Chapter 4

A Hypertext Markup Language Primer

fluency5
WITH INFORMATION TECHNOLOGY   skills, concepts, & capabilities

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 <tags>
- We use the XHTML or the *Extensible Hypertext Markup Language*
- 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*
- 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, and
  - for *italic* text, <i> and </i>
  - for paragraphs, <p> and </p>
- More than one kind of formatting can be used at a time:
<p><i>Veni, Vidi, Vici!</i></p>
produces
Veni, Vidi, Vici!

Tags for Bold and Italic

- 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 */>*

- Examples:
  <hr /> produces a horizontal line

  **
**
  continues the text to the next

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

```html
<html>
  <head> <title> Formatting Examples </title>
  </head>

  <body>
   <p>
   This is <b>bold</b> text.                    <br/>
   This is <big>big</big> text.                 <br/>
   This is <em>emphasized</em> text.            <br/>
   This is <i>italic</i> text.                  <br/>
   This is <small>small</small> text.           <br/>
   This is <strong>strong</strong> text.        <br/>
   This is <sub>subscripted</sub> text.         <br/>
   This is <sup>superscripted</sup> text.       <br/>
   This is <tt>typewriter </tt> text.           <br/>
   This is <q>A short quote </q> text.          <br/>
   This is <blockquote>
        Blockquote text, for quotes longer than
        a line of text text text text text text text
        text text text text text text text text text
        text text text text text text text text text
        text text text text text
        </blockquote>                           <br/>
  </p>
 </body>

</html>
```

The result (inset) window shows:

Formatting Examples

This is **bold** text.
This is big text.
This is *emphasized* text.
This is *italic* text.
This is small text.
This is **strong** text.
This is subscripted text.
This is superscripted text.
This is typewriter text.
This is "A short quote " text.
This is

   Blockquote text, for quotes longer
   than a line of text text text text
   text text text text text text text
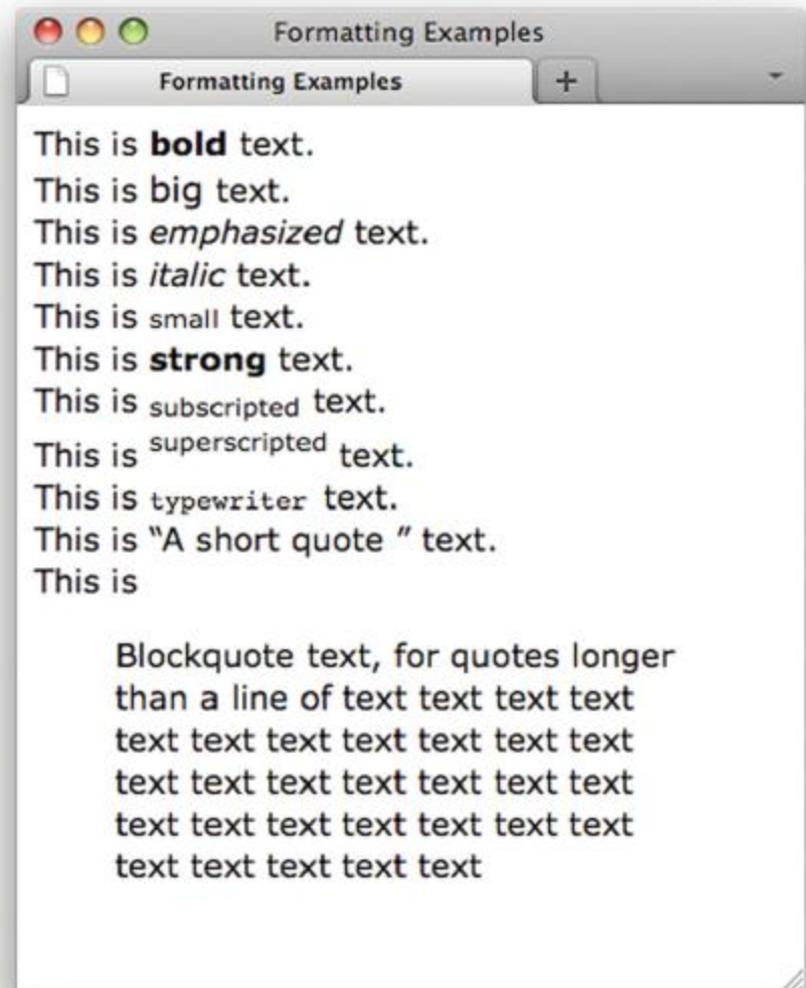   text text text text text text text
   text text text text text text text
   text text text text text

Figure 4.1 HTML source code for formatting tags and the result (inset).

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

```
1  <html xmlns="http://www.w3.org/1999/xhtml">
2    <head>
3      <title>Starter Page</title>
4       preliminary material goes here
5    </head>
6    <body>
7       the main content of the page goes here
8    </body>
9  </html>
```

(a)

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <title>Starter Page</title>
     preliminary material goes here
  </head>
  <body>
     the main content of the page goes here
  </body>
</html>
```

(b)

Figure 4.2 Standard form for an HTML Web page; every page must have these tags nested this way: (a) HTML code as seen in Notepad++ and (b) HTML code as seen in Text Wrangler.

Required Tags

<html xmlns=“http://www.w3.org/1999/xhtml”>

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

Required Tags

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

- The head section contains the beginning material like the title and other information that applies to the whole page

Required Tags

<body>

the main content of the page goes here

</body>

- The body section contains the content of the page.

- This <html> <head> <body> form must *always* be followed

- All of these tags are required

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 www.mozilla.com/en-US/firefox/all.html

Text Editor

- A text editor is a basic way to process text
- Our word processors are called "what-you-see-is-what-you-get" or (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

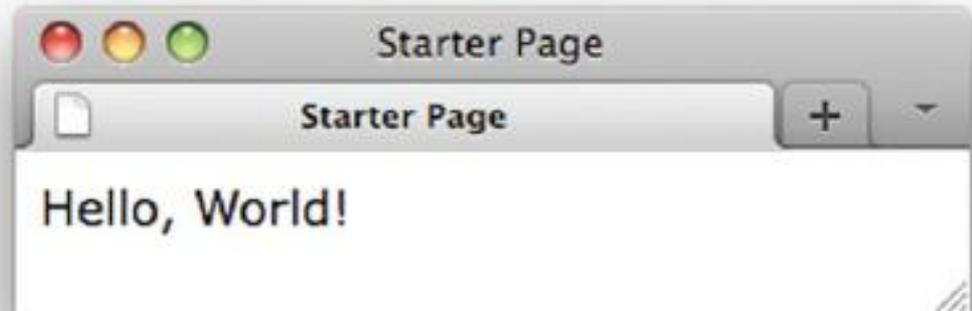- TextWrangler and Notepad++ are better choices

Hello, World!

- To produce your first HTML page, follow these instructions:

  1. In your text editor, open a *New document instance.*
  2. Carefully type in your text (see next slide)
     - Remove the *preliminary material goes here* nothing will replace it, *yet*
     - Replace *the main content of the page goes here* with: <p>Hello, World!</p>
  3. Save the file as starterPage.html
  4. Open the file with the Firefox browser

```html
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
   <title>Starter Page</title>
</head>
<body>
   <p>Hello, World!</p>
</body>
</html>
```


Starter Page

Starter Page

Hello, World!

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

headings are bold and get less "strong" (smaller and perhaps not so bold) as the level number increases.

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

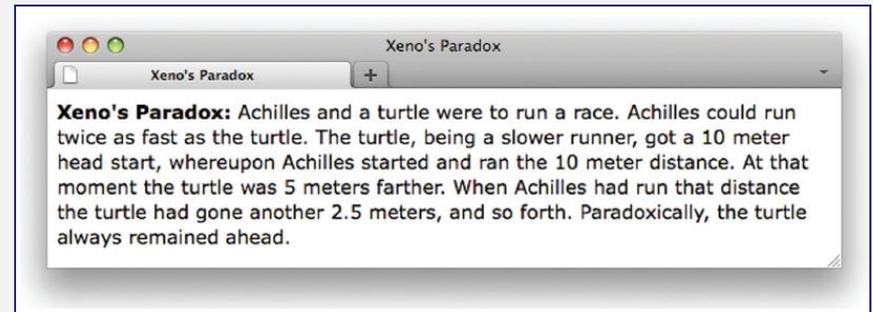- The only white space exception is *preformatted* information contained within <pre> and </pre>
- This information is displayed as it appears
- The width of a line of text is determined by the width of the browser window
  - A narrower or wider browser window makes the lines break in different places

White Space

<p> Xeno's Paradox:

Achilles and a turtle were to run a race. Achilles could

run twice as fast as the turtle. The turtle,

being a slower runner,

got a 10 meter head start, whereupon

Achilles started and ran the 10 meter distance. At that

moment the turtle was 5 meters farther. When Achilles had run

that distance the turtle had gone another 2.5 meters,

and so forth. Paradoxically, the turtle always remained

ahead. </p>

The Escape Symbol

- What if the Web page had to show a math relationship: **0<p>r**

- The browser might misinterpret **<p>** as a paragraph tag

- Using angle brackets as text is prohibited

< displays as <
> displays as >
& displays as &

- To show angle brackets, use an escape symbol (&), followed by an abbreviation, followed by a semicolon

Table 4.1 Basic tags for HTML

Start Tag	End Tag	Meaning	Required
<html xmlns="*fn*">	</html>	XHTML document; first and last tags of XHTML; *fn* = http://www.w3.org/1999/xhtml	✓
<title>	</title>	Title bar text; describes page	✓
<head>	</head>	Preliminary material; e.g., title at start of page	✓
<body>	< /body>	The main part of the page	✓
<p>	</p>	Paragraph	
		Line (horizontal rule)	
<h1>...<h6>	</h1>...</h6>	Headings, six levels	
		Bold	
<i>	< /i>	Italic	
		Anchor reference, *fn* must be a pathname to an HTML file	
		Image source reference, *fn* must be a pathname to .jpg or .gif file	
 		Break, continue text on a new line	

Accent Marks in HTML

- Letters with accent marks also use the escape symbol

- General form is:
  - ampersand, followed by the letter, followed by the name of the accent mark, followed by a semicolon

- The case of the letter is important!
  - é displays as é
  - È displays as È

Table 4.2 Specifying accent marks for Western European language

Symbol	Text	Number	Symbol	Text	Number	Symbol	Text	Number
à	à	à	ê	ê	ê	ô	ô	ô
ä	ä	ä	î	î	î	ù	ù	ù
é	é	é	ó	ó	ó	ā	ã	ã
í	í	í	ø	ø	ø	è	è	è
ò	ò	ò	ü	ü	ü	ì	ì	ì
ö	ö	ö	â	â	â	ñ	ñ	ñ
û	û	û	ç	ç	ç	õ	õ	õ
á	á	á	ë	ë	ë	ú	ú	ú
å	å	å	ï	ï	ï			

Note: For an accent mark on an uppercase letter, make the letter following the & uppercase or find the number for the uppercase letter at www.w3schools.com/tags/ref_entities.asp.

HTML

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
  <title> Twentieth Century Paradoxes </title>
  </head>
  <body>
   <h1>Paradoxes</h1>
   <h2>Russell's Paradox</h2>
    <p> The Twentieth Century logician Bertrand Russell
    introduced a curious paradox: <b>This statement is
    false.</b> The statement can't be true, because it
    claims the converse. However, if it is not true, then
    it's false, just as it says. That makes it true.
    Paradoxically, it seems to be neither true nor false,
    or perhaps both true and false.</p>
    <hr/>
   <h2> Magritte's Paradox</h2>
    <p> The famous Belgian artist Ren&eacute;
    Magritte rendered the idea of Russell's Paradox
    visually in his famous painting <i>Ceci n'est pas une
    pipe</i>. The title translates from French, This Is Not
    A Pipe. The painting shows a pipe with the text
    <i>Ceci n'est pas une pipe</i> below it. Superficially,
    the painting looks like a true statement, since it is a
    <i>picture</i> of the pipe, not an actual pipe. However,
    the assertion is also part of the picture, which seems
    to make it false, because it is clearly a painting of a
    pipe. Paradoxically, the truth seems to depend on
    whether the statement is an assertion about the painting
    or a part of it. But, it's both.</p>
  </body>
</html>
```

- Notice the following:
  - The title is shown on the title bar of the browser window
  - The statement of Russell's Paradox is in bold
  - The HTML source paragraphs are indented more than the <h2> heading lines to make them more readable

HTML

```
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
  <title> Twentieth Century Paradoxes </title>
  </head>
  <body>
  <h1>Paradoxes</h1>
  <h2>Russell's Paradox</h2>
    <p> The Twentieth Century logician Bertrand Russell
    introduced a curious paradox: <b>This statement is
    false.</b> The statement can't be true, because it
    claims the converse. However, if it is not true, then
    it's false, just as it says. That makes it true.
    Paradoxically, it seems to be neither true nor false,
    or perhaps both true and false.</p>
    <hr/>
  <h2> Magritte's Paradox</h2>
    <p> The famous Belgian artist Ren&eacute;
    Magritte rendered the idea of Russell's Paradox
    visually in his famous painting <i>Ceci n'est pas une
    pipe</i>. The title translates from French, This Is Not
    A Pipe. The painting shows a pipe with the text
    <i>Ceci n'est pas une pipe</i> below it. Superficially,
    the painting looks like a true statement, since it is a
    <i>picture</i> of the pipe, not an actual pipe. However,
    the assertion is also part of the picture, which seems
    to make it false, because it is clearly a painting of a
    pipe. Paradoxically, the truth seems to depend on
    whether the statement is an assertion about the painting
    or a part of it. But, it's both.</p>
  </body>
</html>
```

- Notice the following:
  - The line between the two paragraphs crosses the width of the browser window
  - An acute accent is used in Magritte's first name
  - The French phrase from the painting is in italics
  - The word *picture* is in italics for emphasis

```html
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
   <title> Twentieth Century Paradoxes </title>
  </head>
  <body>
   <h1>Paradoxes</h1>
   <h2>Russell's Paradox</h2>
    <p> The Twentieth Century logician Bertrand Russell
    introduced a curious paradox: <b>This statement is
    false.</b> The statement can't be true, because it
    claims the converse. However, if it is not true, then
    it's false, just as it says. That makes it true.
    Paradoxically, it seems to be neither true nor false,
    or perhaps both true and false.</p>
    <hr/>
   <h2> Magritte's Paradox</h2>
    <p> The famous Belgian artist Ren&eacute;
    Magritte rendered the idea of Russell's Paradox
    visually in his famous painting <i>Ceci n'est pas une
    pipe</i>. The title translates from French, This Is Not
    A Pipe. The painting shows a pipe with the text
    <i>Ceci n'est pas une pipe</i> below it. Superficially,
    the painting looks like a true statement, since it is a
    <i>picture</i> of the pipe, not an actual pipe. However,
    the assertion is also part of the picture, which seems
    to make it false, because it is clearly a painting of a
    pipe. Paradoxically, the truth seems to depend on
    whether the statement is an assertion about the painting
    or a part of it. But, it's both.</p>
  </body>
</html>
```
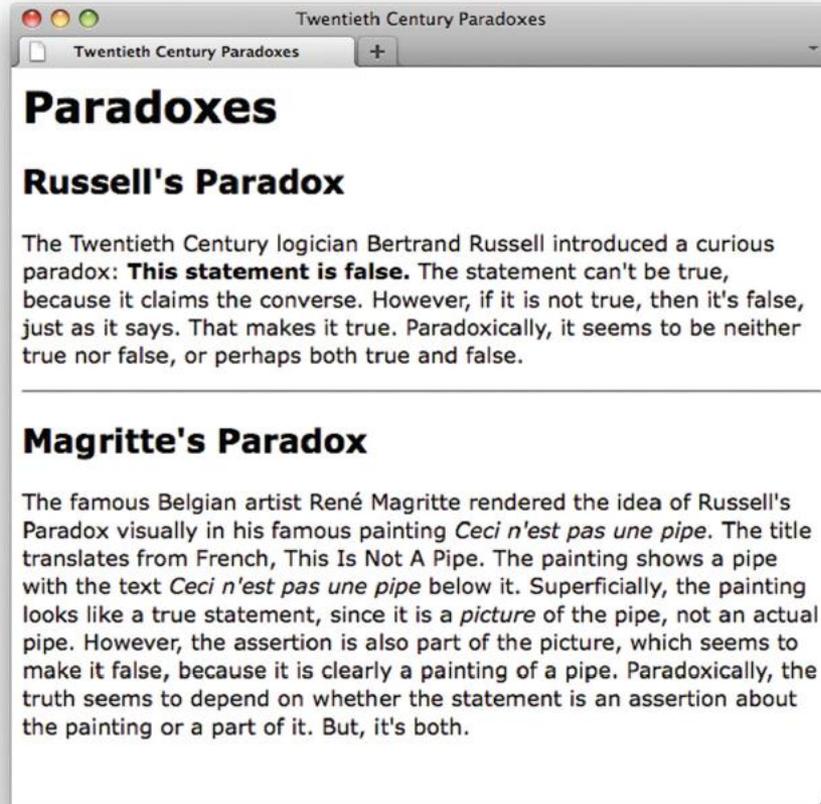

Twentieth Century Paradoxes

Twentieth Century Paradoxes +

Paradoxes

Russell's Paradox

The Twentieth Century logician Bertrand Russell introduced a curious paradox: **This statement is false.** The statement can't be true, because it claims the converse. However, if it is not true, then it's false, just as it says. That makes it true. Paradoxically, it seems to be neither true nor false, or perhaps both true and false.

Magritte's Paradox

The famous Belgian artist René Magritte rendered the idea of Russell's Paradox visually in his famous painting *Ceci n'est pas une pipe*. The title translates from French, This Is Not A Pipe. The painting shows a pipe with the text *Ceci n'est pas une pipe* below it. Superficially, the painting looks like a true statement, since it is a *picture* of the pipe, not an actual pipe. However, the assertion is also part of the picture, which seems to make it false, because it is clearly a painting of a pipe. Paradoxically, the truth seems to depend on whether the statement is an assertion about the painting or a part of it. But, it's both.

Figure 4.3 The HTML for a simple page and the result (inset).

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

Compose and Check

- 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

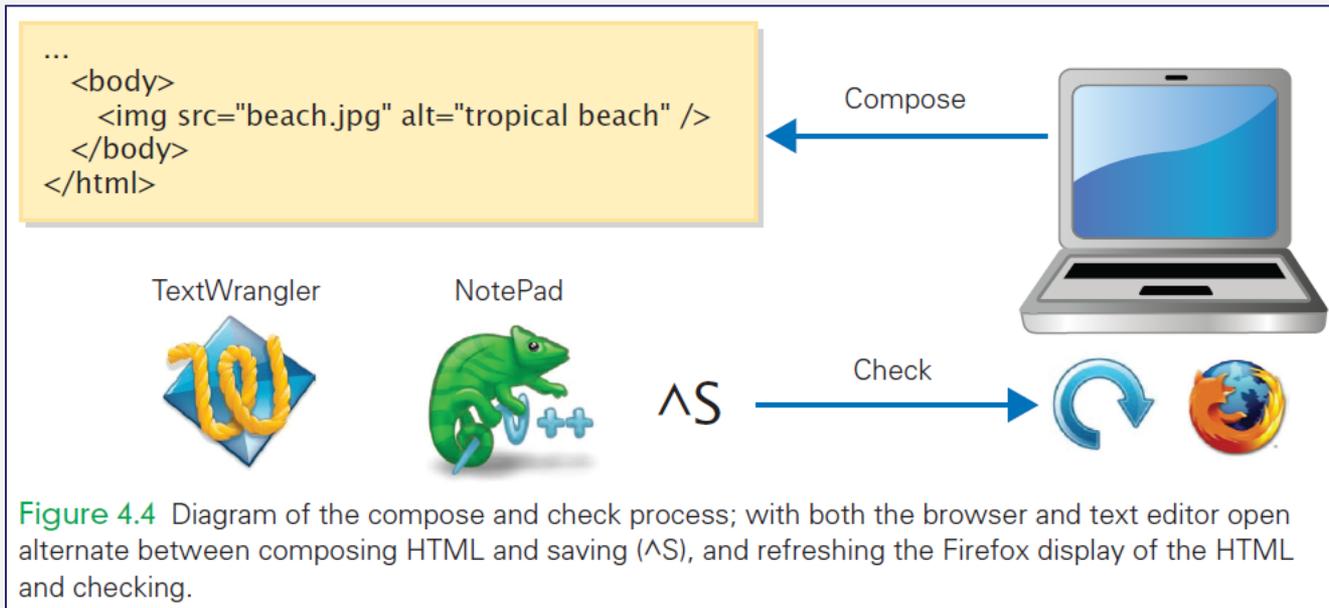- If it is wrong, the service tells you where the mistakes are and what's not proper

Figure 4.4 Diagram of the compose and check process; with both the browser and text editor open alternate between composing HTML and saving (^S), and refreshing the Firefox display of the HTML and checking.

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 Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">

Add Extra Information

- The other line that we need to add is
  \<meta http-equiv="Content-Type" content="text/html; charset=utf-8"/\>
  between \<head\> and \<title\>

- 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"

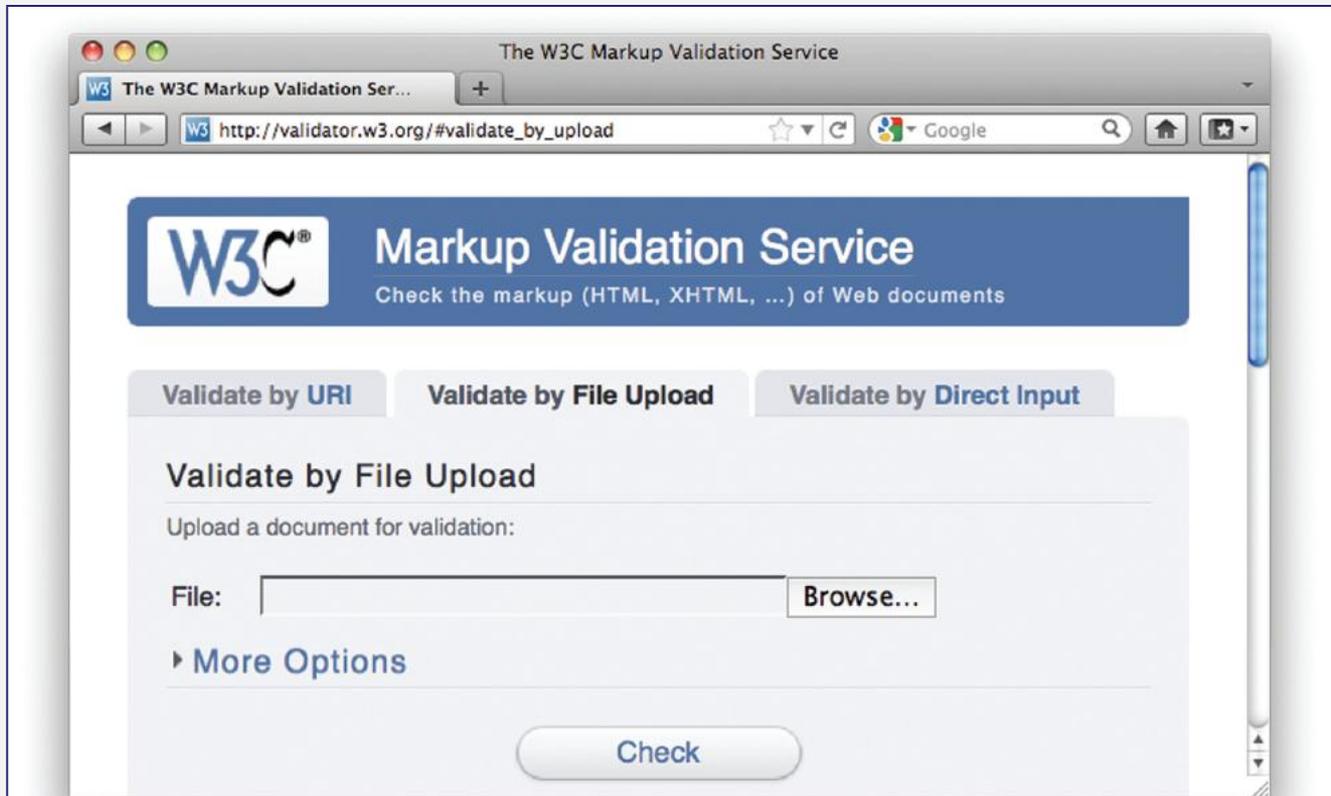- To validate go to the W3C Markup Validation Service at: **validator.w3.org/#validate_by_upload**

Figure 4.5 The W3C Markup Validation Service page; browse for your HTML file and then click *Check*.

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 **** tag.

Bertrand Russell

normal text   hyperlink reference   anchor

Bertrand **Russell**

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/sci/russell.html*