, the ARPANET, connected only a few universities and research centers. Following its inception in 1969, this experimental network grew during the next few years and began using the **Network Control Protocol (NCP)**. In the early days of computing, each computer manufacturer created its own protocol, so computers made by different manufacturers could not be connected to each other. This practice was called **proprietary architecture** or **closed architecture**. NCP was designed so it could be used by any computer manufacturer and was made available to any company that wanted it. This **open architecture** philosophy developed for the evolving ARPANET, which later became the core of the Internet, included the use of a common protocol for all computers connected to the Internet and four key rules for message handling:

- Independent networks should not require any internal changes to be connected to the network.
- Packets that do not arrive at their destinations must be retransmitted from their source network.
- Router computers act as receive-and-forward devices; they do not retain information about the packets that they handle.
- No global control exists over the network.

The open architecture approach has contributed to the success of the Internet because computers manufactured by different companies (Apple, Dell, Hewlett-Packard, Sun, etc.) can be interconnected. The ARPANET and its successor, the Internet, use routers to isolate each LAN or WAN from the other networks to which they are connected. Each LAN or WAN can use its own set of protocols for packet traffic within the LAN or WAN, but must use a router (or similar device) to move packets onto the Internet in its standard format (or protocol). Following these simple rules makes the connections between the interconnected networks operate effectively.

TCP/IP

The Internet uses two main protocols: the **Transmission Control Protocol (TCP)** and the **Internet Protocol (IP)**. Developed by Internet pioneers Vinton Cerf and Robert Kahn, these protocols are the rules that govern how data moves through the Internet and how network connections are established and terminated. The acronym **TCP/IP** is commonly used to refer to the two protocols.

The TCP controls the disassembly of a message or a file into packets before it is transmitted over the Internet, and it controls the reassembly of those packets into their original formats when they reach their destinations. The IP specifies the addressing details for each packet, labeling each with the packet's origination and destination addresses. Soon after the new TCP/IP protocol set was developed, it replaced the NCP that ARPANET originally used.

In addition to its Internet function, TCP/IP is used today in many LANs. The TCP/IP protocol is provided in most personal computer operating systems commonly used today, including Linux, Macintosh, Microsoft Windows, and UNIX.

IP Addressing

The version of IP that has been in use for the past 20 years on the Internet is **Internet Protocol version 4 (IPv4)**. It uses a 32-bit number to identify the computers connected to the Internet. This address is called an **IP address**. Computers do all of their internal calculations using a **base 2** (or **binary**) number system in which each digit is either a 0 or a 1, corresponding to a condition of either off or on. IPv4 uses a 32-bit binary number that allows for more than 4 billion different addresses ($2^{32} = 4{,}294{,}967{,}296$).

When a router breaks a message into packets before sending it onto the Internet, the router marks each packet with both the source IP address and the destination IP address of the message. To make them easier to read, IP numbers (addresses) appear as four numbers separated by periods. This notation system is called **dotted decimal** notation. An IPv4 address is a 32-bit number, so each of the four numbers is an 8-bit number ($4 \times 8 = 32$). In most computer applications, an 8-bit number is called a **byte**; however, in networking applications, an 8-bit number is often called an **octet**. In binary, an octet can have values from 00000000 to 11111111; the decimal equivalents of these binary numbers are 0 and 255, respectively.

Because each of the four parts of a dotted decimal number can range from 0 to 255, IP addresses range from 0.0.0.0 (written in binary as 32 zeros) to 255.255.255.255 (written in binary as 32 ones). Although some people find dotted decimal notation to be confusing at first, most do agree that writing, reading, and remembering a computer's address as 216.115.108.245 is easier than 11011000011100110110110011110101, or its full decimal equivalent, which is 3,631,433,189.

Today, IP addresses are assigned by three not-for-profit organizations: the **American Registry for Internet Numbers (ARIN)**, the **Reséaux IP Européens (RIPE)**, and the **Asia-Pacific Network Information Center (APNIC)**. These registries assign and manage IP addresses for various parts of the world: ARIN for North America, South America, the Caribbean, and sub-Saharan Africa; RIPE for Europe, the Middle East, and the rest of Africa; and APNIC for countries in the Asia-Pacific area. These organizations took over IP address management tasks from the Internet Assigned Numbers Authority (IANA), which performed them under contract with the U.S. government when the Internet was an experimental research project.

You can use the **ARIN Whois** page at the ARIN Web site to search the IP addresses owned by organizations in North America. You can enter an organization name into the search box on the page, then click the Search WHOIS button, and the Whois server returns a list of the IP addresses owned by that organization. For example, performing a search on the word *Carnegie* displays the IP address blocks owned by Carnegie Bank, Carnegie Mellon University, and a number of other organizations whose names begin with Carnegie. You can also enter an IP address and find out who owns that IP address. If you enter "3.0.0.0" (without the quotation marks), you will find that General Electric owns the entire block of IP addresses from 3.0.0.0 to 3.255.255.255. General Electric can use these addresses, which number approximately 16.7 million, for its own computers, or it can lease them to other companies or individuals to whom it provides Internet access services.

In the early days of the Internet, the 4 billion addresses provided by the IPv4 rules certainly seemed to be more addresses than an experimental research network would ever need. However, about 2 billion of those addresses today are either in use or unavailable for use because of the way blocks of addresses were assigned to organizations. The new kinds of devices on the Internet's many networks, such as wireless personal digital assistants and smart phones, promise to keep demand high for IP addresses.

Network engineers have devised a number of stopgap techniques to stretch the supply of IP addresses. One of the most popular techniques is **subnetting**, which is the use of reserved private IP addresses within LANs and WANs to provide additional address space. **Private IP addresses** are a series of IP numbers that are not permitted on packets that travel on the Internet. In subnetting, a computer called a **Network Address Translation (NAT) device** converts those private IP addresses into normal IP addresses when it forwards packets from those computers to the Internet.

The **Internet Engineering Task Force (IETF)** worked on several new protocols that could solve the limited addressing capacity of IPv4, and in 1997, approved **Internet Protocol version 6 (IPv6)** as the protocol that will replace IPv4. The new IP is being implemented gradually because the two protocols are not directly compatible. The process of switching over to IPv6 will take at least another 10 years; however, network engineers have devised ways to run both protocols together on interconnected networks. The major advantage of IPv6 is that it uses a 128-bit number for addresses instead of the 32-bit number used in IPv4. The number of available addresses in IPv6 (2^{128}) is 34 followed by 37 zeros—billions of times larger than the address space of IPv4. The new IP also changes the format of the packet itself. Improvements in networking technologies over the past 20 years have made many of the fields in the IPv4 packet unnecessary. IPv6 eliminates those fields and adds fields for security and other optional information.

IPv6 has a shorthand notation system for expressing addresses, similar to the IPv4 dotted decimal notation system. However, because the IPv6 address space is much larger, its notation system is more complex. The IPv6 notation uses eight groups of 16 bits ($8 \times 16 = 128$). Each group is expressed as four hexadecimal digits and the groups are separated by colons; thus, the notation system is called **colon hexadecimal** or **colon hex**. A **hexadecimal (base 16)** numbering system uses 16 characters (0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a, b, c, d, e, and f). An example of an IPv6 address expressed in this notation is: CD18:0000:0000:AF23:0000:FF9E:61B2:884D. To save space, the zeros can be omitted, which reduces this address to: CD18:::AF23::FF9E:61B2:884D.

Domain Names

The founders of the Internet were concerned that users might find the dotted decimal notation difficult to remember. To make the numbering system easier to use, they created an alternative addressing method that uses words. In this system, an address such as www.course.com is called a domain name. **Domain names** are sets of words that are assigned to specific IP addresses. Domain names can contain two or more word groups separated by periods. The rightmost part of a domain name is the most general. Each part of the domain name becomes more specific as you move to the left.

For example, the domain name www.sandiego.edu contains three parts separated by periods. Beginning at the right, the name "edu" indicates that the computer belongs to an educational institution. The institution, University of San Diego, is identified by the name "sandiego." The "www" indicates that the computer is running software that makes it a part of the World Wide Web. Most, but not all, Web addresses follow this "www" naming convention. For example, the group of computers that operate the **Yahoo! Games** service is named games.yahoo.com.

The rightmost part of a domain name is called a **top-level domain (TLD)**. For many years, these domains have included a group of generic domains—such as .edu, .com, and .org—and a set of country domains. Since 1998, the **Internet Corporation for Assigned Names and Numbers (ICANN)** has had the responsibility of managing domain names and coordinating them with the IP address registrars. ICANN is also responsible for setting standards for the router computers that make up the Internet. Since taking over these responsibilities, ICANN has added a number of new TLDs. Some of these TLDs are **generic top-level domains (gTLDs)**, which are available to specified categories of users. ICANN is itself responsible for the maintenance of gTLDs.Other new domains are **sponsored top-level domains** (**sTLDs**) which are TLDs for which an organization other than ICANN is responsible. The sponsor of a specific sTLD must be a recognized institution that has expertise regarding and is familiar with the community that uses the sTLD. For example, the .aero sTLD is sponsored by SITA, an air transport industry association that has expertise in and is familiar with airlines, airports, and the aerospace industry. Individual countries are permitted to maintain their own TLDs, which their residents can use alone or in combination with other TLDs. For example, the URL of the University of Queensland in Brisbane, Australia is www.uq.edu.au, which combines .edu with .au to indicate that it is an educational institution in Australia. Figure 2-4 presents a list of some commonly used TLDs, the general TLDs added since 2000, and some of the more frequently used country TLDs.

TLD	Use
.com	U.S. commercial
.edu	Four-year educational institution
.gov	U.S. federal government
.mil	U.S. military
.net	U.S. general use
.org	U.S. not-for-profit organization
.us	U.S. general use
.asia	Companies, individuals, and organizations based in Asian–Pacific regions
.biz	Businesses
.info	General use
.name	Individual persons
.pro	Licensed professionals (such as accountants, lawyers, physicians)
.au	Australia
.ca	Canada
.de	Germany
.fi	Finland
.fr	France
.jp	Japan
.se	Sweden
.uk	United Kingdom

Source: Internet Assigned Numbers Authority Root Zone Database, http://www.iana.org/domains/root/db/

FIGURE 2-4 Commonly used domain names

Although these new domain names were chosen after much deliberation and consideration of more than 100 possible new names, many people were highly critical of the selections (see, for example, the **ICANNWatch** Web site). ICANN came under additional fire for acting in ways that many people thought violated the democratic principles on which the organization was founded. You can learn more about these issues on the Web sites of the **Internet Governance Project** and the **Convergence Center**, both at Syracuse University. Increases in the number of TLDs can make it more difficult for companies to protect their corporate and product brand names. You will learn more about these issues in Chapter 7.

Web Page Request and Delivery Protocols

The Web is software that runs on computers that are connected to each other through the Internet. **Web client computers** run software called **Web client software** or **Web browser software**. Examples of popular Web browser software include Google Chrome, Microsoft Internet Explorer, and Mozilla Firefox. Web browser software sends requests for Web page files to other computers, which are called Web servers. A Web server computer runs software called **Web server software**. Web server software receives requests from many different Web clients and responds by sending files back to those Web client computers. Each Web client computer's Web client software renders those files into a Web page. Thus, the purpose of a Web server is to respond to requests for Web pages from Web clients. This combination of client computers running Web client software and server computers running Web server software is an example of a **client/server architecture.**

The set of rules for delivering Web page files over the Internet is in a protocol called the **Hypertext Transfer Protocol (HTTP)**, which was developed by Tim Berners-Lee in 1991. When a user types a domain name (for example, www.yahoo.com) into a Web browser's address bar, the browser sends an HTTP-formatted message to a Web server computer at Yahoo! that stores Web page files. The Web server computer at Yahoo! then responds by sending a set of files (one for the Web page and one for each graphic object, sound, or video clip included on the page) back to the client computer. These files are sent within a message that is HTTP formatted.

To initiate a Web page request using a Web browser, the user types the name of the protocol, followed by the characters "//:" before the domain name. Thus, a user would type http://www.yahoo.com to go to the Yahoo! Web site. Most Web browsers today automatically insert the http:// if the user does not include it. The combination of the protocol name and the domain name is called a **Uniform Resource Locator (URL)** because it lets the user locate a resource (the Web page) on another computer (the Web server).

Electronic Mail Protocols

Electronic mail, or **e-mail**, that is sent across the Internet must also be formatted according to a common set of rules. Most organizations use a client/server structure to handle e-mail. The organization has a computer called an **e-mail server** that is devoted to handling e-mail. Software running on the e-mail server stores and forwards e-mail messages. People in the organization might use a variety of programs, called **e-mail client software**, to read and send e-mail. These programs include **Microsoft Outlook**, **MozillaThunderbird**, **Pegasus Mail**, and many others. The e-mail client software communicates with the e-mail server software on the e-mail server computer to send and receive e-mail messages.

Many people also use e-mail on their computers at home. In most cases, the e-mail servers that handle their messages are operated by the companies that provide their connections to the Internet. An increasing number of people use e-mail services that are offered by Web sites such as **Yahoo! Mail**, Microsoft's **Hotmail**, or Google's **Gmail**. In these cases, the e-mail servers and the e-mail clients are operated by the owners of the Web sites. The individual users only see the e-mail client software (and not the e-mail server software) in their Web browsers when they log on to the Web mail service.

With so many different e-mail client and server software choices, standardization and rules are very important. If e-mail messages did not follow standard rules, an e-mail message created by a person using one e-mail client program could not be read by a person using a different e-mail client program. As you have already learned in this chapter, rules for computer data transmission are called protocols.

SMTP and POP are two common protocols used for sending and retrieving e-mail. **Simple Mail Transfer Protocol (SMTP)** specifies the format of a mail message and describes how mail is to be administered on the e-mail server and transmitted on the Internet. An e-mail client program running on a user's computer can request mail from the organization's e-mail server using the **Post Office Protocol (POP)**. A POP message can tell the e-mail server to send mail to the user's computer and delete it from the e-mail server; send mail to the user's computer and not delete it; or simply ask whether new mail has arrived. The POP provides support for **Multipurpose Internet Mail Extensions (MIME)**, which is a set of rules for handling binary files, such as word-processing documents, spreadsheets, photos, or sound clips that are attached to e-mail messages.

The **Interactive Mail Access Protocol (IMAP)** is a newer e-mail protocol that performs the same basic functions as POP, but includes additional features. For example, IMAP can instruct the e-mail server to send only selected e-mail messages to the client instead of all messages. IMAP also allows the user to view only the header and the e-mail sender's name before deciding to download the entire message. POP requires users to download e-mail messages to their computers before they can search, read, forward, delete, or reply to those messages. IMAP lets users create and manipulate e-mail folders (also called mailboxes) and individual e-mail messages while the messages are still on the e-mail server; that is, the user does not need to download e-mail before working with it.

The tools that IMAP provides are important to the large number of people who access their e-mail from different computers at different times. IMAP lets users manipulate and store their e-mail on the e-mail server and access it from any computer. The main drawback to IMAP is that users' e-mail messages are stored on the e-mail server. As the number of users increases, the size of the e-mail server's disk drives must also increase. In general, server computers use faster (and thus, more expensive) disk drives than desktop computers. Therefore, it is more expensive to provide disk storage space for large quantities of e-mail on a server computer than to provide that same disk space on users' desktop computers. As the price of all disk storage continues to decrease, these cost concerns become less important. You can learn more about IMAP at the University of Washington's **IMAP Connection** Web site.

Unsolicited Commercial E-Mail (UCE, Spam)

Spam, also known as **unsolicited commercial e-mail (UCE)** or **bulk mail**, is electronic junk mail and can include solicitations, advertisements, or e-mail chain letters. The origin of the term spam is generally believed to have come from a song performed by the British comedy troupe, Monty Python, about Hormel's canned meat product, SPAM. In the song, an increasing number of people join in repeating the song's chorus: "Spam spam spam spam, spam spam spam spam, lovely spam, wonderful spam . . ." Just as in the song, e-mail spam is a tiresome repetition of meaningless text that eventually drowns out any other attempt at communication.

Besides wasting people's time and their computer disk space, spam can consume large amounts of Internet capacity. If one person sends a useless e-mail to a million people, that unsolicited mail consumes Internet resources for a few moments that would otherwise be available to other users. Once merely an annoyance, spam has become a major problem for companies. In addition to consuming bandwidth on company networks and space on e-mail servers, spam distracts employees who are trying to do their jobs and requires them to spend time deleting the unwanted messages. A considerable number of spam messages include content that can be offensive to recipients. Some employers worry that their employees might sue them, arguing that the offensive spam they receive while working contributes to a hostile work environment, which can be grounds for harassment allegations. Spam costs businesses more than $40 billion per year in lost productivity and the expenses of dealing with it. You will learn more about spam and how to deal with it in Chapter 8.

MARKUP LANGUAGES AND THE WEB

Web pages can include many elements, such as graphics, photographs, sound clips, and even small programs that run in the Web browser. Each of these elements is stored on the Web server as a separate file. The most important parts of a Web page, however, are the structure of the page and the text that makes up the main part of the page. The page structure and text are stored in a text file that is formatted, or marked up, using a text markup language. A **text markup language** specifies a set of tags that are inserted into the text. These **markup tags**, also called **tags**, provide formatting instructions that Web client software can understand. The Web client software uses those instructions as it renders the text and page elements contained in the other files into the Web page that appears on the screen of the client computer.

The markup language most commonly used on the Web is HTML, which is a subset of a much older and far more complex text markup language called **Standard Generalized Markup Language (SGML)**. Figure 2-5 shows how HTML, XML, and XHTML have descended from the original SGML specification. SGML was used for many years by the publishing industry to create documents that needed to be printed in various formats and that were revised frequently. In addition to its role as a markup language, SGML is a **metalanguage**, which is a language that can be used to define other languages. Another markup language that was derived from SGML for use on the Web is **Extensible Markup Language (XML)**, which is increasingly used to mark up information that companies share with each other over the Internet. The X in XML comes from the word extensible; you might see the word extensible shown as eXtensible. XML is also a meta language because users can create their own markup elements that extend the usefulness of XML (which is why it is called an "extensible" language).

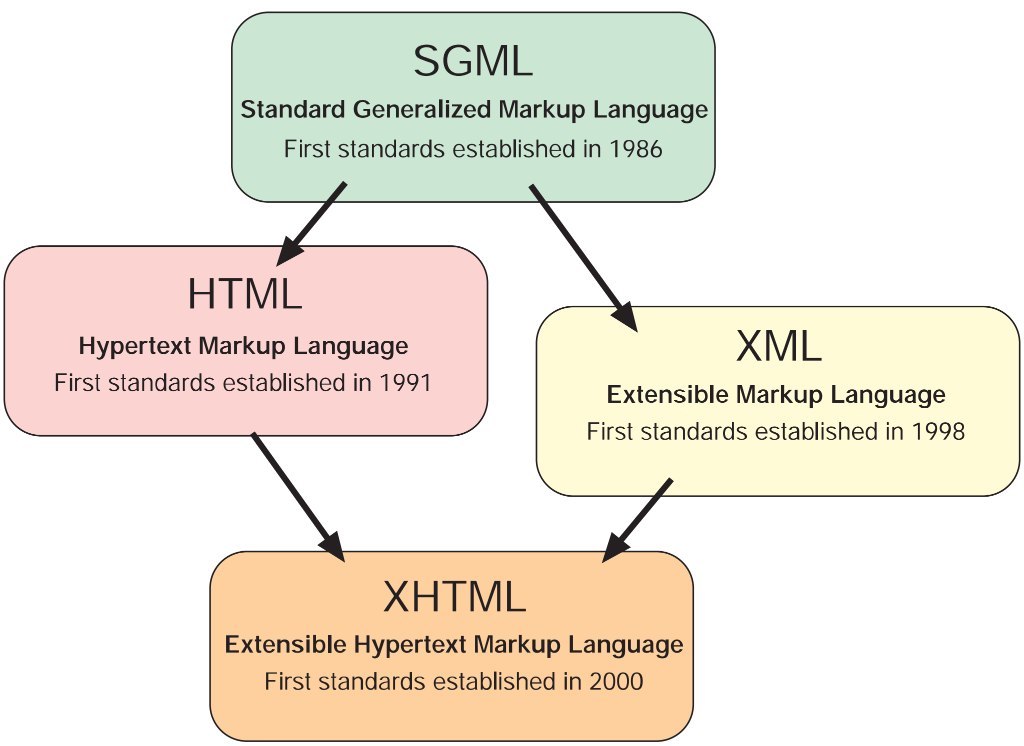


FIGURE 2-5 Development of markup languages

The **World Wide Web Consortium (W3C)**, a not-for-profit group that maintains standards for the Web, presented its first draft form of XML in 1996; the **W3C** issued its first formal version recommendation in 1998. Thus, it is a much newer markup language than HTML. In 2000, the W3C released the first version of a recommendation for a new markup language called **Extensible Hypertext Markup Language (XHTML)**, which is a reformulation of HTML version 4.0 as an XML application. The Online Companion includes a link to the **W3C XHTML Version 1.0 Specification**.

Markup Languages

Since the 1960s, publishers have used markup languages to create documents that can be formatted once, stored electronically, and then printed many times in various layouts that each interpret the formatting differently. U.S. Department of Defense contractors also used early markup languages to create manuals and parts lists for weapons systems. These documents contained many information elements that were often reprinted in different versions and formats. Using electronic document storage and programs that could interpret the formats to produce different layouts saved a tremendous amount of retyping time and cost.

A **Generalized Markup Language (GML)** emerged from these early efforts to create standard formatting styles for electronic documents. In 1986, after many elements of the standard had been in use for years, the **International Organization for Standardization (ISO)** adopted a version of GML called SGML. SGML offers a system of marking up documents that is independent of any software application. Many organizations, such as the Association of American Publishers, Hewlett-Packard, and Kodak, use SGML because they have complex document-management requirements.

SGML is nonproprietary and platform independent and offers user-defined tags. However, it is not well suited to certain tasks, such as the rapid development of Web pages. SGML is costly to set up and maintain, requires the use of expensive software tools, and is hard to learn. Creating document-type definitions in SGML can be expensive and time consuming.

Hypertext Markup Language

HTML includes tags that define the format and style of text elements in an electronic document. HTML also has tags that can create relationships among text elements within one document or among several documents. The text elements that are related to each other are called **hypertext elements**.

HTML is easier to learn and use than SGML. HTML is the prevalent markup language used to create documents on the Web today. The early versions of HTML let Web page designers create text-based electronic documents with headings, title bar titles, bullets, lines, and ordered lists. As the use of HTML and the Web itself grew, HTML creator Berners-Lee turned over the job of maintaining versions of HTML to the W3C. Later versions of HTML included tags for tables, frames, and other features that helped Web designers create more complex page layouts. The W3C maintains detailed information about HTML versions and related topics on its **W3C HTML Working Group** page.

The process for approval of new HTML features takes a long time, so Web browser software developers created some features, called **HTML extensions**, that would only work in their browsers. At various times during the history of HTML, both Microsoft and Netscape enabled their Web browsers to use these HTML extension tags before those tags were approved by the W3C. In some cases, these tags were enabled in one browser and not the other. In other cases, the tags used were never approved by the W3C or were approved in a different form than the one implemented in the Web browser software. Web page designers who wanted to use the latest available tags were often frustrated by this inconsistency. Many of these Web designers had to create separate sets of Web pages for the different types of browsers, which was inefficient and expensive. Most of these tag difference issues were resolved when the W3C issued the specification for HTML version 4.0 in 1997, although enough of them remained to cause regular problems for Web designers.

After HTML 4.0 was finalized in 1999, development on new versions of HTML slowed to a snail's pace. Browser developers worked on adding new features to their software and the W3C directed its efforts to other matters. In 2007, three browser developers (Apple, Opera, and the Mozilla Foundation) began working on an updated version of HTML that would include features such as audio and video within the markup language itself. Audio and video in Web pages have always required the use of add-on software. You can learn more about the current working draft of HTML version 5.0 by visiting the **W3C HTML 5** page.

HTML Tags

An HTML document contains document text and elements. The tags in an HTML document are interpreted by the Web browser and used by it to format the display of the text enclosed by the tags. In HTML, the tags are enclosed in angle brackets (<>). Most HTML tags have an **opening tag** and a **closing tag** that format the text between them. The closing tag is preceded by a slash within the angle brackets (</>). The general form of an HTML element is:

```
<tagname properties>Displayed information affected by tag</tagname>
```

Two good examples of HTML tag pairs are the strong character-formatting tags and the emphasis character-formatting tags. For example, a Web browser reading the following line of text:

```
<strong>A Review of the Book <em>HTML Is Fun!</em></strong>
```

would recognize the and tags as instructions to display the entire line of text in bold and the and tags as instructions to display the text enclosed by those tags in italics. The Web browser would display the text as:

A Review of the Book *HTML Is Fun!*

Some Web browsers allow the user to customize the interpretations of the tags, so that different Web browsers might display the tagged text differently. For example, one Web browser might display text enclosed by strong tags in a blue color instead of displaying the text as bold. Tags are generally written in lowercase letters, however older versions of HTML allowed the use of either case and you might still see Web pages that include uppercase (or mixed case) HTML tags. Although most tags are two-sided (they use both an opening and a closing tag), some are not. Tags that only require opening tags are known as one-sided tags. The tag that creates a line break (</br>) is a common one-sided tag. Some tags, such as the paragraph tag (<p> . . .</p>), are two-sided tags for which the closing tag is optional. Designers sometimes omit the optional closing tags, but this practice is poor markup style.

In a two-sided tag set, the closing tag position is very important. For example, if you were to omit the closing bold tag in the preceding example, any text that followed the line would be bolded. Sometimes an opening tag contains one or more property modifiers that further refine how the tag operates. A tag's property may modify a text display, or it may designate where to find a graphic element. Figure 2-6 (on the next page) shows some sample text marked up with HTML tags and Figure 2-7 (on page 76) shows this text as it appears in a Web browser. The tags in these two figures are among the most common HTML tags in use today on the Web.

```
<html>

    <head>

        <title>HTML Tag Examples</title>

    </head>

    <body>

    <h1>This text is set in Heading one tags</h1>
    <h2>This text is set in Heading two tags</h2>
    <h3>This text is set in Heading three tags</h1>

    <p>
    This text is set within Paragraph tags. It will appear as one paragraph: the
    text will wrap at the end of each line that is rendered in the Web browser no
    matter where the typed text ends. The text inside Paragraph tags is rendered
    without regard to extra spaces typed in the text, such as these:
    Character formatting can also be applied within Paragraph tags. For
    example, <strong>the Strong tags will cause this text to appear bolded in
    most Web browsers</strong> and <em>the emphasis tags will cause this to
    appear italicized in most Web browsers</em>.
    </p>

    <pre>
    HTML includes tags that instruct the Web browser to render the text
    Exactly      the      way      it      is      typed,
    as in this example.
    </pre>

    <p>
    HTML includes tags that instruct the Web browser to place text in bulleted or
    numbered lists:
    </p>

    <ul>
        <li>Bulleted list item one</li>
        <li>Bulleted list item two</li>
        <li>Bulleted list item three</li>
    </ul>

    <ol>
        <li>Numbered list item one</li>
        <li>Numbered list item two</li>
        <li>Numbered list item three</li>
    </ol>

    <p>
    The most important tag in HTML is the Anchor Hypertext Reference tag,
    which is the tag that provides a link to another Web page (or another location
    in the same Web page). For example, the underlined text
    <a href="http://www.w3c.org/">World Wide Web Consortium</a>
    is a link to the not-for-profit organization that develops Web technologies.
    </p>

    </body>

</html>
```

FIGURE 2-6 Text marked up with HTML tags

This text is set in Heading one tags

This text is set in Heading two tags

This text is set in Heading three tags

This text is set within Paragraph tags. It will appear as one paragraph: the text will wrap at the end of each line that is rendered in the Web browser no matter where the typed text ends. The text inside Paragraph tags is rendered without regard to extra spaces typed in the text, such as these: Character formatting can also be applied within Paragraph tags. For example, **the Strong tags will cause this text to appear bolded in most Web browsers** and *the emphasis tags will cause this to appear italicized in most Web browsers*.

```
HTML includes tags that instruct the Web browser to render the text
Exactly        the       way         it        is          typed,
as in this example.
```

HTML includes tags that instruct the Web browser to place text in bulleted or numbered lists:

- Bulleted list item one
- Bulleted list item two
- Bulleted list item three

1. Numbered list item one
2. Numbered list item two
3. Numbered list item three

The most important tag in HTML is the Anchor Hypertext Reference tag, which is the tag that provides a link to another Web page (or another location in the same Web page). For example, the underlined text World Wide Web Consortium is a link to the not-for-profit organization that develops Web technologies.

FIGURE 2-7 Text marked up with HTML tags as it appears in a Web browser

Other frequently used HTML tags (not shown in the figures) let Web designers include graphics on Web pages and format text in the form of tables. The text and HTML tags that form a Web page can be viewed when the page is open in a Web browser by clicking the Page button and selecting View Source in Internet Explorer or by selecting View, Page Source from the command menu in Firefox. A number of good Web sources (such as the **W3C Getting Started with HTML** page) and textbooks are available that describe HTML tags and their uses, and you may wish to consult them for an in-depth look at HTML.

HTML Links

The Web organizes interlinked pages of information residing on sites around the world. Hyperlinks on Web pages form a "web" of those pages. A user can traverse the interwoven pages by clicking hyperlinked text on one page to move to another page in the web of pages. Users can read Web pages in serial order or in whatever order they prefer by following hyperlinks. Figure 2-8 illustrates the differences between reading a paper catalog in a linear way and reading a hypertext catalog in a nonlinear way.

FIGURE 2-8 Linear vs. nonlinear paths through documents

Web sites can use links to direct customers to pages on the company's Web server. The way links lead customers through pages can affect the usefulness of the site and can play a major role in shaping customers' impressions of the company. Two commonly used link structures are linear and hierarchical. A **linear hyperlink structure** resembles conventional paper documents in that the reader begins on the first page and clicks the Next button to move to the next page in a serial fashion. This structure works well when customers fill out forms prior to a purchase or other agreement. In this case, the customer reads and responds to page one, and then moves on to the next page. This process continues until the entire form is completed. The only Web page navigation choices the user typically has are Back and Continue.

Another link arrangement is called a hierarchical structure. In a **hierarchical hyperlink structure**, the Web user opens an introductory page called a **home page** or **start page**. This page contains one or more links to other pages, and those pages, in turn, link to other pages. This hierarchical arrangement resembles an inverted tree in which the root is at the top and the branches are below it. Hierarchical structures are good for leading customers from general topics or products to specific product models and quantities. A company's home page might contain links to help, company history, company officers, order processing, frequently asked questions, and product catalogs. Many sites that use a hierarchical structure include a page on the Web site that contains a map or outline listing of the Web pages in their hierarchical order. This page is called a **site map**. Of course, hybrid designs that combine linear and hierarchical structures are also possible. Figure 2-9 illustrates these three common Web page organization structures.

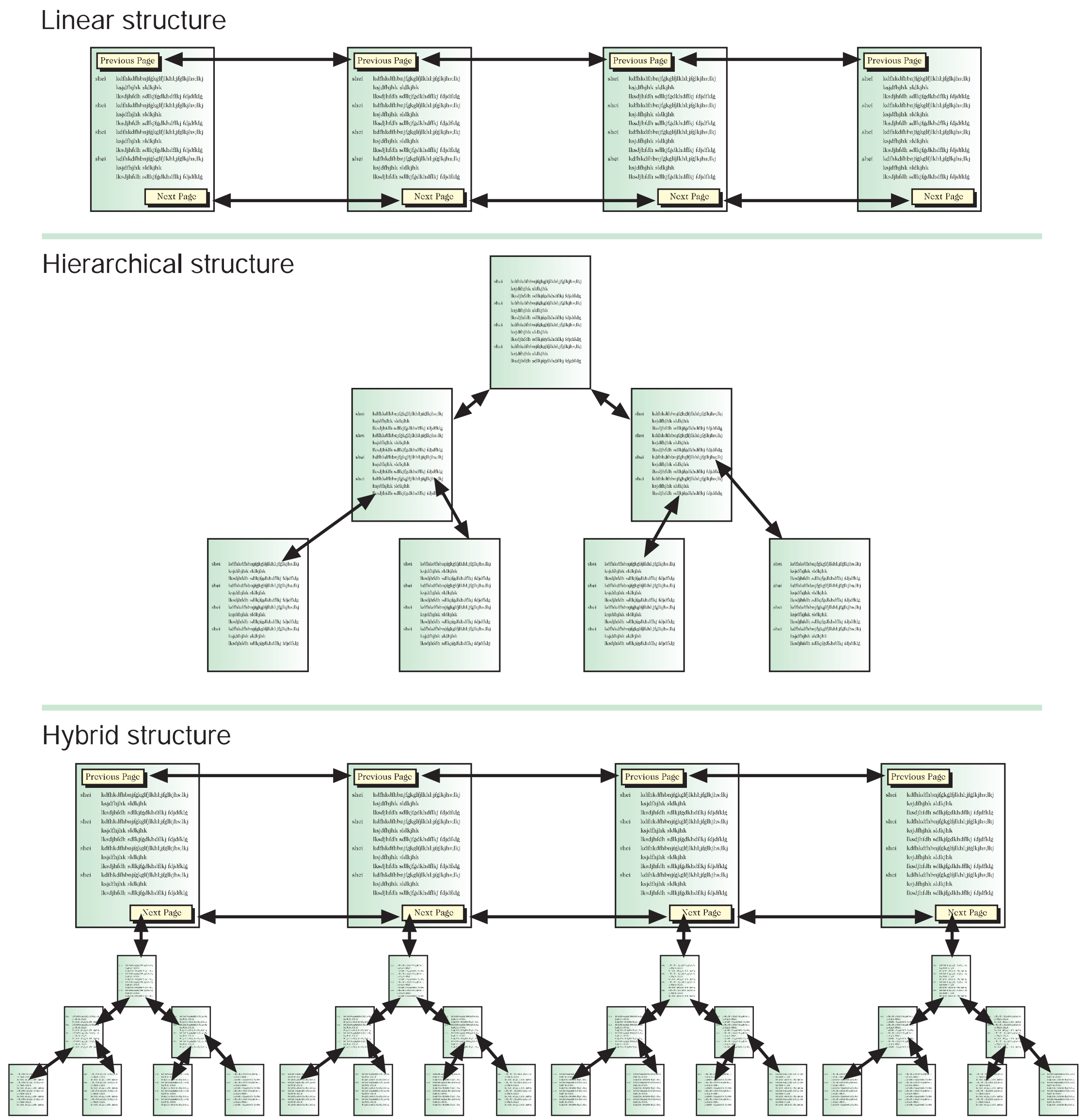


FIGURE 2-9 Three common Web page organization structures

In HTML, hyperlinks are created using the HTML **anchor tag**. Whether you are linking to text within the same document or to a document on a distant computer, the anchor tag has the same basic form:

```
<a href="address">Visible link text</a>
```

Anchor tags have opening and closing tags. The opening tag has a hypertext reference (HREF) property, which specifies the remote or local document's address. Clicking the text following the opening link transfers control to the HREF address, wherever that happens to be. A person creating an electronic résumé on the Web might want to make a university's name and address under the Education heading a hyperlink instead of plain text. Anyone viewing the résumé can click the link, which leads the reader to the university's home page. The following example shows the HTML code to create a hyperlink to another Web server:

```
<a href="http://www.gsu.edu">Georgia State University </a>
```

Similarly, the résumé could include a local link to another part of the same document with the following marked up text:

```
<a href="#references">References are found here</a>
```

In both of these examples, the text between the anchors appears on the Web page as a hyperlink. Most browsers display the link in blue and underline it. In most browser software, the action of moving the mouse pointer over a hyperlink causes the mouse pointer to change from an arrow to a pointing hand.

Scripting Languages and Style Sheets

Versions of HTML released by the W3C after 1997 include an HTML tag called the object tag and include support for **Cascading Style Sheets (CSS)**. Web designers can use the object tag to embed scripting language code on HTML pages. You will learn more about Web page scripting techniques in Chapter 8.

CSS are sets of instructions that give Web developers more control over the format of displayed pages. Similar to document styles in word-processing programs, CSS lets designers define formatting styles that can be applied to multiple Web pages. The set of instructions, called a **style sheet**, is usually stored in a separate file and is referenced using the HTML style tag; however, it can be included as part of a Web page's HTML file. The term *cascading* is used because designers can apply many style sheets to the same Web page, one on top of the other, and the styles from each style sheet flow (or cascade) into the next. For example, a three-stage cascade might include one style sheet with formatting instructions for text within heading 1 tags, a second style sheet with formatting instructions for text within heading 2 tags, and a third style sheet with formatting instructions for text within paragraph tags. A designer who later decides to change the formatting of heading 2 text can just replace the second style sheet with a different one. Those changes would cascade into the third style sheet.

Extensible Markup Language (XML)

As the Web grew, HTML continued to provide a useful tool for Web designers who wanted to create attractive layouts of text and graphics on their pages. However, as companies began to conduct electronic commerce on the Web, the need to present large amounts of data on Web pages also became important. Companies created Web sites that contained lists of inventory items, sales invoices, purchase orders, and other business data. The need to keep these lists updated was also important and posed a new challenge for many Web designers. The tool that had helped these Web designers create useful Web pages, HTML, was not such a good tool for presenting or maintaining information lists.

In the late 1990s, companies began turning to XML to help them maintain Web pages that contained large amounts of data. XML uses paired start and stop tags in much the same way as database software defines a record structure. For example, a company that sells products on the Web might have Web pages that contain descriptions and photos of the products it sells. The Web pages are marked up with HTML tags, but the product information elements themselves, such as prices, identification numbers, and quantities on hand, are marked up with XML tags. The XML document is embedded within the HTML document.

XML includes data-management capabilities that HTML cannot provide. To better understand the strengths of XML and weaknesses of HTML in data-management tasks, consider the simple example of a Web page that includes a list of countries and some basic facts about each country. A Web designer might decide to use HTML tags to show each fact the same way for each country. Each fact would use a different tag. Assume that the Web designer in this case decided to use the HTML heading tags to present the data. Figure 2-10 shows the data and the HTML heading tags for four countries (this is only an example; the actual list would include more than 150 countries). The first item in the list provides the definitions for each tag. Figure 2-11 (on the next page) shows this HTML document as it appears in a Web browser.

```
<html>
   <head>
     <title>Countries</title>
   </head>
   <body>
     <h1>Countries</h1>

     <h2>CountryName</h2>
     <h3>CapitalCity</h3>
     <h4>AreaInSquareKilometers</h4>
     <h5>OfficialLanguage</h5>
     <h6>VotingAge</h6>

     <h2>Argentina</h2>
     <h3>Buenos Aires</h3>
     <h4>2,766,890</h4>
     <h5>Spanish</h5>
     <h6>18</h6>

     <h2>Austria</h2>
     <h3>Vienna</h3>
     <h4>83,858</h4>
     <h5>German</h5>
     <h6>19</h6>

     <h2>Barbados</h2>
     <h3>Bridgetown</h3>
     <h4>430</h4>
     <h5>English</h5>
     <h6>18</h6>

     <h2>Belarus</h2>
     <h3>Minsk</h3>
     <h4>207,600</h4>
     <h5>Byelorussian</h5>
     <h6>18</h6>
   </body>
</html>
```

FIGURE 2-10 Country list data marked up with HTML tags

Countries

CountryName

CapitalCity

AreaInSquareKilometers

OfficialLanguage

VotingAge

Argentina

Buenos Aires

2,766,890

Spanish

18

Austria

Vienna

83,858

German

19

Barbados

Bridgetown

430

English

18

Belarus

Minsk

207,600

Byelorussian

18

FIGURE 2-11 Country list data as it appears in a Web browser

These figures reveal some of the shortcomings of using HTML to present a list of items when the meaning of each item in the list is important. The Web designer in this case used HTML heading tags. HTML has only six levels of heading tags; thus, if the individual items had additional information elements than shown in this example (such as population and continent), this approach would not work at all. The Web designer could use various combinations of text attributes such as size, font, color, bold, or italics to distinguish among items, but none of these tags would convey the meaning of the individual data elements. The only information about the meaning of each country's listing appears in the first list item, which includes the definitions for each element. In the late 1990s, Web professionals began to consider XML as a list-formatting alternative to HTML that would more effectively communicate the meaning of data.

XML differs from HTML in two important respects. First, XML is not a markup language with defined tags. It is a framework within which individuals, companies, and other organizations can create their own sets of tags. Second, XML tags do not specify how text appears on a Web page; the tags convey the meaning (the semantics) of the information included within them. To understand this distinction between appearance and semantics, consider the list of countries example again. In XML, tags can be created for each fact that define the meaning of the fact. Figure 2-12 shows the countries data marked up with XML tags. Some browsers, such as Internet Explorer, can render XML files directly without additional instructions. Figure 2-13 (on the next page) shows the country list XML file as it would appear in an Internet Explorer browser window.

declaration

root element

```
<?xml version="1.0"?>
<CountriesList>
  <Country Name = "Argentina">
    <CapitalCity>Buenos Aires</CapitalCity>
    <AreaInSquareKilometers>2,766,890</AreaInSquareKilometers>
    <OfficialLanguage>Spanish</OfficialLanguage>
    <VotingAge>18</VotingAge>
  </Country>

  <Country Name = "Austria">
    <CapitalCity>Vienna</CapitalCity>
    <AreaInSquareKilometers>83,858</AreaInSquareKilometers>
    <OfficialLanguage>German</OfficialLanguage>
    <VotingAge>19</VotingAge>
  </Country>

  <Country Name = "Barbados">
    <CapitalCity>Bridgetown</CapitalCity>
    <AreaInSquareKilometers>430</AreaInSquareKilometers>
    <OfficialLanguage>English</OfficialLanguage>
    <VotingAge>18</VotingAge>
  </Country>

  <Country Name = "Belarus">
    <CapitalCity>Minsk</CapitalCity>
    <AreaInSquareKilometers>207,600</AreaInSquareKilometers>
    <OfficialLanguage>Byelorussian</OfficialLanguage>
    <VotingAge>18</VotingAge>
  </Country>

</CountriesList>
```

FIGURE 2-12 Country list data marked up with XML tags

```
<?xml version="1.0" ?>
<CountriesList>
- <Country Name="Argentina">
    <CapitalCity>Buenos Aires</CapitalCity>
    <AreaInSquareKilometers>2,766,890</AreaInSquareKilometers>
    <OfficialLanguage>Spanish</OfficialLanguage>
    <VotingAge>18</VotingAge>
  </Country>
- <Country Name="Austria">
    <CapitalCity>Vienna</CapitalCity>
    <AreaInSquareKilometers>83,858</AreaInSquareKilometers>
    <OfficialLanguage>German</OfficialLanguage>
    <VotingAge>19</VotingAge>
  </Country>
- <Country Name="Barbados">
    <CapitalCity>Bridgetown</CapitalCity>
    <AreaInSquareKilometers>430</AreaInSquareKilometers>
    <OfficialLanguage>English</OfficialLanguage>
    <VotingAge>18</VotingAge>
  </Country>
- <Country Name="Belarus">
    <CapitalCity>Minsk</CapitalCity>
    <AreaInSquareKilometers>207,600</AreaInSquareKilometers>
    <OfficialLanguage>Byelorussian</OfficialLanguage>
    <VotingAge>18</VotingAge>
  </Country>
</CountriesList>
```

FIGURE 2-13 Country list data marked up with XML tags as it would appear in Internet Explorer

The first line in the XML file shown in Figures 2-12 and 2-13 is the declaration, which indicates that the file uses version 1.0 of XML. XML markup tags are similar in appearance to SGML markup tags, thus the declaration can help avoid confusion in organizations that use both. The second line and the last line are the root element tags. The root element of an XML file contains all of the other elements in that file and is usually assigned a name that describes the purpose or meaning of the file.

The other elements are called child elements; for example, Country is a child element of CountriesList. Each of the other attributes is, in turn, a child element of the Country element. Unlike an HTML file, when an XML file is displayed in a browser, the tags are visible. The names of these child elements were created specifically for use in this file. If programmers in another organization were to create a file with country information, they might use different names for these elements (for example, “Capital” instead of “CapitalCity”), which would make it difficult for the two organizations to share information. Thus, the greatest strength of XML, that it allows users to define their own tags, is also its greatest weakness.

To overcome that weakness, many companies have agreed to follow common standards for XML tags. These standards, in the form of data-type definitions (DTDs) or XML schemas, are available for a number of industries, including the **ebXML** initiative for electronic commerce standards, the **Extensible Business Reporting Language (XBRL)** for accounting and financial information standards, **LegalXML** for information in the legal profession, and **MathML** for mathematical and scientific information. A number of industry groups have formed to create standard XML tag definitions that can be used by all companies in that industry. **RosettaNet** is an example of such an industry group. In 2001, the W3C released a set of rules for XML document interoperability that many researchers believe will help resolve incompatibilities between different sets of XML tag definitions. A set of XML tag definitions is sometimes called an **XML vocabulary**. Hundreds of publicly defined XML vocabularies have been developed or are currently circulating. You can find links to many of them on the Oasis **Cover Pages: XML Applications and Initiatives** Web page. You can learn more about XML by reading the **W3C XML Pages**.

Although it is possible to display XML files in some Web browsers, XML files are not intended to be displayed in a Web browser. XML files are intended to be translated using another file that contains formatting instructions or to be read by a program. Formatting instructions are often written in the **Extensible Stylesheet Language (XSL)**, and the programs that read or transform XML files are usually written in the Java programming language. These programs, sometimes called **XML parsers**, can format an XML file so it can appear on the screen of a computer, a smart phone, an Internet-capable mobile phone, or some other device. A diagram showing one way that a Web server might process HTTP requests for Web pages generated from an XML database in different formats for different Web browsing devices appears in Figure 2-14.

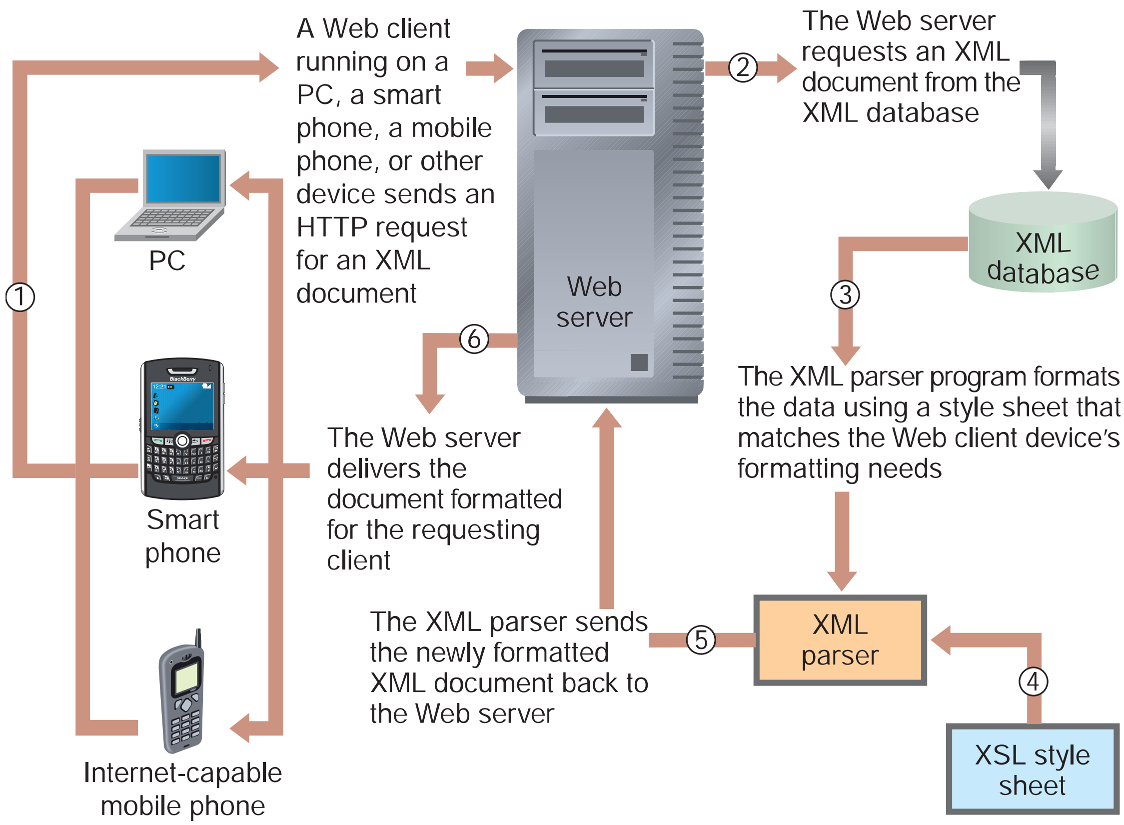


FIGURE 2-14 Processing requests for Web pages from an XML database

HTML and XML Editors

Web designers can create HTML documents in any general-purpose text editor or word processor. However, a special-purpose HTML editor can help Web designers create Web pages much more easily. HTML editors are also included as part of more sophisticated programs that are sometimes called Web site design tools. With these programs, Web designers can create and manage complete Web sites, including features for database access, graphics, and fill-in forms. These programs display the Web page as it will appear in a Web browser in one window and display the HTML-tagged text in another window. The designer can edit in either window and changes are reflected in the other window. For example, the designer can drag and drop objects such as graphics onto the Web browser view page and the program automatically generates the HTML tags to position the graphics.

Web site design programs also include features that allow the designer to create a Web site on a PC and then upload the entire site (HTML documents, graphics files, and so on) to a Web server computer. When the site needs to be edited later, the designer can edit the copy of the site on the PC, then instruct the program to synchronize those changes on the copy of the site that resides on the Web server. The most widely used Web site design tool is **Adobe Dreamweaver**. The Additional Resources section of the Online Companion for this chapter includes links to Web sites that offer downloads of these **HTML Editing and Web Page Design Programs**.

XML files, like HTML files, can be created in any text editor. However, programs designed to make the task of designing and managing XML files easier are also available. These programs provide tag validation and XML creation capabilities in addition to making the job of marking up text with XML tags more efficient. You can find links to these programs' Web sites in the Additional Resources section of the Online Companion under the heading **XML Editing Programs**.

INTRANETS AND EXTRANETS

Not all TCP/IP networks connect to the Internet. Many companies build internets (small "i"), or interconnected networks, that do not extend beyond their organizational boundaries. An **intranet** is an interconnected network (or internet), usually one that uses the TCP/IP protocol set, and does not extend beyond the organization that created it. An **extranet** is an intranet that has been extended to include specific entities outside the boundaries of the organization, such as business partners, customers, or suppliers. Although fax, telephone, e-mail, and overnight express carriers have been the main communications tools for business for many years, extranets can replace many of them at a lower cost.

Public and Private Networks

A **public network** is any computer network or telecommunications network that is available to the public. The Internet is one example of a public network. Although a company can operate its extranet using a public network, very few do because of the high level of security risks. The Internet, as you will learn in later chapters, does not provide a high degree of security in its basic structure.

A **private network** is a private, leased-line connection between two companies that physically connects their intranets to one another. A **leased line** is a permanent telephone connection between two points. Unlike the normal telephone connection you create when you dial a telephone number, a leased line is always active. The advantage of a leased line is security. Only the two parties that lease the line to create the private network have access to the connection. The largest drawback to a private network is cost. Leased lines are expensive. Every pair of companies wanting a private network between them requires a separate line connecting them. For instance, if a company wants to set up an extranet connection over a private network with seven other companies, the company must pay the cost of seven leased lines, one for each company. If the extranet expands to 20 other companies, the extranet-sponsoring company must rent another 13 leased lines. As each new company is added, costs increase by the same amount and soon become prohibitive. Vendors refer to this as a **scaling problem**; that is, increasing the number of leased lines in private networks is difficult, costly, and time consuming. As the number of companies that need to join the extranet increases, other networking options become appealing.

Virtual Private Network (VPN)

A **virtual private network (VPN)** is an extranet that uses public networks and their protocols to send sensitive data to partners, customers, suppliers, and employees using a system called IP tunneling or encapsulation. **IP tunneling** effectively creates a private passageway through the public Internet that provides secure transmission from one computer to another. The virtual passageway is created by VPN software that encrypts the packet content and then places the encrypted packets inside another packet in a process called **encapsulation**. The outer packet is called an **IP wrapper**. The Web server sends the encapsulated packets to their destinations over the Internet. The computer that receives the packet unwraps it and decrypts the message using VPN software that is the same as, or compatible with, the VPN software used to encrypt and encapsulate the packet at the sending end. The "virtual" part of VPN means that the connection seems to be a permanent, internal network connection, but the connection is actually temporary. Each transaction between two intranets using a VPN is created, carries out its work over the Internet, and is then terminated.

VPN software must be installed on the computers at both ends of the transmission. A VPN provides security shells, with the most sensitive data under the tightest control. The VPN is like a separate, covered commuter lane on a highway (the Internet) in which the passengers are protected from being seen by the vehicles traveling in the other lanes. Company employees in remote locations can send sensitive information to company computers using the VPN private tunnels established on the Internet.

Unlike private networks using leased lines, VPNs establish short-term logical connections in real time that are broken once the communication session ends. Establishing VPNs does not require leased lines. The only infrastructure required outside each company's intranet is the Internet.

Extranets are sometimes confused with VPNs. Although a VPN is an extranet, not every extranet is a VPN. You will learn more about VPNs, firewalls, and other network security devices in Chapter 10.

INTERNET CONNECTION OPTIONS

The Internet is a set of interconnected networks. Most organizations have their computers connected to each other using a network. Many families have their home computers connected to each other in a network. Mobile phones are connected to the wireless phone service provider's network. These networks can be connected to the Internet in a number of ways. You will learn about Internet connection options in this section. Companies that provide Internet access to individuals, businesses, and other organizations, called **Internet access providers (IAPs)** or ISPs, usually offer several connection options. Links to the Web sites of companies that offer the various types of Internet connections described in this section appear in the Online Companion under the heading **Internet Service Providers**. This section briefly describes current connection choices and presents their advantages and disadvantages.

Connectivity Overview

ISPs offer several ways to connect to the Internet. The most common connection options are voice-grade telephone lines, various types of broadband connections, leased lines, and wireless. One of the major distinguishing factors between various ISPs and their connection options is the bandwidth they offer. **Bandwidth** is the amount of data that can travel through a communication medium per unit of time. The higher the bandwidth, the faster data files travel and the faster Web pages appear on your screen. Each connection option offers different bandwidths, and each ISP offers varying bandwidths for each connection option. Traffic on the Internet and at your local service provider greatly affects **net bandwidth**, which is the actual speed that information travels. When few people are competing for service from an ISP, net bandwidth approaches the carrier's upper limit. On the other hand, users experience slowdowns during high-traffic periods.

Bandwidth can differ for data traveling to or from the ISP depending on the user's connection type. Connection types include:

- **Symmetric connections** that provide the same bandwidth in both directions.
- **Asymmetric connections** that provide different bandwidths for each direction.

Bandwidth refers to the amount of data that travels and the rate at which it travels. The two bandwidth types in an asymmetric connection are as follows:

- **Upstream bandwidth**, also called **upload bandwidth**, is a measure of the amount of information that can travel from the user to the Internet in a given amount of time.
- **Downstream bandwidth**, also called **download** or **downlink bandwidth**, is a measure of the amount of information that can travel from the Internet to a user in a given amount of time (for example, when a user receives a Web page from a Web server).

Voice-Grade Telephone Connections

The most common way for an individual to connect to an ISP is through a modem connected to your local telephone service provider. **Plain old telephone service (POTS)** uses existing telephone lines and an analog modem to provide a bandwidth of between 28 and 56 Kbps. Some telephone companies offer a higher grade of service called **Digital Subscriber Line (DSL)** protocol. DSL connection methods do not use a modem. They use a piece of networking equipment that is similar to a network switch, but most people call this piece of equipment (incorrectly) a "DSL modem." **Integrated Services Digital Network (ISDN)** was the first technology developed to use the DSL protocol suite and has been available in parts of the United States since 1984. ISDN is more expensive than regular telephone service and offers bandwidths of between 128 and 256 Kbps.

Broadband Connections

Connections that operate at speeds of greater than about 200 Kbps are called **broadband** services. One of the newest technologies that uses the DSL protocol to provide service in the broadband range is **asymmetric digital subscriber line** (**ADSL**, usually abbreviated **DSL**). It provides transmission bandwidths from 100 to 640 Kbps upstream and from 1.5 to 9 Mbps (million bits per second) downstream. For businesses, a **high-speed DSL (HDSL)** connection service can provide more than 768 Kbps of symmetric bandwidth.

Cable modems—connected to the same broadband coaxial cable that serves a television—typically provide transmission speeds between 300 Kbps and 1 Mbps from the client to the server. The downstream transmission rate can be as high as 10 Mbps. In the United States alone, more than 160 million homes have broadband cable service available, and more than 110 million homes subscribe to cable television. The latest estimates indicate that there are more than 11 million cable modem subscribers in the United States. In recent years, DSL monthly fees have been slightly lower than those of cable companies in markets where they compete. Today, about 13 million households have broadband DSL connections. Virtually all companies and organizations of any size have some type of broadband Internet connection.

DSL is a private line with no competing traffic. Unlike DSL, cable modem connection bandwidths vary with the number of other subscribers competing for the shared resource. Transmission speeds can decrease dramatically in heavily subscribed neighborhoods at prime times—in neighborhoods where many people are using cable modems simultaneously.

Connection options based on cable or telephone line connections are wonderful for urban and suburban Web users, but those living in rural areas often have limited telephone service and no cable access at all. The telephone lines used to cover the vast distances between rural customers are usually **voice-grade lines**, which cost less than telephone lines designed to carry data, are made of lower-grade copper, and were never intended to carry data. These lines can carry only limited bandwidth—usually less than 14 Kbps. Telephone companies have wired most urban and suburban areas with **data-grade lines** (made more carefully and of higher-grade copper than voice-grade lines) because the short length of the lines in these areas makes it less expensive to install than in rural areas where connection distances are much longer.

LEARNING FROM FAILURES

NorthPoint Communications

In 1997, Michael Malaga was a successful telecommunications executive with an idea. He wanted to sell broadband Internet access to small businesses in urban areas. DSL technology was just gaining acceptance, and leased telephone lines were available from telephone companies. He wanted to avoid residential customers because they would soon have inexpensive cable modem access to meet their broadband needs. He also wanted to avoid suburban and rural businesses to keep the telephone line leasing costs low (lease charges are higher for longer distances). He and five friends started NorthPoint Communications with $500,000 of their combined savings and raised another $11 million within a few months.

Continued

After six months, the company had raised more money from investors and had acquired 1500 customers, but it was posting a net loss of $30 million. On the strength of its number of customers, the company began the task of raising the $100 million that Malaga estimated it would need to create the network infrastructure.

Independent DSL providers such as NorthPoint were pressed by customers to install service rapidly, but had to rely on local telephone companies to ensure that their lines would support DSL. In many cases, the telephone companies had to install switches and other equipment to make DSL work on a particular line. The telephone companies often were in no rush to do this because they also sold DSL service, and speedy service would be helping a competitor. The delays led to unpredictable installation holdups and many unhappy NorthPoint customers. Customers with problems after the service was installed were often bounced from the telephone company to NorthPoint, without obtaining satisfactory or timely resolutions of their problems.

Although NorthPoint was unable to make its relationship with each customer profitable, Malaga and his team were rapidly raising money in the hot capital markets of the time. The company raised $162 million before its first stock offering in 1999, which brought in an additional $387 million. At that time, the company had 13,000 customers, which means that NorthPoint had raised more than $42,000 from outside investors for each customer. Considering that each customer would generate revenue of about $1,000 per year, the economics of the business did not look good. By the end of 1999, NorthPoint had spent $300 million of the cash it had raised to build its network infrastructure and reported an operating loss of $184 million. At this point, NorthPoint was operating in 28 cities.

During the next year, the company continued to raise additional funds, gain more customers, and lose money on each customer. In August 2000, the telephone company Verizon agreed to purchase 55 percent of the company for $800 million paid in installments. The total funding that NorthPoint had obtained by the end of 2000, including the partial payments received from Verizon, added up to $1.2 billion. By the end of the year, NorthPoint was in 109 cities and needed to spend $66 million in cash per month just to stay in business. Verizon withdrew from the purchase agreement, the stock plunged, and the layoffs began.

NorthPoint filed for bankruptcy in January 2001 and sold its networking hardware to AT&T in March for $135 million. AT&T was not interested in continuing the DSL business (it just wanted the hardware), so NorthPoint's 87,000 small business customers lost their Internet service overnight. In many of the cities that NorthPoint had served, there were no competitors to pick up the service. Because the capital markets of the late 1990s were so eager to invest in anything that appeared to be connected with the Internet, NorthPoint was able to raise incredible amounts of money. However, NorthPoint sold Internet access to customers for less than it cost to provide the service. No amount of investor money could overcome that basic business mistake.

Leased-Line Connections

Large firms with large amounts of Internet traffic can connect to an ISP using higher bandwidth connections that they can lease from telecommunications carriers. These connections use a variety of technologies and are usually classified by the equivalent number of telephone lines they include. (The connection technologies they use were originally developed to carry large numbers of telephone calls.)

A telephone line designed to carry one digital signal is called DS0 (digital signal zero, the name of the signaling format used on those lines) and has a bandwidth of 56 Kbps.

A **T1** line (also called a DS1) carries 24 DS0 lines and operates at 1.544 Mbps. Some telecommunications companies offer **fractional T1**, which provides service speeds of 128 Kbps and upward in 128-Kbps increments. **T3** service (also called DS3) offers 44.736 Mbps (the equivalent of 30 T1 lines or 760 DS0 lines). All of these leased telephone line connections are much more expensive than POTS, ISDN, or DSL connections.

Large organizations that need to connect hundreds or thousands of individual users to the Internet require very high bandwidth. NAPs use T1 and T3 lines. NAPs and the computers that perform routing functions on the Internet backbone also use technologies such as **frame relay** and **asynchronous transfer mode (ATM)** connections and **optical fiber** (instead of copper wire) connections with bandwidths determined by the class of fiber-optic cable used. An OC3 (optical carrier 3) connection provides 156 Mbps, an OC12 provides 622 Mbps, an OC48 provides 2.5 Gbps (gigabits, or 1 billion bits per second), and an OC192 provides 10 Gbps.

Wireless Connections

For many people in rural areas, satellite microwave transmissions have made connections to the Internet possible for the first time. In the first satellite technologies, the customer placed a receiving dish antenna on the roof or in the yard and pointed it at the satellite. The satellite sent microwave transmissions to handle Internet downloads at speeds of around 500 Kbps. Uploads were handled by a POTS modem connection. For Web browsing, this was not too bad, since most of the uploaded messages were small text messages (e-mails and Web page requests). People who wanted to send large e-mail attachments or transfer files over the Internet found the slow upload speeds unsatisfactory.

Today, companies offer satellite Internet connections that do not require a POTS modem connection for uploads. These connections use a microwave transmitter for Internet uploads. This transmitter provides upload speeds as high as 150 Kbps. Initially, the installation charges were much higher than for other residential Internet connection services because a professional installer was needed to carefully aim the transmitter's dish antenna at the satellite. Recently, the accuracy of the antennas improved, and some of these companies now offer a self-installation option that drastically reduces the initial cost. For installations in North America, the antennas must have a clear line of sight into the southwestern sky. This requirement can make these services unusable for many people living in large cities or on the wrong side of an apartment building. In the United States, about 2 million homes are connected to the Internet through a satellite broadband service.

Although satellite connections were the only wireless Internet access media for many years, many types of wireless networks are available now. People today use Internet-capable mobile phones, smart phones, game consoles, and notebook computers equipped with wireless network cards to connect to a variety of wireless networks that, in turn, are connected to the Internet. More than half of U.S. Internet users have used a wireless device to access the Internet.

Bluetooth and Ultra Wideband (UWB)

One of the first wireless protocols, designed for personal use over short distances, is called **Bluetooth**. (The protocol was developed in Norway and is named for Harald Bluetooth, a 10th century Scandinavian king.) Bluetooth operates reliably over distances of up to 35 feet and can be a part of up to 10 networks of eight devices each. It is a low-bandwidth technology, with speeds of up to 722 Kbps. Bluetooth is useful for tasks such as wireless synchronization of laptop computers with desktop computers and wireless printing from laptops or mobile phones. These small Bluetooth networks are called **personal area networks (PANs)** or **piconets**.

One major advantage of Bluetooth technology is that it consumes very little power, which is an important consideration for mobile devices. Another advantage is that Bluetooth devices can discover one another and exchange information automatically. For example, a person using a laptop computer in a temporary office can print to a local Bluetooth-enabled printer without logging in to the network or installing software on either device. The printer and the laptop computer electronically recognize each other as Bluetooth devices and can immediately begin exchanging information.

Another wireless communication technology, **Ultra Wideband (UWB)**, provides wide bandwidth (up to about 480 Mbps in current versions) connections over short distances (30 to 100 feet). UWB was developed for short-range secure communications in military applications during the 1960s. Many observers believe that UWB and other similar technologies will be used in future personal area networking applications such as home media centers (for example, a PC could beam stored video files to a nearby television) and in linking mobile phones to the Internet.

Wireless Ethernet (Wi-Fi)

The most common wireless connection technology for use on LANs is called **Wi-Fi**, **wireless Ethernet**, or **802.11b** (802.11 is the number of the technology's **network specification**, which is the set of rules that equipment connected to the network must follow). Wireless networking specifications are created by the **IEEE** (originally an acronym for an organization named the Institute of Electrical and Electronic Engineers, the letters are now used as the title of the organization and are pronounced eye-triple-E). A computer equipped with a Wi-Fi network card can communicate through a wireless access point connected to a LAN to become a part of that LAN. A **wireless access point (WAP)** is a device that transmits network packets between Wi-Fi-equipped computers and other devices that are within its range. The user must have authorization to connect to the LAN and might be required to perform a login procedure before the laptop can access the LAN through the WAP.

Wi-Fi that uses the 802.11b specification has a potential bandwidth of 11 Mbps and a range of about 300 feet. In actual installations, the achieved bandwidth and range can be dramatically affected by the construction material of the objects (such as walls, floors, doors, and windows) through which the signals must pass. For example, reinforced concrete walls and certain types of tinted glass windows greatly reduce the effective range of Wi-Fi. Despite these limitations, organizations can make Wi-Fi a key element of their LAN structures by installing a number of WAPs throughout their premises.

In 2002, an improved version of Wi-Fi, called **802.11a** (the 802.11b protocol was easier to implement, thus it was introduced first) was introduced. The 802.11a protocol is capable of transmitting data at speeds up to 54 Mbps, but it is not compatible with 802.11b devices. Later in 2002, the **802.11g** protocol, which has the 54-Mbps speed of 802.11a and is compatible with 802.11b devices, was introduced. Because of its compatibility with the many 802.11b devices that were in use, 802.11g was an immediate success.

In 2003, work began on the **802.11n** standard, which was originally planned to be completed in 2006. The expected completion dates were extended to 2007, then 2008, and finally 2010. Although the IEEE continues its work on this standard, drafts of the specification have been published and vendors currently sell Wi-Fi equipment built to the draft standard. The IEEE has promised that any further changes to the specification will not affect the ability of existing "Draft-N" products. The 802.11n wireless networking products provide significantly higher actual bandwidths (80–130 Mbps) than any earlier Wi-Fi standard products. When the 802.11n specification is finalized, the technology should provide bandwidths in the 300–450-Mbps range.

Wi-Fi devices are capable of **roaming**, that is, shifting from one WAP to another, without requiring intervention by the user. Some organizations, including airports, convention centers, and hotels, operate WAPs that are open to the public. These access points are called **hot spots**. Some organizations allow free access to their hot spots, others charge an access fee. A growing number of retail establishments, such as McDonald's, Panera, and Starbucks, offer hot spots. Hotels and office buildings have found that installing a WAP can be cheaper and easier than running network cable, especially in older buildings. Some hotels offer wireless access free, others charge a small fee. Users of fee-based networks authorize a connection charge when they log in. There are Web sites that offer hot spot directories that show hot spots by location, but these sites tend to open and close frequently, so these directories become out of date rather quickly. The best way to find hot spots (or a hot spot directory) is to use your favorite search engine.

Some communities have installed wireless networks that can be accessed from anywhere in the area. For example, the city of Grand Haven, Michigan, installed a metropolitan area Wi-Fi network. Grand Haven is a growing town on the shores of Lake Michigan. The company that built the network, Ottawa Wireless, sells network access to residents and businesses throughout the area. The company offers access not only on land, but on boats up to 15 miles out on Lake Michigan. Several small company owners use this network to conduct their online business while sailing!

Fixed-Point Wireless

In a growing number of rural areas that do not have cable TV service or telephone lines with the high-grade wires necessary to provide Internet bandwidths, some small companies have begun to offer fixed-point wireless service as an inexpensive alternative to satellite service. One version of **fixed-point wireless** uses a system of repeaters to forward a radio signal from the ISP to customers. The **repeaters** are transmitter–receiver devices (also called **transceivers**) that receive the signal and then retransmit it toward users' roof-mounted antennas and to the next repeater, which receives the signal and passes it on to the next repeater, which can be up to 20 miles away. The users' antennas are connected to a device that converts the radio signals into Wi-Fi packets that are sent to the users' computers or wireless LANs. Another version of fixed-point wireless directly transmits Wi-Fi packets through hundreds, or even thousands, of short-range transceivers that are located close to each other. This approach is called **mesh routing**. As Wi-Fi technologies improve, the number and variety of options for wireless connections to the Internet should continue to increase.

Cellular Telephone Networks

By the end of 2010, industry experts estimate that about 6 billion mobile phones will be in use around the world. These phones are sometimes called cellular (or cell) phones because they broadcast signals to (and receive signals from) antennas that are placed about 3 miles apart in a grid, and the hexagonal area that each antenna covers within this grid is called a cell.

Although mobile phones were originally designed to handle voice communications, they have always been able to transmit data. However, their data transmission speeds were very low, ranging from 10 to 384 Kbps. Several changes in mobile phone technology have increased the speeds in today's most capable mobile phones to 2 Mbps. The devices that combine the latest technologies available today are called **third-generation (3G) mobile phones**. Many mobile phones have a small screen and can be used to send and receive short text messages using a protocol called **short message service (SMS)**. As you learned at the beginning of this chapter, Internet-enabled mobile phones and smart phones are very popular in highly developed countries as convenient ways to stay connected while traveling. But more importantly, mobile phones are giving large numbers of people in developing countries their first access to the online world.

Many companies have seen great business potential for these wireless networks and the devices connected to them. They use the term **mobile commerce** or **m-commerce** to describe the kinds of resources people might want to access (and pay for) using wireless devices. You will learn more about revenue models that use wireless technologies in Chapter 3 and cost-reduction strategies that use wireless technologies in Chapter 5. Chapter 6 includes an overview of the development of mobile commerce to date and an outline of future directions. In Chapter 11, you will learn how some companies are using these technologies to process online payments for goods and services. Figure 2-15 summarizes speed and cost information for the most commonly available wired and wireless options for connecting a home or business to the Internet.

Service	Upstream Speed (Kbps)	Downstream Speed (Kbps)	Capacity (Number of Simultaneous Users)	One-time Startup Costs	Continuing Monthly Costs
Residential-Small Business Services					
POTS	28–56	28–56	1	$0–$20	$9–$20
Cellular 3G network	10–800	10–2000	1	$50–120	$50–150
ISDN	128–256	128–256	1–3	$60–$300	$50–$90
ADSL	100–640	500–9000	4–20	$50–$100	$200–$500
Cable	300–1500	500–10,000	4–10	$0–$100	$100–$300
Satellite	125–150	400–500	1–3	$0–$800	$40–$100
Fixed-point wireless	250–1500	500–3000	1–4	$0–$350	$50–$150
Business Services					
Leased digital line (DS0)	64	64	1–10	$50–$200	$40–$150
Fractional T1 leased line	128–1544	128–1544	5–180	$50–$800	$100–$1000
Fixed-point wireless	500–10,000	500–10,000	5–1000	$0–$500	$300–$5000
T1 leased line	1544	1544	100–200	$100–$2000	$600–$1600
T3 leased line	44,700	44,700	1000–10,000	$1000–$9000	$5000–$12,000
Large Organization					
OC3 leased line	156,000	156,000	1000–50,000	$3000–$12,000	$9000–$22,000
OC12 leased line	622,000	622,000	Backbone	Negotiated	$25,000–$100,000
OC48 leased line	2,500,000	2,500,000	Backbone	Negotiated	Negotiated
OC192 leased line	10,000,000	10,000,000	Backbone	Negotiated	Negotiated

FIGURE 2-15 Internet connection options

INTERNET2 AND THE SEMANTIC WEB

At the high end of the bandwidth spectrum, a group of network research scientists from nearly 200 universities and a number of major corporations joined together in 1996 to recapture the original enthusiasm of the ARPANET with an advanced research network called **Internet2**. When the National Science Foundation turned over the Internet backbone to commercial interests in 1995, many scientists felt that they had lost a large, living laboratory. **Internet2** is the replacement for that laboratory. An experimental test bed for new networking technologies that is separate from the original Internet, Internet2 has achieved bandwidths of 10 Gbps and more on parts of its network.

Internet2 is also used by universities to conduct large collaborative research projects that require several supercomputers connected at very fast speeds, or that use multiple video feeds—things that would be impossible on the Internet given its lower bandwidth limits. For example, doctors at medical schools that are members of Internet2 regularly use its technology to do live videoconference consultations during complex surgeries. Internet2 serves as a proving ground for new technologies and applications of those technologies that will eventually find their way to the Internet. In 2008, CERN (the birthplace of the original Web in Switzerland) began using Internet2 to share data generated by its new particle accelerator with a research network of 70 U.S. universities. Every few weeks, each university downloads about two terabytes (a terabyte is one thousand gigabytes) of data within a four-hour time period.

The Internet2 project is focused mainly on technology development. In contrast, Tim Berners-Lee began a project in 2001 that has a goal of blending technologies and information into a next-generation Web. This **Semantic Web** project envisions words on Web pages being tagged (using XML) with their meanings. The Web would become a huge machine-readable database. People could use intelligent programs called **software agents** to read the XML tags to determine the meaning of the words in their contexts. For example, a software agent given the instruction to find an airline ticket with certain terms (date, cities, cost limit) would launch a search on the Web and return with an electronic ticket that meets the criteria. Instead of a user having to visit several Web sites to gather information, compare prices and itineraries, and make a decision, the software agent would automatically do the searching, comparing, and purchasing.

The key elements that must be added to Web standards so that software agents can perform these functions include XML, a resource description framework, and an ontology. You have already seen how XML tags can describe the semantics of data elements. A **resource description framework (RDF)** is a set of standards for XML syntax. It would function as a dictionary for all XML tags used on the Web. An **ontology** is a set of standards that defines, in detail, the relationships among RDF standards and specific XML tags within a particular knowledge domain. For example, the ontology for cooking would include concepts such as ingredients, utensils, and ovens; however, it would also include rules and behavioral expectations, such as that ingredients can be mixed using utensils, that the resulting product can be eaten by people, and that ovens generate heat within a confined area. Ontologies and the RDF would provide the intelligence about the knowledge domain so that software agents could make decisions as humans would.

The development of the Semantic Web is expected to take many years. The first step in this project is to develop ontologies for specific subjects. Thus far, several areas of scientific inquiry have begun developing ontologies that will become the building blocks of the Semantic Web in their areas. Biology, genomics, and medicine have all made progress toward specific ontologies. These fields can benefit greatly from a tool like the Semantic Web, which can increase the speed with which research results, experimental data, and new procedures can be made available to all researchers in the field. Thus, these fields have a high incentive to collaborate on the hard work involved in creating ontologies. Other sciences, such as climatology, hydrology, and oceanography have similar incentives (as many researchers around the world work on common problems such as global warming) and are working on ontologies for their disciplines. The government of the United Kingdom is also developing an ontology for data it collects with the hope that it will be useful to a wide range of researchers.

Although many researchers involved in the Semantic Web project have expressed frustration at its slow progress, a number of important users of the Semantic Web have developed important ontologies that will allow the project to continue moving forward. You can learn more about the current status of this project by following the link in the Online Companion to the **W3C Semantic Web** pages.