HTML

Basic theory about HTML

What is the World Wide Web?

- The World Wide Web (WWW) is most often called the Web
- The Web is a network of computers all over the world
- All the computers in the Web can communicate with each other.
- All the computers use a communication standard called HTTP (Hypertext Transfer Protocol)

How does the WWW work?

- Web information is stored in documents called Web pages
- Web pages are text files stored on computers called Web servers
- Computers reading the Web pages are called Web clients
- Web clients view the pages with a program called a Web browser
- Popular browsers are: Internet Explorer, Netscape Navigator/Communicator, Firefox, Safari, Mozilla, Konqueror, and Opera
- Other browsers are: Omniweb, iCab, etc.

How does the browser fetch pages?

- A browser fetches a Web page from a server by sending a request
- A request is a standard HTTP request containing a page address
- A page address looks like this: http://www.someone.com/page.html
- A page address is a kind of URL (Uniform Resource Locator)

How does the browser display pages?

- All Web pages are ordinary text files
- All Web pages contain display instructions
- The browser displays the page by reading these instructions.
- The most common display instructions are called HTML tags
- HTML tags look like this:
  This is a Paragraph

Who makes the Web standards?

- The Web standards are *not* made up by Netscape or Microsoft
- The rule-making body of the Web is the W3C
- W3C stands for the World Wide Web Consortium
- W3C puts together specifications for Web standards
- The most essential Web standards are HTML, CSS and XML
- The latest HTML standard is XHTML 1.0

What is an HTML File?

- HTML stands for Hypertext Markup Language
- An HTML file is a text file containing small markup tags
- The markup tags tell the Web browser how to display the page
- An HTML file must have an htm or html file extension
  - .html is preferred
  - .htm extensions are used by servers on very old operating systems that can only handle "8+3" names (eight characters, dot, three characters)
- An HTML file can be created using a simple text editor
  - Formatted text, such as Microsoft Word's .doc files, cannot be used in HTML files

HTML Tags

- HTML tags are used to mark up HTML elements
- HTML tags are surrounded by angle brackets, < and >
- Most HTML tags come in pairs, like and
- The tags in a pair are the start tag and the end tag
- The text between the start and end tags is the element content
- The tags act as containers (they contain the element content), and should be properly nested
- HTML tags are not case sensitive; means the same as <B>
- XHTML tags *are* case sensitive and must be *lower case*
  - To ease the conversion from HTML to XHTML, it is better to use lowercase tags

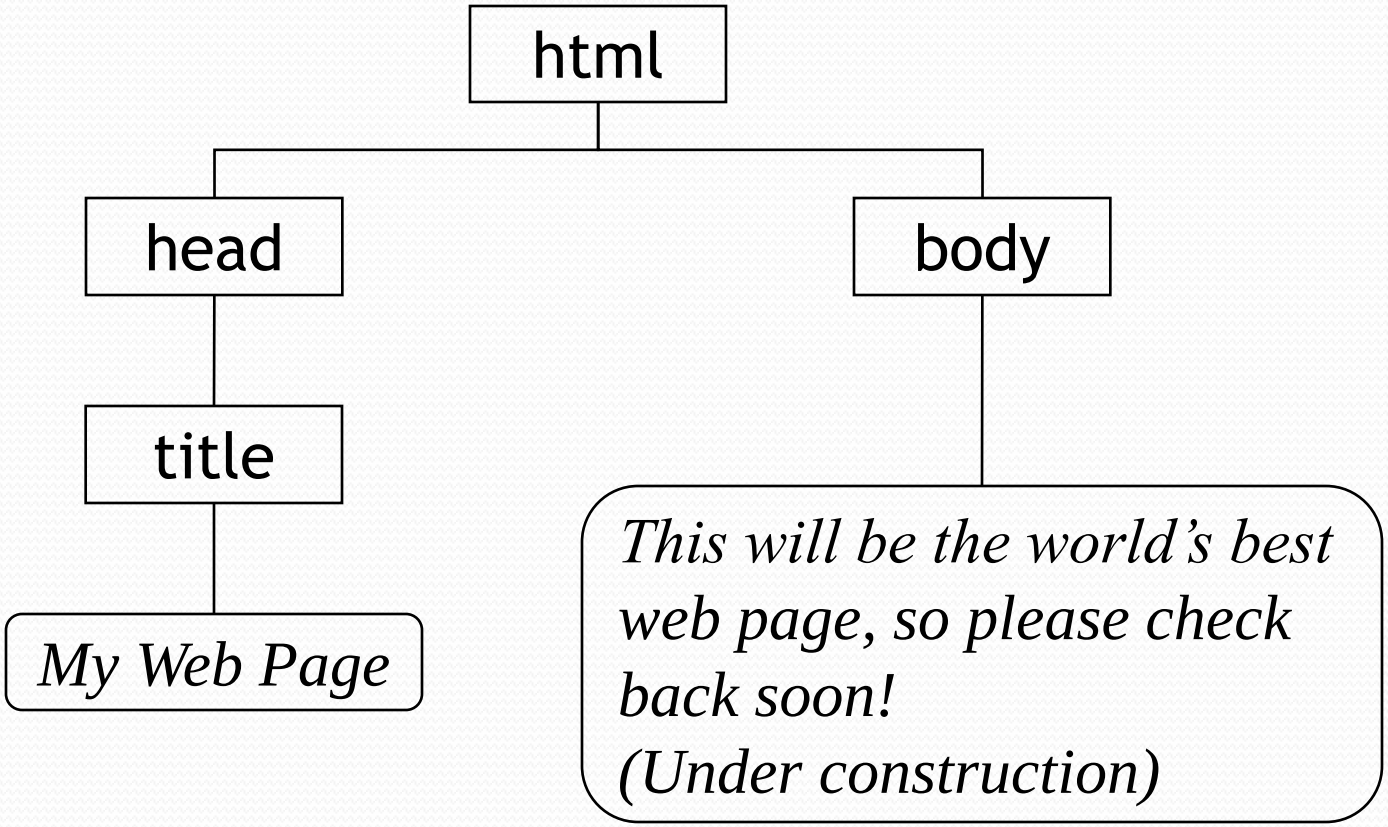
Structure of an HTML document

- An HTML document is contained within <html> tags
  - It consists of a <head> and a <body>, in that order
  - The <head> typically contains a <title>, which is used as the title of the browser window
  - Almost all other content goes in the <body>

- Hence, a fairly minimal HTML document looks like this:

```
<html>
  <head>
    <title>My Title</title>
  </head>
  <body>
     Hello, World!
  </body>
</html>
```

HTML documents are trees



Text in HTML

- Anything in the body of an HTML document, unless marked otherwise, is text
- You can make text *italic* by surrounding it with `<i>` and `</i>` tags
- You can make text **boldface** by surrounding it with `<b>` and `</b>` tags
- You can put headers in your document with `<h1>`, `<h2>`, `<h3>`, `<h4>`, `<h5>`, or `<h6>` tags (and the corresponding end tag, `</h1>` through `</h6>`)
  - `<h1>` is quite large; `<h6>` is very small
  - Each header goes on a line by itself

Whitespace

- Whitespace is any non-printing characters (space, tab, newline, and a few others)
- HTML treats all whitespace as word separators, and automatically *flows* text from one line to the next, depending on the width of the page
- To group text into paragraphs, with a blank line between paragraphs, enclose each paragraph in `<p>` and `</p>` tags
- To force HTML to use whitespace exactly as you wrote it, enclose your text in `<pre>` and `</pre>` tags ("pre" stands for "preformatted")
  - `<pre>` also uses a monospace font
  - `<pre>` is handy for displaying programs

Lists

- Two of the kinds of lists in HTML are ordered, <ol> to , and unordered, <ul> to
- Ordered lists typically use numbers: 1, 2, 3, …
- Unordered lists typically use bullets (•)
- The elements of a list (either kind) are surrounded by <li> and

- Example:

```
The four main food groups are:
<ul>
    <li>Sugar</li>
    <li>Chips</li>
    <li>Caffeine</li>
    <li>Chocolate</li>
</ul>
```

Attributes

- Some markup tags may contain attributes of the form *name*="*value*" to provide additional information
- Example: To have an ordered list with letters A, B, C, … instead of numbers, use <ol type="A"> to
  - For lowercase letters, use type="a"
  - For Roman numerals, use type="I"
  - For lowercase Roman numerals, use type="i"
  - In this example, type is an attribute

Links

- To link to another page, enclose the link text in to
  - Example: I'm taking Dr. Dave's CIT597 course this semester.
  - Link text will automatically be underlined and blue (or purple if recently visited)
- To link to another part of the same page,
  - Insert a named anchor: References
  - And link to it with: My references
- To link to a named anchor from a different page, use My references

Images

- Images (pictures) are not part of an HTML page; the HTML just tells where to find the image
- To add an image to a page, use:
  `<img src="URL" alt="text description" width="150" height="100">`
  - The `src` attribute is required; the others are optional
  - Attributes may be in any order
  - The *URL* may refer to any `.gif`, `.jpg`, or `.png` file
  - Other graphic formats are not recognized
  - The `alt` attribute provides a text representation of the image if the actual image is not downloaded
  - The `height` and `width` attributes, if included, will improve the display as the page is being downloaded
    - If `height` or `width` is incorrect, the image will be distorted
  - There is no `</img>` end tag, because `<img>` is *not* a container

Tables

- Tables are used to organize information in two dimensions (rows and columns)
- A <table> contains one or more table rows, <tr>
- Each table row contains one or more table data cells, <td>, or table header cells, <th>
  - The difference between <td> and <th> cells is just formatting--text in <th> cells is boldface and centered
- Each table row should contain the same number of table cells
- To put borders around every cell, add the attribute border="1" to the <table> start tag

Example table

```
<table border="1">
  <tr>
    <th>Name</th>  <th>Phone</th>
  </tr>
  <tr>
    <td>Dick</td>  <td>555-1234</td>
  </tr>
  <tr>
    <td>Jane</td>  <td>555-2345</td>
  </tr>
  <tr>
    <td>Sally</td>  <td>555-3456</td>
  </tr>
</table>
```

Name	Phone
Dick	555-1234
Jane	555-2345
Sally	555-3456

More about tables

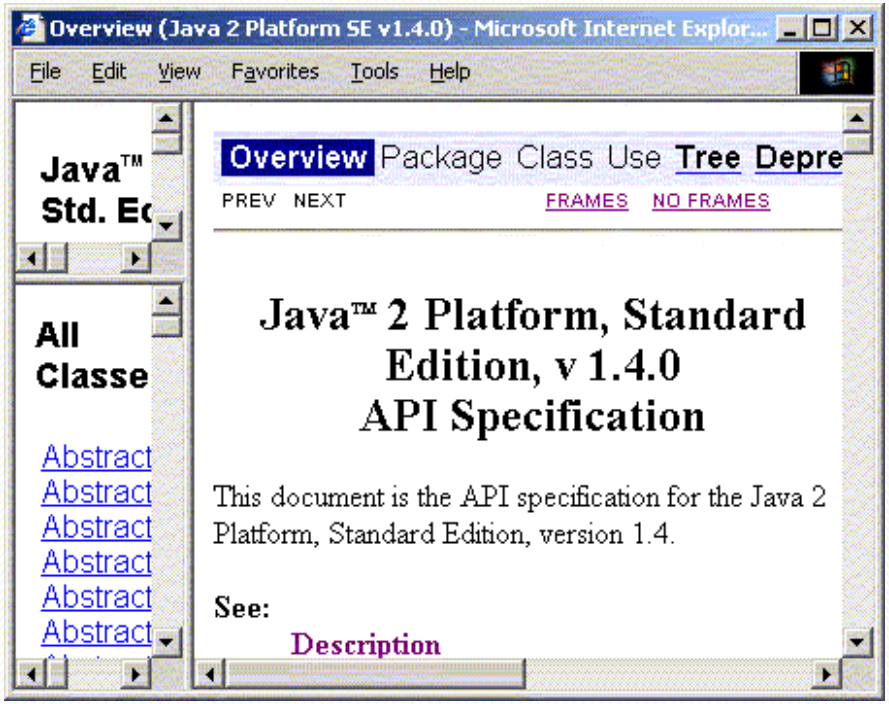
- Tables, with or without borders, are excellent for arranging things in rows and columns
  - Wider borders can be set with border="*n*"
  - Text in cells is less crowded if you add the attribute cellpadding="*n*" to the <table> start tag
- Tables can be nested within tables, to any (reasonable) depth
  - This is very convenient but gets confusing
- Tables, rows, or individual cells may be set to any background color (with bgcolor="*color*")
  - Columns have to be colored one cell at a time
  - You can also add bgcolor="*color*" to the <body> start tag

Entities

- Certain characters, such as <, have special meaning in HTML
- To put these characters into HTML without any special meaning, we have to use entities
- Here are some of the most common entities:
  - < represents <
  - > represents >
  - & represents &
  - ' represents '
  - " represents "
  -   represents a "nonbreaking space"--one that HTML does *not* treat as whitespace

Frames

- Frames are a way of breaking a browser window up into "panes," and putting a separate HTML page into each pane
  - The Java API is an example of a good use of frames



Framesets

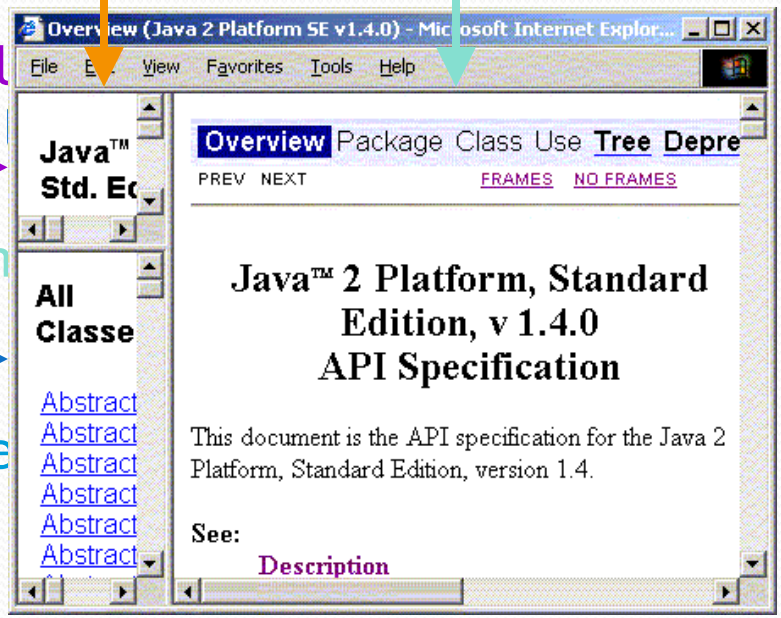
- Frames are enclosed within a frameset
- Replace <body>…</body> with <frameset>…</frameset>
  - Within the <frameset> start tag, use the attributes:
    - rows=*row_height_value_list*
    - cols=*col_width_value_list*
  - The value lists are comma-separated lists of values, where a value is any of:
    - *value*% – that percent of the height or width
    - *value* – that height or width in pixels (usually a bad idea)
    - * – everything left over (use only once)
- Example: <frameset cols="20%,80%">

Adding frames to a frameset

- Put as many <frame> tags within a <frameset> as there are rows or columns
  - <frame> is not a container, so there is no </frame> end tag
- Each <frame> should have this attribute:
  - src=*URL* – tells what page to load
- Some optional tags include:
  - scrolling="yes|no|auto" (default is "auto")
  - noresize
- Within a <frameset> you can also put <noframes>*Text to display if no frames*</noframes>

Example: The Java API

```
<HTML>
  <HEAD>
    <TITLE>Java 2 Platform SE v1.4.0</TITLE>
  </HEAD>
  <FRAMESET cols="20%,80%">
    <FRAMESET rows="30%,70%">
      <FRAME src="overview-frame.html
      <FRAME src="allclasses-frame.html
    </FRAMESET>
    <FRAME src="overview-summary.htm
  </FRAMESET>
  <NOFRAMES>
    <H2>If you see this, you have frame
  </NOFRAMES>
</HTML>
```



The rest of HTML

- HTML is a large markup language, with a lot of options
  - None of it is really complicated
  - I've covered only enough to get you started
  - You should study one or more of the tutorials
  - Your browser's View -> Source command is a great way to see how things are done in HTML
  - HTML sometimes has other things mixed in
  - There is no such "thing" as DHTML (Dynamic HTML)
    - DHTML is simply HTML with several other technologies mixed in, such as forms and JavaScript, some of which we will cover
    - If something on an HTML page doesn't look like HTML, it probably isn't--so don't worry about it for now

Vocabulary

- WWW: World Wide Web
- W3C: World Wide Web Consortium
- HTML: Hypertext Markup Language
- URL: Uniform Resource Locator