Lecture 2b: Markup languages and XHTML basics
Structure vs style

XHTML document by itself has no style elements

We will learn first how to create structurally correct web pages and then later on learn how to style them (Make 'em purdy!)
What is a markup language?

Historically

A collection of detailed stylistic instructions written on a manuscript to be typeset.

Example:
Is that really what you want? She said incredulously.

Here the underline can be used to tell the typesetter that the word “really” should be italicized.
Markup Languages

Computer science sense:

It is a set of designations (tags) to be used on a document to indicate how different sections of the document should be presented, structured or interpreted.

HTML Example:

```html
<table width="300px" border="1">
  <tr>
    <td><b>Year</b></td>
    <td><b>Budget</b></td>
  </tr>
  <tr>
    <td>1975</td>
    <td>23.7M</td>
  </tr>
  <tr>
    <td>1976</td>
    <td>25.8M</td>
  </tr>
</table>
```
Examples of Markup Languages

- LaTeX
- HTML-HyperTextMarkup Language
- XHTML-eXtensibleHyperTextMarkup Language
Meta-Markup Languages

Meta-markup language is a set of rules for creating a markup language

- SGML - Standard Generalized Markup Language
- XML - eXtensible Markup Language
What is HTML?

HTML stands for Hyper Text Markup Language

• A simple SGML application used to markup text documents into Web pages

• Used only a small subset of SGML’s capabilities

Note: HTML is a markup language SGML is meta-markup language! HTML is not a programming language.

• Specifies meaning and structure of content

• Not designed to specify presentation

• Created by Tim Berners-Lee in 1991
HTML Documents

• HTML documents describe web pages
• HTML documents contain HTML tags and plain text
• HTML documents are also called web pages
• Browsers interpret the HTML documents and use the tags to give the text structure and style
Problems with HTML: Chaos of Competition

• Late 1990’s Netscape and Microsoft battled it out for browser market dominance in the **Browser Wars**

• Proprietary HTML tags and incompatible implementations of HTML arose

• Intertwining of presentation layer and structure layer (eg font and center tags)

• HTML processors did not enforce HTML rules
What is XHTML?

**EXtensible Hyper Text Markup Language**

- An XML application for the structure and meaning of Web documents
- XHTML 1.0 keeps the vocab of HTML 4.01 but gains the stricter syntactic rules of XML
- Offers more consistency in structure than HTML
- Latest version: XHTML 1.1
Hello World!

```xml
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" 
 "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">

<!-- This is our first XHTML example -->

<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Hello World Example</title>
  </head>
  
  <body>
    <p>Hello world!</p>
  </body>
</html>
```
Every XHTML document must begin with an **xml declaration** element

```xml
<?xml version="1.0" encoding="utf-8"?>
```

Then immediately after must be the **SGML DOCTYPE command**

```xml
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
 "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
```
The actual web page information structure

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

<head>
  <title>Some title</title>
</head>

<body>
  <!-- This is where all your markup goes -->
</body>

</html>
We call **html** the **root element** of the XHTML document.

It is the marks the start of the XHTML part of the document.

All other tags are inside the `<html> </html>` tags.
XHTML <head>

• An opening tag for the header section
  <head>

• It is closed by
  </head>

Header section may contain:

Scripts
Instructions for the browser where to find style sheets
Meta data

Header section contains the invisible content of the web page

The following tags are permitted inside head section:
<title>, <base>, <link>, <meta>, <script>, and <style>
The title element is **mandatory** inside the header section.

```html
<title>Titleof MyPage</title>
```

The title element:
- defines a title in the browser toolbar
- provides a title for the page when it is added to favorites
- displays a title for the page in search-engine results
XHTML  \texttt{<body>} tag

Starts the main part of the web page.

The text between \texttt{<body>} and \texttt{</body>} is the visible page content.
XHTML Tags

• XHTML tags are keywords surrounded by angle brackets like `<html>`
• XHTML must come in pairs like `<b>` and `</b>`. The only exceptions are empty tags which close themselves:

  eg The new line tag `<br />`

• The first tag in a pair is the start tag, the second tag is the end tag
• XHTML tags are always in lowercase
XHTML Attributes

• An tag can have zero or more attributes which
  – clarify, describe or modify it
  – provide additional information separate from its content

• Each attribute is indicated by a name-value pair

Attribute names must be in lower case

Attribute values must be quoted:

<eltname att1="value1" att2 ="value2"> Stuff </eltname>

Example:

<a href="www.math.ucla.edu/~virtanen">My home page!</a>
Tags are fundamental syntactical units used to specify categories of content

Syntax of an element:

\[ \text{<element}_\text{name}> \text{ content (optional) } \text{</element}_\text{name}> \]

Not all elements have content.

Empty Elements Must Also Be Closed
Examples of XHTML Elements

With content and no attributes:

<p>Welcome to my page!</p>

Without content (empty element):

After this sentence is a new line <br/>

Without content (empty element) but with attributes

<img src="mountain.jpg" alt="A mountain" />

With an attribute and content:

<a href="mypage.html">My Page</a>
Nested HTML Elements

- Most XHTML elements can be nested (can contain other XHTML elements).
- XHTML documents consist of nested XHTML elements.
- Elements must be nested correctly!

**Correct:**
```html
<ul>
  <li>
  </li>
</ul>
```

**Incorrect:**
```html
<ul>
  <li>
  </li>
</ul>
```

Outer elements are called parent elements and inner ones are child elements.
XHTML Comments

Syntax:

<!--Insert comment here -->

• Browsers ignore them

• Can be spread over as many lines as needed
VALIDATION

Valid XHTML means that it conforms to the W3C standards

• All elements are closed
• All elements are properly nested
• All tags in lowercase
• All attribute values are in quotes etc.

Always use http://validator.w3.org/ to check your documents

The w3c validator is also your debugger
To practice writing web pages...

You don't need an HTML editor

- You don't need a web server
- You don't need a web site

At home you can use notepad to write a test web page

Important: When using notepad make extension .html and save as all files and not as .txt file.

Do not use word or some complicated text editor!

I recommend notepad++

Never use dreamweaver or some other automated web page program!
Optional mini assignment

This is not due!

Write a hello world web page using notepad at home

While you can practice some things at home. I highly recommend writing the homework assignments at the PIC lab. At home you cannot properly test the web page and any directories/files you use will have incorrect paths and the web page you write at home will not work at the PIC lab.
When you save an HTML file, you can use either the .htm or the .html file extension. The extension .htm is a relic from the past, when the software only allowed three letters in file extensions.

With new software it is perfectly safe to use .html.

**In this class we will use .html extension unless otherwise specified**