HTML Overview

With an emphasis on XHTML

What is HTML?

- Stands for HyperText Markup Language
- A client-side technology (i.e. runs on a user's computer)
- HTML has a specific set of tags that allow:
  - the structure of a document to be described
    - e.g., `<h1>` - large heading element
  - Content to be added to a page
    - e.g. `<img>` - image element
  - links to other documents on web defined
    - e.g, `<a>` - hypertext anchor element (i.e. link)
- A "small" group of browser supported "tags" that are interpreted upon page loading (small implies < 100)
- Inherently "stupid" (i.e. is unable to make decisions, use logic, etc)
- Loose and easy with rules of formatting (browser will make logical assumptions)

HTML: Example

```
<html>
<head>
     <title> My Page </title>
</head>
<body>
     <h1> My Page </h1>
     <p> This is great </p>
 </body>
 </html>
```

Uses Tags:

- html
- head
- title
- body
- h1
- p

Tags and Elements

- The core building block of HTML is the **tag**, which marks each element in a document.
- HTML tags are used to mark up a page.
- Tags start and end with angle brackets (e.g., `<h1>`)
- Need a start and and an end tag to mark up a region. End tags have an additional "/".

```
<h1> My Page </h1>
```

- The two tags define named *elements.* Above is an `<h1>` element.
- `<head>` element gives information about page.
- `<body>` element contains content of page.

Two-Sided Tags

- Tags can be two-sided or one-sided.
- A **two-sided tag** is a tag that contains some document content. General syntax for a two-sided tag:

`<element>content</element>`

- The information or text between the element tags is interpreted/parsed and displayed by the browser

- Examples: `<h1>…` `<h6>,` `<a>,` `<b>,` `<u>` etc.

One-Sided Tags

- A one-sided tag contains no document text content
- Are sometimes called empty elements.
- General syntax for a one-sided tag:

<element /\>

- The /> will be explained later in the lecture
- One-sided tags tend to be for page formatting or for non-text page content
- Examples: `<br/>, <hr/>, <img/>`

Elements and Attributes

- Elements can also have attributes, giving additional information.

```
<body bgcolor="blue"> ..</body>
<table width="610"> ..</table>
<a href="http://opentech2.durhamcollege.ca">link</a>
```

- Last example defines anchor with hyperlink to another page.

Adding Multimedia

- Images and other multimedia elements easily added to the page, e.g.,

```
<img src="mypic.jpg" alt="me" />
<img src="images/logo.gif" alt="Company Logo" />
<img src="http://www.myserver.ca/mycar.jpg" alt="My Car" />
```

- The included image is indicated using a URL.
  - The first two examples are relative URLs
  - The third is an absolute URL
- Strictly you must use an "`alt`" attribute giving a text description of the file, for folks who cannot access the image.
- More recent versions of HTML use the `<object>` tag to incorporate any media object, rather than the `<img>` tag.

HTML Special Character Entities

- HTML comments

```
<!-- everything inside these tags is
  ignored by the browser -->
```

- Special Characters
  - Certain characters are special to the browser or are not part of a standard keyboard setup, HTML therefore has special character codes

© is `©` ¥ is `¥` `&` is the &

< is `<` £ is `£`

> is `>` € is `€`

- For a more complete listing, check out: http://www.w3schools.com/html/html_entities.asp

HTML Page Structure

- You can’t use any tags anywhere on the page!
- The rules for which structures are allowed are well defined (as for programming languages).
- References are available
- These state which elements are allowed within which other elements.
- They may distinguish *block level* and *inline* elements - block level ones apply to long sections.

HTML/Web Resources

- There is not an HTML textbook for this course
- There are several online resources on the web (too many to count)
- Instructor's preference http://www.w3schools.com/
- Feel free to try to find your own

What is XML?

- The eXtensible Markup Language
- Looks like HTML (in the fact that there are tags)
- Designed to be structured
  - Strict rules about how data can be formatted
- These rules apply to how the tags can be written and are supported (tighter rules)
- The rules are laid out in a XML standards Document Type Definition (DTD)

Document Type Definition

- Defines the rules for a specific document type
- Including:
  - Acceptable tag names
  - Tag structure (what tags can exist in other tags)
  - Acceptable attributes for tags (including data type for the attributes)

What is XHTML then?

- XHTML is a family of current and future document types and modules that reproduce, subset, and extend HTML 4.0 HTML family document types are XML based, and ultimately are designed to work in conjunction with XML-based user agents.
- Looks exactly like HTML (because it is)
- The X deals instead with formalizing the way in which HTML is implemented (to deal with future support of tags within browsers)

Why XHTML?

- Alternate ways of accessing the Internet are constantly being introduced. The XHTML family is designed with general user agent interoperability in mind.
  - Ultimately, it will be possible to develop XHTML-conforming content that is usable by any XHTML-conforming user agent.
- WHAT DOES THAT MEAN?
- Motivation: intended to alleviate browser issues of the late 1990's and enhance portability in the future
- Plus, it is worth marks on assignments in this course

HTML vs. XHTML

```
<ul>
    <li>HTML allows <b><i>improper nesting</b></i></li>.
    <li>HTML allows start tags, without end tags, like
            the <br> tag.<!-- NOTE missing closing li -->
    <li>HTML allows <span color=#9900CC>attribute
        values</span> without quotes</li>
    <LI>HTML is case-insensitive</li>
</ul>


<ul>
    <li>XHTML requires <b><i>proper nesting</i></b>.</li>
    <li>XHTML requires empty tags to be identified with
        a trailing slash, as in <br/>.</li>
    <li>XHTML requires <span color="#9900CC">quoted
attribute
        values</span>.</li>
    <li>XHTML is case-sensitive (lower case is the XHTML
            standard)</li>
</ul>
```

XHTML Validation

- There are tools that allow you to validate or check your HTML page.
- For them to work you must make it clear which version of XHTML you are writing!
- You can do that by adding the *document type declaration* to top of your page. E.g.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

Character encodings

- You will also need to state what character encoding and language you are using
- For the language you can add the following at the top of the document (below the DOCTYPE declaration) as an attribute for the <html> tag:

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

- For the character set in the <head> of the document you should add:

```
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
```

  - N.B. (UTF-8 is a common character set for web implementations)

Minimum XHTML Valid Template

- Therefore, for this course you will be require to have these following lines at the top of all of your pages

```
<!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">
<head>
   <meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
   <title>A title Specific to the Page</title>
</head>
<body>
   <!-- this is where the main page content will show up -->
   <!- please note the tabbing, ignored by the browser but appreciated
   by your instructor -->
</body>
</html>
```

- The closing `</body>` and `</html>` at the bottom make the rest of the file XHTML compliant

XHTML Validation

- The World Wide Web Consortium (W3C) offers an online tool that will validate HTML to see if it complies to XHTML rules
- URL http://validator.w3.org/
- “Transitional” is a little looser than “Strict”
  - Note: we will be using “transitional” for the first few assignments, but will be going “strict” later in this course.