



PIC 40A

Lecture 2b: Markup languages and HTML basics

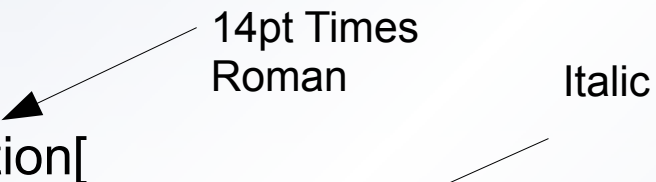
Markup Languages

What is a markup language?

Historically

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

Example:


]Introduction[

In mathematics]Lie Algebra[is an algebraic structure whose main use is studying geometric objects, such as Lie groups and differentiable manifolds.

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:

```
<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 **H**yper **T**ext **M**arkup **L**anguage

- 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

HTML Timeline and versions

1980 Tim Berners Lee (physicist) working at CERN introduced a system called ENQUIRE to share documents. This system was much like a modern wiki.

1989 He wrote a memo detailing internet based hypertext system.

1990 (late) He wrote HTML and wrote the browser and server software.

HTML 2.0 November 1995

HTML 3.2 January 1997

HTML 4.0 December 1997

HTML 4.01 December 1999

HTML 5 January 2008 (working draft)

HTML5 target date for W3C recommendation is end of 2014.

What is XHTML?

EXtensible **H**yper **T**ext **M**arkup **L**anguage

- An XML application for the structure and meaning of Web documents
- XHTML 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

HTML5 VS. XHTML

The good points of HTML5

- It is the new standard endorsed by W3C.
- It is supported by all new browsers and will eventually be widely supported.
- It is backwards compatible with older HTML4 and XHTML.
- It introduces some simplifications and new and exciting elements for audio and video.

The bad points of HTML5

- It abandons the stricter rules of XHTML.
- Browsers do not enforce HTML rules.
- Leads to bad web pages
- Validating HTML5 documents is harder.

HTML5 VS. XHTML

Since XHTML rules and syntax are valid in HTML5 we will study HTML5 from the XHTML point of view.

The material and basic syntax is presented as XHTML.

We will also study the new HTML5 elements.

We will validate documents as XHTML documents. More on this later.

Hello World!

```
<!DOCTYPE html>

<html lang="en">
  <head>
    <!--Title is optional in HTML5 -->
    <title>Hello World Example</title>
    <meta charset = "utf-8" />
  </head>

  <body>
    <p>Hello world!</p>
  </body>
</html>
```

HTML Document Structure

- Every HTML document must begin with the **SGML DOCTYPE** command

```
<!DOCTYPE html>
```

This specifies what particular rule set this document confirms to. In this instance no additional ruleset is specified HTML5 is implied.

The actual web page information structure

```
<html lang="en">
  <head>
    <title>Some title</title>
    <meta charset = "utf-8" />
  </head>

  <body>
    <!-- Visible content -->
  </body>
</html>
```

<html> tag

```
<html lang="en">
```

We call `html` the **root element** of the HTML document

It marks the start of the HTML part of the document

All other tags are inside the `<html>` `</html>` tags

HTML <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>`, `<meta>`, `<script>`, and `<style>`

HTML <title>

The title element is mandatory for XHTML and optional (but highly recommended) for HTML 5

It goes inside the header section.

```
<title>Title of 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

HTML `<body>` tag

Starts the main part of the web page.

The text between `<body>` and `</body>` is the visible page content

XHTML Hello world

```
<?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>
```

XHTML Hello world

The main difference is in the declaration of the document.

The head and the body section are almost identical.

The only real difference is that in the head section HTML 5 documents should have the meta element:

```
<meta charset="utf-8">
```

Whereas XHTML documents may omit it.

XHTML documents need a title, HTML 5 documents do not.

Additionally XHTML documents may contain a reference to a namespace in the <html> tag:

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

Note about XHTML tags and syntax

All of XHTML tags carry over to HTML5 so as we learn XHTML tags we are also learning HTML5 tags.

There are a few additional HTML5 tags that we will learn later.

There are some XHTML tags that used to be popular, but are not used much in HTML5.

XHTML syntax is stricter than HTML5 syntax.

Example:

XHTML requires all tags to be lowercase.

Generally writing HTML within the XHTML framework will lead to better web documents.

XHTML Tags

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

eg The new line tag

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

XHTML Elements

Tags are fundamental syntactical units used to specify categories of **content**

Element consists of the opening and closing tag and any content in between.

Syntax of an **element** :

```
<element_name> content(optional) </element_name>
```

Not all elements have content.

```
<element></element>
```

Empty Elements Must Also Be Closed

`
` is an example of a self closed element.

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>
```

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

```

```

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:

```
<ul>  
  <li>  
  </li>  
</ul>
```

Incorrect:

```
<ul>  
  <li>  
</ul>  
  </li>
```

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"

VALIDATION

Since we will write HTML5 documents, validation is tricky.

Essentially we will have to tell the validator to check our document as an XHTML document.

We will only do this for homework 1. Download an XHTML template from our examples page, then copy and paste your HTML5 body into the XHTML documents body section.

Then go to W3C and validate.

Later on we will have to use HTML5 validation. This has the drawback that it is inaccurate and is still in experimental stages.

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.

.HTM or .HTML File Extension?

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

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!)