Web Programming Step by Step

Chapter 2
HTML Basics

2.1: Basic HTML

Hypertext Markup Language (HTML) (2.1.1)

- describes the *content* and structure of information on a web page
  - not the same as the *presentation* (appearance on screen)
- surrounds text content with opening and closing **tags**
- each tag's name is called an **element**
  - syntax: <*element*> *content* </*element*>
  - example: `<p>This is a paragraph</p>`
- most whitespace is insignificant in HTML
  (it gets ignored or collapsed into a single space)

XHTML

- a newer version of HTML, standardized in 2000
- uses a markup format called **XML** (XML + HTML = XHTML)
- though the browser will accept some malformed HTML, we'll write "strict" XHTML that complies to the official web standards
- a strict XHTML page uses some different syntax and tags

Structure of an XHTML page (2.1.2)

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

<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    information about the page
  </head>

  <body>
    page contents
  </body>
</html>
```

- the **header** describes the page and the **body** contains the page's contents
- an HTML page is saved into a file ending with extension `.html`

Page title: `<title>`

describes the title of the web page

```HTML
<title>Chapter 2: HTML Basics</title>
```

- placed within the `head` of the page
- displayed in the web browser's title bar and when bookmarking the page

Paragraph: <p> (2.1.3)

paragraphs of text (block)

```HTML
<p>You're not your job.
You're not how much money you have in the bank.
You're not the car you drive.   You're not the contents
of your wallet. You're not your         khakis.  You're
   the all-singing, all-dancing crap of the world.</p>
```

output

You're not your job. You're not how much money you have in the bank. You're not the car you drive. You're not the contents of your wallet. You're not your khakis. You're the all-singing, all-dancing crap of the world.

- placed within the `body` of the page
- more paragraph examples

Headings: <h1>, <h2>, ..., <h6>

headings to separate major areas of the page (block)

```HTML
<h1>University of Whoville</h1>
<h2>Department of Computer Science</h2>
<h3>Sponsored by Micro$oft</h3>
```

output

University of Whoville

Department of Computer Science

Sponsored by Micro$oft

- More heading examples

Horizontal rule: <hr>

a horizontal line to visually separate sections of a page (block)

```
<p>First paragraph</p>
<hr />
<p>Second paragraph</p>
```
HTML

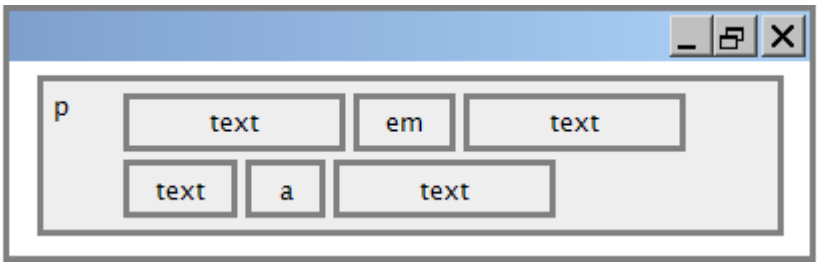
First paragraph

Second paragraph

output

- should be immediately closed with />

Block and inline elements (explanation)



- **block** elements contain an entire large region of content
  - examples: paragraphs, lists, table cells
  - the browser places a margin of whitespace between block elements for separation
- **inline** elements affect a small amount of content
  - examples: bold text, code fragments, images
  - the browser allows many inline elements to appear on the same line
  - must be nested inside a block element

More about HTML tags

- some tags can contain additional information called **attributes**
  - syntax: <*element attribute*="*value*" *attribute*="*value*"> *content* </*element*>
  - example: `<a href="page2.html">Next page</a>`
- some tags don't contain content; can be opened and closed in one tag
  - syntax: <*element attribute*="*value*" *attribute*="*value*" />
  - example: `<hr />`
  - example: `<img src="bunny.jpg" alt="pic from Easter" />`

Links: <a> (2.1.4)

links, or "anchors", to other pages (inline)

```
<p>
  Search
  <a href="http://www.google.com/">Google</a>
  now!
</p>
```

HTML

Search Google now!

output

- uses the `href` attribute to specify the destination URL
- anchors are inline elements, so they must be placed inside a block element such as a `p` or `h1`

More about anchors

```
<p><a href="1-internet.html">Lecture Notes 1</a></p>
<p><a href="http://www.google.com/"
title="Search">Google</a></p>
```
HTML

Lecture Notes 1

Google

output

- types of URLs that can appear in anchors:
  - **absolute** (to another web site)
  - **relative** (to another page on this web site)
- specify a tooltip with the `title` attribute
- more anchor examples

- to make links that open in new windows, we'll need to learn Javascript (later)

Nesting tags

Bad:

```
<p>
  <a href="1-internet.html">Lecture Notes 1
</p>
<p>
  This text also links to Lecture Notes 1</a>
</p>
```
HTML

- tags must be correctly nested
- (a closing tag must match the most recently opened tag)
- the browser may render it correctly anyway, but it is invalid XHTML

Images: <img>

inserts a graphical image into the page (inline)

```HTML
<img src="images/gollum.jpg" alt="Gollum from LOTR" />
```

output

- the `src` attribute specifies the image URL
- XHTML also requires an `alt` attribute describing the image

More about images

```HTML
<a href="http://theonering.net/">
  <img src="images/gandalf.jpg" alt="Gandalf from LOTR"
       title="You shall not pass!" />
</a>
```

output

- if placed inside an `a` anchor, the image will become a link
- the `title` attribute specifies an optional tooltip

Line break:

forces a line break in the middle of a block element (inline)

```
<p>Teddy said it was a hat, <br /> So I put it on.</p>
<p>Now Daddy's sayin', <br /> Where
the heck's the toilet plunger gone?</p>
```
HTML

Teddy said it was a hat,
So I put it on.

Now Daddy's sayin',
Where the heck's the toilet plunger gone?

output

- `br` should be immediately closed with `/>`
- `br` should not be used to separate paragraphs or used multiple times in a row to create spacing

Comments: <!-- ... -->

comments to document your HTML file or "comment out" text

```
<!-- My web page, by Suzy Student
CSE 190 D, Spring 2048 -->
<p>CSE courses are <!-- NOT --> a lot of fun!</p>
```
HTML

CSE courses are a lot of fun!

output

- many web pages are not thoroughly commented (or at all)
- comments are still useful for disabling sections of a page
- comments cannot be nested and cannot contain a `--`

Phrase elements : ,

em: emphasized text (usually rendered in italic)
strong: strongly emphasized text (usually rendered in bold)

```
<p>
  HTML is <em>really</em>,
  <strong>REALLY</strong> fun!
</p>
```
HTML

HTML is *really*, **REALLY** fun!

output

- as usual, the tags must be properly nested for a valid page

2.2: More HTML Elements

- 2.1: Basic HTML
- **2.2: More HTML Elements**
- 2.3: Web Standards

Unordered list: <ul>, <li> (2.2.1)

ul represents a bulleted list of items (block)
li represents a single item within the list (block)

```HTML
<ul>
  <li>No shoes</li>
  <li>No shirt</li>
  <li>No problem!</li>
</ul>
```

- No shoes
- No shirt
- No problem!

output

More about unordered lists

- a list can contain other lists:

```HTML
<ul>
  <li>Simpsons:
    <ul>
      <li>Homer</li>
      <li>Marge</li>
    </ul>
  </li>
  <li>Family Guy:
    <ul>
      <li>Peter</li>
      <li>Lois</li>
    </ul>
  </li>
</ul>
```

- Simpsons:
  - Homer
  - Marge
- Family Guy:
  - Peter
  - Lois

output

Ordered list:

ol represents a numbered list of items (block)

```HTML
<p>RIAA business model:</p>
<ol>
  <li>Sue customers for copying music</li>
  <li>???</li>
  <li>Profit!</li>
</ol>
```

RIAA business model:

1. Sue customers for copying music
2. ???
3. Profit!

output

- we can make lists with letters or Roman numerals using CSS (later)

Common error: Not closing a list

```HTML
<ul>
  <li>No shoes</li>
  <li>No shirt</li>
  <li>No problem!</li>

<p>Paragraph after list...</p>
```

- No shoes
- No shirt
- No problem!

  Paragraph after list...

output

- if you leave a list open, subsequent contents will be indented

Common Error: Improper nested list placement

```
<ul>
  <li>Simpsons:</li>
    <ul>
    <li>Bart</li>
    <li>Lisa</li>
    </ul>
  </li>
  <li>Family Guy:
    <ul>
    <li>Peter</li>
    <li>Lois</li>
    </ul>
</ul>
```

HTML

- closing the outer `li` too early (or not at all) will render correctly in most browsers, but it is incorrect XHTML

Definition list: <dl>, <dt>, <dd>

`dl` represents a list of definitions of terms (block)
`dt` represents each term, and `dd` its definition

```
<dl>
  <dt>newbie</dt> <dd>one who does not have mad skills</dd>
  <dt>own</dt> <dd>to soundly defeat
    (e.g. I owned that newbie!)</dd>
  <dt>frag</dt> <dd>a kill in a shooting game</dd>
</dl>
```

HTML

newbie
: one who does not have mad skills

own
: to soundly defeat (e.g. I owned that newbie!)

frag
: a kill in a shooting game

output

Quotations: <blockquote> (2.2.3)

a lengthy quotation (block)

```
<p>As Lincoln said in his famous Gettysburg Address:</p>
<blockquote>
  <p>Fourscore and seven years ago, our fathers brought forth
    on this continent a new nation, conceived in liberty, and
    dedicated to the proposition that all men are created equal.</p>
</blockquote>
```
HTML

As Lincoln said in his famous Gettysburg Address:

> Fourscore and seven years ago, our fathers brought forth on this continent a new nation, conceived in liberty, and dedicated to the proposition that all men are created equal.

output

Inline quotations: <q>

a short quotation (inline)

```
<p>Quoth the Raven, <q>Nevermore.</q></p>
```
HTML

Quoth the Raven, “Nevermore.”

output

- Why not just write the following?
  `<p>Quoth the Raven, "Nevermore."</p>`

We don't use " marks for two reasons:

1. XHTML shouldn't contain literal quotation mark characters; they should be written as `&quot;`
2. using <q> allows us to apply CSS styles to quotations (seen later)

HTML Character Entities (2.2.4)

a way of representing any Unicode character within a web page

character(s)	entity
< >	< >
é è ñ	é è ñ
™ ©	™ ©
π δ Δ	π δ Δ
И	И
" &	" &

- Complete list of HTML entities
- How would you display the text `&` on a web page?

HTML-encoding text

```HTML
&lt;p&gt;
  &lt;a href=&quot;http://google.com/search?q=marty&amp;ie=utf-8&amp;aq=t&quot;&gt;
    Search Google for Marty
  &lt;/a&gt;
&lt;/p&gt;
```

```output
<p> <a href="http://google.com/search?q=marty&ie=utf-8&aq=t"> Search Google for Marty </a> </p>
```

- To display the link text in a web page, its special characters must be encoded as shown above

Computer code: <code>

code: a short section of computer code (usually rendered in a fixed-width font)

```
<p>
  The <code>ul</code> and <code>ol</code>
  tags make lists.
</p>
```
HTML

The `ul` and `ol` tags make lists.

output

Preformatted text: <pre>

a large section of pre-formatted text (block)

```
<pre>
         Steve Jobs speaks loudly
            reality distortion
           Apple fans bow down
</pre>
```
HTML

```
         Steve Jobs speaks loudly
            reality distortion
           Apple fans bow down
```
output

- displayed with exactly the whitespace / line breaks given in the text
- shown in a fixed-width font by default
- how would it look if we had instead enclosed it in `code` tags?

Using `pre` and `code` together

```
<pre><code>
    public static void main(String[] args) {
        System.out.println("Hello, world!");
    }
</code></pre>
```
HTML

```
    public static void main(String[] args) {
        System.out.println("Hello, world!");
    }
```
output

- When showing a large section of computer code, enclose it in a `pre` to preserve whitespace and a `code` to describe the semantics of the content

2.3: Web Standards

- 2.1: Basic HTML
- 2.2: More HTML Elements
- **2.3: Web Standards**

Web Standards (2.3.1)

- It is important to write proper XHTML code and follow proper syntax.
- Why use XHTML and web standards?
  - more rigid and structured language
  - more interoperable across different web browsers
  - more likely that our pages will display correctly in the future
  - can be interchanged with other XML data: SVG (graphics), MathML, MusicML, etc.

W3C XHTML Validator (2.3.2)

```
<p>
  <a href="http://validator.w3.org/check/referer">
    <img src="http://www.w3.org/Icons/valid-xhtml11"
    alt="Validate" />
  </a>
</p>
```

HTML

output

- validator.w3.org
- checks your HTML code to make sure it meets the official strict XHTML specifications
- more picky than the web browser, which may render malformed XHTML correctly

Web page metadata: <meta> (2.3.3)

information about your page (for a browser, search engine, etc.)

```HTML
<meta name="description"
 content="Authors' web site for Building Java Programs." />
<meta name="keywords" content="java, textbook" />
<meta http-equiv="Content-Type"
 content="text/html; charset=iso-8859-1" />
```

- placed in the `head` of your XHTML page
- `meta` tags often have both the `name` and `content` attributes
  - some `meta` tags use the `http-equiv` attribute instead of `name`

meta element to aid browser / web server

```HTML
<meta http-equiv="Content-Type"
 content="type of document (character encoding)" />
<meta http-equiv="refresh"
 content="how often to refresh the page (seconds)" />
</head>
```

- using the `Content-Type` gets rid of the W3C "tentatively valid" warning

  ```
  <meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" />
  ```

- the `meta refresh` tag can also redirect from one page to another:

  ```
  <meta http-equiv="refresh" content="5;url=http://www.bjp.com" />
  ```

  - why would we want to do this? (example)

`meta` element to describe the page

```HTML
<head>
<meta name="author"
 content="web page's author" />
<meta name="revised"
 content="web page version and/or last modification date" />
<meta name="generator"
 content="the software used to create the page" />
</head>
```

- many WYSIWYG HTML editors (FrontPage, PageMaker, etc.) place their names in the `meta generator` tag (why?)

`meta` element to aid search engines

```HTML
<head>
<meta name="description"
 content="how you want search engines to display your page" />
<meta name="keywords"
 content="words to associate with your page (comma-separated)" />
</head>
```

- these are suggestions to search engines about how to index your page
- the search engine may choose to ignore them (why?)