Introduction to HTML

G6921 and G6931

Web Technologies

Dr. Séamus Lawless

Housekeeping

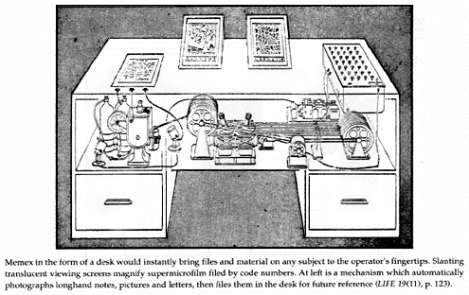
- Course Structure
  - 1) ~~Intro to the Web~~
  - 2) HTML
  - 3) HTML and CSS – Essay Information Session
  - 4) Intro to Databases
  - 5) Intro to Databases
  - 6) PHP and MySQL
  - 7) Reading Week
  - 8) PHP and MySQL
  - 9) PHP and XML
  - 10) CMS
  - 11) Analytics
  - 12) Visualisation

Housekeeping

- You all have a Laptop?
- Can you all get connected to the Internet?
- Starter HTML Page that I emailed?
- Download the Latest Version of Chrome (HTML5)
- Personal Webspace/Sandbox
- Assessment
  - Owen will be along at 11 to fill you in….
  - Continuous Assessment
    - Keep the files you create!

Be Prepared…

Hypertext



Memex in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference (*LIFE* 19(11), p. 123).

- Term "Hypertext" was coined by Ted Nelson in 1963
- Project Xanadu

The World Wide Web

- Initially developed by Tim-Berners Lee
  - The Web had a more pragmatic beginning
- Constructed in a simple manner enabling users to easily write documents
  - Documentation Available
  - Development of Mosaic
- Allowed information on the Internet to be linked together through "hyperlinks"

Where does HTML come in?

From Computer Desktop Encyclopedia
Reproduced with permission

- HTML is the language that allows users to easily author pages that can be placed on the WWW
- Supports the uni-directional linking of these pages
- Mosaic, one of the first browsers popularised the web and started people thinking about adding images, sound and (later) video

What is HTML?

- HyperText Markup Language or HTML
  - Application of SGML
- A document formatting language with the capability for hypertext links
- Became the primary publishing format for the WWW which supports:
  - Publishing in a platform independent format
  - Creating links to related works from your document
  - The including of graphics and multimedia data with your document

HTML through the years…

Year	Standard
1991	HTML
1994	HTML 2
1996	CSS + Javascript
1997	HTML 4
1998	CSS 2
2000	XHTML 1
2002	Tableless Web Design
2005	AJAX
2009	HTML 5

Structure

- Why is Document Structure Important?
  - Ensures that every user will see all of your content the way you intended
  - If you say something on the web, you simply cannot be sure how and when your content will be accessed
    - Browser
    - Language
    - Accessibility Issues

What is Markup?

- **Markup** enhances content
  - Annotations
  - Comments
  - Revisions
  - Display Instructions
  - Links…….
- **Markup** can take the form of tags, which are inserted in a document.

Tags in HTML

- A tag is a text element (a name) flanked by a less-than sign (<) and greater-than sign (>).
  - For example, "*<H1>*".
- HTML tags are not case-sensitive, meaning that it doesn't matter if they are entered in uppercase or lowercase letters.
- HTML tags work in pairs*, and affect whichever elements they are surrounding.
- Example

  Digital Humanities is very interesting stuff...

* In the vast majority of cases

Tags

- HTML tags can be nested within one another
- Overlapping tags are not allowed
- Example…
  - <u>Goodbye</u>, cruel World
  - is displayed as
  - *<u>Goodbye</u>, cruel World*
- However, you cannot have…
  - ~~<u>Goodbye, cruel World</u>~~

Tag Attributes

- An attribute is an element found within the opening tag that lets you define additional properties.
- Attributes are most often displayed as a name=value pair
  - some attributes may be defined only by their name
- Example
  - Example paragraph

HTML Document Layout

- As we have seen, a HTML document begins with the tag *<HTML>* and ends with the tag *</HTML>*.
- It also contains a *header* describing the title of the page…
- …and a *body* where the page's main content is located.
- The header is delimited by the tags
  - *<HEAD>* and *</HEAD>*
- The body is delimited by the tags
  - *<BODY>*and *</BODY>*.

HTML Document Layout

```
<HTML>
    <HEAD>
        <TITLE>Page title</TITLE>
    </HEAD>

    <BODY>
        Page content
    </BODY>
<HTML>
```

DOCTYPE

- The DOCTYPE declaration was introduced for validation of SGML documents and carried over to HTML
- It points to a Document Type Definition (DTD), which describes the syntax that the markup language adheres to.

Doctype Definitions

<!DOCTYPE root-element PUBLIC "FPI" ["URI"]>

or

<!DOCTYPE root-element SYSTEM "URI">

Version	Declaration
HTML 2.0	<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 4.0//EN">
HTML 3.2	<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
HTML 4.01	•**Strict**: <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Strict//EN" "http://www.w3.org/TR/html4/strict.dtd"> •**Transitional**: <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd"> •**Frameset**: <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN" "http://www.w3.org/TR/html4/frameset.dtd">
XHTML 1.0	•**Strict**: <!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd"> •**Transitional**: <!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-loose.dtd"> •**Frameset**: <!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">
XHTML 1.1	<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">

Doctype Definitions

- Web browsers typically internally implement a rendering engine specific to HTML rather than relying on an external DTD
- You won't need to worry about that for creating new HTML Documents
- Doctype Definition in HTML5 is…
- <!DOCTYPE html>

HTML Document Layout

```
<!DOCTYPE html>
<HTML>
        <HEAD>
                <TITLE>Page title</TITLE>
        </HEAD>

        <BODY>
                Page content
        </BODY>
<HTML>
```

Document Character Set

- To promote interoperability you must specify its *document character set.*
- The character set used dictates what characters can be part of a HTML document
  - This could include the Latin letter A - "ā", the Cyrillic letter I - "И", the Chinese character meaning "water " – "水", etc.
- The ASCII character set is not sufficient for a global information system such as the WWW, so HTML uses the much more complete *Universal Character Set (UCS)*

Character Encoding

- HTML must also be encoded so that it can be transmitted across the internet
- You must specify in your HTML the character encoding that was used
  - also known as the charset
- A simple one-byte-per-character encoding technique is not sufficient for text strings containing complex symbols (like chinese)

Character Encoding

- Commonly used character encodings on the Web include:
  - ISO-8859-1 (also referred to as "Latin-1"; usable for most Western European languages)
  - ISO-8859-5 (which supports Cyrillic)
  - SHIFT_JIS (a Japanese encoding)
  - EUC-JP (another Japanese encoding)
  - UTF-8 (an encoding of ISO 10646 using a different number of bytes for different characters).

Specifying the Charset

- The character encoding, or charset, used to transmit can by specified by the server
- However, to address server or configuration limitations, HTML documents may include explicit information about the document's character encoding
- Example…
  - <META http-equiv="Content-Type" content="text/html; charset=UTF-8">

HTML Document Layout

```
<!DOCTYPE html>
<HTML>
        <HEAD>
                <TITLE>Page title</TITLE>
                <META charset="UTF-8">
        </HEAD>

        <BODY>
                Page content
        </BODY>
<HTML>
```

Spaces, Line Breaks and Tabs

- Spaces, tabs, and line breaks are disregarded in HTML documents
  - The idea is to allow HTML to be indented to make it more readable, without changing how the content appears in the browser.
- Example…

```
<p>Example
of HTML
Code</p>
```

or

```
<p>Example of HTML Code</p>
```

Spaces, Line Breaks and Tabs

- However, HTML has ways of expressly defining each of these layout elements:
- **Non-breaking space**: A space which cannot be broken up by the end of a line. Its HTML encoding is * *
- **Manual line break**: An explicit line break. Its HTML encoding is *
*
  - *
* for compliance with XHTML.

Useful Tags

- Document Structure
  - <html>
  - <head>
  - <title>
  - <body>
- Document Formatting
  - <p>
  -
  -
  - <h1> - <h6>
  - <hr>
- Lists
  -
  -
  - <dl>
  -
- Text Formatting
  - <b> or
  - <i> or
  - <u>
  - <font>
  - <sub> and <sup>

Hyperlinks

- <a> - The Anchor Tag
- One of the main reasons all this exists, and the real power of the WWW
- Links are defined with the <a> tag
- The most important attribute of the anchor tag is **href**
  - This is where you specify the URL of a link
- Example…
  - This a link to the Journal

Semantic Tags

- HTML5 has added improved semantic tags
  - <section>
  - <header>
    - <hgroup>
  - <nav>
  - <article>
  - <aside>
  - <figure>
    - <figcaption>
  - <footer>
  - <time>

Semantic Tags

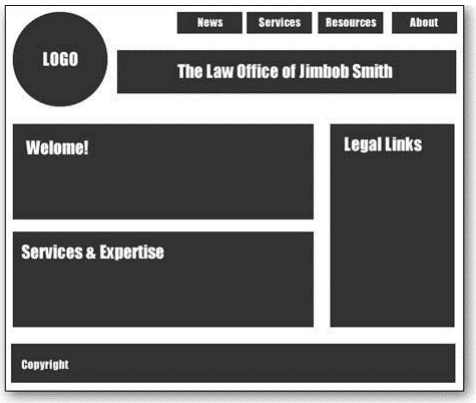


Figure 1 - General browser layout for fictitious home page

Semantic Tags

Content	Relevant HTML5 Element
Logo	<header>
Name of lawyer	<header>
Navigation	<nav>
Welcome message	<section><article>
Featured services	<section><article>
List of legal resources	<aside>
Copyright and legal notices	<footer>

Semantic Tags

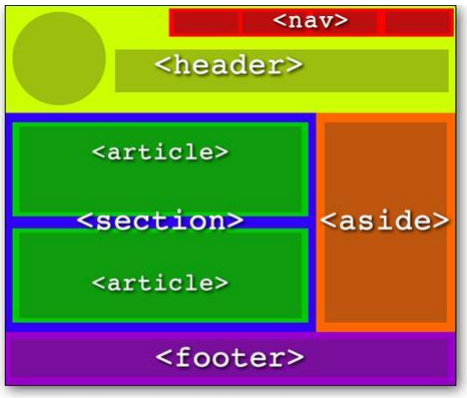


Figure 2 - New HTML5 semantic elements applied to layout for fictitious site

Comments

- Information can be added to a web page without being shown on the screen, with a special set of tags called *comment tags*.
  - Can be used to comment on the text or on the HTML code itself
- Example…
  - <!-- This is a comment -->

Example

- Suppose the following is the desired result:

Introduction

This article is a review of the book *Dietary Preferences of Penguins*, by **Alice Jones** and **Bill Smith**. Jones and Smith's controversial work makes three hard-to-swallow claims about penguins:

- First, that penguins actually prefer tropical foods such as bananas and pineapple to their traditional diet of fish
- Second, that tropical foods give penguins an odour that makes them unattractive to their traditional predators

Example

- Start with the document's raw text:

Introduction

This article is a review of the book Dietary Preferences of Penguins, by Alice Jones and Bill Smith. Jones and Smith's controversial work makes two hard-to-swallow claims about penguins:

First, that penguins actually prefer tropical foods such as bananas and pineapple to their traditional diet of fish

Second, that tropical foods give penguins an odor that makes them unattractive to their traditional predators

Example

- Add markup *tags*, which provide additional information about the text:
  - Formatting information (<i> for italic, for lists)
  - Meaning of the text (<h1> means top-level heading)
  - Additional content to display (e.g., <img>)

Example

```
<h1>Introduction</h1>
<p>
This article is a review of the book <i>Dietary Preferences of Penguins</i>, by <b>Alice Jones</b> and <b>Bill Smith</b>. Jones and Smith's controversial work makes three hard-to-swallow claims about penguins:
</p>
<ul>
        <li>
        First, that penguins actually prefer tropical foods such as bananas and pineapple to their traditional diet of fish
        </li>
        <li>
        Second, that tropical foods give penguins an odour that makes them unattractive to their traditional predators
        </li>
</ul>
```

Character References

- A given character encoding may not be able to express all characters of the document character set.
  - i.e. UTF-8 cannot express all of the UCS
- In such cases, HTML authors can use character references
- Character references in HTML may appear in two forms:
  - Numeric character references
  - Character entity references.

Numeric Character References

- *Numeric character references* specify the code position of a character in the document character set
  - &#*D*; - where *D* is a decimal number, refers to the ISO 10646 decimal character number *D*.
  - &#x*H*; - where *H* is a hexadecimal number, refers to the ISO 10646 hex character number *H*.
- Example…
  - å (in decimal) represents the letter "å"
  - å (in hexadecimal) represents the same letter.

Character Entity References

- Character entity references use symbolic names so that authors need not remember code positions.
  - å refers to our previous example – "å"
  - Character entity references are case-sensitive.
- HTML does not define a character entity reference for every character in the document character set.
- Are frequently used to represent characters that are markup sensitive in certain context
  - & → & (ampersand)
  - < → < (less-than sign)
  - > → > (greater-than sign)
  - " → " (quotation mark)
  - ' → ' (apostrophe)

HTML Metadata

- The <meta> tag provides metadata about the HTML document.
  - This metadata will not be displayed on the page, but will be machine parsable.
- The metadata can be used by browsers (how to display content or reload page), search engines (keywords), or other web services.
- <meta> elements are typically used to specify page description, keywords, author of the document, last modified etc.
  - We have already used it to define the charset!

HTML Metadata

- The <meta> tag always goes inside the head element.

```
<head>
  <meta name="description" content="MPhil  Class" />
  <meta name="keywords" content="HTML,CSS" />
  <meta name="author" content="Seamus Lawless" />
  <meta charset="UTF-8" />
</head>
```

New with HTML5

- 2D drawing API with the canvas element.
- Easy playing of video and audio with the video and audio elements.
- Drag & drop API in combination with a draggable attribute.
- Canvas
- …..

New with HTML5

- Interactive Multimedia - Canvas
- http://craftymind.com/factory/html5video/CanvasVideo.html
- http://hakim.se/experiments/html5/trail/03/
- http://hakim.se/experiments/html5/blob/03/
- http://andrew-hoyer.com/experiments/cloth/

HTML5 and XML

- HTML5 is backwards compatible with XHTML
- Below is an example that conforms to the XML syntax of HTML5

```
<?xml version="1.0" encoding="UTF-8"?>
<html xmlns="http://www.w3.org/1999/xhtml">
    <head>
        <title>Example document</title>
    </head>
    <body>
        <p>Example paragraph</p>
    </body>
</html>
```

Local v Remote

- "Local" refers to something that is stored on your computer
- "Remote" refers to something that is stored on another machine
- A remote machine could be a web server or even someone else's personal computer

Client Server

- **Client–server** is a computing model where tasks or workloads are divided between
  - providers of a resource or service, called **servers**
  - service requesters, called **clients**