A CONDENSED INTRODUCTION TO HTML/XHTML

(TEXT CHAPTER 2)

Origins and Evolution of HTML

- HTML was defined with SGML
- Original intent of HTML: for general layout control of documents that could be displayed by a wide variety of computers
- Recent versions:
 - HTML 4.0 1997
 - » Introduced many new features and deprecated many older features
 - HTML 4.01 1999 A cleanup of 4.0
 - XHTML 1.0 2000
 - » 4.01 redefined using XML, instead of SGML
 - XHTML 1.1 2001
 - » Modularized 1.0, and drops frames

Origins and Evolution of HTML

(continued)

- Reasons to use XHTML vs. HTML:
 - 1. HTML has lax syntax rules, leading to sloppy and ambiguous documents
 - XHTML syntax is more strict, leading to *clean* and *clear* documents in a standard form
 - 2. HTML processors do not even enforce the few syntax rule that do exist in HTML
 - 3. The syntactic correctness of XHTML documents can be validated
- However ... (don't take them yet!)

The Opposite Side: HTML vs. XHTML

- Although the value of the consistent and coherent syntax rules of XHTML were widely recognized and accepted, but its draconian error handling was not so accepted, thus XHTML documents are served as HTML a compromise.
- Therefore, in 2009, work on XHTML 2.0 was stopped; W3C took over development of HTML5
- In our textbook, HTML5 is readopted, but with XHTML syntax rules (which has advantage)
 - So in my lectures

Basic Syntax

- Elements (as containers of contents) are defined by tags (markers)
 - Tag format:
 - » Opening tag: <name>
 - » Closing tag: </name>
 - The opening tag and its closing tag together specify a container for the *content* they enclose

Basic Syntax (continued)

- Not all tags have content
 - If a tag has no content, its form is <name />
- The container and its content together are called an element
- If a tag has attributes, they appear between its name and the right bracket of the opening tag
- Comment form: <!-- ... -->
- Browsers ignore comments, unrecognizable tags, line breaks, multiple spaces, and tabs
- Tags are only <u>suggestions</u> to the browser, even if they are recognized by the browser
- In XHTML, element & attribute names must be lowercase letters
- In HTML, can be any combination of uppercase & lowercase

Standard HTML Document Structure

• Every HTML document must begin with:

```
<!DOCTYPE html>
```

- <html>, <head>, <title>, and <body> are required
- The whole document must have <html> as the root
- html in an XHTML doc "must" have the xmlns attribute:

```
<html xmlns = "http://www.w3.org/1999/xhtml">
```

- A document consists of a head and a body
- The <title> tag is to give the document a title, normally displayed in the browser's window title bar (at the top of the display)
- The meta tag is used to provide the character set used

```
<meta charset = "utf-8" />
```

Basic Text Markup

- Text is normally placed in paragraph elements
- Paragraph Elements
 - The tag breaks the current line and inserts a blank line the new line gets the beginning of the content of the paragraph
 - The browser puts as many words of content as will fit in each line

```
• e.g.:
```

- Line breaks
 - The visual effect of the
br /> tag is same as that of , except for the blank line
- Example of paragraphs and line breaks

```
On the plains of hesitation  bleach the bones of countless millions  <br/> <br/> who, at the dawn of victory <br/> > sat down to wait, and waiting, died.
```

Typical display of this text:

On the plains of hesitation

bleach the bones of countless millions

who, at the dawn of victory sat down to wait, and waiting, died.

To Preserve whitespace, simply use elements

Headings

- Six sizes, 1 6, specified with <h1> to <h6>
- 1, 2, and 3 use font sizes that are larger than the default font size
- 4 uses the **default size**
- 5 and 6 use smaller font sizes

```
<!-- headings.html
  An example to illustrate headings
   -->
<html xmlns = "http://www.w3.org/1999/xhtml">
 <head> <title> Headings </title>
 </head>
 <body>
  <hi>Aidan's Airplanes (hı) </hi>
  <h2> The best in used airplanes (h2) </h2>
  <h3> "We've got them by the hangarful" (h3) </h3>
  <h4> We're the guys to see for a good used airplane (h4) </h4>
  <h5> We offer great prices on great planes (h5) </h5>
  <h6> No returns, no guarantees, no refunds, all sales are final (h6) </h6>
 </body>
</html>
```

```
Aidan's Airplanes (h1)

The best in used airplanes (h2)

"We've got them by the hangarful" (h3)

We're the guys to see for a good used airplane (h4)

We offer great prices on great planes (h5)

No returns, no guarantees, no retunds, all sales are final! (h6)
```

- Blockquotes
 - Content of <blockquote>
 - To set a block of text off from the **normal flow** and appearance of text
 - Browsers often indent, and sometimes italicize
- Font Styles and Sizes (can be nested)
 - Boldface
 - Italics
 - Larger <big>
 - Smaller <small>
 - Monospace <tt> or <code> (often displayed in Courier)
- To preserve whitespace: , otherwise,

```
The <big> sleet <big> in <big> <i> Crete </i> <br/> <br/> in <br/> in <br/> in </big> the street
```

The sleet in Crete

lies completely in the street

- » These tags are not affected if they appear in the content of a <blockquote>, unless there is a conflict (e.g., italics)
- Superscripts and subscripts
 - » Subscripts with <sub>
 - » Superscripts with <sup>

Example: x₂³

Display: x_2^3

- Inline versus block elements (What are they?)
 - Block elements CANNOT be nested in inline elements like <h2>

- The above font size and font style stuff can be done with style sheets, but these tags are not yet deprecated
- Character Entities

<u>Char.</u>		Entity Meaning
&	&	Ampersand
<	<	Less than
>	>	Greater than
"	"	Double quote
,	'	Single quote
$\frac{1}{4}$	¼	One quarter
1/2	½	One half
3/4	¾	Three quarters
0	°	Degree
(spa	ce)	Non-breaking space
@	©	Copyright
€	€	Euro

- Horizontal rules: <hr /> draws a line across the display, after a line break
- The meta element (for search engines) used to provide additional information about a document, with attributes, but no content!
- E.g.,
 <head>
 <meta name="description" content="Free Web tutorials" />
 <meta name="keywords" content="HTML,CSS,XML,JavaScript" />
 <meta name="author" content="Hege Refsnes" />

<meta http-equiv="Content-Type" content="text/html;charset=ISO-8859-1" />

</head>

Images

- GIF (Graphic Interchange Format)
 - 8-bit color (256 different colors)
- JPEG (Joint Photographic Experts Group)
 - 24-bit color (16 million different colors)
- Both use compression, but JPEG compression is better
- Images are inserted into a document with the tag with the src attribute
 - The alt attribute is required by XHTML for purposes of:
 - 1.Non-graphical browsers
 - 2. Browsers with images turned off

```
<img src = "comets.jpg" alt = "Picture of comets" />
```

- The tag has 30 different attributes(!), including width and height (in pixels)
 - <u>Alphabetical Tag List of HTML5</u> (a nice tool to quickly find details of each tag!!)
- Portable Network Graphics (PNG)
 - Relatively new, big than JPEG files, but not data loss
 - Should eventually replace both gif and jpeg

Images (continued): An Example – Image.html

```
<!-- image.html
      An example to illustrate an image -->
 <html xmlns = "http://www.w3.org/1999/xhtml">
   <head> <title> Images </title>
   </head>
   <body>
     <h1> Aidan's Airplanes </h1>
     <h2> The best in used airplanes </h2>
     <h3> "We've got them by the hangarful"
\langle h2 \rangle / \text{Special of the month } \langle /h2 \rangle
>
• 1960 Cessna 210 <br />
   577 hours since major engine overhaul
  <br />
   1022 hours since prop overhaul
• <br /><br />
       <imq src = "c210new.jpg" alt = "Picture of a Cessna 210"/>
       <br />
       Buy this fine airplane today at a
       remarkably low price <br />
       Call 999-555-1111 today!
     <q\>
   </body>
 </html>
```

2-17

Images (continued)

Aidan's Airplanes

The best in used airplanes

"We've got them by the hangarful"

Special of the month

1960 Cessna 210 577 hours since major engine overhaul 1022 hours since prop overhaul



Buy this fine airplane today at a remarkably low price Call 999-555-1111 today!

Hypertext Links

- Hypertext is the essence of the Web!
- A link is specified with the href (hypertext reference)
 attribute of <a> (the anchor tag)
 - The content of <a> is the visual link in the document
 - If the target is a whole document (another one), the target need not be specified in the target document as being the target
- Note: Relative addressing of targets is easier to maintain and more portable than absolute addressing!

Hypertext Links: an Example

<!-- link.html An example to illustrate a link <head> <title> Links </title> </head> <body> <h1> Aidan's Airplanes </h1> <h2> The best in used airplanes </h2> <h3> "We've got them by the hangarful" </h3> <h2> Special of the month </h2> 1960 Cessna 210
 Information on the Cessna 210 </body> </html>

Hypertext Links (continued)

- If the target is not at the beginning of a doc, it must be marked with a **label**
- **Target labels** can be defined in many different tags with the **id** attribute, as in <h1 **id** = "**baskets**"> Baskets </h1>
- The link to an id must be preceded by a pound sign (#); if the id is in the same document, this target can simply be

```
<a href = "#baskets"> What about baskets? </a>
```

- If the target is <u>in a different document</u>, the <u>document reference</u> must be included Baskets
- **Style note**: a link should **blend in** with the surrounding text, so *reading without* taking the link should not be made less pleasant
- Links can be an image (instead of textual names), for example,

 Info on C210

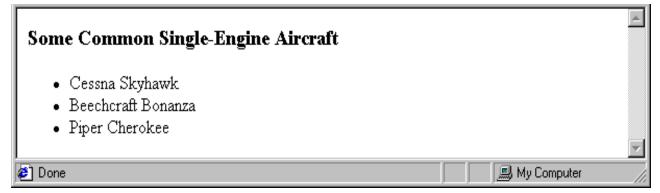
Lists: 3 types of lists

- Unordered lists
- The list is the content of the
 tag
- List elements are the content of the tag

```
<h3> Some Common Single-Engine Aircraft </h3>

    Cessna Skyhawk 
    Beechcraft Bonanza 
    Piper Cherokee 

    Aircraft 
    Common Single-Engine Aircraft 
    Chall
```



- Ordered lists
 - The list is the content of the
 tag
 - Each item in the display is preceded by a sequence value

Lists (continued)

```
<h3> Cessna 210 Engine Starting Instructions
</h3>

 Set mixture to rich 
 Set propeller to high RPM 
 Set ignition switch to "BOTH" 
 Set auxiliary fuel pump switch to
    "LOW PRIME" 
 When fuel pressure reaches 2 to 2.5
    PSI, push starter button 
Cessna 210 Engine Starting Instructions

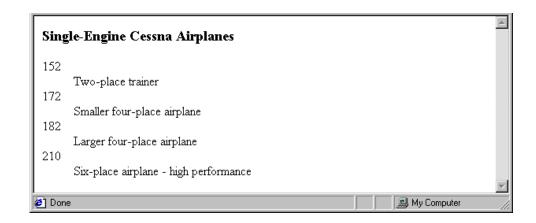
    Set mixture to rich

                                  Set propeller to high RPM
                                  Set ignition switch to "BOTH"
                                  Set auxiliary fuel pump switch to "LOW PRIME"
                               5. When fuel pressure reaches 2 to 2.5 PSI, push starter button
                                                                                💻 My Computer
                             ଌ 🕽 Done
```

- Nested lists
 - Any type of list can be nested inside any type of list
 - The nested list must be in a list item

Lists (continued)

- Definition lists (for glossaries, etc.)
 - List is the content of the <d1> tag
 - Terms being defined are the content of the <dt> tag
 - The definitions themselves are the content of the <dd> tag



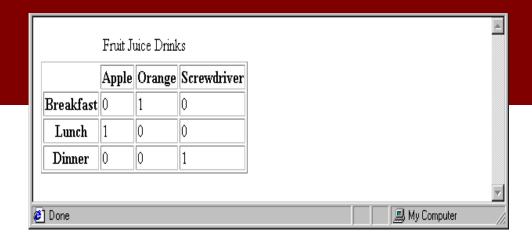
Tables

- A table is a <u>matrix of cells</u>, each possibly having content
- The cells can include almost any element
- Some cells have row or column labels and some have data
- A table is specified as the content of a tag
- A **border** attribute in the tag specifies a border between the cells
- If border is set to "border", the browser's <u>default width</u> border is used
- The border attr. can be set to a number, which will be the border width
- Without the border attribute, the table will have **no border lines**!
- Tables are given titles with the <caption> tag, which must immediately follows

- Each row of a table is specified as the content of a
 tag
- The heading cells are specified as the content of a tag
- The data cells are specified as the content of a tag

```
<caption> Fruit Juice Drinks </caption>

   Apple 
   Orange 
   Screwdriver 
  Breakfast 
  0 
  1 
  0 
  Lunch 
  1 
  0 
  0
```



- "Supper cells" may take up space of multiple cells
 - If so, the colspan attribute must be set in the tag to specify that the label must span some number of columns
- .
- Fruit Juice Drinks
- •



• Similarly the rowspan attribute allows covering of multi-rows

```
  Fruit Juice Drinks
  Apple 
  Orange 
  Screwdriver
```

- The <u>align</u> attribute controls the horizontal placement of cell contents
 - Values are left (default), right, and center
 - align is an attribute of >, >, and elements
- The valign attribute controls the vertical placement
 - Values are top, bottom, and middle or center (default)
 - valign is an attribute of and elements
 - → The cellspacing attribute of is used to specify the distance between cells in a table
- The cellpadding attribute of is used to specify the spacing between cell content and the inner wall of the cell

Colorado is a state of contrasts. The eastern half is a mostly treeless prairie. On the prairie, trees grow only in the Platte and Arkansas river valleys, with a few found along some other small streams. The forested Rocky Mountains rise abruptly from the high plains about midway from east to west and cover most of the western half of the state. There are 54 mountains in Colorado that top 14,000 feet.

South Dakota is somewhat similar to Colorado in that it is a mostly treeless prairie in the east, but has a range of forested mountains in the west. But in South Dakota, the mountains, named the Black Hills, lie only in the far western part of the state and rise to only a little over 7500 feet. However, they are still the highest mountains east of the Rockies in the U.S. The famous Mount Rushmore is nestled in the middle of the Black Hills.

Table (continued): New Features in HTML5

- Table Sections

- Header, body, and footer, which are the elements of table: thead, tbody, and tfoot
- If a document has *multiple* thody elements, they are separated by thicker horizontal lines

- Uses of Tables

- In the past, tables were used to align elements in rows and columns for general (page) layout
- That use of tables is now frowned upon!
- Nowadays the pros tend to use Cascading Style Sheets to place elements in rows and columns as general layout
- Use tables only when the information is naturally tabular!

Forms

- A form is the usual way for server to receive data from browser
- HTML has tags to create a collection of objects that accomplish this <u>information gathering</u>
 - The objects are called *widgets* (e.g., radio buttons and checkboxes)
- When the Submit button of a form is clicked, the form's values are sent to the server
- All of the widgets are defined in the content of a <form> tag
 - The only required attribute of <form> is action, which specifies the URL of the application that is to be called when the *Submit* button is clicked
 - E.g.: action = "http://www.cs.ucp.edu/survey.php"
 - » If the form has no action, the value of action is an empty string

- The method attribute of <form> specifies the method of transferring the form data to the server, such as get and post
- The enctype attribute
 - As a quick example, the following form-data will be sent unencoded:

```
<form action="form_action.asp" method="get" enctype="text/plain">
First name: <input type="text" name="fname" /><br />
Last name: <input type="text" name="lname" /><br />
  <input type="submit" value="Submit" />
  </form>
```

- The enctype attribute specifies how form-data should be encoded before sending it to the server.
- By default, enctype="application/x-www-form-urlencoded". This means that all characters are encoded before they are sent (<u>spaces</u> are converted to "+" symbols, and <u>special characters</u> converted to <u>ASCII Hexadecimal</u> values).

Widgets

- -Many are created with the <input> tag
 - » The type attribute of <input> specifies the kind of widget being created

1.Text

- Creates a horizontal text box for text input
- Default size is 20; it can be changed with the size attribute
- If more characters entered than will fit, the box is scrolled (shifted) left
- If you don't allow the user to type more characters than will fit, set maxlength, which causes excess input to be ignored

```
<input type = "text" name = "Phone" size = "12" >
```

» Password – just like text except asterisks are displayed, rather than the input characters

- 2. Checkboxes to collect multiple choice input
 - » Every checkbox <u>requires a value attribute</u>, which is the widget's value in the form data when the checkbox is 'checked'
 - A unchecked checkbox contributes no value to the form data
 - » By default, no checkbox is initially 'checked'
 - » To initialize a checkbox to 'checked', the checked attribute must be set to "checked"

[example on next slide]

```
Grocery Checklist
<form action = "">
  >
  <input type = "checkbox" name ="groceries"</pre>
          value = "milk" checked = "checked">
   Milk
  <input type = "checkbox" name ="groceries"</pre>
          value = "bread">
   Bread
  <input type = "checkbox" name = "groceries"</pre>
          value= "eggs">
   Eggs
  </form>
             Grocery Checklist

☑ Milk □ Bread □ Eggs
```

- 3. Radio Buttons collections of checkboxes in which only one button can be 'checked' at a time
 - Every button in a radio button group MUST have the same
 name

• Example on next slide

3. Radio Buttons (continued)

```
- If no button in a group is 'pressed', the browser often 'presses' the first one
Age Category
<form action = "">
  >
  <input type = "radio" name = "age"</pre>
   value = "under20" checked = "checked"> 0-19
  <input type = "radio" name = "age"</pre>
          value = "20-35" > 20-35
  <input type = "radio" name = "age"</pre>
          value = "36-50" > 36-50
  <input type = "radio" name = "age"</pre>
          value = "over50"> Over 50
  </form>
            ⊙ 0-19 ○ 20-35 ○ 36-50 ○ Over 50
```

4. Menus - created with <select> tag

- There are two kinds of menus, those that behave like *checkboxes* and those that behave like *radio buttons* (the default)
 - Checkboxes-like menus must include the multiple attribute set to "multiple"
- The <u>name</u> attribute of <select> is required
- The <u>size</u> attribute of <select> can be included to specify the number of menu items to be <u>displayed</u> (the default is 1)
 - If size > 1 or multiple is specified, the menu is displayed as a "pop-up" menu (otherwise "drop-down")

4. Menus (continued)

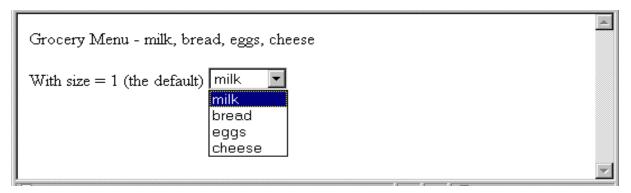
- Each item of a menu is specified with an <option> tag, whose pure text content is the value of the item
- An <option> tag can include the selected attribute, which when assigned "selected" specifies that the item is pre-selected

```
Grocery Menu - milk, bread, eggs, cheese
<form action = "">
 >
    With size = 1 (the default)
    <select name = "groceries">
      <option> milk </option>
      <option> bread </option>
     <option> eggs </option>
     <option> cheese </option>
    </select>
  </form>
```

- Widgets (continued)



After clicking the menu:

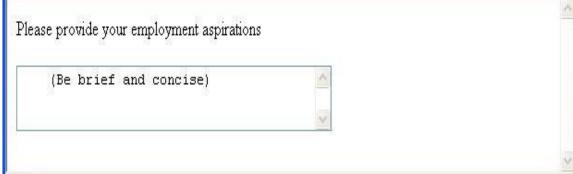


After changing size to 2:



5. Text areas - created with <textarea>

- Usually for bigger chunk of text, using rows and cols attributes to specify the size of the text area
- Default text can be included as the content of <textarea>
- Scrolling is implicit if the area is overfilled



6. Reset and Submit buttons

– Both are created with <input>

```
<input type = "reset" value = "Reset Form">
<input type = "submit" value = "Submit Form">
```

Submit has two actions:

- 1. Encodes the data of the form
- 2. <u>Requests</u> that the server to execute the server-resident program specified as the value of the action attribute of <form>
- A Submit button is required in every form

Next we will introduce a few new tags added by HTML5

HTML5 ADDED TAGS

SOUTHERN ILLINOIS UNIVERSITY

- The audio Element
 - Prior to HTML5, a plug-in was required to play sound while a document was being displayed
 - Audio encoding algorithms are called *audio* codecs e.g., MP3, Vorbis
 - Coded audio data is stored in containers—e.g., Ogg, MP3, and Wav (file name extension indicates the container, not the audio code)
 - Vorbis code is stored in Ogg containers
 - MP3 code is stored in MP3 containers
 - Way code is stored in Way containers

- The audio Element (continued)
 - General syntax:

- Browser chooses the first audio file it can play and skips the content; if none, it displays the content
- Different browsers have different audio capabilities
- The *controls* attribute, which is set to "controls", creates a start/stop button, a clock, a progress slider, total time of the file, and a volume slider

- The video Element
 - Prior to HTML5, there was no standard way to play video clips while a document was being displayed
 - Video codecs are stored in containers
 - Video codecs:

H.264 (MPEG-4 AVC) – can be stored in an MPEG-4 container

Theora – can be stored in any container

VP8—can be stored in any container

- Different browsers support different video codecs
 - -The width and height attributes set the screen size
 - -The autoplay attribute, set to "autoplay", specifies that the video should play as soon as it is ready
 - -The preload attribute, set to "preload", specifies that the video should be loaded as soon as the document is loaded
 - -The controls attribute, set to "controls", is like that of the audio element

- Organizational Elements
 - Header Elements

- Footer Elements
 - footer a container for footer information (within <body> elem)

```
<footer>
  &copy; The Podunk Press, 2012
  <br />
  Editor in Chief: Squeak Martin
</footer>
```

- -The section Element a container for sections
- -The article Element a container for self-contained part of a document (from another source)
- The aside Element a container for tangential info
- The nav Element navigation sections (list of links)

Example of <nav>

```
<nav>
 <a href="/html/">HTML</a> |
 <a href="/css/">CSS</a> |
 <a href="/js/">JavaScript</a> |
 <a href="/jquery/">jQuery</a>
</nav>
```

- The time Element

- For putting a time stamp on a document
- Two parts, text and machine-readable (datetime)
- The machine-readable part: given by the datetime attribute (optional)
 - Date part: 4-digit year, a dash, 2-digit month, a dash, 2-digit day of the month ("2012-08-29")
 - Time (optional) format: T09:00
- The text part: given as the content of <time>

```
<time datetime = "2012-08-29T09:00">
   August 8, 2012 9:00 am
</time>
```

- The two parts need not specify the same date
- Deficiencies:
 - 1. Dates prior to the Christian era are not possible
 - 2. No approximations

Syntactic Differences: XHTML vs. HTML

- Case sensitivity (see notes page)
- Closing tags
- Quoted attribute values
- Explicit attribute values
- id and name attributes
- Element nesting