



computer science illuminated

The World Wide Web

Nell Dale & John Lewis

(adaptation by Michael Goldwasser)



The World Wide Web

- The Web is an infrastructure of distributed information combined with software that uses networks as a vehicle to exchange that information
- A **Web page** is a document that contains or references various kinds of data, such as text, images, graphics, and programs
- Web pages also contain **links** to other Web pages so that the user can “move around” as desired



The World Wide Web

- A **Web site** is a collection of related Web pages
- The Internet makes the communication possible, but the Web makes that communication easy, more productive, and more enjoyable



Web Browser

- A **browser** is a software tool that issues the request for the Web page we want and displays it when it arrives
- We often talk about “visiting” a Web site, as if we were going there
 - In truth, we actually specify the information we want, and it is brought to us
 - The concept of visiting a site is understandable in that we often don’t know what’s at a particular site until we “go to it” and see



Web Browser

- The computer that is set up to respond to Web requests is called a **Web server**
- A Web address is the core part of a **Uniform Resource Locator**, or **URL**, which uniquely identifies the page you want out of all of the pages stored anywhere in the world



Web Browser

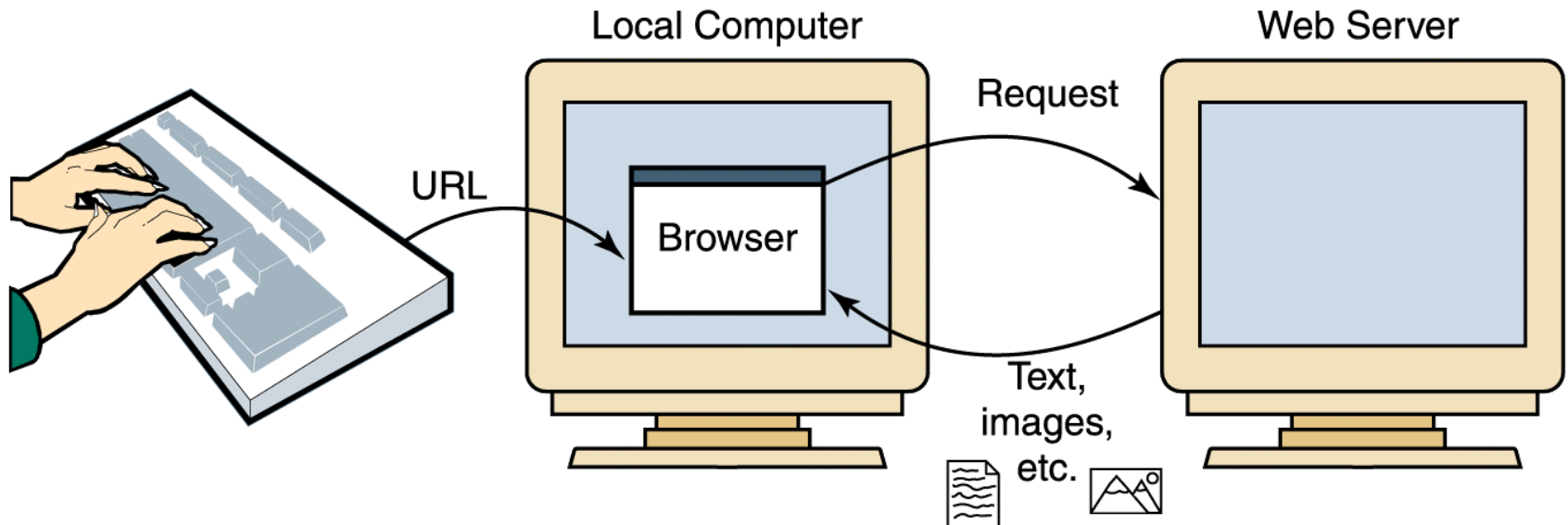


Figure 16.2 A browser retrieving a Web page



HTML

- Web pages are created (or built) using a language called the **Hypertext Markup Language**, or **HTML**
- The term **markup language** comes from the fact that the primary elements of the language take the form of **tags** that we insert into a document to annotate the information stored there



HTML

Final Report
European Conference on Expert Systems

boldface

Center

Submitted by Justin Parker

First of all, our thanks go out to the following sponsors for their support of the conference and its supplemental activities.

Allied Interactive
Sybernetics, Inc.

Dynamic Solutions of New Jersey

make these bullets

The conference was a great success. It ran a full four days, including workshops and special sessions. Subjective feedback from conference attendees was largely positive, and financially the revenues resulted in a surplus of over \$10,000.

Figure 16.2
A marked-up document

HTML



Figure 16.3
The Student Dynamics Web
page as displayed in Netscape
Navigator



HTML

```
<HTML>
  <HEAD>
    <TITLE>Student Dynamics</TITLE>
  </HEAD>
  <BODY>
    <CENTER><IMG SRC="stuDynamics.gif"></CENTER>
    <HR>
    <CENTER><I>A student-based community organization.</I></CENTER>
    <HR>
    <P>Plan to participate in our upcoming events:</P>
    <UL>
      <LI>Visit Children's Hospital (March 23)</LI>
      <LI>Help clean renovated school grounds (April 4)</LI>
      <LI>Continue the <a href="outreach.html">K-12 outreach
        program.</a> <IMG SRC="updated.gif"></LI>
    </UL>
    <P><B><I>You should be involved!</B></I> Help our active and
      energetic team make a difference in the lives of people. Our
      <a href="execBoard.html">executive board</a> is always willing
      to answer any questions you may have.</P>
    <P>How can we help you? What suggestions do you have regarding
      new activities? <a href="suggestions.html">Let us know!</a></P>
    <P>We are always in need of donations: equipment, money, or
      (better yet) you!</P>
    <CENTER><H3>Join us for our Spring Picnic in May!</H3></CENTER>
  </BODY>
</HTML>
```

Figure 16.4
The HTML document
defining the Student
Dynamics Web page



HTML

- Tags are enclosed in angle brackets (`< . . . >`)
- Words such as HEAD, TITLE, and BODY are called elements and specify the type of the tag
- Tags are often used in pairs, with a start tag such as `<BODY>` and a corresponding end tag with a / before the element name, such as `</BODY>`



HTML

- The browser determines how the page should be displayed based on the tags
 - Ignores the way we format the HTML document using carriage returns, extra spaces, and blank lines
 - Takes into account the width and height of the browser window
 - When you resize the browser window, the contents of the Web page are reformatted to fit the new size



Basic HTML Formatting

- The paragraph tags (`<P> . . . </P>`) specify text that should be treated as a separate paragraph
- The center tags (`<CENTER> . . . </CENTER>`) indicate that the enclosed information should be centered in the browser window



Basic HTML Formatting

- The B, I, and U elements are used to indicate that the enclosed text should be bold, italic, or underlined, respectively
- The <HR> tag inserts a horizontal rule (that is, a line) across the page



Basic HTML Formatting

- We often have cause to display a list of items
 - The UL element stands for an unordered list, and the LI element represents a list item
- Several elements are used to define headings in a document
 - There are six predefined heading elements defined in HTML: H1, H2, H3, H4, H5, and H6



Images and Links

- Many tags can contain attributes that indicate additional details about the information or how the enclosed information should be displayed
 - An image can be incorporated into a Web page using the IMG element, which takes an attribute that identifies the image file to display
 - ``



Images and Links (cont.)

- A link is specified using the element A, which stands for anchor
- The tag includes an attribute called HREF that specifies the URL of the destination document.
- For example

```
<A HREF = "http://duke.csc.villanova.edu/docs/">  
Documentation Central!</A>
```



Interactive Web Pages

- When HTML was first developed, there was no way to interact with the information and pictures presented in a Web page
- As users have clamored for a more dynamic web, new technologies were developed to accommodate these requests
- Some (but not all) of the new ideas were offshoots of the newly developed Java programming language



Java Applets

- A **Java applet** is a program that is designed to be embedded into an HTML document and transferred over the Web to someone who wants to run the program
 - An applet is embedded into an HTML document using the `APPLET` tag

```
<APPLET code="MyApplet.class" width=250  
height=150 ></APPLET>
```



Java Applets

- Java programs are compiled into Bytecode, a low-level representation of a program that is not the machine code for any particular type of CPU
- Java applets are restricted as to what they can do
 - The Java language has a carefully constructed security model
 - An applet, for instance, cannot access any local files or change any system settings



Java Server Pages

- A Java Server Page, or JSP, is a Web page that has **JSP scriptlets** embedded in them
- A scriptlet is a small piece of executable code intertwined among regular HTML content



Java Server Pages

- A JSP scriptlet is encased in special tags beginning with `<%` and ending with `%>`
- Imagine JSP scriptlets as having the expressive power of a full programming language

```
<H3>  
<%  
out.println ("hello there");  
%>  
</H3>
```



Java Server Pages

- Note that JSPs are executed on the server side where the Web page resides
- By the time it arrives at your computer, all active processing has taken place, producing a static (though dynamically created) Web page
- JSPs are particularly good for coordinating the interaction between a Web page and an underlying database



XML

- HTML has a predefined set of tags and each tag has its own meaning
- There is nothing about HTML tags that describes the true content of a document
- The **Extensible Markup Language**, or **XML**, allows the creator of a document to describe its contents by defining his or her own set of tags



XML

- XML is a metalanguage
- A **metalanguage** is a language for talking about, or defining, other languages



XML

```
<?xml version="1.0" ?>
<!DOCTYPE books SYSTEM "books.dtd">
<books>
  <book>
    <title>The Hobbit</title>
    <authors>
      <author>J.R.R. Tolkien</author>
    </authors>
    <publisher>Ballantine</publisher>
    <pages>287</pages>
    <isbn>0-345-27257-9</isbn>
    <price currency="USD">7.95</price>
  </book>
  <book>
    <title>A Beginner's Guide to Bass Fishing</title>
    <authors>
      <author>J. T. Angler</author>
      <author>Ross G. Clearwater</author>
    </authors>
    <publisher>Quantas Publishing</publisher>
    <pages>750</pages>
    <isbn>0-781-40211-7</isbn>
    <price currency="USD">24.00</price>
  </book>
</books>
```

- Like HTML, an XML document is made up of tagged data

Figure 16.5 An XML document containing data about books



XML

- **Document Type Definition (DTD):** a specification of the organization of the document
- The structure of a particular XML document is described by its corresponding DTD document

```
<!ELEMENT books (book*) >
<!ELEMENT book (title, authors, publisher, pages, isbn, price)>
<!ELEMENT authors (author+)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT publisher (#PCDATA)>
<!ELEMENT pages (#PCDATA)>
<!ELEMENT isbn (#PCDATA)>
<!ELEMENT price (#PCDATA)>
<!ATTLIST price currency CDATA #REQUIRED>
```

Figure 16.6 The DTD document corresponding to the XML books document



XML

- XML represents a standard format for organizing data without tying it to any particular type of output
- **Extensible Stylesheet Language (or XSL):** A language for defining transformations from XML documents to other output formats

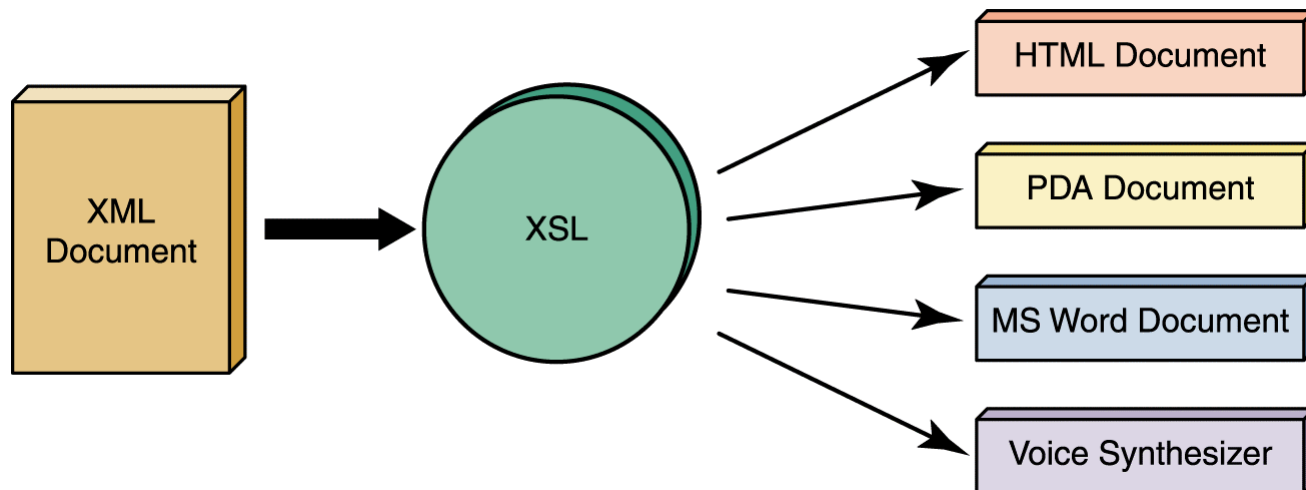


Figure 16.7
An XML document
can be transformed
into many output
formats



Ethical Issues: Cookies

- Internet cookies are very small text files (usually about 50 to 150 bytes) that are downloaded from a Web server to a Web browser
- Cookies were intended to allow a server to maintain information about those that visit it
 - For example, the cookie could contain a user ID and password for a Web site so that the information would not have to be reentered every time the Web site is visited



Ethical Issues: Cookies

- Companies, called targeted marketing companies, have evolved that sell cookie services
- These companies pay Web sites to send them copies of the cookies for each transaction and thus build up a profile of a user's Internet habits
- This information is used for marketing and advertising purposes.
- Some people like the idea of having marketing and advertising targeted to their interests
- Others feel that this is an invasion of privacy