

B211 Internet Computing

HTML and XHTML

B211 Week 7 Lecture 2

1

Lecture Outline

- Development of HTML Specifications
- HTML and XHTML
- Cascading Style Sheets (CSS)
- Other work by W3C on HTML

B211 Week 7 Lecture 2

2

In this lecture...

- In this lecture, I will assume you have gone through the early tutorial exercises and know basically the structure of an HTML (HyperText Mark-up Language) document.
- The purpose of this lecture (like all lectures in the current topic) is not to teach you to code, but to show you
 - the developments of the language(s), and
 - the technical implications of using the language for implementation.

B211 Week 7 Lecture 2

3

Early Ideas on HTML

- As mentioned in the lectures in Topic 2 *The World Wide Web*, the concept of hypertext through HTML is one of the key foundation behind the WWW.
- The idea of hypertext is not new - it dates all the way back to manuscripts in medieval times.
 - manuscripts contain pictures, and cross-references to other manuscripts.

B211 Week 7 Lecture 2

4

Early Hypertext in Computers

- Vannevar Bush described the idea of a computer-aided cross-referencing system name *memex*, in an article in 1940's.
- In 1960's, Ted Nelson coined the term *hypertext*. Nelson spent his life promoting a global hypertext system called *Xanadu*.
- In 1987, Bill Atkinson released *HyperCard*, the first widely-used computer hypertext system, found on the Apple Macintosh.
 - Many imitations followed

The Birth of HTML

- In 1989, Tim Berners-Lee created HTML as part of a linked information systems in CERN
 - to support data exchange between researchers in different locations.
 - He also created HTTP at the same time as the network transfer protocol to transfer HTML documents
- Refer to concepts of the World-Wide-Web in Topic 2 lectures.

Versions of HTML

- HTML 1.0
 - the original version defined by Berners-Lee.
- HTML 2.0
 - Released 1995 by IETF's HTML Working Group
- HTML 3.2
 - Released Jan 1997 by W3C - W3C took over from IETF for most specifications of the WWW.
 - Included tables, applets, super/subscripts, etc.

Versions of HTML (cont'd)

- HTML 4.0
 - Release December 1997, revised April 1998.
 - Adds support for style sheets, internationalization features, accessibility features
- HTML 4.01
 - Released December 1999
 - A few bug fixes from version 4.0
- Each specification recommends backward compatibility with previous versions.

Versions of HTML (cont'd)

- Updates in from one version to the next are usually driven by developments of non-standard HTML “extensions” popular browsers have chosen to implement

The eXtensible HTML (XHTML)

- widely promoted by W3C as the future of HTML
 - XHTML 1.0 released in August 1999
 - Current version XHTML 1.1 released May 2001 - the modular version of XHTML 1.0.
- Aims to create standards that provide richer Web pages on an ever increasing range of browser platforms
 - eg. cell phones, televisions, cars, wallet sized wireless communicators, kiosks, and desktops.
- Includes markup tags for things like vector graphics, multimedia, math, electronic commerce and more.

XHTML (cont'd)

- Reformulating HTML as an XML application
 - more about what this means in the next lecture on XML
- All XHTML compliant documents are also HTML 4 compliant.
 - But not the other way around!

Flavours of XHTML

- Three flavours of XHTML documents:
 1. **XHTML Transitional** - general access web pages
 2. **XHTML Strict** - very strict structural mark-ups, clean of tags associated with layout. Used in conjunction with Cascading Style Sheets (CSS) for the proper layout.
 3. **XHTML Frameset** - web pages with frames
- The flavour of XHTML is specified using the `<!DOCTYPE>` tag at the beginning of the document
 - allows software to process it in different ways if the software deems it necessary.

HTML: Structural or Presentation Language?

- HTML was designed as a structural language
 - It was meant to indicate the *structure* of a page, rather than how to process the page.
 - Eg. <p> is suppose to mean “ this is the start of a paragraph”, rather than “put in a new line”
- But with the presentation being so important, more and more features of formatting are going into HTML
 - There is major disagreement as to whether the web should be a *presentation mechanism* or *an area for information only*.
 - Some argue that presentation IS information in a lot of cases.

Cascading Style Sheets (CSS)

- To preserve HTML as a structural language, specifications of HTML (and XHTML) have tried to isolate the structural components from the layout components.
- The idea of using style sheets was introduced starting from HTML 4.0
 - All layout features (eg. colors, fonts, margins) of an HTML page is suppose to be put in a separate style sheet file.
 - Structurally clean HTML pages can be linked to different style sheets.

CSS (cont'd)

- CSS level 1 (CSS1) released in December 1996.
- CSS level 2 (CSS2) released in May 1998.
- W3C currently working on CSS3.

HTML Validation Service

- W3C has an HTML validation service at <http://validator.w3.org/>
- HTML developers can put in the URL of their web pages, and the service will validate the web pages according to W3C's HTML specifications
 - reports all errors and non-conforming usage of tags
- There is also a CSS validator service at <http://jigsaw.w3.org/css-validator/>

HTML Tidy

- Popular W3C-endorsed open-source software for fixing up HTML
 - <http://www.w3.org/People/Raggett/tidy/>
- Where possible, TIDY corrects markup code in a way that matches the observed rendering in popular browsers.
- Can convert presentation markup to CSS rules.

HTML Tidy (cont'd)

- Knows about dynamic web pages such as ASP and PHP and can be taught about new tags, eg. from Cold Fusion.
- Good at cleaning up the HTML exported from Microsoft Word.
- Makes it trivial to roll content over to XHTML.
- Available on most platforms and has been integrated into numerous authoring environments.

Example Other Work by W3C on HTML

- Document Object Model (DOM)
 - Specifications of how an HTML document is to be represented as an object.
 - Eg. the object hierarchy used in JavaScript follows DOM.
- Internationalization
- Accessibility
- XForms
 - the next generation of HTML forms
 - parallels the development of XHTML, but specifically for XHTML forms

ISO HTML

- International Standards Organization (ISO) also defined a version of HTML as ISO/IEC 15445, published in May 2000.
- The specifications is a subset of HTML 4.0
- Intended for experts in SGML (Standard Generalized Mark-up Language), a very technically rigorous mark-up language
 - Eg. in ISO HTML, the <h3> tag cannot occur after <h1> if there is no intervening <h2> tag.
 - more on SGML in the next lecture on XML (eXtensible Mark-up Language)

Further Information

- W3C maintains a comprehensive set of information and links about HTML. Detailed elaboration of all material in this lecture can be found at

<http://www.w3.org/MarkUp/>

- Blank Page -

- Blank Page -

- Blank Page -