

HTML and XHTML  
2015/09/14

## Tutorial 9

### Working with XHTML

New Perspectives on HTML and XHTML, Comprehensive

## Objectives

- Describe the history and theory of XHTML
- Understand the rules for creating valid XHTML documents
- Apply a DTD to an XHTML document
- Understand how to apply the XHTML namespace
- Test an XHTML document under the transitional DTD
- Test an XHTML document under the strict DTD
- Explore the use of character and parsed character data

New Perspectives on HTML and XHTML, Comprehensive 2

## Introducing XHTML

- SGML (Standard Generalized Markup Language)
  - Device-independent and system-independent
  - Introduced in the 1980s
  - Not intended for the World Wide Web
- HTML
  - Standards get confusing among browsers
  - Can be applied inconsistently

New Perspectives on HTML and XHTML, Comprehensive 3

## Introducing XHTML

Version	Date Released	Description
XHTML 1.0	2001	This version is a reformulation of HTML 4.01 in XML and combines the strength of HTML 4.0 with the power of XML. XHTML brings the rigor of XML to Web pages and provides standards for more robust Web content on a wide range of browser platforms.
XHTML 1.1	2002	A minor update to XHTML 1.0 that allows for modularity and simplifies writing extensions to the language.
XHTML 2.0	not yet released	The latest version, designed to remove most of the presentational features left in HTML. XHTML 2.0 is not backward-compatible with XHTML 1.1.
XHTML 5.0	not yet released	A version of HTML 5.0 written under the specifications of XML; unlike XHTML 2.0, XHTML 5.0 will be backward-compatible with XHTML 1.1.

New Perspectives on HTML and XHTML, Comprehensive 4

## 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

New Perspectives on HTML and XHTML, Comprehensive 5

## 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)

New Perspectives on HTML and XHTML, Comprehensive 6

## 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

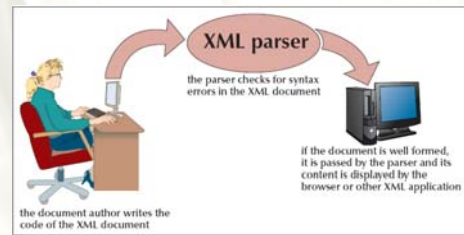
- For XHTML documents, use the declaration:
 

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

## Introducing XHTML

- XML (Extensible Markup Language)
  - Used to create markup languages
- XML documents must be evaluated with an **XML parser**
- An XML document with correct syntax is a **well-formed** document
- A well-formed document with correct content and structure is a **valid** document
- DTD** specifies correct content and structure

## Introducing XHTML



## Creating a Well-Formed Document

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

## Creating a Well-Formed Document

- XHTML documents must also include a single root element that contains all other elements
  - For XHTML, that root element is the `html` element
- Attribute minimization** is when some attributes lack attribute values
  - XHTML doesn't allow attribute minimization

## Attribute minimization in HTML and XHTML

HTML	XHTML
compact	compact="compact"
checked	checked="checked"
declare	declare="declare"
readonly	readonly="readonly"
disabled	disabled="disabled"
selected	selected="selected"
defer	defer="defer"
ismap	ismap="ismap"
nohref	nohref="nohref"
noshade	noshade="noshade"
nowrap	nowrap="nowrap"
multiple	multiple="multiple"
noresize	noresize="noresize"

New Perspectives on HTML and XHTML, Comprehensive

13

## DTDs

- **Transitional**: supports many of the presentational features of HTML, including the deprecated elements and attributes. Best used for older documents that contain deprecated features.
- **Frameset**: used for documents containing frames, and also supports deprecated elements and attributes
- **Strict**: does not allow any presentational features or deprecated HTML elements and attributes. Does not support frames or inline frames. It is best used for documents that need to strictly conform to the latest standards

New Perspectives on HTML and XHTML, Comprehensive

14

## Creating Valid XHTML Documents

- The DTD used depends on the content of the document and the needs of your users
- To support old browsers, use the **transitional** DTD
- To support old browsers in a framed Web site, use the **frameset** DTD
- To support more current browsers and to weed out any use of deprecated features, use the **strict** DTD

New Perspectives on HTML and XHTML, Comprehensive

15

## Creating a Valid Document

- Elements **not** allowed under the strict DTD:
  - applet
  - basefont
  - center
  - dir
  - font
  - frame
  - frameset
  - iframe
  - isindex
  - menu
  - noframes
  - s
  - strike
  - u

New Perspectives on HTML and XHTML, Comprehensive

16

## Creating a Valid Document

- Some attributes are restricted, while others are required in XHTML

New Perspectives on HTML and XHTML, Comprehensive

17

## Attributes Prohibited in the Strict DTD

Element	Prohibited Attributes
a	target
area	target
base	target
body	alink, bgcolor, link, text, vlink
br	clear
caption	align
div	align
dl	compact
form	name, target
hr	align
hr	align, noshade, size, width
img	align, border, bgcolor, name, vspace
input	align
li	type, value
link	target
map	name
object	align, border, bgcolor, vspace
ol	compact, start
pre	align
pre	width
script	language
table	align, bgcolor
td	bgcolor, height, nowrap, width
tr	bgcolor, height, nowrap, width
tr	bgcolor
ul	type, compact

New Perspectives on HTML and XHTML, Comprehensive

18

## Required XHTML Attributes

Element	Required Attributes
applet	height, width
area	alt
base	href
basefont	size
bdo	dir
form	action
img	alt, src
map	id
meta	content
optgroup	label
param	name
script	type
style	type
textarea	cols, rows

New Perspectives on HTML and XHTML, Comprehensive

19

## Inserting the DOCTYPE Declaration

- You can also add the DOCTYPE declaration, which tells XML parsers what DTD is associated with the document

```
<!DOCTYPE root type "id" "url">
```

New Perspectives on HTML and XHTML, Comprehensive

20

## The XHTML Namespace

- A namespace is a unique identifier for elements and attributes originating from a particular document type (like XHTML or MathML)
- Two types of namespaces:
  - Default: applied to a root element and any element within it
  - `<root xmlns="namespace">`

New Perspectives on HTML and XHTML, Comprehensive

21

## The XHTML Namespace

- Local: applies to only select elements
  - Each element in the local namespace is marked by a prefix attached to the element name
  - `xmlns:prefix="namespace"`
  - Identify any element belonging to that namespace by modifying the element name in the tag:
    - `prefix:element`

New Perspectives on HTML and XHTML, Comprehensive

22

## Setting the XHTML Namespace

- To set XHTML as the default namespace for a document, add the xmlns attribute to the html element with the following value:
  - `<html xmlns=http://www.w3.org/1999/xhtml>`

New Perspectives on HTML and XHTML, Comprehensive

23

## Testing an XHTML Document

- To test your document, you need to send the file to an XML parser
  - Several are available on the Web
- Sometimes the same mistake results in several errors noted in the report
  - Fixing one mistake can solve several errors

New Perspectives on HTML and XHTML, Comprehensive

24

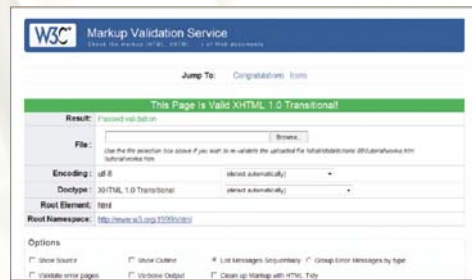
# Class Notes: Topic 9 – Working with XHTML

## Testing an XHTML Document

```
<div id="main">

<p class="firstp">
Welcome to wizard works, your one-stop shopping source for discount
fireworks on the web! We maintain a select variety of consumer fireworks
on our web site year round. You'll find the fireworks you want with the
speed and convenience of the Internet, and you'll always have the best
show in the neighborhood when you shop at wizard works.</p>
<p>
We supply quality fireworks to buy online at wholesale prices, our
professional fireworks display team is available to design and organize
your displays, or if you want to organize your own fireworks party, our
experienced fireworks designers have selected the best fireworks for the
occasion, we've made it easy for you to buy on-line for delivery direct
to your home.</p>
<p>
We also create customized fireworks and displays for special needs. Please
contact one of our design engineers to see if we can create a display for
your special event. Be sure to order special event fireworks early to ensure
on-time delivery.</p>
```

## Successful Validation Under XHTML Transitional



## Testing an XHTML Document

- To test under another DTD, you'll need to change the DOCTYPE declaration

```
<?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">
<head>
```

## Using Style Sheets and XHTML

- **Parsed character data (PCDATA)** is text parsed by a browser or parser
- **Unparsed character data (CDATA)** is text not processed by the browser or parser
  - A CDATA section marks a block of text as CDATA so that parsers ignore any text within it