# SGG 3643
# Computer Programming III

## Text Formatting in HTML

Ivin Amri Musliman

# About this Chapter

- Deals with tags that affect the appearance of your Web pages, including

  - appearance of individual characters

  - how to display special characters

  - create preformatted text

  - align text horizontally

  - change the font and font size

  - other text elements: line breaks, rule lines, addresses, and quotations

# About this Chapter

- Also, it considers
  - HTML extensions
  - when to use various tags, especially when there are alternative approaches to achieving the same effect

# HTML Document: Basic Structure

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 2.0//EN"
          "http://www.w3.org/TR/xhtml2/DTD/xhtml2.dtd">
<html xmlns="http://www.w3.org/2002/06/xhtml2">
<head>
<title>A study of population dynamics</title>
        ... other head elements...
</head>
<body>
        ... document body...
</body>
</html>
```

# Line breaks, spaces in your file

- Line breaks, tabs and consecutive empty spaces, inserted in the HTML code of your page, **do not modify the layout** of the web page in the browser.

- However, it will be **easier for you to understand** your HTML code.

# Textual structure of a [scientific] document

- Summary
- Title
- Information about the author, imprint
- Table of contents
- headings
- paragraphs, line breaks
- fonts
- colors
- lists (enumeration)
- tables

# Table of content

- No direct support by HTML, therefore:

- Tool supported or manual generation of hyperlinks

- General form:

```
<a
href="http://www.gismngt/format/format.htm#standa
rds"> ... Standards... </a>
```

# Information about the author, imprint

- No direct support given by HTML

- General recommendation:
  - Name of contact person
  - Terms of use, Copyright and trademark information, deny of liability

How to Suggest a Site | Company Info | Privacy Policy | Terms of Service | Jobs | Advertise With Us | Help
Copyright © 2004 Yahoo! Inc. All rights reserved. Copyright Notice.

Manage Your Profile | Contact Us | Subscribe to Microsoft This Week! | Careers

©2004 Microsoft Corporation. All rights reserved. Terms Of Use | Trademarks | Privacy Statement | Legal

# Headings

- Headings for hierarchical document structure are defined with the <h1> to <h6> tags. <h1> defines the largest heading. <h6> defines the smallest heading.

- HTML automatically adds an extra blank line before and after a heading.

```
<h1>This is a heading</h1>
<h2>This is a heading</h2>
<h3>This is a heading</h3>
<h4>This is a heading</h4>
<h5>This is a heading</h5>
<h6>This is a heading</h6>
```

# Paragraphs

- Paragraphs are defined with the <p> tag.

- HTML automatically adds an extra blank line before and after a paragraph.

```
<p>This is a paragraph.</p>
<p>This is another paragraph.</p>
```

# Comments

- The comment tag is used to insert a comment in the HTML source code. A comment will be ignored by the browser. You can use comments to explain your code, which can help you when you edit the source code at a later date.

```
<!-- This is comment <br /> inside HTML source.-->
```

- Comments are sometimes used for dynamic generation or modification of web pages by web authoring tools, i.e. Server Side Includes.
  à Be careful with comments not generated by you!

# Line breaks: **<br />**

- is used when you want to end a line, but don't want to start a new paragraph - similar to Shift-Enter in Microsoft Word.
- forces a line break wherever you place it.
- The <br /> tag is an empty tag. It has no closing tag.
- If appearing within a paragraph, <br /> doesn't logically create another paragraph.

```
<p>This is a line break<br /> inside a
paragraph.</p>
```

# Character Styles

- affect the appearance of individual characters or words within a larger block (such as a paragraph or heading)

- Two main types, logical and physical

  - Logical styles deal with how text is <u>used</u>

  - Physical styles deal with how text should <u>appear</u>

- Logical styles are more in the spirit of HTML, but physical styles still seem to be used much more

- Each style has a start and end tag: <xx>...</xx>

# Character Styles: Logical Styles

- Emphasis: <em> (Often implemented as italic)

- Strong: <strong> (Often boldface; [> <em>])

- Code: <code> (Fixed-width font, say, Courier)

- Sample: <samp> (Sample text--similar to <code>)

- Keyboard: <kebd> (Text to be entered by user)

- Variable: <var> (Name of replaceable "variable")

- Definition: <dfn> (for word being defined--v. 3.2)

- Cite: <cite> (A short quote or citation)

# Character Styles: Logical Styles II

- Abbreviation: <abbr>

- Acronym: <acronym>

# Character Styles: Physical Styles

- Bold: `<b>`    [HTML v. 2.0]
- Italic: `<i>`    [v. 2.0]
- Teletype writer: `<tt>` [v. 2.0]  (Monospaced font)
- Underline: `<u>` [v. 3.2, but deprecated in v. 4.0]
- Strike-thru: `<s>` [v. 3.2, but deprecated in v. 4.0]
- Big: `<big>` [v. 3.2]  (Bigger than surrounding text)
- Small:  `<small>` [v. 3.2]  (Smaller than surrounding text)
- Subscript:  `<sub>` [v. 3.2]
- Superscript:  `<sup>` [v. 3.2]

# Formatted text

```
<b>This text is bold</b><br />
<strong>This text is strong.</strong><br />
<big>This text is big.</big><br />
<em>This text is emphasized.</em><br />
<i>This text is italic.</i><br />
<small>This text is small.</small><br />
This text contains <sub>subscripts.</sub><br />
This text contains <sup>superscripts.</sup>
```

**This text is bold**

**This text is strong**

**This text is big**

*This text is emphasized*

*This text is italic*

This text is small

This text contains subscript

This text contains superscript

# Inserted / deleted text

- Most browsers will overstrike deleted text and underline inserted text.

- Some older browsers will display deleted or inserted text as plain text.

This is deleted text.

This is inserted text.

# Character Styles: Physical Styles

- You can combine styles, say, to form bold-italic text
- Some browsers may not give you exactly what you are requesting, both in nested styles and in individual styles.

- Generally spoken: In all media you should avoid combination of such styles!

# Citations, Quotations, and Definition Tags

- <address> Defines an address element

- <bdo> Defines the text direction

- <blockquote> Defines a long quotation

- <q> Defines a short quotation

- <cite> Defines a citation

- <dfn> Defines a definition term

# Quotation:  <blockquote> and <cite>

- <cite> … </cite> highlights small quotes
- <blockquote> … </blockquote>: for larger quotes that are to be set off from the surrounding text, perhaps with indenting
  - Can use <br> to break across lines
  - Avoid using <blockquote> within paragraphs.
- The q element does not render as anything special.

```
Here comes a long quotation:
<blockquote>
This is a long quotation. This is a long quotation.
This is a long quotation. This is a long quotation.
This is a long quotation.
</blockquote>

Here comes a short quotation:
<q>
This is a short quotation
</q>
```

Here comes a long quotation:

> This is a long quotation. This is a long quotation. This is a long quotation. This is a long quotation. This is a long quotation.

Here comes a short quotation: " This is a short quotation "

# "Computer output" tags

- <code> Defines computer code text

- <kbd> Defines keyboard text

- <samp> Defines sample computer code

- <tt> Defines teletype text

- <var> Defines a variable

- <pre> Defines preformatted text

# Preformatted Text: <pre>…</pre>

- Presented the way it appears in the HTML document

- Usually presented in a monospaced font

- Excellent for programming examples or for visually aligned character pictures

- OK for simple tables.

- Can embed character styles & links, but not headings or paragraphs

- Avoid tabs - use spaces instead

# Preformatted text

- The pre tag is good for displaying computer code.

```
<pre>
This is
preformatted text.
It preserves       both spaces
and line breaks.
</pre>
<p>The pre tag is good for displaying computer
code:</p>
<pre>
for i = 1 to 10
     print i
next i
</pre>
```

```
This is
preformatted text.
It preserves       both spaces
and line breaks.

The pre tag is good for displaying computer code:

for i = 1 to 10
     print i
next i
```

# Horizontal Rule:  <hr />

- Draws a horizontal line
- Permits attributes [v. 3.2], but deprecated in v. 4.0
- Size="2"   [Value is **thickness** in pixels]
- Width="80%"  or  Width="50"  [Percentage of line, number of pixels, respectively]   Which is better?
- Align="Center"  [The default.  Can use Left or Right]
- Noshade  [Draw as a plain line, without 3-D effect]
    - In XHTML, you say noshade="noshade"

# Address: <address>

- Used for signature-like entities on Web pages, often at the bottom (following `<hr />`)

- Often displayed in italics

- You can make the entity as long as you wish, adding line breaks as needed with `<br />`

- Question: Given that you can use `<i>` and avoid this tag, why do they have it in the language?

# Special Characters

- How does one get characters that aren't on a standard keyboard to display on a page?
  - For example, ©, ä, Ø

- How does one get characters displayed that may be in an HTML document but are not usually displayed ("<" ">" "&")?
  - For example, how would you get the following sentence displayed on a Web page?

    I like using the <br> tag but not the <!-- tag.

# Special Characters

- HTML defines character entities [called *escape sequences* in most other contexts] for specifying characters

- Two forms of character entities

  - named entities (more convenient, but only used for the most common symbols)

  - numbered entities (less convenient, but can be used with most special symbols)

- You must use entities for <, >, &, etc.

- A character entity has three parts: an ampersand (&), an entity name or a # and an entity number, and finally a semicolon (;).

# Special Characters: Named Entities

- Begin with an & and end with a ;
- Case sensitive [all lower case]
- Examples: &agrave; &quot; &copy; (...right)
- Examples:   &lt; &gt; &times;
- Examples: &deg; (degree sign) &divide;


- Question: How would one code a sign for "two slides back"?

# Non-breaking Space

- The most common character entity in HTML is the non-breaking space.

- Normally HTML will truncate spaces in your text. If you write 10 spaces in your text HTML will remove 9 of them.
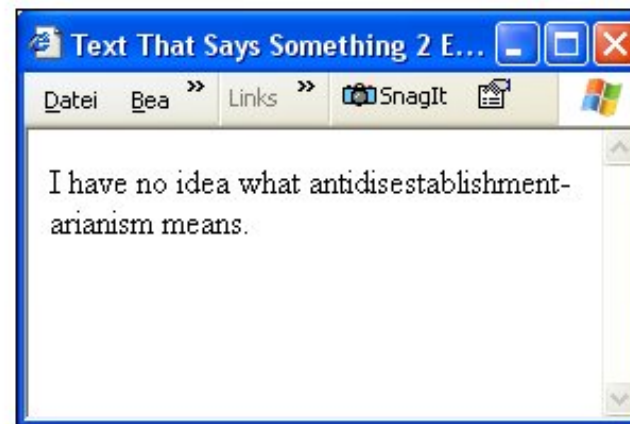
  -> To add spaces to your text, use the ` ` character entity.

# Soft Hyphens with &shy;

- Soft Hyphens are used to indicate a point in a word where you would like it to be split on to two lines if that is necessary. It simply makes for a nicer appearance when space is limited such as when you have text in a thin column.

<p>I have no idea what antidisestablishment&shy;arianism means.</p>

# Named Entities

| Result | Entity Name | Entity Number |
|--------|-------------|---------------|
| non-breaking space |   |   |
| < less than | &lt; | &#60; |
| > greater than | &gt; | &#62; |
| & ampersand | &amp; | &#38; |
| " quotation mark | &quot; | &#34; |
| ' apostrophe | | &#39; |

http://www.w3.org/MarkUp/Guide/Advanced.html

# Named Entities

| Result | Entity Name | Entity Number |
|---|---|---|
| ¢ cent | &cent; | &#162; |
| £ pound | &pound; | &#163; |
| ¥ yen | &yen; | &#165; |
| § section | &sect | &#167; |
| © copyright | &copy; | &#169; |
| ® registered trademark | | |
| | &reg; | &#174; |
| × multiplication | &times; | &#215; |
| ÷ division | &divide; | &#247; |

# Special Characters: Numbered Entities

- Begin with an &# and end with a ;

- Uses a unique numeric code in between

- Codes correspond with ISO-Latin-1 characters

- HTML 4.0 proposes Unicode as an option

- Examples:  &#60; [&lt;]   &#38; [&amp;]

- Example:  &#123;braces&#125;

  – Of course, an easier way to say that is:  {braces}

- Example:  r&eacute;sum&#233;  [résumé]

# Named vs. Numbered Entities

- To display a less than sign in an HTML document we can use **&lt;** or **&#60;**

- The advantage of using a name instead of a number is that a name is easier to remember. The disadvantage: not all browsers support the newest entity names, while the support for entity numbers is very good in almost all browsers.

# Preventing Spam Mail

- HTML-Maillinks like

  ```
  <a href="mailto:name@xyz.de">name@xyz.de</a>
  ```

  can be used by spammers to find more email addresses.

- Some prevention can be achieved writing:
  ```
  <a
  href="&#109;&#97;&#105;&#108;&#116;&#111;&#58;&#110;&#97;&#109;&#101;&#64;&#102;&#105;&#114;&#109;&#97;&#46;&#100;&#101;">&#110;&#97;&#109;&#101;&#64;&#102;&#105;&#114;&#109;&#97;&#46;&#100;&#101;</a>
  ```

Source: http://www.2andfro.de/mailto-encoder/download.htm

# Horizontal Text Alignment

- Left, Center, Right  [Starting with v. 3.2]
- Deprecated - Version 4.0 still supports it, but doing it in style sheets preferred
- Can use the Align attribute in headings or paragraphs:
  - <p  align="center"> …
  - <h4  align="right"> …
- <center> </center> also works to center

# Horizontal Text Alignment: Blocks of Elements

- <div> … </div> around a block of text, say, a header and several paragraphs, creates a "division"

- The <div> tag supports several attributes, including align. It applies to all parts of the division

- <div> works on items that don't allow an "align" attribute

- "Align" attributes on entities within <div> override the div's align attribute for those entities

- *The <div> tag is important in combination with style sheets.*

# Fonts and Font Sizes

- The <font>…</font> tags were introduced in
  v. 3.2 but **deprecated** in v. 4.0 (style sheets)

  – Original attribute was size

  – Later, face (for "typeface") was added

  – Color was also added

- Size attribute can use absolute & relative sizes

  – Absolute:  <font  size="7">  (go from 1 - 3 - 7)

  – Relative:  use +/- within the quotes

# Fonts and Font Sizes

- The deprecated <basefont> tag can set a default size. Subsequently, the relative changes can be in terms of the basefont size.

- Color: <font color="red">Apple</font>

- <u>Face</u> attribute introduced officially in v. 3.2
  - Works in v. 4.0, but style sheet specification preferred

- Value is a prioritized list of typeface names, e.g.,

  <font face="Futura,Helvetica">...

# <blink> .. </blink>

- This is a browser specific extension
- It is used quite a lot on the Web
- Not all versions of browsers cause the text to blink
- It's a good idea that doesn't really work
- *I suggest you don't use it.  Ever.*

# <nobr> ... </nobr>

- Opposite of <br>
- Tells the browser *not* to do a line break on the text between the symbols
- Use it on things you want to appear all on the same line
- Keep the non-breaking text short if possible (why?)

# <wbr>

- The word-break tag normally would be placed within <nobr>…</nobr>

- The tag is ignored if the text would normally fit on the line

- *None of <nobr>, </nobr>, <wbr> is in the HTML 3.2 or 4.0 standard. They are accepted by IE and Netscape and they appear in XHTML 1.0*