**3 XHTML Elements and Rules**

XHTML Elements and Rules

At the very beginning of an XHTML document, you should place the following:  
  
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">  
  
The first line is an XML declaration that indicates to the browser the version of XML used for the document's code; that is, it is an XML document that uses custom tags as defined by the W3C's XHTML specification along with the 8-bit Unicode character set, the one most commonly used today. If a browser is going to be able to recognize and correctly display the tags contained in the document, it needs to know about the XHTML specification. The second and third lines in the code above give this information to the browser. These lines (which may be written all on one line) make up what is called a Document Type Definition (DTD). As the name implies, the DTD specifies what type of document this is (in this case an XHTML document). The DOCTYPE declaration lets designers specify the DTD file that has the definitions of all the tags and attributes used within the XHTML document.  
  
The next line in the code involves the <html> tag. It becomes:  
  
<html xmlns="http://www.w3.org/1999/xhtml">  
  
The xmlns stands for XML namespace, which helps the browser keep track of what all the custom XML tags mean, especially if the same name is used for different tags. A namespace lessens the confusion by providing the browser with information on the meaning of the tags in the document. By adding the above xmlns information to the <html> tag, we are telling the browser that all tags contained within the <html> element belong to the XHTML namespace, as defined by the W3C and located at the given URL. It should be noted that our textbook gives the namespace as:  
  
<html xmlns="http://www.w3.org/TR/xhtml1">  
  
The validation services do not currently accept this as a valid namespace.

At this point, you might ask the question: "What about the millions of existing web pages that follow the old rules? Will they suddenly cease to work?" Because there is such a large installed base of web pages created under the old rules of HTML version 4 and before, it is likely that the major PC browsers will continue to recognize old HTML code and display it properly. But as browsers are designed for other new devices, it's likely that only XHTML-based web pages will work on them. Many web designers are already using XHTML rules for their new web pages and rewriting their previous web pages to be compatible with XHTML. Fortunately the gap between XHTML and HTML is not that great. I must confess though that because of time constraints, many of the web pages and examples given in the course content modules are in done HTML4, NOT in XHTML.

For the remainder of your XHTML document, as long as you adhere to the following rules, you'll be writing proper XHTML:

* XHTML documents must be well-formed
* XHTML documents must have a <title> element within the header section
* Elements must nest and not overlap
* All element and attribute names must be in lowercase
* All tags must close
* All attributes must have values enclosed within quotes

**XHTML documents must be well-formed:**

With the exception of a frameset page, a well-formed XHTML document has an <html> root element that contains two embedded, non-overlapping sections, namely "head" and "body". So, the shell of an XHTML document would look like this (with .... referring to the content of each section):  
  
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">  
<html xmlns="http://www.w3.org/1999/xhtml">  
<head>  
....  
....  
</head>  
<body>  
....  
....  
</body>  
</html>

**XHTML documents must have a <title> element within the header section:**

The <title> element has been around since HTML 2.0, and although recommended for any web page, it is not required by the browser. If omitted, the viewer will see 'untitled' in the browser's title bar. In XHTML, the <title> element is required. For example:  
  
<head>  
<title>Steve Rubin's Homepage</title>  
</head>

**Elements must nest and not overlap:**

Most web browsers don't care if you overlap elements by writing code such as the following, in which the <i> tag starts within a paragraph element and ends outside it:  
  
<p>The following word is written in italics: <i>italics</p></i>  
  
Technically, overlapping tags are illegal (albeit widely tolerated by web browsers) in HTML as well. However, as mentioned previously, XHTML documents must be well-formed. One day soon Web browsers will enforce the XHTML standard and will not "guess" what is meant when the XHTML statement syntax is wrong. If you always write your closing element (and section) tags in reverse of the order in which you wrote the opening tags, your elements (and sections) will not overlap. The following is correct:  
  
<p>The following word is written in italics: <i>italics</i></p>

**All element and attribute names must be in lowercase:**

HTML, unlike XHTML, is case insensitive. For example <BODY> is the same as <body> within an HTML document. As a result, many authors try to make the code within their web page files more readable by writing element names in all capital letters to set HTML tags off from text content and attributes (written in lowercase). For example, the following would be correct using HTML:  
  
<BODY BGCOLOR="yellow">  
<P>This is a paragraph</P>  
  
Although easy to pick out, none of the above tags are valid in XHTML. Because XHTML is case-sensitive and the XHTML DTD contains elements in lowercase, you must write XHTML tags in lowercase. Otherwise, the browser will ignore the undefined tag. The following would be correct using XHTML:  
  
<body bgcolor="yellow">  
<p>This is a paragraph</p>  
  
In XHTML, although you must write attribute names in lowercase, you can write attribute values in either upper or lowercase. For example, the following is valid in XHTML, even though I don't recommend it.  
  
<body bgcolor="YELLOW">

**All tags must close:**

Many authors have used the <p> tag to separate paragraphs when using HTML. The <p> tag essentially inserts a blank line and moves to the left margin. For example:  
  
This is the first paragraph<p>  
This is the second paragraph<p>  
This is the third paragraph  
  
This practice is NOT valid in XHTML. You must close each paragraph container that you open with a <p> tag. The following would be correct in XHTML.  
  
<p>This is the first paragraph</p>  
<p>This is the second paragraph</p>  
<p>This is the third paragraph</p>  
  
In fact all container tags, such as the <p> tag must have a corresponding closing tag, such as </p>. This is true for all container tags, such as list (<ul></ul>), list item (<li></li>), table (<table></table>), etc.   
  
In HTML, it is possible and more syntactically correct to replace the <p> tag with the <br> tag. For example:  
  
This is the first paragraph<br><br>  
This is the second paragraph<br><br>  
This is the third paragraph  
  
In XHTML, you must terminate empty tags such as the <br> tag with a space followed by a forward slash ("/"). For Example: <br />  
  
This requirement to close empty tags also extends to all other tags, which would not normally close in HTML. For example:  
  
<img src="xyz.gif" />  
<meta name="keywords" content="Steve Rubin, web design, critical thinking" />  
<bgsound="xyz.wav" />

**All attributes must have values enclosed within quotes:**

You will find plenty of tags within HTML documents where the author assigned numeric (and sometimes single-word text) values to attributes without enclosing the value in quotes. For example:  
  
<body bgcolor=blue>  
<img src=xyz.gif width=200 height=300>  
<map name=europe>  
  
All of the above would be incorrect in XHTML. XHTML requires that all attribute values (both numeric and text) be enclosed within quotes. For Example:  
  
<body bgcolor="blue">  
<img src="xyz.gif" width="200" height="300" />  
<map name="europe">  
  
In addition, XHTML requires that all attributes have a value. HTML lets you use a shortcut called *attribute minimization* when using an attribute that can have only one possible value. When allowed, attribute minimization lets you omit the attribute's value and specify only its name within the HTML tag. For example, the following is correct in HTML:  
  
<input type="checkbox" name="choice" value="europe" checked>  
  
The above example is incorrect in XHTML. In XHTML, it would be correctly coded as follows:  
  
<input type="checkbox" name="choice" value="europe" checked="checked" />  
  
Thus, in XHTML, we are required to state each attribute's value explicitly (and within quotes), even when only one value is possible.

**Summary:**

To repeat, when I talk about HTML in this course, I mean HTML version 4, which is the current version of the HTML language. XHTML is the newest version of HTML, essentially, HTML version 5. XHTML stands for Extensible Hypertext Markup Language. XHTML has much stricter syntax than HTML4, but is otherwise virtually identical.

NOTE: Feel free to use either HTML or XHTML in your Homework Assignment Web Pages, but I encourage you to only use XHTML in most of your Homework Assignment Web Pages, if not in all of them.  
  
NOTE: The examples presented from this point on in the course content modules will reflect XHTML.  
  
A full discussion of XHTML is given in our textbook in Chapter 16.  
  
You can convert your HTML code to XHTML or correct your XHTML code, using this [W3C service](http://cgi.w3.org/cgi-bin/tidy), called tidy. Before using tidy, I suggest inserting at the top of your HTML file the following:  
  
<?xml version="1.0" encoding="UTF-8"?>  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"  
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">  
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">  
  
You should validate your XHTML code at <http://validator.w3.org/>.   
  
To check whether you have broken links (especially non-clickable links), go to [Link Checker from W3C](http://validator.czweb.org/link-checker.php).  
  
Check here for [CSS Validator](http://jigsaw.w3.org/css-validator/).

**Strict XHTML:**

For this course, I recommend using a a *transitional* XHTML DOCTYPE. If you want the most rigorous XHTML validation then you should use a *strict* XHTML DOCTYPE by using the following:  
  
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">