SIENA*college*
Computer Science

Computer Science 180
Web Design
Siena College
Fall 2011

Topic Notes: Basic HTML

Our first topic, and your first lab, involves the basics of the language that underlies the World Wide Web: the *HyperText Markup Language (HTML)*.

Why a Markup Language?

Before we look at some specifics of HTML, let's consider the idea of a *markup language* and why it matters.

The term comes from the days of editors manually "marking up" manuscripts with revisions and instructions for formatting when the text would later be professionally typeset.

The idea of a markup language today is similar in that it, in part, gives instructions for the presentation of text.

On the web: Wikipedia Article "Markup language" at
`http://en.wikipedia.org/wiki/Markup_Language`

As you know, web pages are much more than collections of text. Like documents you may have created using a word processor, there is additional information about fonts, font sizes and styles and colors, organization into paragraphs, probably with section headers, maybe some tables or even graphics.

Most likely, the documents you have written have been developed in a *What You See Is What You Get (WYSIWYG)* word processor. You choose which fonts to use where and make other formatting decisions using menu items and buttons within the program. You certainly expect that all of this formatting is retained when you save your document and reopen it or print it.

This means your word processor must have some internal representation, beyond the text itself, to be able to keep track of all of this formatting and to remember it. This requires a formally-specified *file format* that the program can use to write the document to a file and be able to read it back in to recreate the document.

There are a number of file formats used for word processors. Files of a given format are often given a particular *file extension*. Common ones for word processors are `.doc`, `.docx`, `.odt`.

In many ways, HTML is another file format with goals similar to the word processor formats. HTML files typically have the file extension `.html` or `.htm`. In fact, most modern word processors have an option to save your document in HTML format.

Unlike many word processor file formats, an HTML file is just text. Special text patterns, called *elements* (which are made up of or delimited by *tags*), are embedded within the readable text to indicate details of formatting.

HTML was designed for and is primarily used to represent formatted text for presentation in web browsers. In general, the HTML gives guidelines for how to format the file's contents, but the browser has some flexibility about exactly how it is displayed.

Viewing Page Source

Unlike many markup languages and file formats, the HTML source used to render a page is not hidden or "secret" in any way. The HTML is transmitted from a web server to your browser for display. While most browser users rarely if ever care to do so, browsers to allow you to view the HTML source of pages they are displaying. The actual menu item and/or keystroke are browser-specific.

Basic HTML Elements

We will start by building a simple HTML file using some of the most basic HTML elements.

There are many online references for HTML elements, including a complete and well-organized list on Wikipedia:

On the web: Wikipedia Article "HTML element" at
`http://en.wikipedia.org/wiki/HTML_element`

We will use most of these at some point during the semester, but for now, we focus on just a handful that will allow us to build some simple, but complete, web pages.

To get started, we will build a simple web page using several basic HTML elements:

- The entire document is contained inside of an `<html>` ... `</html>` pair of tags. This is the *root element* of the HTML document.
- The header information is contained in a `<head>` ... `</head>` pair.
- The main content of the document is contained in a `<body>` ... `</body>` pair.
- The document's title (typically displayed in the browser window's title bar) is specified inside of a `<title>` ... `</title>` pair.
- Paragraphs of text to be displayed are contained in a `<p>` ... `</p>` pair.
- A line break within a paragraph can be forced with the `<br>` tag.
- Section headers of different sizes are contained in tag pairs `<h1>` ... `</h1>` for the largest headers, `<h2>` ... `</h2>` for the next largest, on down to `<h6>` ... `</h6>` for the smallest.
- Ordered (numbered) lists are contained in the tag pair `<ol>` ... `</ol>`, while unordered (bullet) lists are contained in a `<ul>` ... `</ul>` pair. List items in either type of list are contained in a `<li>` ... `</li>` pair.

- Some important formatting elements:
  - `<center>` ... `</center>` will center justify content.
  - `<hr>` will draw a horizontal line.
  - `<pre>` ... `</pre>` will display preformatted text. Whereas whitespace and line breaks in the HTML source usually have no meaning for the rendering of the content in a browser, whitespace and line breaks are preserved in text contained in this tag pair.
  - `<b>` ... `</b>` will request a bold font for the text contained between the pair.
  - `<i>` ... `</i>` will request an italic font for the text contained between the pair.
  - `<tt>` ... `</tt>` will request a fixed-width (typewriter) font for the text contained between the pair.

- An HTML *comment* – text that is included in the HTML source but which should not be displayed by the browser looks like this:

```
<!-- This is an HTML comment -->
```

- What would a web page be without graphics in addition to text? The `<img>` tag inserts an image. The following will retrieve an image in a file named "`myimage.jpg`" from the same web server and folder on the web server as the page in which it is specified:

```
<img src="myimage.jpg" alt="my image">
```

  The `src=` and `alt=` parts are called *attributes* of the tag. Many tags, including the ones we have seen already, allow specification of a variety of attributes. The text in the `alt` attribute is what should be displayed if the graphic cannot be loaded or displayed for any reason.

- Finally, an *anchor* element is what puts the "HT" in "HTML". This is how a hyperlink is specified. The following will create a hyperlink in a document such that when the text "My favorite class" is clicked, the browser will retrieve and display the document at `http://courses.teresco.org/cs180_f11`.

```
<a href="http://courses.teresco.org/cs180_f11">My favorite class</a>
```

There are other basic elements that we will consider later, but these should get you going.