Information Technology
Inside and Outside
- David Cyganski & John A. Orr

Part I. Introduction

2. The World Wide Web: A Unique Product of the Information Age

Hoon-Jae Lee
hjlee@dongseo.ac.kr
http://cg.dongseo.ac.kr/~hjlee

2. The World Wide Web: A Unique Product of the Information Age

- The ways in which the Web is a fundamentally new and different information medium, with major advantages over all other such systems;
- The fundamentals of the underlying operations and facilities that make the Web possible and make it work;
- Principles of Web browsers and Web servers;
- The significance and basic operation of a hypertext markup language (HTML); and
- The significance of the Java programming language, and its relation to the Web.

2.3 What Is the Web, and Why Was It Created? (1)

- The Web was developed originally to solve a very specific problem; fortunately, it turned out to be useful for many other problems and needs as well. It was created to facilitate communications among nuclear physicists located throughout the world. These physicists make use of data gathered at specialized facilities, such as CERN (European Particle Physics laboratory) in Geneva, Switzerland, and LANL (Los Alamos National Labs) in the United States, to name two.
- The first new communication tool to benefit these physicists in their work was ***e-mail***. Originally, e-mail only provided for the sending of what is often called ***raw text*** between correspondents.
- In 1991, a system known as **MIME** (Multi purpose Internet Mail Extensions) was introduced to allow e-mail users to make ***attachments*** of additional documents to an e-mail.

2.4 The Origin of the Web

- ***Killer apps*** : software applications that have taken on importance due to broad acceptance and wide applicability.
- Typically the killer app list includes
  - The Visicalc spreadsheet (Software Arts, May 1979)
    - ✓ **A lone programmer, *Dan Bricklin*, envisioned and wrote this first spreadsheet program while he was still a student at Harvard University. While the spreadsheet began humbly as an Apple II program written in the BASIC programming language, it ultimately resulted, indirectly, in the success of the Apple II computer platform, spawning several generations of successful businesses based upon spread sheets. Furthermore, it forever changed the way in which business people plan, review, and present their work.**
  - The WordStar word processor (MicroPro, June 1979)
  - The dBase (Aston-Tate, 1980) database system for personal computers.

2.4 The Origin of the Web (2)

- The ***WWW*** was the brainchild of ***Tim Berners-Lee***, an Oxford University graduate working at the CERN facility, who is now Director of the World Wide Web Consortium(W3C).
- ***Berners-Lee*** undertook the task of finding a simple, extendible, and distributed approach to solving the communication problems of the particle physicists described above.
  - ✓ **The actual observed working structure of the organization is a multiply connected ``web'' whose interconnections evolve with time. In this environment, a new person arriving, or someone taking on a new task, is normally given a few hints as to who would be useful people to talk to. Information about what facilities exist and how to find out about them travels in the corridor gossip and occasional news letters, and the details about what is required to be done spread in a similar way. All things considered, the result is remarkably successful, despite occasional misunderstandings and duplicated effort.**
- In 1992, the CERN team introduced the first version of ***a Web browser***, along with the required Web server software.

2.4 The Origin of the Web (3)

- An Internet ***Web server*** is a software application that executes on a computer connected to the Internet and offers a service to other computers on the Internet. The purpose of a Web server is to receive requests for information from Web browsers, and to send, via the Internet, the requested documents
- The ***Web browser*** is an application that presents a user interface for requesting and viewing these documents. Chances are that you are familiar with one or more browsers, which include ***Netscape Navigator,Netscape Communicator, and Microsoft Internet Explorer***.
- The ***Mosaic Web browser*** introduced extensive graphical rendering and text font and format flexibility. Mosaic 1.0 for the UNIX Xwindows was introduced in April of 1993, and it ignited the interest of computer users throughout the world. By late 1993, Version 2.0 had been released, and versions for the PC and Apple McIntosh were introduced.

2.4 The Origin of the Web (4)

- The lead programmer in the effort to create ***Mosaic, Marc Andreessen***, soon left NCSA along with several others from the team to form the company ***Netscape Communications***. Here they developed the ***Netscape Navigator Web browser*** that was to catapult the Web to fame. Since then, the Web has become an essential tool on computers in homes, businesses, and schools.

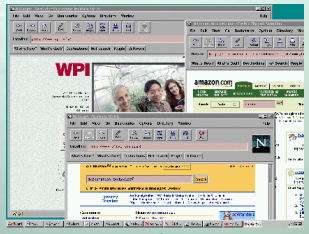


Figure 2.1: Screen image of a typical Web browser session.

2.5 How the Web Solves Our Document Distribution Problem

- ***Hyper-TextMarkup Language (HTML) editor*** :
  - a specialized editing program, or
  - a special format conversion system in a word processor.
  - be able to insert pictures and graphs within the textual flow of the document
- the document's ***Universal Resource Locator(URL)*** **address**
  - the URL identifies your document uniquely among all other documents in the world.
  - **New observations of neutrino mass oscillations: http://www.nlab.org/masstrans/Dec99/exp23anal.html**
    - ✓ **In this case, there ought not to be many complaints about wasted download time, and those who have an interest can simply enter (by cutting and pasting with a mouse pointer) the stated URL into a Web browser to see th full information in context. If one of your colleagues thinks that some of her friends ought to look at it, she merely sends a similar e-mail and includes your URL.**
- **HTML editor** to create a ***hypertext link*** (or what some people like to call a ***hot button***) in their document.

2.5 How the Web Solves Our Document Distribution Problem(2)

- ***Search engines***, which we can access at special URLs (*Web sites*).
  - become known to us through advertising in magazines, books, and other Web sites, and
  - give us access to large, indexed sets of data that have been gathered and are continuously being updated by the operators of the search service providers.
  - *The data is gathered by software that automatically scans every existing Web site and catalogs key words that appear in the associated Web documents.*
  - Example, ``***neutrino oscillation***'' might ask the search engine to generate a list of all Web sites that contain that phrase.

2.6 How the Web Was Won

- The World Wide Web represents a sophisticated system for information capture and delivery

- **2.6.1 The Universal Resource Locator (URL)**
  - similar to the ISBN number assigned to all published books.
  - not only a unique name for the document,
  - a set of directions for obtaining a copy of that document
  - A key idea behind the Web is that ***the URL is a single, universal address mechanism*** that denotes an information service or document type, the location of the hosting server within the Internet, and the location of that information within the host, and may also actually include information to be forwarded to that server for processing.
  - An example of using a URL to access a *service* was seen in our previous example, when we used a search engine service located at one URL to locate data at other URLs.

2.6 How the Web Was Won (2)

- **2.6.2 Virtual Path Addressing**
  - ***File name & a path name*** for your document.
  - The ***path name*** serves two roles: an ***address*** and a ***hierarchical***

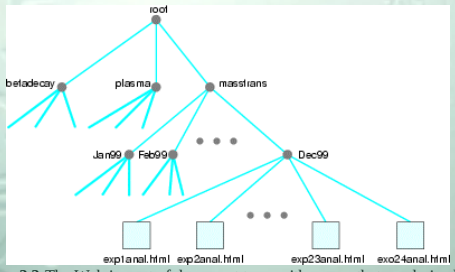


Figure 2.2: The Web is a set of document trees with names that are derived from the name of the tree (set by the Web server host computer and Web server) and a path down the branches to a specific document.

2.6 How the Web Was Won (3)

- **2.6.2 Virtual Path Addressing(2)**
  - In Figure 2.2, the file ***``exp23anal.html''*** obtains the path name: ***"masstrans/Dec99/exp23anal.html"*** your document.
  - ***``http:''*** indicates that a ***hyper text transfer protocol*** server program will be used and that it is located on a computer to be found at the Internet location designated by ***``www.nlab.org.''***
  - The flexibility of the system used actually goes well beyond the level portrayed above. The path addressing scheme which is allowed is actually a ***virtual path*** scheme. What this means is that there need not really be any hierarchy of directories (or folders) in the computer's permanent storage area corresponding to those portrayed by the path name in the URL.

2.6 How the Web Was Won (4)

2.6.3 The Hypertext Markup Language (HTML)

- ***Hypertext markup language,*** HTML, as a universal, simple language for formatting, embedding of graphics, and hypertextual linking of documents.
- With HTML the menu becomes the **document** and **vice versa.**
- ***Markup languages*** was incorporated in HTML.
  - ✓ This was derived from pioneering work that culminated in the creation of ***SGML (Standard Generalized Markup Language).***
  - ✓ ***SGML*** is **a language that can be used to describe other languages that, in turn, describe documents**.
  - ✓ In fact, HTML is defined by an SGML description.
- HTML inherited from this work is that a document structure can be defined in terms of appearance, without specific reference to the display device characteristics.
  - ✓ This went to the heart of problems that had been previously encountered when people who did not share the same computer types, software types, display types, and so forth, tried to share documents.

2.6 How the Web Was Won (5)

2.6.3 The Hypertext Markup Language (HTML) (2)

- For example, consider that I may write a document on a computer that presents the text screen in a format of 132 characters per line, with 80 lines per visible page, using 8 different font styles.
- You, on the other hand, may have a computer and display configured for at most 80 characters per line, 25 lines per page, and 3 font styles.
- These differences can make my document appear very disorganized and confusing on your computer; my longer lines may be chopped off or continued on the next line, in which case the alignment of tables, for example, is completely upset. In addition, important distinctions that I encoded using special fonts may be lost.
- HTML provides a means to describe a document so that text and other document components are ***``reflowed''*** in an appropriate way to fit the display area.
- ***Reflow*** changes the size or shape of the window. The fact that the content remains unconfused over large variations of window sizes is due to the intelligent interpretation of the markup information in HTML.

2.6 How the Web Was Won (6)

2.6.4 Viewer Extensions

- Web browser software incorporates ***helper applications*** and ***plug-ins*** that can be associated with particular file content types, so that the types of documents that a browser can display may be extended easily.
- ***Postscript*** was a language developed by the Adobe corporation to describe exactly how a printer should create a printed copy of that document. Wishing not to waste paper, we sometimes want to see a postscript-described document on our computer screen. Several software applications, called **postscript viewers**, are available to render a postscript document on our display.
- A related capability of newer Web browsers is that of accepting a special kind of viewer known as a ***plug-in***
- For example, a ***spreadsheet plug-in*** would allow us to see a spreadsheet within a text document, alongside the text and graphs that describe the spreadsheet's contents.

2.6 How the Web Was Won (7)

- **2.6. 5 A Network of Distributed Servers**
  - Probably the greatest innovation of the World Wide Web concept was that of a ***loosely interconnected network*** of independent content servers distributed throughout the world's Internet.
  - a bottleneck for the flow of data
  - Instead, **with the World Wide Web, every computer can be a server.** If we are using a professional Web server service and don't like the content space limits or other rules, we can buy a computer and put up our own Web site. When we run out of space, we can simply buy more disk space. With this system there is no single bottleneck, as the servers are as distributed as are the users.

2. 7 The Success of the World Wide Web

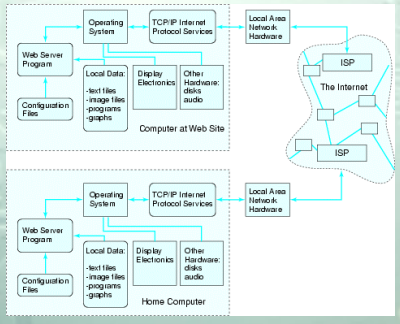
- Fundamentally, the Web provided a new capability that was seen as being both ***useful*** and ***widely applicable***.
- Summarize the capabilities provided by the mechanisms of the Web:
  - 1. The ability to access remote information instantly and conveniently;
  - 2. The potential for every user to be a world wide publisher;
  - 3. The ability to incorporate formatted text, images and, later, interactive components, permitting artistic expression. This was important in attracting widespread participation by individuals and by the commercial sector, which found the Web to be a ready-made canvas for high quality advertising;
  - 4. Multimedia communications capability all channeled through a single device, your computer;
  - 5. Hyperlinks that enable the user to pursue desired topics immediately without regard to location of the source material;
  - 6. Powerful search capabilities to locate desired information anywhere on the Web; and
  - 7. Flexibility and upgrade capability--the fact that different computer architectures running different operating systems can simply display the same information and be upgraded via various plug-ins to deal with new data types.

2. 7 The Success of the World Wide Web (2)

- What were the keys to the success of the World Wide Web concept beyond the notion of useful capabilities as summarized above?
  - The Web uses open (public and unprotected) specifications, and well-constructed and documented protocol standards; thus, a freely distributed sample implementation allowed rapid emulation, deployment, experimentation, and improvement;
  - The Web coopted existing services (e.g., ftp, gopher, and telnet) by supporting their protocols within a new user-friendly context; and
  - While it was believed by some industry pundits that the Web would die under the weight of its distributed nature (``no one in charge,'' ``no obvious way to find anything''), the availability of advertising space (on a huge number of Web pages each of which drew the attention of people who had the resources needed to use the Web and hence probably more resources to spend) led to the funding of information search engines that made the Web work regardless of the number of participants.

2.8 The Structure of the Web

- **Figure 2.3:**Some of the systems that make up the World Wide Web.



2.8 The Structure of the Web (2)

- Web is really a ***system of systems***. Here is a very short and simplified list of the systems that make it possible.
  - **Individual computers** of many makes, models, generations, capabilities and using different and incompatible software operating systems[2.2] such as **PCs, Macs, work stations, Web-TVs, Network Computers** that host the client (Web browser) and server programs.
  - **Electronic document storage**, sometimes just **simple text files** on disk, may include **graphics, audio, video, and executable programs** providing services such as databases.
  - **Means for capture or creation of sounds** (such as audio recording equipment or music synthesis software), **images** (such as electronic cameras, software for drawing and painting**), and video** (such as digital video cameras, virtual reality modeling, rendering and animation software).
  - A computer program called a ***Web server***, which accepts requests for linked files and fulfills these requests by delivering the data stream back to the client browser.

2.8 The Structure of the Web (3)

- Web is really a *system of systems*. Here is a very short and simplified list of the systems that make it possible (2).
  - A computer program called a ***Web browser***, which displays files in the desired format, provides links to other files, and manages details such as indices and print requests.
  - **Local area networks (LANs)** that are used to interconnect many computers within a room or within an entire company. These LANs provide a conduit for information to travel from one computer to the next within the LAN and a means for information to be delivered to and from a connection with an **Internet Service Provider (ISP).**
  - **Physical global communications links** (based on wire and fiber optic based cables and the equipment that places information signals on one end and retrieves them from the other) supplied by telephone companies or companies specializing in delivery of computer data;
  - **Equipment and software** creating organized collections of computer networks and global connections that allow information to be routed properly between the many ISPs and computers and are known collectively as the Internet;

2.8 The Structure of the Web (4)

- Examples of these protocols at work in the system of systems we have been discussing include.
  - **The MIME system** of naming and organizing files so that file locations maybe uniquely identified (Internet addresses, the domain name system, and URLs);
  - **The operation of LANs and ISPs**, founded upon the availability of software that can execute on the connected computers that implement globally understood protocols for the transmission and reassembly of information (such as TCP/IP and HTTP);

2.9 Technologies That Enhance the Power of the Web

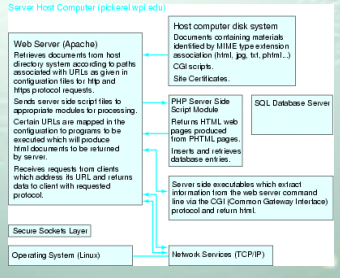
2.9.1 Dynamic, Active, and Interactive Web Pages

- A dynamic Web page is distinguished from a static one in that there is no one unchanging document that you and others will obtain when your browser downloads a document from a certain URL. An example of a dynamic Web page would be one that provides a weather map with the most current radar image, temperature, winds, and weather observations for a given location. These dynamic Web pages are made possible through an early extension of the Web concept known as the ***server side executable*** or sometimes ***server side script***.
- The active Web page approach has **several virtues:**
  - You can interact with the displayed information at much greater speeds than you could have if the program were located at the Web Server location. That is because your computer can react to your inputs nearly instantly. As you will see later in this book, it takes time for information to travel a given distance, and, if we are sharing the equipment and wiring of the Internet with millions of other users,this time delay can become quite large.
  - If hundreds of users are interested in the same information,the server computer's computational resources are similarly divided.By down loading the program that generates the multimedia to your computer, the remotely located server offloads its responsibility for generating that content for your computer.

2.9 Technologies That Enhance the Power of the Web(2)

2.9.1 Dynamic, Active, and Interactive Web Pages (2)

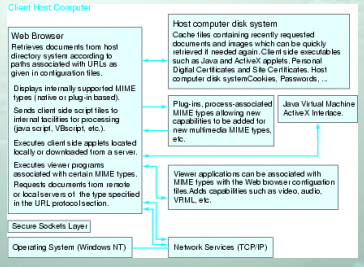
Figure 2.4: Some of the elements that provide the services and data that are delivered by a Web server.



2.9 Technologies That Enhance the Power of the Web(3)

2.9.1 Dynamic, Active, and Interactive Web Pages (3)

Figure 2.5: Some of the elements that provide the services and data that are delivered by a Web browser.



2.9 Technologies That Enhance the Power of the Web(4)

2.9.2 Java and the Web

- A great boon to the concept of the active page was the development of a computer language and computer platform technology known as ***Java***.
- Java was announced by **Sun Microsystems at the Sun World '95 conference in May 1995.** The initial idea that had grown into Java with Sun Microsystems was that a new kind of programming language and system technology was needed if the many pieces of electronics in a home or business were to be someday interconnected for the purposes of sharing information and central control.
- That is, if we ever expected to be able to call our **VCR** on the phone and program it with our **cellular telephone's keypad**, these devices would have to have an entire infrastructure and design philosophy in common.
- Java technology was intended to provide just such a framework and related system construction tools.
- Thus, **Netscape**, the largest web browser and Web server provider at that time, announced their adoption of **Java technology** into their systems on the same day its availability was announced by **Sun**.

2.9 Technologies That Enhance the Power of the Web(5)

2.9.2 Java and the Web(2)

- Java was meant to provide, among other attributes, a means to eliminate the need to know the ultimate destination of our program. That is, the intention was to translate the source code into form that could be transported to any computer and be executed.
- Java revived an old notion, connected with a programming language and compilation system known as UCSD Pascal and the P-code system developed in **the late 1970s by Kenneth Bowles.**
- The idea is not to compile source code to target a specific computer, but to compile it for use with no real computer!
- A program written with the Java language is generally converted into executable software for the ***JavaVirtual Machine*** **(JVM),** a ***simulated*** computer that can be simulated on any other computer platform.
- Downloadable programs that run within the context of your Web browser are referred to as ***``Java Applets.''***