

Web Technology & Programming

XHTML

LECTURER: KRUNAL PATEL
ICCT ENGG. COLLEGE
NEW VALLABH VIDYANAGAR

Introduction

What is XHTML?

- **XHTML** stands for Extensible Hypertext Markup Language
 - XHTML is aimed to replace HTML
 - XHTML is almost identical to HTML 4.01
 - XHTML is a stricter and cleaner version of HTML
- **XML** (Extensible Markup Language) is a markup language designed for describing *data*
 - XHTML is HTML redefined as an XML application
 - XHTML is a “bridge” between HTML and XML

The problem with HTML

- HTML started out as a way of describing the structure of documents, with tags to indicate headers, paragraphs, and the like...
- Because people wanted to control the appearance of documents, HTML acquired tags to control fonts, alignment, etc.
- The result is a markup language that does both, but isn't very good at either

HTML vs. XML

- XML looks a lot like HTML, but--
 - ✓ HTML uses a fixed set of tags
 - ✓ HTML is designed to display data to humans
 - ✓ Browsers are very tolerant of errors in HTML
 - ✓ All browsers can display HTML
 - ✓ With XML you make up your own tags (and define what they mean in a separate document)
 - ✓ XML is designed to describe data to computers
 - ✓ XML documents must be well-formed (syntactically correct)
 - ✓ Most modern browsers can display XML

From HTML to XHTML, I

- XHTML elements must be properly nested
`<i>bold and italic</i>` is *wrong*
- XHTML documents must be well-formed
`<html>
<head> ... </head>
<body> ... </body>
</html>`
- *Tag names* must be in *lowercase*
- *All XHTML elements must be closed*
 - If an HTML tag is not a container, close it like this:
`
, <hr />, <image src="smile.gif" />`
 - Note: Some browsers require a space before the `/`

From HTML to XHTML, II

- Attribute names must also be in lower case
 - Example: `<table width="100%">`
- Attribute values must be quoted
 - Example: `<table width="100%">`
- Attribute minimization is forbidden
 - Example: `<frame noresize="noresize">`,
cannot be abbreviated to `<frame noresize>`
- The **id** attribute replaces the **name** attribute
 - Wrong: ``
 - Right: ``
 - Best: ``

SGML and DTDs

- SGML stands for “Standard Generalized Markup Language”
- HTML, XHTML, XML and many other markup languages are defined in SGML
- A DTD, or “Document Type Definition” describes the syntax to use for the current document
- There are three different DTDs for XHTML--you can pick the one you want
- These DTDs are public and on the web.
- You must start your XHTML document with a reference to one of these DTDs

DOCTYPE declaration, I

- Every XHTML document must begin with one of the **DOCTYPE** declarations (DTDs):
 - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`
 - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">`
 - `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">`

DOCTYPE declaration, II

- The three main DTDs are as follows:
 - **Strict**
 - Use for really clean markup, with no display information (no font, color, or size information)
 - Use with CSS (Cascading Style Sheets) if you want to define how the document should look
 - **Transitional**
 - Use with standard HTML and/or with CSS
 - Allows deprecated HTML elements
 - **Frameset**
 - Use if your document uses HTML frames

Transitional documents

- The transitional document type definition specifies an XHTML document that uses some deprecated elements or attributes such as the **** tag, and can also be used for documents that will be displayed in browsers that do not support CSS.

Transitional documents

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

`<html>`

`<head>`

`<title>Transitional document type</title>`

`</head>`

`<body>`

This is a transitional HTML document

`</body>`

`</html>`

Frameset documents

- The frameset document type definition specifies an XHTML document that uses frames to display one or more pages at the same time in the browser window and can also include deprecated elements or attributes (like the transitional document type definition).
- But unlike the transitional document type definition, the frameset document type definition uses the **<frameset>** tag instead of the **<body>** tag.

Frameset documents

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

`<html>`

`<head> <title>Frameset document type</title> </head>`

`<frameset cols="50%, 50%">`

`<frame src="frame1.html" /> <frame
src="frame2.html" />`

`</frameset>`

`</html>`

Strict documents

- The Strict document type definition specifies an XHTML document that includes no deprecated elements or attributes such as the **** tag, and no frames.
- `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`
`<html>`
`<head> <title>Strict document type</title>`
`</head>`
`<body> <h1>Content</h1>`
`</body>`
`</html>`

Creating an XHTML Document

- The first line of an XHTML document should contain a declaration indicating that the document adheres to the rules and syntax of XML
- XML (and thus XHTML) documents are based on a character set
- A character set is a set of abstract symbols matched to code numbers

Character Sets

- Universal Character Set (UCS)
- Unicode
- *Character encoding* is the process in which bytes are translated back into characters (when a document is sent across the Internet)

Adding an xml Declaration

- To declare that a document is written in XML, enter the following as the first line of the file:

*<?xml version="value" encoding="type"
standalone="type"?>*

- Where the version attribute indicates the XML version of the document, the encoding attribute specifies the character encoding, and the standalone attribute indicates whether the document contains references to an external DTD.

Adding an xml Declaration

- For XHTML documents, use the declaration:

*<?xml version="1.0" encoding="UTF-8" standalone="no"
?>*

An XHTML Example

- `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">`
`<html>`
 `<head>`
 `<title>A simple document</title>`
 `</head>`
 `<body>`
 `<p>A simple paragraph.</p>`
 `</body>`
`</html>`

Example

```
<? xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
    "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml"
    xml:lang="en" lang="en">
  <head>
    <title>My First Web Page</title>
  </head>
  <body style="background-color: cyan">
    <p>Hello everyone!</p>
    <p>My Name is (put your name here) and today is (put in the date).
    HTML is wack! XHTML is most certainly not wack! HA HA HA HA
    HA HA HA HA HA HA HA HA HA HA HA HA HA HA!</p>
  </body>
</html>
```

XHTML Basic – Modules and Elements

Module	Elements included
Structure	<body>, <head>, <html>, <title>
Text	<abbr>, <acronym>, <address>, <blockquote>, , <cite>, <code>, <dfn>, <div>, , <h1>, <h2>, <h3>, <h4>, <h5>, <h6>, <kbd>, <p>, <pre>, <q>, <samp>, , <string>, <var>
Hypertext	<a>
List	<dl>, <dt>, <dd>, , ,
Basic Forms	<form>, <input>, <label>, <select>, <option>, <textarea>
Basic Tables	<caption>, <table>, <td>, <tr>, <th>
Image	
Object	<object>, <param>
Meta Information	<meta>
Link	<link>
Base	<base>

Rules for well-formed XHTML documents

Rule	Incorrect	Correct
Element names must be lowercase	<code><P>This is a paragraph.</P></code>	<code><p>This is a paragraph.</p></code>
Elements must be properly nested	<code><p>This text is bold.</p></code>	<code><p>This text is bold.</p></code>
All elements must be closed	<code><p>This is the first paragraph. <p>This is the second paragraph.</code>	<code><p>This is the first paragraph. </p><p>This is the second paragraph.</p></code>
Empty elements must be terminated	<code>This is a line break
</code>	<code>This is a line break
</code>
Attribute names must be lowercase	<code><td ALIGN="right"></code>	<code><td align="right"></code>
Attribute values must be quoted	<code><table width=620></code>	<code><table width="620"></code>
Attributes must have values	<code><option selected></code>	<code><option selected="selected"></code>

Tips for Converting old HTML Code to XHTML

- Include an xml declaration in the first line of your file so that your document can be accessed by XML parsers
- Add a DOCTYPE declaration for one of the XHTML DTDs and check your document for well-formedness and validity whenever you make a change to the code
- Add the XHTML default namespace to the html element of your document
- Make sure that all element and attribute names are in lowercase letters and that all attribute values are placed in quotes
- Make sure that all empty elements are entered as one-sided tags. Look especially for improper syntax in the img, hr, and br elements
- Make sure that all two-sided tags are properly closed. Old HTML code often does not have closing tags for the p element

Tips for Converting old HTML Code to XHTML

- Make sure that all inline images contain the alt attribute
- Look for deprecated attributes such as align, bgcolor, and background, and replace them with the float (or text-align), background-color, and background-image styles
- Replace the name attribute with the id attribute
- Fix all instances of attribute minimization
- Replace the use of the font element with either the span element or with a style that applies the same formatting specified by the font element
- Replace the use of the width attribute in the td or th element with the width style