



**INTERNET & WEB**  
**APPLICATION DEVELOPMENT**  
**SWE 444**

Fall Semester 2008-2009 (081)

**Module 2 (III): XHTML**

**Dr. El-Sayed El-Alfy**

Computer Science Department  
King Fahd University of Petroleum and Minerals  
alfy@kfupm.edu.sa

## Objectives/Outline

### • Objectives

- Learn what and why is XHTML
- Learn the differences between HTML and XHTML
- Learn how to create XHTML documents

### • Outline

- Introduction
- Why XHTML?
- Differences from HTML
- Anatomy of XHTML document
- XHTML DTDs
- XHTML Validation

## Introduction

- XHTML stands for eXtensible HyperText Markup Language
- XHTML is aimed to replace HTML
  - XHTML is almost identical to HTML 4.01 but a stricter and cleaner version
  - XHTML is a reformulation of HTML into a language that conforms to the XML 1.0 Recommendation
- XHTML is a bridge between HTML and XML
  - XHTML Family document types are all XML-based, and ultimately are designed to work in conjunction with XML-based user agents

## Introduction (cont.)

- XHTML 1.0
  - separation of document structure from presentation
  - issues concerning accessibility and internationalization
  - the three DTD offerings (strict, transitional, and frameset)
- XHTML 1.1 (modular XHTML)
  - Small devices (like mobile devices) cannot support all XHTML 1.0 functions.
  - XHTML 1.1 divides the specification into modules with limited functionality.
  - Small browsers can reduce their complexity by supporting only selected modules (but once a module has been chosen, all of its features must be supported).
  - XHTML 1.1 is a strict language and is not backward compatible with HTML 4.
- XHTML 2.0
  - A next generation markup language.
  - The functionality is expected to remain similar to XHTML 1.1, but not intended to be backward compatible with HTML 4, XHTML 1.0 and XHTML 1.1

## Why XHTML?

- Many pages on the WWW contain "bad" HTML
  - e.g. not well-formed documents
- Need to separate document structure and document formatting (styling)
  - XHTML phases out (deprecate) formatting information from HTML and makes place for CSS
- Conformance with XML syntax
  - XHTML pages can be read by all XML enabled devices
- XHTML is a combination of HTML and XML
  - Like XML, everything has to be marked up correctly, which results in "well-formed" documents
  - XHTML consists of all the **elements** in HTML 4.01 combined with the **syntax** of XML

## Differences from HTML

- The most important differences are:
  - All tag and attribute names must be in lowercase
  - All elements must be closed
  - All elements must be properly nested
  - All documents must be well-formed
  - All values of attributes must be quoted, e.g. width="50"
  - Attribute *minimization* is forbidden
  - The *id* attribute replaces the *name* attribute
  - All documents must have DTD declaration
  - The XHTML DTD defines mandatory elements

## Lowercase

- because XHTML documents are XML applications
- XML is case-sensitive
- Tags like `<br>` and `<BR>` are interpreted as different tags

## Elements Must Be closed

- Non-empty elements must have an end tag, e.g.
  - `<li>A list item </li>`
- Empty elements must also be closed by having an end tag or by ending the start tag with `/>`, e.g.
  - ``
  - Line break `<br />`
  - Horizontal rule `<hr />`

\* Note that some browsers require a space before the `/`

## Properly Nested

- In HTML some elements can be improperly nested within each other, e.g.
  - `<b><i>`This text is bold and italic`</b></i>`
- In XHTML all elements must be properly nested within each other, e.g.
  - `<b><i>`This text is bold and italic`</i></b>`
- A common mistake in nested lists, is to forget that the inside list must be within an li element

## Well-Formed Documents

- All XHTML elements must be nested within the `<html>` root element
- All other elements can have sub-elements (children). Sub-elements must be in pairs and correctly nested within their parent element.
- The basic document structure is

```
<html >  
<head> . . . </head>  
<body> . . . </body>  
</html >
```

## Attribute Minimization is Forbidden

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"

### Examples

- `<frame noresize>` not allowed
- `<frame noresize="noresize">` correct
- `<input type="radio" name="stuff" CHECKED>` not allowed
- `<input type="radio" name="stuff" checked="checked" />` correct

## The id replaces the name attribute

- HTML 4.01 defines a *name* attribute for the elements *a*, *applet*, *frame*, *iframe*, *img*, and *map*
- In XHTML the *name* attribute is deprecated and replaced with the *id* attribute
  - ``
- Both *name* and *id* attributes are designed to be used as fragment identifiers
  - there can only be a single attribute of type *id* per element.

## Mandatory XHTML Elements

- An XHTML document may have an optional (but recommended) xml declaration, which has three attributes:
  - `<?xml version="1.0" encoding="UTF-8" standalone="yes"?>`
  - The version attribute is required
  - The encoding attribute specifies the character encoding the document uses. The Unicode Transformation Format (UTF) is the default in XML
  - The standalone attribute says whether a document uses an external DTD (a grammar for a class of documents)
- An XHTML document must have three main parts:
  - A DOCTYPE declaration
  - A head
  - A body

## XHTML Document Template

```
<?xml version="1.0" encoding="UTF-8"?>
<!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">
<head>
<title>... </title>
</head>
<body> ... </body>
</html>
```

## XHTML DTD

- The DOCTYPE declaration is used to indicate the DTD that is used by an XHTML document
  - Gives a reference to one of the publicly available DTDs on the web
- An XHTML DTD describes in precise the allowed syntax and grammar of XHTML markup.
- There are currently 3 XHTML 1.0 document types:
  - STRICT
  - TRANSITIONAL
  - FRAMESET
- These document types are distinguished in part by the degree to which they accept or do not accept deprecated HTML elements

## XHTML DTD (cont.)

- XHTML 1.0 Strict
  - Use this when you want really clean markup, free of presentational clutter. Use this together with Cascading Style Sheets.
- XHTML 1.0 Transitional
  - Use this when you need to take advantage of HTML's presentational features and when you want to support browsers that don't understand Cascading Style Sheets.
- XHTML 1.0 Frameset
  - Use this when you want to use HTML Frames to partition the browser window into two or more frames.

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



## XHTML Validation

- An XHTML document is validated against a Document Type Definition (DTD)
- W3C Validation Tool
  - <http://validator.w3.org/> an HTML form for checking (but not fixing) HTML and XHTML documents
- Dave Raggett's HTML TIDY
  - <http://www.w3.org/People/Raggett/tidy/> a free UNIX tool for checking and cleaning up HTML pages
- XHTML Tag List
  - [http://www.w3schools.com/xhtml/xhtml\\_reference.asp](http://www.w3schools.com/xhtml/xhtml_reference.asp)
- XHTML Attributes
  - [http://www.w3schools.com/xhtml/xhtml\\_standardattributes.asp](http://www.w3schools.com/xhtml/xhtml_standardattributes.asp)

## Q & A



## References

- Some useful links with examples and other resources:
  - W3C at <http://www.w3.org/xhtml/>
  - W3Schools at <http://www.w3schools.com/>
    - <http://www.w3schools.com/xhtml/>
  - [http://en.wikipedia.org/wiki/List\\_of\\_document\\_markup\\_languages](http://en.wikipedia.org/wiki/List_of_document_markup_languages)
  - *Internet and World Wide Web How to Program*, 4/e, H. M. Deitel, P. J. Deitel, and A. B. Goldberg, Pearson Education Inc., 2008. Chapter 4.
  - *Web Development and Design Foundations with XHTML*, 4/e, Pearson Education Inc. 2009. Chapter 2.