Introduction, Internet and Web Basics
XHTML and HTML

January 30, 2008
Harvard University
Division of Continuing Education
Extension School

Course Web Site: http://cscie12.dce.harvard.edu/
Copyright 1998-2008 David P. Heitmeyer
Instructor email: david_heimeyer@harvard.edu
Course staff email: cscie12@fas.harvard.edu

---

CSCI E-12, Fundamentals of Web Site Development

The Course

Course Syllabus | Course Schedule

- Markup (XHTML, HTML)
- Style (CSS)
- Images
- Forms (Interactivity)
- Building a Site
- Javascript and Ajax
- Video and Multimedia
- Dynamic Sites
- Hypertext Transfer Protocol (HTTP)
- Apache Web Server
- Webmaster Tools

Assignments

1. Web Basics, XHTML and HTML
2. XHTML and HTML; Using SSI; Getting Started with CSS
3. CSS
4. Forms, Javascript and Ajax
5. Dynamic Pages
6. HTTP and Apache Web Server
Texts

In addition to the texts, there will be online readings assigned and online references cited.

Required texts:


Texts are available through:

- Amazon.com
- Barnes & Noble
- bookpool.com

Recommended for those who are not yet familiar working with Unix or Linux:


Goals of CSCI E-12

This course is designed to give students a thorough overview of web development. Students explore the vocabulary, tools, and standards that prevail in the field today, and learn how the various components - including CSS, Javascript, multimedia, scripting languages, content management systems, web/applications servers, and databases - fit together.

For the final project, each student will have the opportunity to produce an interactive web site on the topic of their choice.

The course will give students a strong foundation for more specific web development study in the future, whether that be in programming, database administration, server administration, security, interface design, or multimedia development.
The International Network (Internet)

Image by Matt Britt (used with permission under Creative Commons Attribution 2.5)

"A model of Internet topology using k-shell decomposition"

The Internet in Pictures
Some Internet Numbers

- Nearly 490 million Internet hosts as of July 2007, ISC Internet Domain Survey
- 1.26 billion Internet users world-wide (19% of world population), Minwatts Marketing Group

<table>
<thead>
<tr>
<th>Region</th>
<th>Internet Penetration</th>
</tr>
</thead>
<tbody>
<tr>
<td>North America</td>
<td>71%</td>
</tr>
<tr>
<td>Oceania/Australia</td>
<td>69%</td>
</tr>
<tr>
<td>Europe</td>
<td>45%</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>22%</td>
</tr>
<tr>
<td>Middle East</td>
<td>17%</td>
</tr>
<tr>
<td>Asia</td>
<td>12%</td>
</tr>
<tr>
<td>Africa</td>
<td>9%</td>
</tr>
</tbody>
</table>

- Half of all Americans have broadband at home, Pew Internet and American Life Project (September 2007)

The Internet: Uses

- Email (SMTP, POP, IMAP)
- News Groups
- FTP
- Gopher (!)
- Web

Gopher

gopher://home.jumpjet.info/
The World Wide Web

The irony is that in all its various guises -- commerce, research, and surfing -- the Web is already so much a part of our lives that familiarity has blunted our perception of the Web itself.

Tim Berners-Lee in Weaving the Web

History

- History of Internet and WWW: The Roads and Crossroads of Internet History by Gregory R. Gromov
- A Little History of the World Wide Web, W3C
- How it all started, W3C 10-year Anniversary

The Web in Pictures

Image from Opte Project and is used under the Creative Commons 1.0 license.

Graph made from TouchGraph GoogleBrowser.
Some Web Numbers

45 billion pages as of January 2008 (http://www.worldwidewebsize.com/)

Nielsen NetRatings for Home Web Use, December 2007:

<table>
<thead>
<tr>
<th>Sessions/Visits per Person per Month</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domains Visited per Person per Month</td>
<td>69</td>
</tr>
<tr>
<td>Web Pages per Person per Month</td>
<td>1,564</td>
</tr>
<tr>
<td>Time Spent During Surfing Session</td>
<td>0:56:28</td>
</tr>
</tbody>
</table>

Top Sites for the United States

Top Sites, United States from Alexa: The Web Information Company

- Google
- Yahoo!
- MySpace
- YouTube
- Facebook
- Windows Live
- eBay
- Wikipedia
- Microsoft Network (MSN)
- Amazon
- AOL
- Amazon.com
- Blogger.com
- Megaupload
- CNN - Cable News Network
- Internet Movie Database
- Photobucket image hosting and photo sharing
- Digg
- Craigslist.org
- Flickr
- comcast.net
Features of the World Wide Web

- HyperText Information System
- Cross-Platform
- Distributed
- Approximately 70 million active sites as of December 2007.
- Open Standards and Open Source
  - TCP/IP, HTTP, HTML, CSS
  - Apache HTTP Server, Mosaic, Netscape, Firefox, JavaScript, Perl, PHP, etc.
- Web Browser: provides a single interface to many services
  - Banking, molecular biology, information search and retrieval, etc.
- Dynamic, Interactive, Evolving
- "Web 2.0"
  - Collective Intelligence
  - Data
  - Rich User Experience

Nature of the Web

- A VARIETY of
  - Platforms
  - Connection Speeds
  - Displays
  - Web Browsers (HTTP Clients)
  - Browser Settings
  - Devices
  - Users
  - Languages
  - Other Sites
- Open Standards (and Open Source)

We cannot command nature except by obeying her.

Francis Bacon
HTTP Clients and HTTP Servers

client-server computing

The interaction between two programs when they communicate across a network. A program at one site sends a request to a program at another site and awaits a response. The requesting program is called a client; the program satisfying the request is called the server.

(definition from The Internet Book, 2nd edition by Douglas E. Comer)

Client-Server Architecture from Webopedia, http://www.webopedia.com/

Uniform Resource Identifier (URI)

URI, Uniform Resource Identifier
URL, Uniform Resource Locator
URN, Uniform Resource Name

For those who truly wish to find out more of the details, see Untangle URIs, URLs, and URNs by Dan Connolly

Anatomy of a URI

http://cscie12.doe.harvard.edu/lecture_notes/20080130/toc.html#slide10

- **Scheme**
  - http

- Common schemes: http, ftp, mailto, file, rtsp

- **Host**
  - http://cscie12.doe.harvard.edu

- **Port**
  - http://cscie12.doe.harvard.edu:80

- **Path**

- **Fragment Identifier**
  - http://cscie12.doe.harvard.edu/lecture_notes/20080130/toc.html#slide10
XHTML: a simple example

```xml
<?xml version="1.0"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html>
<head>
<title>My Schools</title>
</head>
<body>
<h1>My Schools</h1>
<ul>
<li>
<a href="http://www.harvard.edu/">Harvard University</a>
<br/>
<img src="images/veritas.gif" alt="Harvard Shield" height="84" width="72"/>
</li>
<li>
<a href="http://www.ku.edu/">University of Kansas</a>
<br/>
<img src="images/KUSeal.gif" alt="University of Kansas Seal" height="73" width="72"/>
</li>
</ul>
</body>
</html>
```

(X)HTML Document Structure

- Document Type Declaration
  - html
    - head
    - body

A Tree

A "tree" structure view of XHTML produced by Amaya, the open source Web editor/browser from the W3C.
Components of XHTML Elements

<table>
<thead>
<tr>
<th>Start Tag</th>
<th>Element Name</th>
<th>Attribute and Value Pairs</th>
<th>Content</th>
<th>End Tag</th>
</tr>
</thead>
</table>

A Hypertext Link

Example 1.1

Example 1.1 Source:
```
<a href="http://www.harvard.edu/">Harvard</a>
```

Example 1.1 Rendered:
Harvard

Structure and Style

HyperText Markup Language
- Primary purpose: structure and content

Cascading Style Sheets (CSS)
- Primary purpose: style, presentation, appearance

Maintain this "separation of concerns" -- between structure and style.
XHTML: a simple example with CSS

View the example.

And here is the stylesheet (simple-style.css):

```xml
body { margin-left: 10%; margin-top: 1em; margin-right: 10%; background: #ff9; }

h1 { font-family: Arial, sans-serif; color: #600; border-bottom: thin dotted black; }

li { margin-top: 1em; font-size: 1.25em; }

ul { list-style-type: none; }

a { text-decoration: none; }

a:link, a:visited { color: blue; }

a:hover { color: white; text-decoration: underline; background: blue; }

img { border: thin solid black; }
```

Harvard College Admissions

With CSS disabled:
Examples from: CSS Zen Garden

css Zen Garden: The Beauty in CSS Design. A demonstration of what can be accomplished visually through CSS-based design.
Well-formed (X)HTML and Valid (X)HTML

- well-formed
- valid

Well-formed

Required for well-formed XML

- elements must be properly nested
- elements must have a start and end tag
- element names case-sensitive
- attribute values must be enclosed in quote marks
- attributes may not be repeated

Good practices for HTML

- Choose lower case conventions for element names and attributes.
- Use double quotes for all attribute values.
- Close tags even if they are optional in HTML.
- Use new lines and indentation (spaces or tabs) for readability.

Valid

Well-formed + Conforms to DTD = Valid

HyperText Markup Language

W3C HyperText Markup Language Home Page

- HTML 5 Working Draft (WHATWG)
  January 2008
- XHTML 2.0 Working Draft
  July 2008
- XHTML 1.1, Module-based XHTML
  May 2001
- XHTML 1.0, a reformulation of HTML 4.01 into XML 1.0
  January 2000
- HTML 4.01
  December 1999
- HTML 4.0
  December 1997
- HTML 3.2
  January 1997
- HTML 2.0
  November 1995
SGML, XML, HTML, and XHTML

Defining Markup Languages
- Standard Generalized Markup Language (SGML)
- Extensible Markup Language (XML)

Specific Markup Languages
- Hypertext Markup Language (HTML)
- Extensible HyperText Markup Language (XHTML)

Relationships
- SGML
  - XML is a simplified subset of SGML
  - HTML is an application of SGML
  - XHTML is an application of XML

Document Type Declaration and Document Type Definition (DTD)

The Document Type Declaration for an XHTML 1.0 strict document is:

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

A closer look at the components follows:

- html
  - The name of the root element.
- PUBLIC
  - An indication that the DTD is a public standard. Private standards would use "SYSTEM".
- "-//W3C//DTD XHTML 1.0 Strict//EN"
  - The formal public identifier (FPI) for the DTD (Document Type Declaration). This identifier can be used to locate the parser to a local copy of the DTD.
- W3C
  - The owner of the DTD (in this case the W3C, the World Wide Web Consortium).
- DTD
  - The type of document that is referenced (in this case a Document Type Definition, DTD).
- XHTML 1.0 Strict
  - The name of the document that the public identifier references.
- EN
  - The language identifier (in this case, "EN" = English). Note that the language code is case-sensitive.
- "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd"
  - The URI at which the parser can locate the referenced Document Type Declaration. If the parser has the DTD available in a local library, the URI is not necessary.
Common Document Type Declarations and DTDs

Some Document Type Declarations for HTML documents. Remember that the HTML document must conform to the rules of the Document Type Definition that is referenced in the Document Type Declaration:

- XHTML 1.1
  
  `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11-flat.dtd">

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

- HTML 4.01 Strict
  
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html401/strict.dtd">

- HTML 4.01 Transitional
  
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html401/loose.dtd">

- HTML 4.01 Frameset
  
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN" "http://www.w3.org/TR/html401/frameset.dtd">

- HTML 4.0 Transitional
  
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">

- HTML 4.0 Strict
  
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0//EN" "http://www.w3.org/TR/html4/strict.dtd">

- HTML 4.0 Frameset
  
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Frameset//EN" "http://www.w3.org/TR/html4/frameset.dtd">

- HTML 3.2
  
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 3.2//EN" "http://www.w3.org/TR/html32.dtd">

- HTML 2.0
  
  `<!DOCTYPE html PUBLIC "-//IETF//DTD HTML 2.0//EN" "http://www.w3.org/MarkUp/html-spec/html.dtd">

Specifications and Implementations
XHTML: a closer look

XHTML 1.0 Strict Elements

XHTML 1.0 Strict documentation produced from DTD

- Structures
  - body, head, html, title
- Text
  - abbr, acronym, address, blockquote, br, cite, code, dfn, div, em, hr, h1, h2, h3, h4, h5, h6, kbd, p, pre, q, samp, strong, var
- Heading
  - h1, h2, h3, h4, h5, h6
- Block
  - address, blockquote, div, p, pre
- Inline
  - abbr, acronym, br, cite, code, dfn, em, hr, kbd, q, samp, strong, var
- Hypertext
  - a
- List
  - ol, ul, dl, dt, dd, li
- Text Extension
  - b, big, br, i, small, sub, sup, tt
- Edit
  - del, ins
- Bi-direction Text
  - bdo
- Form
  - form, fieldset, input, label, select, option, textarea, button, legend, optgroup
  - Form
    - form, fieldset
  - Fieldset
    - input, label, select, textarea, button
- Table
  - table, td, th, tr, caption, col, colgroup, tbody, thead, tfoot
- Image
  - img
- Server-side Image Map
- Object
  - object, param
- Metainformation
  - meta
- Scripting
  - script, noscript
- Style Sheet
  - style
- Link
  - link
- Base
  - base
Benefits of Web Standards

- Improved Accessibility
  - People (Section 508, WAI)
  - Machines
    - Search Engines
    - Devices
- Stability
  Forward-compatible and backward-compatible.
- Separation of Structure and Style
  - lighter, cleaner pages
  - maintenance
  - redesign
- Validation

What do GE, IBM, Library of Congress, EDS, Stanford, AGFA, Abbott, and Princeton have in common?
They adhere to Web standards.

Web Standards Project
The Web Standards Project is a grassroots coalition fighting for standards which ensure simple, affordable access to web technologies for all.

What are Web Standards and why should I use them?

Pragmatic Steps: Software Tools

- HTTP Client (Web Browser)
- SSH or Client
- SFTP Client
- Text/HTML Editor or Authoring Package
- Graphics Program
- Validators and Checkers
HTTP Clients

Browser Statistics

Firefox (Mozilla)
- Web Developer Extension for Firefox
- Firebug Extension for Firefox

Internet Explorer (Microsoft)

Opera (Opera Software)

Safari (Apple)
- Safari available for Mac and Windows.

Netscape Navigator (AOL/Netscape)

SSH Clients and SFTP Clients

- PuTTY, http://www.chiark.greenend.org.uk/~sgtatham/putty/
- OpenSSH, http://www.openssh.org/
Text or HTML Editor

Windows
- Dreamweaver
- Adobe GoLive
- Notepad++ or Wordpad or Notepad

Mac / Apple
- BBEdit
- BareBones Software
  http://www.barebones.com/
- Dreamweaver
- Adobe GoLive
- SimpleText

Linux / UNIX
- Emacs
  http://www.gnu.org/software/emacs/emacs.html
  (UNIX, Windows, Macintosh)
  - nXML mode for Emacs

(X)HTML Well-formedness and Validation Resources

- W3C HTML Validation Service
  http://validator.w3.org/
- HTML Tidy
  http://www.w3.org/People/Raggett/tidy/
  minerva tidy -help