

Tutorial

The Internet, the World Wide Web, Library Web Browsers, and Library Web Servers

Jian-Zhong (Joe) Zhou

This article first examines the difference between two very familiar and sometimes synonymous terms, the Internet and the Web. The article then explains the relationship between the Web's protocol HTTP and other high-level Internet protocols, such as Telnet and FTP, as well as provides a brief history of Web development. Next, the article analyzes the mechanism in which a Web browser (client) "talks" to a Web server on the Internet. Finally, the article studies the market growth for Web browsers and Web servers between 1993 and 1999. Two statistical sources were used in the Web market analysis: a survey conducted by the University of Delaware Libraries for the 122 members of the Association of Research Libraries, and the data for the entire Web industry from different Web survey agencies.

Many librarians are now dealing with the Internet and the Web on a daily basis. While the Web is sometimes synonymous with the Internet in many people's minds, the two terms are quite distinct, and they refer to different but related concepts in the modern computerized telecommunication system.

The Internet is nothing more than many small computer networks that have been wired together and allow electronic information to be sent from one network to the next around the world. A piece of data from

Beijing, China may traverse more than a dozen networks while making its way to Washington, D.C. We can compare the Internet to the Great Wall of China, which was built in the Qin dynasty around the third century B.C. by connecting many existing short defense walls built by previous feudal states. The Great Wall not only served as a national defense system for ancient China, but also as a fast military communication system. A border alarm was raised by means of smoke signals by day, and beacon fires at night, ignited by burning a mixture of wolf dung, sulfur, and saltpeter. The alarm signal could be relayed over many beacon-fire towers from the western end of the Great Wall to the eastern end (4,500 miles away) within a day. This was considered light speed two thousand years ago. However, while the Great Wall transferred the message in a linear mode, the Internet is a multidimensional network.

The Web is a late-comer to the Internet, one of the many types of high-level data exchange protocols on the Internet. Before the Web, there was Telnet, the traditional command-driven style of interaction. There was FTP, a file transfer protocol useful for retrieving information from large file archives. There was Usenet, a communal bulletin board and news system. There was also e-mail for individual information exchange, and e-mail lists, for one-to-many broadcasts. In addition, there was Gopher, a campus-wide information system shared among universities and research institutions, and WAIS, a powerful search and retrieval system developed by Thinking Machines, Inc. In 1990 Tim Berners-Lee and Robert Cailliau at CERN (www.cern.ch), the European Laboratory for Particle Physics, created a new information system called "World Wide Web" (WWW). Designed to help the CERN scientists with the increasingly confusing task of exchanging information on the

Internet, the Web system was to act as a unifying force, a system that would seamlessly bind all file-protocols into a single point of access. Instead of having to invoke different programs to retrieve information via various protocols, users would be able to use a single program, called a "browser," and allow it to handle all the details of retrieving and displaying information. In December 1993 WWW received the IMA award, and in 1995 Berners-Lee and Cailliau received the Association for Computing (ACM) Software System Award for its development.

The Web is best known for its ability to combine text with graphics and other multimedia on the Internet. In addition, the Web has some other key features that make it stand out from earlier Internet information exchange protocols. Since the Web is a late-comer to the Internet, it has to be compatible backwards with other communications protocols in addition to its native language, HyperText Transfer Protocol (HTTP). Among the foreign languages spoken by Web browsers are Telnet, FTP, and other high-level communication protocols mentioned earlier. This support for foreign protocols lets people use a single piece of software, the Web browser, to access information without worrying about shifting from protocol to protocol and software incompatibility.

Despite different high-level protocols including HTTP for the Web, there is one thing in common for all parts of the Internet—TCP/IP, the lower level of the Internet protocol. TCP/IP is responsible for establishing the connection between two computers on the Internet and guarantees that the data can be sent and received intact. The format and content of the data are left for high-level communication protocols to manage, among which the Web is the best known one. At the TCP/IP level all computers "are created equal." Two computers establish a connection and start to

Joe Zhou (joezhou@udel.edu) is Associate Librarian at the University of Delaware Library, Newark.

communicate. In reality, however, most conversations are asymmetric. The end user's machine (the client) usually sends a short request for information, and the remote machine (the server) answers with a long-winded response. The media is the Internet. The common language on the Internet can be the Web or any other high-level protocols.

On the Web, the client is the Web browser; it handles the user's request for a document. The first Web browser, NCSA Mosaic, developed by the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign, was released in mid-November 1993 for Unix, Windows, and Macintosh platforms. Version 3.0 of NCSA Mosaic is available at www.ncsa.uiuc.edu/SDG/Software/Mosaic. Both source code and binaries are free for academic use. Mosaic lost market share to Netscape after its key developer left NCSA and joined Netscape. Even after Mosaic introduced an innovative 32-bit version in early 1997, which can perform feats that other major browsers had not even thought of back then, Mosaic remained out of the major browsers' market.

The two most widely-used browsers today are Microsoft's Internet Explorer (IE) and Netscape's Navigator (part of the Netscape Communicator suite). Recent Web browser surveys conducted by different Internet survey companies such as www.zonaresearch.com/browserstudy, www.psrinc.com/Trends.htm, and www.statmarket.com all indicate that IE is the market leader with more than 60 percent market share, leaving Navigator with between 35 percent and 40 percent. In 1995 IE had only 1 percent share versus Navigator's more than 90 percent, an unimaginable rise critics have attributed to Microsoft's strategy of bundling the browser with its near-monopoly Windows operating system. However, a survey

conducted in December 1998 by the University of Delaware Library of 122 members of the Association of Research Libraries (ARL) showed that Netscape still remained the market leader among big academic libraries. More than 90 percent of ARL libraries supported Netscape, and about 50 percent also supported IE. Most ARL libraries supported both browsers, and unlike the browser industry survey mentioned earlier, in which only one product can be picked as the primary browser, the sum of the percentages for the ARL survey was greater than 100 percent. The main function of the Web browser is to request a document available from a specific server through the Internet using the information in the document's URL. The server on a remote machine returns the document usually physically stored on one of the server's disks. With the use of Common Gateway Interface (CGI), the documents do not have to be static. Rather, they can be synthesized at the point of being requested by CGI scripts running on the server's side of the connection. In some database-driven Web servers that make the core of today's e-commerce, the documents provided may never exist as physical files but are generated as needed from database records.

The Web server can be run on almost any computer, and server software is available for almost all operating systems, such as Unix, Windows 95/98/NT, Macintosh, and OS/2. According to the University of Delaware Library's 1998 survey of Internet Web servers among ARL member libraries, more than 32 percent of ARL libraries chose Apache as their Web server software, followed by the Netscape series at 29.32 percent, NCSA HTTPd at 11.28 percent, and Microsoft Internet Information Server (IIS) at 7.52 percent. In July 1999 the author checked the Netcraft survey at www.netcraft.com/Survey. The top three Web serv-

er software programs for more than 6.5 million Web sites are Apache (56.35 percent), Microsoft-IIS (22.33 percent), and Netscape (5.65 percent). The Netcraft survey also provides the historical market share information of major Web servers since August 1995.

NCSA HTTPd was the first Web server software released, about the same time as the release of Mosaic in 1993. However, it slipped from the number-one position with more than 90 percent market share in 1993, and almost 60 percent in 1995, to less than 1 percent in July 1999. It is no longer supported by NCSA, however, HTTPd remains a popular choice for Web servers due to its small size, fast performance, and solid collection of features. The "inertia effect" of the existing sites (if it runs well, why bother to change?) will likely keep NCSA on the major Web server software list for some time. NCSA is free, but available only for the Unix platform. It is available from <http://hoohoo.ncsa.uiuc.edu>. However, when the author visited the site in July 1999, the following message appeared on the main page: "THE NCSA HTTPd IS NO LONGER UNDER DEVELOPMENT. It is an unsupported product. We recommend that you check out the Apache server, instead of installing our server."

Most people who use only Web browsers may have heard of Apache only as an Indian nation or a military helicopter, not the most popular Web server software with more than 50 percent market share. It was first introduced as a set of fixes or "patches" to the NCSA HTTPd. Apache 1.0 was released in December 1995 as open-source server software by a group of webmasters who named themselves the Apache Group. Open-source means the source code is available and freely distributed, and it is the key to Apache's attractiveness and popularity. The Apache Group members were NSCA users

who decided to coordinate development work on the server software after NSCA stopped. In July 1999 the Apache Group announced that it was establishing a more formal organization called the Apache Software Foundation (ASF). In the future, the ASF (www.apache.org) will monitor development of the free software, but it will remain a "not-for-profit" foundation. Apache is high-end, enterprise-level server software and can be run on OS/2, Unix (including Linux), and Windows platforms, but a Mac version is still not available.

The Netscape series includes Netscape-Enterprise, Netscape-FastTrack, Netscape-Commerce, and Netscape-Communication. Enterprise is a high-end, enterprise-level server while FastTrack serves as an entry-level server for small workgroups. Netscape supports both the Unix and the Windows NT platforms. The other major commercial Web server, Microsoft Internet Information Server (IIS), as of 1999, is only available for

the Windows platform. However, one advantage of IIS over Netscape is that it can be downloaded for free as part of the Windows Option Pack. In addition, IIS can handle MS Office documents very well. While both the Microsoft and Netscape brand names are well recognized by millions of end users, a name alone does not necessarily equate to large market share, nor does a deep pocket. Apache remains the top Web server despite intense competition. One of the keys to Apache's success, in addition to its outstanding performance, lies in its open-source code movement and active user support on a wide basis.

The Web server of choice for the Macintosh platforms is WebStar. However, due to the limitations of the operating system networking software, the performance of Macintosh-based servers has not been great. WebStar can be downloaded as a free evaluation release from www.starnine.com/webstar.

The Web server market is dynam-

ic and competition intense. There are more than sixty Web server products on the top list (of Web servers with more than one thousand Web sites) as of July 1999, and newcomers are being added frequently.

Acknowledgments

The author thanks Peter Liu, Head of the Systems Department at the University of Delaware Library, for providing the Web survey data of ARL libraries. After this article was submitted, the survey data was published by ARL in 1999 as SPEC Kit 246: Web Page Development and Management.

The author also wants to thank his dear wife Min Yang for her technical assistance. Min is Webmaster and System Administrator for the Web site at A. I. duPont Nemours Foundation and Hospital for Children, <http://kidshealth.org>.