

The Structural Layer: XML and (X)HTML

Web Design in a Nutshell, Third Edition

Chapters 7, 8 and 9

Anatomy of Web Pages

- XML — eXtensible Markup Language
- HTML+XML = XHTML
- Document Structure
 - <html>
 - <head>information about the document</head>
 - <body>content that is displayed</body>
 - </html>

HTML: Then and Now

- HTML is an application of the Standard Generalized Markup Language
- Intended to represent simple document structure
- Uses hypertext to link related topics
- Designed for use over the Internet
- The Web has outgrown the capabilities of HTML

XML — eXtensible Markup Language

- A markup language is a structured language that lets you identify common sections of a document such as headings, paragraphs, and lists
- XML is a **w3c** standard for text document markup
- XML is not really a language, but rather a set of rules for creating other markup languages.
- Languages written according to XML syntax are called XML applications

XML — Four Components

- A document marked up in an XML language, such as an XHTML document
- An optional Document Type Definition (DTD) or schema that defines the elements and rules for use in that XML language
- Style sheets for presentation instructions
- Parsers that interpret the XML document, e.g. a web browser

XML — Application Examples

- XHTML — eXtensible HyperText Markup Language
- RSS — Really Simple Syndication
- SVG — Scalable Vector Graphics
- SMIL — Synchronized Multimedia Integration Language

XML + HTML = XHTML

- HTML was intended only to express logical document structure, not display characteristics
- Well-structured XHTML documents are the cornerstone of standards-compliant web design
- A semantically marked up document ensures accessibility in the widest range of browsing environments — from web browsers to cell phones to screen readers
- Structure documents logically
- Keep presentation separate from structure

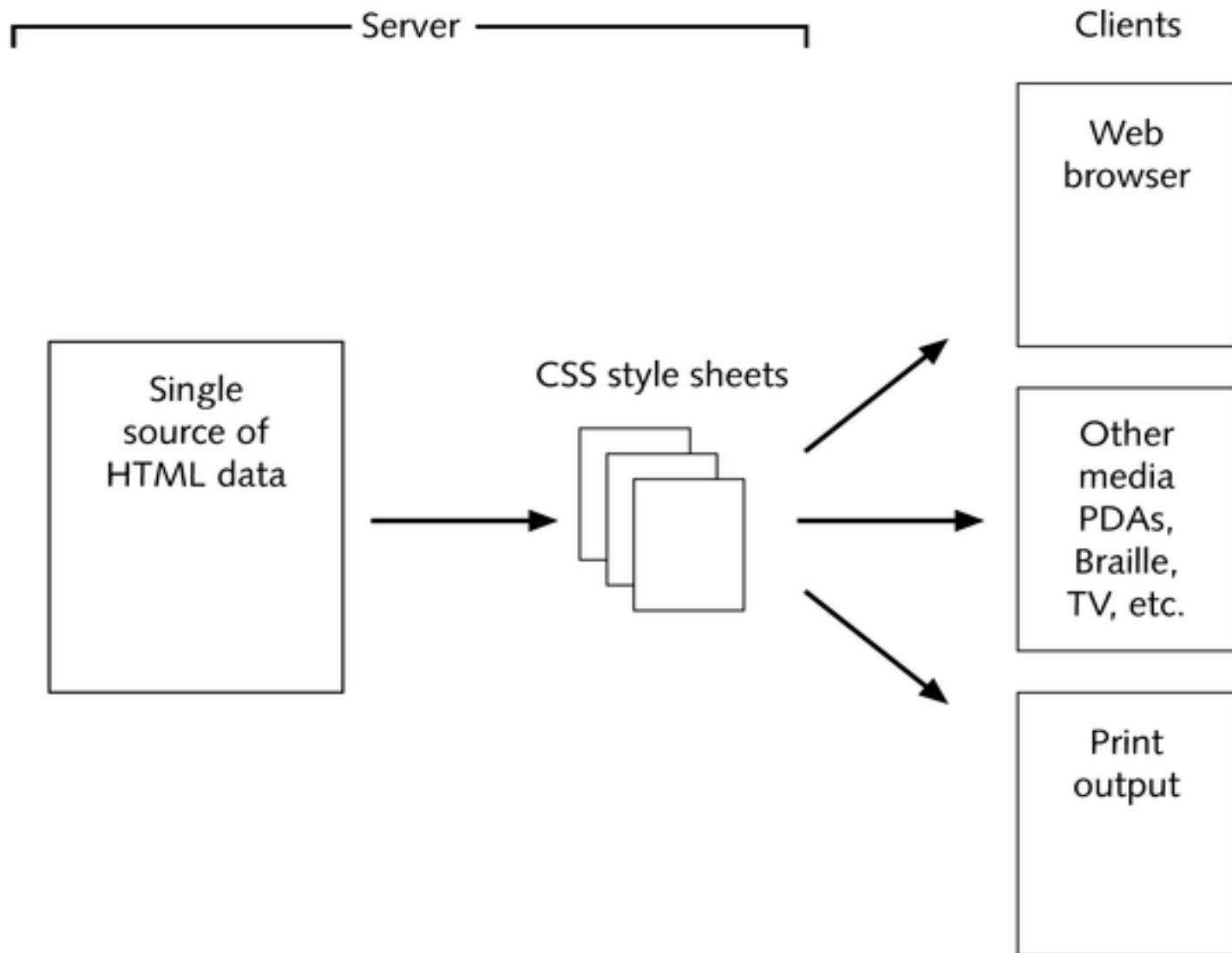


Figure 1-1 Formatting data for multiple destinations

XHTML Documents

- ASCII (plain text) or Unicode (UTF-8) documents marked up with tags indicating elements or other necessary declarations, e.g. paragraphs
- Can be versions 1.0 or 1.1 in strict, transitional and frameset flavors
- We will be using XHTML version 1.0 transitional, which means it's a step up from a regular HTML document, but conforms to the most basic XHTML standard

XHTML Elements and Attributes

- Elements are denoted in the document by special tags

`<p>`This is a paragraph.`</p>`

- Attributes clarify or modify an element's actions

`<p class="paragraph">`This is a paragraph.`</p>`

- In XHTML, all elements and attribute names must be lower case and have opening and closing tags
- Attribute values must be enclosed in quotes and explicitly specified
- Nested elements must be properly nested within one another

XHTML Tips and Tricks

- Choose elements that accurately and meaningfully describe the content and avoid choosing them based simply on how they look in the browser
- Write compliant, valid documents
- Follow proper conventions in naming documents
 - 🌐 Avoid spaces and special characters
 - 🌐 Use .html or .htm to denote xhtml documents
 - 🌐 use all lowercase names as servers can be case sensitive

XHTML Document Structure

Every valid XHTML document has a nested structure consisting of three basic sections:

- **html** — This is the root element for the document
- **head** — header elements, including the title, style, and script information, and meta information, such as the language of the document, authors, etc., contain information about the document
- **body** — The part of the document that is seen in a browser or other device

XHTML — HTML or root element

Every HTML page starts and ends with `<html></html>` tags

- Valid XHTML documents include DOCTYPE and xml sections specifying what kind of document it is and the language it was written in:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"  
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">  
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
```

```
</html>
```

XHTML — Head

Every valid HTML page has a `<head></head>` tagged section containing some of these items:

- **title** — the required name of the document and also what is displayed as the name of the current window
- **meta** — including the character set of the document, e.g. iso-8859-1, keywords for search engines, author information, etc.; not required, but recommended
- **style, script**, etc. non-visual elements to modify the page's display or behavioral characteristics

XHTML — Body

Every HTML page has a `<body></body>` tagged section containing what people actually see in their browser. These can include these elements:

- **Generic** — div (block) or span (inline) elements
- **Block** — h# (heading), p (paragraph), pre (preformatted text), etc.
- **Inline** — em (emphasized text), strong (strongly emphasized text), etc.
- **lists** — ordered or unordered

XHTML — Generic Elements

Div and span tags let authors create custom elements that don't fit into other categories, such as headers or paragraphs, or emphasized or strongly emphasized text:

- **Div** — used to identify generic block-level divisions; by default behaves similar to heading and paragraph elements, but is designed to be *tweaked* by stylesheets
- **Span** — used to identify inline elements; by default it has no presentation appearance

XHTML — Text Elements

These elements can contain inline elements, but not other block elements:

- **Headings** — Should be used in order with most important ideas formatted as h1 and least h6;
- **Paragraphs** — The most basic element of any text document; can contain inline elements, but not other block elements
- **Lists** — unordered (bullets), ordered (numbered) or definition (used for term and definition pairs)

What you should know

- XML means **eXtensible Markup Language**
 - RSS, which stands for **Really Simple Syndication**, is an XML application
 - XML is not really a language, but rather a set of rules for creating other markup languages.
 - XHTML, which stands for **eXtensible HyperText Markup Language** is another XML application
-

What you should know

- An element is specified by special tags in the document, e.g. **p**, **div**, etc.
 - An attribute clarifies or modifies an elements actions, e.g. **style="font-weight: bold;"**
 - Every valid XHTML documents consists of the root or **html** section, a **head** section and a **body** section
 - A **Doctype** and **language** must be specified
-

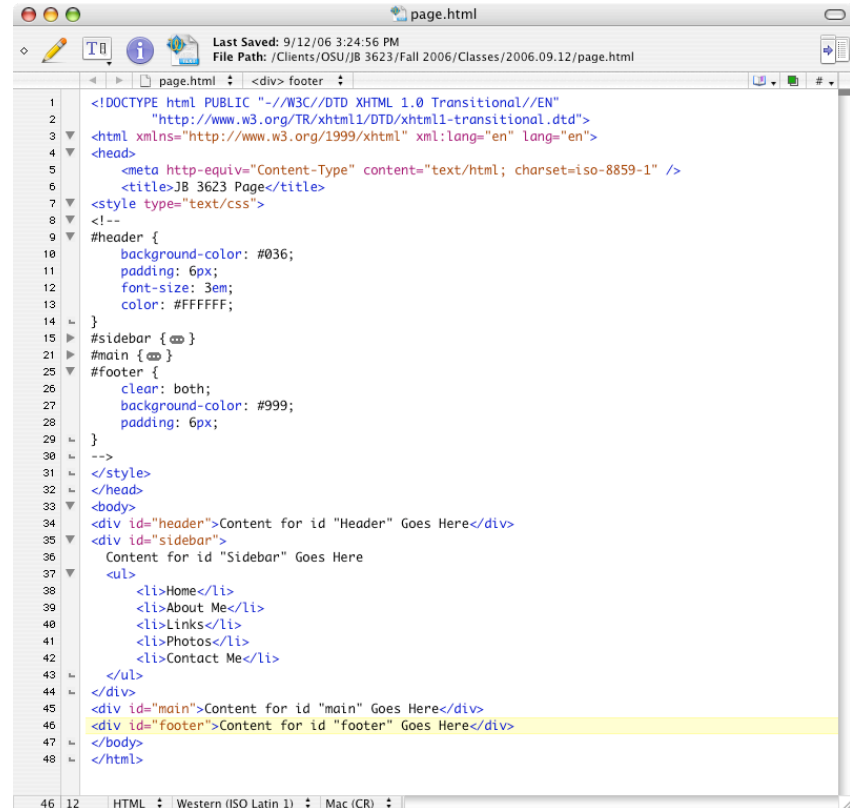
What you should know

- The **title** element is part of the **head** section and defines a browser window's name and also what is saved when a page is bookmarked
 - Every valid web page begins with an **html** tag and ends with an **html** tag, e.g. `<html></html>`
 - The parts between the **body** tags are what people see in their browsers
-

TextWrangler



We'll Use TextWrangler for editing our xhtml content



TextWrangler



- Located in Applications folder and it should be added to your dock
- Baby brother of BBEdit, by Barebones Software, <http://www.barebones.com/>
- A good generic text editor with color coding for a variety of languages, including xhtml
- It also includes a spell checker for your documents under the text menu
- You can load a page open in TextWrangler in your browser by dragging the icon in the TextWrangler window to your browser