Chapter 4
A Hypertext Markup Language Primer

fluency5

Learning Objectives
- Know the meaning of and use hypertext terms
- Use HTML tags to structure a document
- Use HTML tag attributes
- Use HTML tags to link to other files
- Explain the differences between absolute and relative pathnames
- Use Cascading Style Sheets to style a Web page
- Use HTML to encode lists and tables

Web Pages
- Web pages are created, stored, and sent in encoded form
- A browser converts them to what we see on the screen
- Hypertext Markup Language (HTML) is the main language used to define how a Web page should look
- Features like background color, font, and layout are specified in HTML

Marking Up with HTML
- The words on a Web page are embellished by hidden formatting <tag>
- There is XHTML or the Extensible Hypertext Markup Language
- We use HTML
- XHTML tags are also HTML tags, but not vice versa
- There are some parts of the original HTML that are not part of XHTML

Formatting with Tags
- Tags are words or abbreviations enclosed in angle brackets, < and >
- Many tags come in pairs
- The second of the pair comes with a slash:
  <title>Fluency</title>
- In XHTML, the tags must be lowercase
  <TITLE>, <Title>, and <TITLE> are illegal
- HTML is not so restrictive
  The tag pair surrounds the text to be formatted like parentheses

Formatting with Tags
- <title>Serena Williams</title>
- These tags can be read as "this is where the title starts" and "this is where the title ends"
- <title> is referred to as the start or open tag
  </title> is the end or close tag
- The title appears on the title bar of the browser
Tags for Bold and Italic

- HTML has tags:
  - for bold text, `<b>` and `</b>`
  - for italic text, `<i>` and `</i>`
  - for paragraphs, `<p>` and `</p>`
- More than one kind of formatting can be used at a time:
  `<p><b><i>Veni, Vidi, Vici</i></b></p>` produces
  Veni, Vidi, Vici

- It doesn’t matter in which order you start the tags: italic follows bold, or bold follows italic
- You get the same result
- The rule is to make sure the tags “nest” correctly… they should mirror each other
- All the tags between a starting tag and its ending tag should be matched

Singleton Tags

- A few tags are not paired
- They do not have a matching ending tag
- For those tags, the closing angle bracket `>` of the singleton tag is replaced by `/>` in XHTML, not needed in HTML
- Examples in HTML:
  `<hr/>` produces a horizontal line
  `<br/>` continues the text to the next line
- These are `<hr/>` and `<br/>` in XHTML

More Formatting Tags

- Documents can be formatted in many ways
- Each formatting feature requires its own tag
- Programmers and Web designers need to remember a few common tags
- Uncommon tags can be looked up: for example, at: www.w3schools.com/tags/default.asp

Required Tags

- Every Web page is composed of a `head` and a `body`
- There are three HTML tags required for every Web page:
  - `<head>` and `</head>` enclose the head
  - `<body>` and `</body>` enclose the body
  - `<html>` and `</html>` to enclose those two parts
Required Tags

<!-- The text following the letters html:
   -- the dialect is XHTML
   -- the part inside of the quotes must be written
      exactly as given

<head>
<title>Starter</title>
<other stuff goes here... that will come later
</head>

<body>
the main content of the page goes here
</body>

Configure Your Computer for Writing HTML

• Check that two programs are installed:
  -- A browser (check for Firefox)
  -- A text editor (Notepad++ for Windows or Text Wrangler for Macs)

• Both programs are free
• These programs are preferred for technical reasons

Firefox

• Firefox is a free open source browser
• Open source means that the program
code is publicly available, and any
programmer can contribute improvements
to it
• Firefox is the browser referenced
throughout this book
• It is available at
Text Editor

- A text editor is a basic way to process text
- Our word processors are called "what-you-see-is-what-you-get" (WYSIWYG)
- Word processors include many application-specific information in their files
- This information confuses browsers

Text Editor

- You must use a text editor to write HTML
- Text editors do not include this extra information, browsers like their files!
- Browsers want Web pages written in ASCII characters only
- Think of ASCII as the normal keyboard characters with "nothing strange"

Text Editor

- Text editors figure out what language you are writing in and color code your HTML to make it easier to read
- Operating systems come with text editors installed
  - TextEdit can be found on the Mac
  - Notepad comes with Windows
- TextWrangler and Notepad++ are better choices

Hello, World!

- To produce your first HTML page, follow these instructions:
  - In your text editor, open a New document instance.
  - Carefully type in your text (see next slide)
    - Remove the preliminary material given here
    - Replace the main content of the page goes here with:
      ```html
      <p>Hello, World!</p>
      ```
  - Save the file as starterPage.html
  - Open the file with the Firefox browser

Open with Double-Click

- As HTML is written, files must be opened in two applications:
  - the text editor, to make changes
  - the browser, to see the changes made
- Double-click on the file to open it with the default application (your browser)
Save This Page

- All HTML files have the same structure as the starterPage.html file just created.
- Use it as a template for future HTML coding.
- Set up a new folder to keep your HTML files in.
- Using your new page as a template ensure that all pages will have the correct form.

Headings in HTML

- Documents tend to have headings, subheadings.
- HTML provides several levels of heading tags:
  - `<h1>` and `<h1>` level one
  - `<h2>` and `<h2>` level two
  - ...
  - `<h6>` and `<h6>` level six.
- Headings display content on a new line.

HTML Format Versus Display Format

- HTML source code tells the browser how to produce the formatted page based on the meanings of the tags.
- The source’s form is unimportant.
- HTML is written in a structured format to make it easier for people to understand.
- Indenting is frequently used to emphasize the tags’ meanings.

White Space

- Spaces that have been inserted for readability are called **white space**.
- White space is created with spaces, tabs, and new lines (return or enter).
- HTML ignores white space.
- The browser turns a **sequence** of white space characters into a single space.

White Space

- The only white space exception is **preformatted** information contained within `<pre>` and `</pre>`.
- This information is displayed as it appears.
- The width of a line of text is determined by the width of the browser window.
- A narrower or wider browser window makes the lines break in different places.

White Space

```html
<p>4th Lane's Paradise; 3rd St.
Athens and a turtle wade in a race.
Athens could
run as fast as the turtle. The turtle,
being a slower animal,
got a 10-meter head start, whereas
Athens started and ran the 10-meter
distance. At that
time the turtle was 5 meters farther.
When Athens hit the finish line,
the turtle had given another
2 meters.
and so both. Peripatos, the turtle
always remembered
about. <p>
```
The Escape Symbol

- What if the Web page had to show a math relationship: $0 < p \leq r$
- The browser might misinterpret <p> as a paragraph tag
- Using angle brackets as text is prohibited
- To show angle brackets, use an escape symbol (&), followed by an abbreviation, followed by a semicolon

Accent Marks in HTML

- Letters with accent marks also use the escape symbol
- General form is:
  - &amp;aeacute; displays as á
  - &amp;egrave; displays as ë

HTML

- Notice the following:
  - The title is shown on the title bar of the browser window
  - The statement of Pender's Paradox is in bold
  - The HTML source paragraphs are indented more than the HTML heading lines to make them more readable

HTML

- Notice the following:
  - The line between the header paragraphs crosses the width of the browser window
  - An acute accent is used in Marville's first name
  - The French phrase has the painting in its focal point
  - The word picture is in italics for emphasis
Compose and Check

• Most often Web pages are created all at once—both content and form
• It is smart to check your typing and your tagging often
  — Assume a page is okay
  — Add a few more tags, then the page is wrong
  — It must be the last tags added that have the error

• A productive way to work is to keep two windows open:
  — your text editor
  — your browser
• After writing a few HTML formatting tags, save the file
• Check the result in the browser by a Reload or Refresh of the source
• Repeat

Markup Validation Service

• Another way to limit the mistakes you make is to have it automatically validated
• This service checks to make sure your XHTML is correct
• If it is wrong, the service tells you where the mistakes are and what’s not proper

Add Extra Information

• To benefit from the automatic checking service, you need to add three more lines to the starterPage.html
• These lines are not required for the file to be a proper XHTML page, but they are needed by the checking service
• Before the <html>... > tag, add the lines:
  "<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
• Go to www.w3.org/DX/2003/04/xhtml1-light for exact format"
Add Extra Information

- The other line that we need to add is:
  ```html
  <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">
  ```

- This code specifies that the character encoding for the Web page will be UTF-8, or Unicode Translation Format for bytes.

- This Unicode representation will be explained in Chapter 7.

Check My Work

- With the three lines added, the validation service is ready to be used.

- Normally, validation doesn't happen until the HTML page is finished and stable.

- During "compose and check," the validation occurs at a "stopping place."

- To validate go to the W3C Markup Validation Service:
  validator.w3.org/#!validate_by_upload

Check My Work

- If a green banner comes back: the NHTML checks out.

- If a red banner comes back, it will have a list of errors with it and an explanation of what's wrong.

- It's common to have a lot of errors at the start.

Marking Links with Anchor Tags

- Two Sides of a Link, making **hyperlinks**.

- When a user clicks a hyperlink, the browser loads a new Web page.

- There are two parts to a hyperlink:
  - the highlighted text in the current document, which is called the **anchor text**.
  - the address of the other Web page, called the **hyperlink reference**.

Marking Links with Anchor Tags

- Both parts of the hyperlink are specified in the **anchor tag**:
  - Begin with `<a` and make sure there's a space after the `a` but not before it, `a` is for anchor.
  - Give the hyperlink reference using the `href` attribute `href="filename"`, making sure to include the double quotes.
  - Close the anchor tag with the `>` symbol.
  - Give the anchor text, which will be highlighted when it is displayed by the browser.
  - End the hyperlink with the `</a>` tag.
Absolute Pathnames (URLs)

- In these anchor tags, the hyperlink reference is an **entire URL**
  - The Web browser needs to know how/where to find the page
- A URL is made from:
  - a protocol specification, **http://**
  - a domain or IP address, **www.bioz.com**
  - a path to the file, **/bios/scirussell.html**