Chapter Goals

• Compare and contrast the Internet and the World Wide Web
• Describe general Web processing
• Write basic HTML documents
• Describe several specific HTML tags and their purposes
Chapter Goals

• Describe the processing of Java applets and Java server pages
• Compare and contrast HTML and XML
• Define basic XML documents and their corresponding DTDs
• Explain how XML documents are viewed
The World Wide Web

- **The Web** An infrastructure of distributed information combined with software that uses networks as a vehicle to exchange that information.

- **Web page** A document that contains or references various kinds of data, such as text, images, graphics, and programs.

- **Links** A connection between one web page and another that can be used “move around” as desired.
The World Wide Web

• **Website**  A collection of related web pages

• The Internet makes the communication possible, but the Web makes that communication easy, more productive, and more enjoyable
Search Engines

• **Search Engine** A website that helps you find other websites
  – For example, Yahoo and Google are search engines
  – You enter keywords and the search engine produces a list if links to potentially useful sites

• There are two types of searches
  – Keyword searches
  – Concept-based searches
Instant Messaging

- **Instant messaging** (IM) An application that allows people to send and receive messages in real time
  - Both sender and receiver must have an IM running
  - Most IM applications use a proprietary protocol that dictates the precise format and structure of the messages that are sent across the network to the receiver.
  - Instant messages are not secure
Cookies

• **Cookie** A small text file that a web server stores on your local computer’s hard disk
  – A cookie contains information about your visit to the site
  – Cookies can be used
    • to determine number of unique visitors to the site
    • to customize the site for your future visits
    • to implement shopping carts that can be maintained from visit to visit
  – Cookies are not dangerous
Web Browser

- **Browser** 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 website, 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

• **Web server**  The computer that is set up to respond to web requests

• **Web address**  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.
Final Report
European Conference on Expert Systems
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

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 16.3
The Student Dynamics Web page as displayed in Netscape Navigator
Figure 16.4
The HTML document defining the Student Dynamics Web page
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>
The browser determines how the page should be displayed based on the tags.

The browser:

- 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.
- Reformats the contents to fit your browser window.
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.
• 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
    – `<IMG SRC = "myPicture.gif">`
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

• Many of the new ideas were offshoots of the newly developed Java programming language
Java Applets

- **Java applet** 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

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

• A browser has a built-in interpreter that executes the applet, allowing the user to interact with it.
  – Consider the difficulties inherent in this situation
  – How can we execute a program that was written on one type of computer on possibly many other types of computers?
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

• **Scriptlet** 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

```html
<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
• 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

• **Metalanguage**  A language for talking about, or defining, other languages

• XML is a metalanguage
XML

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

```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>
```

Figure 16.5 An XML document containing data about books.
• **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

```xml
<!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