The Internet

Chapters 1 & 2

1. Introduction
   - Who's Involved
   - Layers
2. WWW
   - Languages
   - Framework
3. HTML
   - Inline Elements
4. Standards

- **Internet** - set of computer networks that uses many different technologies, e.g. e-mail, web pages, chat, file transfer...
- **WWW** - the world wide web is a collection of web pages. The WWW $\neq$ Internet, rather, WWW $\subset$ Internet.
- **ARPANET** - predecessor of the Internet used by the United States Department of Defense in the late 1960's. Goals were to send messages and transfer files. Had several unique features:
  - Sending data as "packets"
  - Ability to have network subsets
  - Able to dynamically add/remove machines
  - Open standards
  - Lack of centralized control (i.e. not one crucial server)
- **Protocol** - a standard for how information is received and sent.

- **Internet Engineering Task Force (IETF)** - creates and defines protocol standards.
- **Internet Corporation for Assigned Names and Numbers (ICANN)** - controls IP addresses and top level domain names.
- **World Wide Web Consortium (W3C)** - provides *recommended* web standards.

- The Internet works in layers. Data is sent through layers of software and hardware. Uses the OSI model: Application ↔ Presentation ↔ Session ↔ Transport ↔ Network ↔ Data Link ↔ Physical.
- **Internet Protocol (IP)** - specifies how data is sent across the Internet. Only provides device to device routing and addressing.
- Every device connected directly to the Internet has a unique IP address.
- **IPv4 Addresses** - a 32 bit number broken up into four 8-bit numbers for readability. For example, 141.233.181.107 is the Linux labs uws07 machine. All 4 parts are between 0 and 255.
- **IPv6 Addresses** - a 128 bit number broken up into four 16-bit numbers. For example, fe80::2e27:d7ff:fe27 is the Linux labs uws07.
- A device can generally connect to itself using **localhost** or 127.0.0.1.

- **Transmission Control Protocol (TCP)** - a communication standard implemented on top of IP.
- TCP can guarantee non-corrupted and in-order delivery of packets (numbering packets and sending acknowledgements).
- **Multiplexing** - multiple programs using the same IP address. Accomplishes this using *ports*.
- **User Datagram Protocol (UDP)** - simpler protocol than TCP but missing guaranteed in-order delivery. Useful for streaming video/audio.

- **Server** - a device connected to the Internet waiting for connections.
- Sample server software include Apache, Microsoft Internet Information Server, Microsoft Exchange Server...
- **Web Browsers** - software that retreives and displays web pages.
  - If protocols are followed, you should not worry about which browser is being used to access your web site.
  - Some browsers are not able to view certain content. Chrome, Internet Explorer and Firefox are the three most common as of today.

- **Domain Name System (DNS)** - a translation from a plain text name to an IP address. No one wants to remember 74.125.225.52 but everyone can remember www.google.com (host name).
- A known and trusted set of root DNS servers provide the domain name → IP address translation. As of 2013, there were 13 root servers.
- **Uniform Resource Locator (URL)** - used to request a particular resource from a web site.
- A sample URL:
  http://www.uwosh.edu/faculty_staff/krohne/index.php
  - **Protocol** - http
  - **Host Name** - www.uwosh.edu
  - **Path** - faculty_/krohne/index.php

- http://www.uwosh.edu:80/computer_science/faculty-staff#krohn
- :80 is the port number. If omitted, 80 is assumed.
- #krohn is an anchor.
- https://www.google.com/search?q=web+software
- q=web+software is a query string. It's a set of parameters passed to a web site. q is the parameter, it's value is "web+software".

- **Hypertext Transfer Protocol (HTTP)** - a protocol on top of TCP for viewing and retreiving web pages. Commands include GET, POST, PUT and HEAD.
- HTTP is *stateless*. There is no persistent connection between client and server.
- **Status Code** - the server sends back the requested document and a code. 200 (OK), 301 (moved), 403 (forbidden), 404 (not found), etc. 404 pages don't have to be bland: They can be fun.
- **MIME type** - an Internet media type. Two-part identifier separated by a slash. Many more in the book.

MIME	Extension
text/html	.html
image/gif	.gif
image/jpeg	.jpg

- **Hypertext Markup Language (HTML)** - used for writing web pages.
- **Extensible HyperText Markup Language (XHTML)** - used for writing better web pages. Elements are properly nested, elements are closed, certain tags are mandatory.
- **Cascading Style Sheets (CSS)** - stylistic information for web pages. Deals with appearance, layout and presentation.
- **PHP Hypertext Processor (PHP)** - allows web server to generate pages dynamically, one of many server side languages.
- **JavaScript (JS)** - interactive and programmable web pages.
- **Extensible Markup Language (XML)** - organizes and formats data.
- **Asynchronous JavaScript and XML (Ajax)** - provides much faster web applications with extensive user/server interaction.
- **Structured Query Language** - interact with databases.

- **Linux-Apache-MySQL-PHP (LAMP)** - a standard installation of a web server with basic software to run it.
- This is the installation used on webdev.cs.uwosh.edu.
- You are welcome to install LAMP on your own Linux machine for testing purposes but all work must be submitted to the webdev server.
- WAMP is a Windows based installations similar to LAMP. You can find it at http://www.wampserver.com/en/.

- HTML describes the contents of web pages. CSS describes the appearance. Scripts describe the behavior.
- Current version is HTML5.
- Be sure your browsers are up-to-date as older browsers may not be able to handle HTML5.
- Most web servers assume **index.html** or some variant of it is the "main" page. If a page is not specified, the index page is shown. www.uwosh.edu is really http://www.uwosh.edu/index.html.

- An HTML document has the extension .html.
- The *content* is the information you want displayed.
- The *markup* consists of tags that tell the browser how to structure and display the information.
- Tags consist of lowercase names surrounded by angle brackets. For example, . Most tags come in pairs, e.g. and .
- **Always** include end tags.

Example

The following is called an *element*.

 Here is my content!

Page Structure

A basic HTML page will always contain the following tags.

Example

```
<!DOCTYPE html>
<html>
    <head>
        <title>Title of Page</title>
        <!– Header information contains general information about page –>
    </head>
    <body>
        Content of page.
    </body>
</html>
```

Block Elements

- **Block Element** - represents a significant element of a page and contains a large amount of content. Displayed in a browser with a line break and vertical margins above and below it.
- **Do not choose tags based on how it will look**. Choosing tags based on the content is called *semantic HTML*. Top level headings should be h1, nested headings should be h2 and so on.
- Several block elements located inside the *body* section include:

Element	Description	Syntax
p	Paragraph	Content
h1 to h6	Headings	Content
hr	Horizontal Rule (line)	<hr />
	Comments	<!– Comment here –>

- **Inline Element** - generally links or images. Must be inside a block element. Cannot contain a block element. Displayed on same line as surrounding content.
- Several common inline elements located inside the *body* section include:

Element	Description	Syntax
img	Images	
a	Anchor/Link	Content
br	Line Break	
em	Emphasis	Content
strong	Strong Emphasis	Content

- img contains **attributes** src and alt. Attributes are written in lowercase with a value in quotation marks.
- **Absolute URL** - A URL that is fully specified. For example, http://www.uwosh.edu/img/parts/wordmark.png.
- **Relative URL** - A URL that is relative. If modifying http://www.uwosh.edu/index.html, one can use the relative: img/parts/wordmark.png

Lists

- HTML has three types of lists: unordered lists, ordered lists and definition lists.

Element	Description	Syntax
ul	Unordered List	List Item ...
ol	Ordered List	List Item ...
dl	Definition List	<dl><dt>Term</dt> <dd>Definition</dd></dl>

Tables

- table, tr and td display the content of the table.

Element	Description	Syntax
table	Table	<table>...</table>
tr	Row	<tr>...</tr>
td	Cell	<td>...</td>
th	Heading	<th>...</th>
caption	Caption	<caption>...</caption>

- To span across multiple rows or columns, use:
  <td rowspan="rows" colspan="columns">
- **Do not use tables for layout.**

Element	Description	Syntax
q	Quotation	<q>Content</q>
blockquote	Bigger Quotation	<blockquote>Content</blockquote>
code	Computer Code	<code>Content</code>
pre	Pre-formatted Text	<pre>Content</pre>
sup	Superscript	^{Content}
sub	Subscript	_{Content}
abbr	Abbreviation	<abbr title="long title">short</abbr>

- In order to display <p> on a webpage, you'd need to use *character entity references*. A large list can be found at http://dev.w3.org/html5/html-author/charref
- In order to display a <, you would need to use <

Multimedia

- HTML5 introduced tags for audio and video. Much easier to use than previous versions.
- Most browsers will be able to display common audio/video formats. However, there is no format that will always play on all browsers. To use audio:
  <audio src="URL">text</audio>
- Some common attributes include:

Name	Value	Details
autoplay	"autoplay"	clips starts immediately
controls	"controls"	shows play, pause, stop buttons
loop	"loop"	clips plays on a loop
src	URL	location of audio file

Multimedia

- To use video:
  <video src="URL" width="400" height="200"
  controls="controls">text</video>
- Some common attributes include:

Name	Value	Details
autoplay	"autoplay"	clips starts immediately
controls	"controls"	shows play, pause, stop buttons
loop	"loop"	clips plays on a loop
src	URL	location of audio file
muted	"muted"	mutes audio

- **Web Standards** - general group of formal standards and technical specifications for various aspects of the World Wide Web.
- Several groups publish standards including W3C, IETF, ISO, Unicode Consortium, etc.
- Following standards is the best way to ensure as many browsers as possible view the page correctly.
- Your code must meet W3C specifications. You can validate your code at http://validator.w3.org/
- Similar to the Java compiler, errors may cascade. As with Java, fix the first error and revalidate.
- Be sure to add the following to the bottom of all of your pages:

  Validate Me

Metadata

- You should use metadata to describe your page.
- Metadata is placed in the head section.
- All of the metadata details can be found at http://www.w3schools.com/tags/tag_meta.asp