Housekeeping

- Minerva was down
- Assignment 1 - updated answers.html
- Teaching Assistants

XHTML: a simple example

<!-- HTML: a simple example -->

<!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>
<link rel="stylesheet" href="simple-style.css" type="text/css"/>
</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.ukans.edu/">University of Kansas</a>
<br/><img src="images/KUSeal.gif" alt="University of Kansas Seal" height="73" width="72"/></li>
</ul>

And here is the stylesheet:

body {
  margin-left: 10%;
  margin-top: 2em;
  margin-right: 10%;
  background-color: #339922;
}

h1 {
  font-family: sans-serif;
  color: #990000;
}

li {
  margin-top: 0.5em;

a {
  text-decoration: none;

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

  a:hover {
    color: white;
    background-color: blue;
  }
}
Well-formed (X)HTML and Valid (X)HTML

Well-formed XHTML

- 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

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

XHTML: Content vs. Appearance

Extensible HyperText Markup Language

- Primary purpose: content

Cascading Style Sheets (CSS)

- Primary purpose: style and appearance

This is really more of a "separation of concerns" -- let each one do that they do best.
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.
HTTP Clients

- Mozilla Firefox
  - Web Developer Extension for Firefox
- Microsoft Internet Explorer
- Opera
- Safari (Mac)
- Netscape Navigator/Communicator
- Mozilla Camino (Mac)
- Lynx
- Amaya

HyperText Markup Language

W3C HyperText Markup Language Home Page

Specifications from the W3C

- XHTML 1.1, Module-based XHTML (May 2001)
- XHTML 1.0, a reformulation of HTML 4.0 into XML 1.0, Strict and Transitional (January 2000)
- HTML 4.01, Strict, Transitional and Frameset (December 1999)
- HTML 4.0, Strict, Transitional and Frameset (December 1997)
- HTML 3.2 (January 1997)
- HTML 2.0 (November 1995)

Strict, Transitional ("Loose"), and Frameset

- Strict
- Loose
- Frameset

General Recommendation: XHTML 1.0 (Strict or Transitional)

Document Type Declaration and Document Type Definition (DTD)

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

```
<!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:

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

- 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. 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.
Common Document Type Declarations for XHTML and HTML

- **XHTML 1.1**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
  "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">`
  ```
- **XHTML 1.0 Transitional**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">`
  ```
- **XHTML 1.0 Strict**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`
  ```
- **HTML 4.01 Frameset**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Frameset//EN"
  "http://www.w3.org/TR/html401/frameset.dtd">`
  ```
- **HTML 4.01 Strict**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Strict//EN"
  "http://www.w3.org/TR/html40/strict.dtd">`
  ```
- **HTML 4.01 Transitional**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"
  "http://www.w3.org/TR/html401/loose.dtd">`
  ```
- **HTML 4.0 Frameset**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0 Frameset//EN"
  "http://www.w3.org/TR/html40/frameset.dtd">`
  ```
- **HTML 4.0**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.0//EN"
  "http://www.w3.org/TR/html4/strict.dtd">`
  ```
- **HTML 3.2**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 3.2//EN"
  "http://www.w3.org/TR/HTML.dtd">`
  ```
- **HTML 2.0**
  ```xml```
  `<!DOCTYPE html PUBLIC "-//IETF//DTD HTML 2.0//EN"
  "http://www3.org/MarkUp/html-spec/html.dtd">`
  ```

XHTML 1.0

XHTML 1.0 is the Extensible HyperText Markup Language (Second Edition) from the W3C. XHTML 1.0 is a reformulation of HTML 4.0 in XML 1.0. It comes in three "flavors": strict, transitional, and frameset.

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

Documentation for XHTML 1.0

- **XHTML 1.0 The Extensible HyperText Markup Language (Second Edition) (W3C)**
- **XHTML 1.0 Strict DTD (W3C)**
- **XHTML 1.0 Annotated DTD (W3C)**
- **Documentation Derived from the DTD**
  - **XHTML 1.0 Strict documentation**
  - **XHTML 1.0 Strict documentation (DTD Parse)**

XHTML 1.0 Elements

- **Structural**
  - `body`, `head`, `html`, `title`
- **Text**
  - `abbr`, `acronym`, `address`, `blockquote`, `br`, `cite`, `code`, `dfn`, `div`, `em`, `h1`, `h2`, `h3`, `h4`, `h5`, `h6`, `kbd`, `p`, `pre`, `q`, `samp`, `span`, `strong`, `var`
  - **Heading**
    - `h1`, `h2`, `h3`, `h4`, `h5`, `h6`
  - **Block**
    - `address`, `blockquote`, `div`, `p`, `pre`
  - **Inline**
    - `abbr`, `acronym`, `br`, `cite`, `code`, `dfn`, `em`, `kbd`, `q`, `samp`, `span`, `strong`, `var`
  - **Flow**
    - **Heading**, **Block**, **Inline**
- **Hypertext**
  - `a`
- **List**
  - `dl`, `dt`, `dd`, `ol`, `ul`, `li`
- **Applet**, `deprecated`.  
  - `applet`, `param`
- **Text Extensions**
  - **Presentation**
    - `b`, `big`, `br`, `i`, `small`, `sup`, `sub`, `tt`
  - **Edit**
    - `del`, `ins`
Bi-directional text
• bdo

Forms
• Basic Forms
  • form, input, label, select, option, textarea
• Forms
  • form, input, select, option, textarea, button, fieldset, label, legend, optgroup

Tables
• Basic Tables
  • caption, table, td, th, tr
• Tables
  • caption, table, td, th, tr, col, colgroup, tbody, thead, tfoot

Image
• img

Client-side Imagemap
• area, map

Object
• object, param

Frames
• frameset, frame, noframes

Iframe
• iframe

Metainformation
• meta

Script
• script, noscript

Style
• style

---

Structural
html, head, body, title

Basic HTML Page

---

Reading the DTD

Three main things to be concerned with:
1. Content Model
2. Attribute List
3. Expanding defined entities (e.g. %Block, %Inline)

What are the attributes and content model for "body"?
• body, W3C annotated DTD
• body, DTD Parse

Some notations important for reading DTDs:
*, asterisk
zero or more
+, plus
one or more
( ), parentheses
 grouping
|, pipe
 or

Reading the XHTML Specification

Start with the HTML 4.01 Specification
XHTML in particular...
Hypertext

The `a` element.

**Example 2.1**

**Example 2.1 Source:**

```html
<ul>
  <li><a href="http://cscie12.dce.harvard.edu">CSCIE 12 Web Site</a></li>
  <li><a href="mailto:david_heitmeyer@harvard.edu">david_heitmeyer@harvard.edu</a></li>
  <li><a href="slide2.html">Slide 2</a></li>
  <li><a href="http://www.w3.org/" title="World Wide Web Consortium: Leading the Web to its Full Potential">W3C</a></li>
  <li><a href="http://extension.harvard.edu/">Extension School</a></li>
</ul>
```

**Example 2.1 Rendered:**

- CSCIE 12 Web Site
- david_heitmeyer@harvard.edu
- Slide 2
- W3C
- Extension School

---

**An aside: Absolute and Relative Locations**

Absolute, or fully-qualified, URIs specify the complete information.

**Example 2.2**

**Example 2.2 Source:**

```html
```

**Example 2.2 Rendered:**

Diplomacy of Lewis and Clark stressed in exhibit

Relative, or partial, URIs specify partial information. The information not given is inferred from the current location.

**Example 2.3**

**Example 2.3 Source:**

```html
<a href="slide1.html">Slide 1</a>
```

**Example 2.3 Rendered:**

Slide 1

Is this relative or absolute? Scheme, host, and port would be taken from current location, but path is
Example 2.4

*Example 2.4 Source:*

```html
<a href="/copyright.html">copyright information</a>
```

*Example 2.4 Rendered:*

copyright information

---

Aside: Relative Paths to Parent Locations

- `../` refers to the parent directory
- `./` refers to current directory

Example 2.5

*Example 2.5 Source:*

```html
<a href="../index.html">Up a level</a>
```

*Example 2.5 Rendered:*

Up a level

---

Text: Heading

- h1
- h2
- h3
- h4
- h5
- h6

Example 2.6

*Example 2.6 Source:*

```html
<h3>A Third Level Heading</h3>
<p>Lorem ipsum dolor sit amet, consectetuer adipiscing elit...</p>
<h4>A Fourth Level Heading</h4>
<p>Lorem ipsum dolor sit amet, consectetuer adipiscing elit...</p>
<h5>A Fifth Level Heading</h5>
<p>Lorem ipsum dolor sit amet, consectetuer adipiscing elit...</p>
<h6>A Sixth Level Heading</h6>
<p>Lorem ipsum dolor sit amet, consectetuer adipiscing elit...</p>
```

*Example 2.6 Rendered:*

**A Third Level Heading**

Lorem ipsum dolor sit amet, consectetuer adipiscing elit...

**A Fourth Level Heading**

Lorem ipsum dolor sit amet, consectetuer adipiscing elit...

**A Fifth Level Heading**

Lorem ipsum dolor sit amet, consectetuer adipiscing elit...

**A Sixth Level Heading**

Lorem ipsum dolor sit amet, consectetuer adipiscing elit...
Text: Block

- div
- p
- address
- blockquote
- pre

Example 2.7

Example 2.7 Source:

```html
<div style="text-align: center; background-color: yellow; color: red; font-family: helvetica, sans-serif;" >
Division (div) elements are block-level and will be very useful when we discuss stylesheets.</div>
<p>Paragraph (p) elements should be used to denote paragraphs. They are commonly mis-used to introduce vertical whitespace.</p>
```

Example 2.7 Rendered:
Division (div) elements are block-level and will be very useful when we discuss stylesheets.

Paragraph (p) elements should be used to denote paragraphs. They are commonly mis-used to introduce vertical whitespace.

Example 2.8

Example 2.8 Source:

```html
<p>In his <em>I Have a Dream</em> speech delivered in August 1963, Martin Luther King Jr. said: </p>
<blockquote cite="http://www.americanrhetoric.com/speeches/Ihaveadream.htm">
<p>I have a dream that one day this nation will rise up and live out the true meaning of its creed: We hold these truths to be self-evident that all men are created equal.</p>
<p>I have a dream that one day on the red hills of Georgia the sons of former slaves and the sons of former slave owners will be able to sit down together at the table of brotherhood.</p>
<p>I have a dream that one day even the state of Mississippi, a state sweltering with the heat of injustice, sweltering with the heat of oppression, will be transformed into an oasis of freedom and justice.</p>
<p>I have a dream that my four little children will one day live in a nation where they will not be judged by the color of their skin but by the content of their character. I have a <em>dream</em> today!</p>
<p>I have a dream that one day, down in Alabama, with its vicious racists, with its governor having his lips dripping with the words of interposition and nullification; one day right there in Alabama little black boys and black girls will be able to join hands with little white boys and white girls as sisters and brothers. I have a <em>dream</em> today!</p>
<p>I have a dream that one day every valley shall be exalted, and every hill and mountain shall be made low, the rough places will be made plain, and the crooked places will be made straight, and the glory of the Lord shall be revealed and all flesh shall see it together.</p>
</blockquote>
```

Example 2.8 Rendered:
In his <em>I Have a Dream</em> speech delivered in August 1963, Martin Luther King Jr. said:

I have a dream that one day this nation will rise up and live out the true meaning of its creed: We hold these truths to be self-evident that all men are created equal.

I have a dream that one day on the red hills of Georgia the sons of former slaves and the sons of former slave owners will be able to sit down together at the table of brotherhood.

I have a dream that one day even the state of Mississippi, a state sweltering with the heat of injustice, sweltering with the heat of oppression, will be transformed into an oasis of freedom and justice.

I have a dream that my four little children will one day live in a nation where they will not be judged by the color of their skin but by the content of their character. I have a <em>dream</em> today!

I have a dream that one day, down in Alabama, with its vicious racists, with its governor having his lips dripping with the words of interposition and nullification; one day right there in Alabama little black boys and black girls will be able to join hands with little white boys and white girls as sisters and brothers. I have a <em>dream</em> today!

I have a dream that one day every valley shall be exalted, and every hill and mountain shall be made low, the rough places will be made plain, and the crooked places will be made straight, and the glory of the Lord shall be revealed and all flesh shall see it together.

pre: where whitespace is important!

<table>
<thead>
<tr>
<th>Stock Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
</tr>
<tr>
<td>RA</td>
</tr>
</tbody>
</table>

Example 2.9

Example 2.9 Source:

```html
<p>The address of the Science Center is: 1 Oxford St., Cambridge, Massachusetts, 01460</p>
```

Example 2.9 Rendered:
The address of the Science Center is:
1 Oxford St., Cambridge, Massachusetts, 01460

pre: where whitespace is important!
Web clients and servers communicate via HTTP.

Example 2.11
Example 2.11 Source:
```
<p>
<acronym title="National Aeronatics and Space Administration">NASA</acronym> was founded in 1958. </p>
```

Example 2.11 Rendered:

NASA was founded in 1958.

Example 2.12
Example 2.12 Source:
```
<p>
<span style="color: white; background: red;" >span elements</span> are useful in CSS. They are an <em>inline</em> partner with the block level <strong>div</strong> elements. </p>
```

Example 2.12 Rendered:

span elements are useful in CSS. They are an inline partner with the block level div elements.

Example 2.13
Example 2.13 Source:
```
<p>
<strong>Strong text</strong> and <b>bold text</b> should not be confused. Likewise, <em>emphasized text</em> should not be confused with <i>italicized text</i>.
</p>
```

Example 2.13 Rendered:

Strong text and bold text should not be confused. Likewise, emphasized text should not be confused with italicized text.

Example 2.14
Example 2.14 Source:
```
<div>
Whitespace, including spaces, carriage returns, and line feeds, are generally ignored in XHTML. If you need a line break, you can use the br element.
</div>
```

Example 2.14 Rendered:

Whitespace, including spaces, carriage returns, and line feeds, are generally ignored in XHTML. If you need a line break, you can use the br element.
Lists

- **Example 2.15**
  
  **Example 2.15 Source:**

  ```html
  <ul>
  <li>Tea</li>
  <ul>
  <li>Kenyan</li>
  <li>Sikkim Temi</li>
  <li>Himalayan</li>
  </ul>
  </li>
  <li>Potato Chips</li>
  <ul>
  <li>Dirty's</li>
  <li>Art's and Mary's</li>
  <li>Tim's Cascade</li>
  </ul>
  </ul>
  ```

  **Example 2.15 Rendered:**

  - Tea
    - Kenyan
    - Sikkim Temi
    - Himalayan
  - Potato Chips
    - Dirty's
    - Art's and Mary's
    - Tim's Cascade

- **Example 2.16**

  **Example 2.16 Source:**

  ```html
  <ol>
  <li>Boil water</li>
  <li>Measure tea (approximately 1 tsp. per 6 oz. cup)</li>
  <li>Steep tea for 3 to 5 minutes</li>
  <li>Enjoy!</li>
  </ol>
  ```

  **Example 2.16 Rendered:**

  1. Boil water
  2. Measure tea (approximately 1 tsp. per 6 oz. cup)
  3. Steep tea for 3 to 5 minutes
  4. Enjoy!

- **Example 2.17**

  **Example 2.17 Source:**

  ```html
  <div>
  <dl>
  <dt>bread</dt>
  <dd>a usually baked and leavened food made of a mixture whose basic constituent is flour or meal</dd>
  <dt>butter</dt>
  <dd>a solid emulsion of fat globules, air, and water made by churning milk or cream and used as food</dd>
  </dl>
  </div>
  ```

  **Example 2.17 Rendered:**

  - bread
    - a usually baked and leavened food made of a mixture whose basic constituent is flour or meal
  - butter
    - a solid emulsion of fat globules, air, and water made by churning milk or cream and used as food
Image

HTML documents do not contain the images themselves, but merely contain references to the images to be displayed. Common image file types are:

- GIF, Graphics Interchange Format, "gif"
- JPEG, Joint Photographic Experts Group, "jpg", "jpeg"
- PNG, Portable Network Graphics, "png"

Example 2.18
Example 2.18 Source:

```html
<img src="http://cscie12.dce.harvard.edu/images/group.png" alt="Harvard University Extension School Logo" height="65" width="220" />
```

Example 2.18 Rendered:

Harvard University Extension School

Example 2.19
Example 2.19 Source:

```
<a href="http://extension.harvard.edu/"/>
```

Tables

Tables are great for data.

Tables are often co-opted for page layout purposes.

Basic Tables

- `table`
- `tr`
- `td`
- `th`
- `caption`

Example 2.20
Example 2.20 Source:

```html
<table>
  <caption>A table</caption>
  <tr>
    <td>row 1 column 1</td>
    <td>row 1 column 2</td>
    <td>row 1 column 3</td>
  </tr>
  <tr>
    <td>row 2 column 1</td>
    <td>row 2 column 2</td>
    <td>row 2 column 3</td>
  </tr>
  <tr>
    <td>row 3 column 1</td>
    <td>row 3 column 2</td>
    <td>row 3 column 3</td>
  </tr>
</table>
```

Example 2.20 Rendered:

A table

row 1 column 1 row 1 column 2 row 1 column 3
row 2 column 1 row 2 column 2 row 2 column 3
row 3 column 1 row 3 column 2 row 3 column 3
**Comments**

Example 2.21

Example 2.21 Source:

```html
<p>You can make comments in HTML. <!-- Browsers will not display comments --> Comments are very useful. You should use them liberally. <!-- Comments can be seen in the source of the page though! --></p>
```

Example 2.21 Rendered:

You can make comments in HTML. Comments are very useful. You should use them liberally.

---

**XHTML/HTML Character Entities**

XHTML/HTML Character entities can be defined by

- name (&name;)
- numeric value (&#nnn;)

**Character Entities Defined for XML/SGML**

Critical character entities are:

- `&gt;`
- `&lt;`
- `&amp;`
- `&quot;`
- `&apos;`

**Character Entities defined specifically for XHTML/HTML**

And Because we have deficient input devices...

Copyright symbol:

- `&copy;`
- `&copy;`

List of XHTML 1.0 Entity Sets and Character Entities

---

**Pragmatic Items**

**User directories**

Web documents for each user are kept in the user's home directory, in a directory named `public_html` by convention. As an example, for the user `jharvard` whose home directory is `/home/courses/j/h/jharvard`

URI  | File
---   | ---

**Document Root**

The Web documents are typically kept under a single directory, traditionally named `htdocs`. The full path to this directory is called the "document root" of the Web server, for example, `/var/www/html`.

URI  | File
---   | ---

---

**URI to Filename Mapping**

**User directories**

Web documents for each user are kept in the user's home directory, in a directory named `public_html` by convention. As an example, for the user `jharvard` whose home directory is `/home/courses/j/h/jharvard`

URI  | File
---   | ---

**Document Root**

The Web documents are typically kept under a single directory, traditionally named `htdocs`. The full path to this directory is called the "document root" of the Web server, for example, `/var/www/html`.

URI  | File
---   | ---

---
Directory Requests and index.html

Directory without index.html. Note that this assumes the directory permissions are set correctly. If not, you may see Forbidden.

Example 2.22
Example 2.22 Source:

Example 2.22 Rendered:
Images from the CSCIE12 Web Site

Directory with index.html:

Example 2.23
Example 2.23 Source:

Example 2.23 Rendered:
Harvard Academics (index.html)

Software Tools

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

SSH Clients and SFTP Clients

- SecureCR and Secure FX from Van Dyke Technologies
- PuTTY
- Open SSH
- SSH Secure Shell

Text or HTML Editor

Windows

- Your favorite text editor
- Dreamweaver
- Adobe GoLive
- Editpad or Wordpad or Notepad

Macintosh

- Your favorite text editor
- BBEdit by BareBones Software
- Dreamweaver
- Adobe GoLive
- SimpleText

Linux/UNIX

- Your favorite text editor
- Emacs
  Note that Emacs is available for UNIX, Windows, Macintosh
(X)HTML Well-formedness and Validation Resources

- W3C HTML Validation Service
- HTML Tidy

```
minerva% tidy -help
```

HTML Tidy

```
- http://www.w3.org/People/Raggett/tidy/
- http://tidy.sourceforge.net/
```

My favorite tidy options:

- `--wrap 0`
- `--indent`
- `--asxhtml`
- `--e`
- `--m`

```
minerva% tidy -h
```

```
tidy [option...] [file...] [option...] [file...]
Utility to clean up and pretty print HTML/XHTML/XML
see http://tidy.sourceforge.net/
```

Options for HTML Tidy for Linux/x86 released on 1st February 2004:

File manipulation
-----------------
- `-out` or `-o <file>` specify the output markup file
- `-config <file>` set configuration options from the specified <file>
- `-f <file>` write errors to the specified <file>
- `-modify` or `-m` modify the original input files

Processing directives
---------------------
- `-indent` or `-i` indent element content
- `-wrap <column>` wrap text at the specified <column> (default is 68)
- `-clean` or `-c` replace FONT, NOBR and CENTER tags by CSS
- `-bare` or `-b` strip out smart quotes and em dashes, etc.
- `-numeric` or `-n` output numeric rather than named entities
- `-errors` or `-e` only show errors
- `-quiet` or `-q` suppress nonessential output
- `-omit` omit optional end tags
- `-xml` specify the input is well formed XML
- `-asxml` convert HTML to well formed XHTML
- `-asxhtml` force XHTML to well formed HTML
- `-access <level>` do additional accessibility checks (<level> = 1, 2, 3)

Character encodings
-------------------
- `-raw` output values above 127 without conversion to entities
- `-ascii` use US-ASCII for output, ISO-8859-1 for input
- `-latin0` use US-ASCII for output, ISO-8859-1 for input
- `-latin1` use ISO-8859-1 for both input and output
- `-iso2022` use ISO-2022 for both input and output
- `-utf8` use UTF-8 for both input and output
- `-mac` use MacRoman for input, US-ASCII for output
- `-win1252` use Windows-1252 for input, US-ASCII for output
- `-ibm858` use IBM-858 (CP850+Euro) for input, US-ASCII for output
- `-utf16le` use UTF-16LE for both input and output
- `-utf16be` use UTF-16BE for both input and output
- `-utf16` use UTF-16 for both input and output

Miscellaneous
-------------
- `-version` or `-v` show the version of Tidy
- `-help`, `-h` or `-?` list the command line options
- `-help-config` list all configuration options
- `-show-config` list the current configuration settings

Use `--blah blarg` for any configuration option "blah" with argument "blarg"

Input/Output default to stdin/stdout respectively

- Use `tidy -f errs.txt -imu foo.html` for further info on HTML see http://www.w3.org/MarkUp
Download/Viewing Source of a Web Page from another site.

- Browser: "View Source" or "Save"

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

- Mozilla Firefox
  - Web Developer Extension for Firefox
- Netscape Navigator/Communicator
- Microsoft Internet Explorer
- Opera
- Safari (Apple)
- Lynx
- Amaya

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
- Editpad or Wordpad or Notepad

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

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

(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

HTML Tidy

http://www.w3.org/People/Raggett/tidy/

```
minerva% tidy -h
```

tidy [option...] [file...]
Utility to clean up and pretty print HTML/XHTML/XML
see http://www.w3.org/People/Raggett/tidy/

Options for HTML Tidy for Linux/x86 released on 1st September 2002:

Processing directives

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-indent</td>
<td>to indent element content</td>
</tr>
<tr>
<td>-omit</td>
<td>to omit optional end tags</td>
</tr>
<tr>
<td>-wrap</td>
<td>to wrap text at the specified &lt;column&gt; (default is 68)</td>
</tr>
<tr>
<td>-upper</td>
<td>to force tags to upper case (default is lower case)</td>
</tr>
<tr>
<td>-clean</td>
<td>to replace FONT, NOBR and CENTER tags by CSS</td>
</tr>
<tr>
<td>-bare</td>
<td>to strip out smart quotes and em dashes, etc.</td>
</tr>
<tr>
<td>-numeric</td>
<td>to output numeric rather than named entities</td>
</tr>
<tr>
<td>-errors</td>
<td>to only show errors</td>
</tr>
<tr>
<td>-quiet</td>
<td>to suppress nonessential output</td>
</tr>
<tr>
<td>-xml</td>
<td>to specify the input is well formed XML</td>
</tr>
<tr>
<td>-asxml</td>
<td>to convert HTML to well formed XHTML</td>
</tr>
<tr>
<td>-axhtml</td>
<td>to force XHTML to well formed HTML</td>
</tr>
<tr>
<td>-slides</td>
<td>to burst into slides on #2 elements</td>
</tr>
<tr>
<td>-access</td>
<td>to do additional accessibility checks</td>
</tr>
</tbody>
</table>

You can also use --blah for any configuration option blah

Character encodings

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-raw</td>
<td>to output values above 127 without conversion to entities</td>
</tr>
<tr>
<td>-ascii</td>
<td>to use US-ASCII for output, ISO-8859-1 for input</td>
</tr>
<tr>
<td>-lat1</td>
<td>to use ISO-8859-1 for both input and output</td>
</tr>
<tr>
<td>-iso2022</td>
<td>to use ISO-2022 for both input and output</td>
</tr>
<tr>
<td>-utf8</td>
<td>to use UTF-8 for both input and output</td>
</tr>
<tr>
<td>-mac</td>
<td>to use MacRoman for input, US-ASCII for output</td>
</tr>
<tr>
<td>-win1252</td>
<td>to use Windows-1252 for input, US-ASCII for output</td>
</tr>
</tbody>
</table>

File manipulation

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>-config</td>
<td>to set configuration options from the specified &lt;file&gt;</td>
</tr>
<tr>
<td>-f</td>
<td>to write errors to the specified &lt;file&gt;</td>
</tr>
<tr>
<td>-m</td>
<td>to modify the original input files</td>
</tr>
<tr>
<td>-version</td>
<td>to show the version of Tidy</td>
</tr>
<tr>
<td>-help</td>
<td>to list the command line options</td>
</tr>
<tr>
<td>-show-config</td>
<td>to list all configuration options</td>
</tr>
<tr>
<td>-show-config</td>
<td>to list the current configuration settings</td>
</tr>
</tbody>
</table>

You can also use --blah for any configuration option blah

Input/Output default to stdin/stdout respectively

Single letter options apart from -f may be combined as in: tidy -f errs.txt -imu foo.html
For further info on HTML see http://www.w3.org/MarkUp
Your Minerva Account

1. Hostname: minerva.dce.harvard.edu
2. ssh minerva.dce.harvard.edu
3. Username: first letter of first name concatenated with up to the first 7 letters (only count "word" characters) of your last name, all lower-case
4. Password: up to the first 8 characters of your last name, all lower-case

minerva% passwd

URI to Filename Mapping

<table>
<thead>
<tr>
<th>User directories</th>
<th>URI</th>
<th>File</th>
</tr>
</thead>
<tbody>
<tr>
<td>htdocs/www/htdocs</td>
<td>URI</td>
<td>File</td>
</tr>
</tbody>
</table>

Absolute and Relative Locations

Absolute

Absolute, or fully-qualified, URIs specify the complete information.

Example 2.24

Example 2.24 Source:

```html
<a href="http://cscie12.dce.harvard.edu/lecture/2/notes/slide1.html" >Slide 1</a>
```

Example 2.24 Rendered:

Slide 1

Example 2.25

Example 2.25 Source:

```html
<a href="http://www.whitehouse.gov" >White House</a>
```

Example 2.25 Rendered:

White House

Relative

Relative, or partial, URIs specify partial information. The information not given is assumed from the current location.

Example 2.26

Example 2.26 Source:

```html
<a href="slide1.html" >Slide 1</a>
```

Example 2.26 Rendered:

Slide 1
Location Shortcuts: ../ and ./

- ../
- ./

Example 2.27

Example 2.27 Source:

```html
<a href="../index.html">Up a level</a>
```

Example 2.27 Rendered:

Up a level

Directory Requests and index.html

Directory without `index.html`. Note that this assumes the directory permissions are set correctly. If not, you may see Forbidden.

Example 2.28

Example 2.28 Source:

```html
<a href="http://cscie12.dce.harvard.edu/images/">Images from the CSCIE12 Web Site</a>
```

Example 2.28 Rendered:

Images from the CSCIE12 Web Site

Directory with `index.html`:

Example 2.29

Example 2.29 Source:

```html
<a href="http://www.harvard.edu/academics">Harvard Academics</a>
```

Example 2.29 Rendered:

Harvard Academics (index.html)

Download/Viewing Source of a Web Page from another site.

- Browser: "View Source" or "Save"

Online Web Development Resources

Some of my favorites:

- World Wide Web Consortium
  http://www.w3c.org/
- O'Reilly Web and Internet Center
  http://web.oreilly.com/

Table of Contents | All Slides | Link List | Examples | CSCI E-12