Unit 2

1.2 Introduction to XHTML

1.2 XHTML

XHTML is a stricter and cleaner version of HTML. In this section you will learn the difference between HTML and XHTML.

1.2.1 What Is XHTML?

- XHTML stands for E**X**tensible **H**yper**T**ext **M**arkup **L**anguage
- XHTML is almost identical to HTML 4.01
- XHTML is a stricter and cleaner version of HTML
- XHTML is HTML defined as an XML application
- XHTML is a W3C Recommendation

1.2.2 All Browsers Support XHTML

XHTML is compatible with HTML 4.01. All browsers support XHTML.

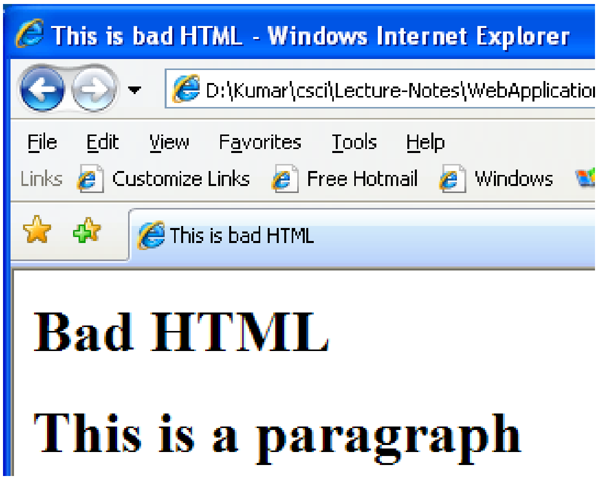
1.2.3 Why you should use XHTML

XHTML is a combination of HTML and XML (E**X**tensible **M**arkup **L**anguage) and consists of all the elements in HTML 4.01, combined with the strict syntax of XML.

Many pages on the internet contain "bad" HTML. The following HTML code (in Example 1.10 will work just fine if you view it in a browser (even if it does NOT follow the HTML rules):

```
<html>
<head>
<title>This is bad HTML</title>
<body>
<h1>Bad HTML
<p>This is a paragraph
</body>
```

Example 1.2-1


This is bad HTML - Windows Internet Explorer
D:\Kumar\csci\Lecture-Notes\WebApplicatio
File
Edit
View
Favorites
Tools
Help
Links
Customize Links
Free Hotmail
Windows
This is bad HTML

Bad HTML

This is a paragraph

Figure1.2-1

XML on the other hand is a markup language where everything must be marked up correctly, which results in "well-formed" documents.
XML is designed to describe data, and HTML is designed to display data.

Today's market consists of different browser technologies, some browsers run on computers, and some browsers run on mobile phones or other small devices. The last-mentioned do not have the resources or power to interpret a "bad" markup language. Therefore - by combining the strengths of HTML and XML, W3C recommended a markup language that is useful now and in the future - XHTML.

1.2.4 Important Differences

The important differences compared to HTML are:

- XHTML elements must be properly nested
- XHTML elements must always be closed
- XHTML elements must be in lowercase
- XHTML documents must have one root element

1.2.4.1 XHTML Elements Must be Properly Nested.

In HTML, some elements can be improperly nested within each other, like this:

<b><i>This text is bold and italic**</b></i>**

In XHTML, all elements must be properly nested within each other, like this:

<b><i>This text is bold and italic**</i></b>**

Note: A common mistake with nested lists, is to forget that the inside list must be within **<li>** and **</li>** tags.

This is Wrong

```
<ul>
  <li>Coffee</li>
  <li>Tea
    <ul>
      <li>Black tea</li>
      <li>Green tea</li>
    </ul>
  <li>Milk</li>
</ul>
```

Example 1.2-2a

This is Correct

```
<ul>
  <li>Coffee</li>
  <li>Tea
    <ul>
      <li>Black tea</li>
      <li>Green tea</li>
    </ul>
  </li>
  <li>Milk</li>
</ul>
```

Example 1.2-2b

Notice that a **</li>** tag is inserted after the **</ul>** tag in the "correct" code example.

1.2.4.2 XHTML Elements Must Always Be Closed

Non-empty elements must have a closing tag.

This is wrong:	This is Correct
<p>This is a paragraph	<p>This is a paragraph</p>
<p>This is another paragraph	<p>This is another paragraph</p>

1.2.4.3 Empty Elements Must Also Be Closed

Empty elements must also be closed.

This is wrong:
A break: <br>
A horizontal rule: <hr>
An image: <img src="happy.gif" alt="Happy face">

This is correct:
A break: <br />
A horizontal rule: <hr />
An image: <img src="happy.gif" alt="Happy face" />

1.2.4.4 XHTML Elements Must Be In Lower Case

Tag names and attributes must be in lower case.

This is wrong:	**This is correct**
<BODY>	<body>
<P>This is a paragraph</P>	<p>This is a paragraph</p>
</BODY>	</body>

1.2.4.5 XHTML Documents Must Have One Root Element

All XHTML elements must be nested within the <html> root element. Child elements must be in pairs and correctly nested within their parent element.

The basic document structure is:

```
<html>
<head> … </head>
<body> … </body>
</html>
```

1.2.5 XHTML Syntax

Here are some more XHTML syntax rules :

- Attribute names must be in lower case
- Attribute values must be quoted
- Attribute minimization is forbidden
- The id attribute replaces the name attribute
- The XHTML DTD defines mandatory elements

1.2.5.1 XHTML Attribute Names must be in lower case

Attribute Names Must Be In Lower Case

This is wrong	**This is correct**
<table WIDTH="100%">	<table width="100%“>

1.2.5.2 XHTML Attribute values must be quoted.

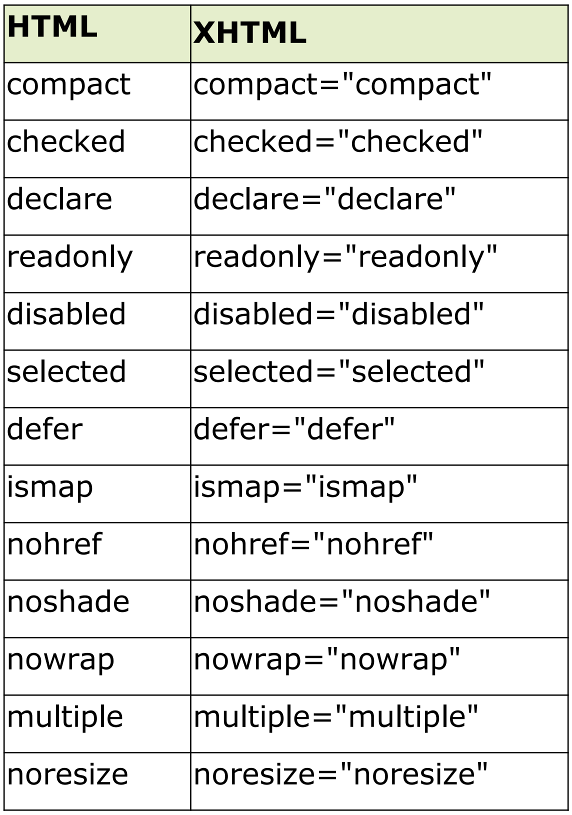
Attribute Values Must Be Quoted

This is wrong	**This is correct**
<table width=100%>	<table width="100%">

1.2.5.3 Attribute Minimization Is Forbidden

This is wrong	This is correct
<input checked>	<input checked="checked" />
<input readonly>	<input readonly="readonly" />
<input disabled>	<input disabled="disabled" />
<option selected>	<option selected="selected" />
<frame noresize>	<frame noresize="noresize" />

The chart in Figure 1.2-1 shows a list of the minimized attributes in HTML and how they should be written in 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"

Figure1.2-1

1.2.5.4 The Lang Attribute

The **lang** attribute applies to almost every XHTML element**. It specifies the language of the content within an element**. If you use the lang attribute in an element, you must also add the x**ml:lang** attribute, like this:

```
<div lang="it" xml:lang="it">Ciao bella!</div>
```

1.2.5.5 Mandatory XHTML Elements

All XHTML documents must have a DOCTYPE declaration. The html, head, title, and body elements must be present.

This is an XHTML document with a minimum of required tags where it starts with a DOCTYPE (see details about DOCTYPE next page).

```
<!DOCTYPE Doctype goes here>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>Title goes here</title>
</head>

<body>
</body>

</html>
```

Note: The DOCTYPE declaration is not a part of the XHTML document itself. It is not an XHTML element. You will learn more about the XHTML DOCTYPE in the next chapter.

Note: The **xmlns** attribute in <html>, specifies the xml namespace for a document, and is required in XHTML documents. However, the HTML validator at w3.org does not complain when the xmlns attribute is missing. This is because the namespace "xmlns=http://www.w3.org/1999/xhtml" is default, and will be added to the <html> tag even if you do not include it.

1.2.6 XHTML DTD (Document Type Definition)

DTD is the first item of code that appears on a web page. Each doctype declaration identifies whether it's strict, transitional or frameset. For example, the following string is from the DTD code**: "-//W3C//DTD XHTML 1.0 Transitional//EN."** Located between the "W3C" and "EN" is the type of DTD, which is XHTML 1.0 Transitional. For strict DTD or frameset DTD, replace the word "transitional" with "Strict" or "Frameset." Your visual studio development tool automatically insert them at the beginning of an HTML document. **The DTD contains instructions for the browser about how to display the web pages**.

1.2.6.1 Strict

The **Strict** document type declaration informs Web browsers that Web pages contain only acceptable XHTML markup language where they are marked up cleanly, free of presentational clutter. Use of either deprecated elements or deprecated attributes will prevent Web pages from loading properly. This is because those elements and attributes are no longer used in XHTML. For instance, font styles such as <u></u> for underlining text, are acceptable when using HTML 4.01, but not recognized with XHTML. Underlining text can only be done in the cascading style sheet (CSS). **You use the strict DTD together with cascading style sheets, because it doesn’t allow attributes like "bgcolor" to be set for the <body> tag, etc.**

The strict DTD looks like this:

<!DOCTYPE html
PUBLIC "-//W3C//DTD XHTML 1.0 **Strict**//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">

1.2.6.2 Transitional

The transitional document type declaration tells Web browsers the Web pages contain markup language from both XHTML and HTML 4.01. This simply means that the obsolete underline tag that you can't use with strict DTD, you can use it with transitional DTD. The underline tag along with

other deprecated elements and attributes will be recognized when the browser goes to load the page. The transitional document type declaration is most often used to avoid compatibility issues that may occur within a browser. This is how the transitional DTD is written:

The transitional DTD looks like this:

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

1.2.6.3 Frameset

If you want your Web pages to contain frames, the frameset DTD is the best suited for that purpose. Other than the use of frames, the frameset DTD is the same as the transitional DTD. This is how the frameset DTD is written:

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

1.2.6.5 An XHTML Example

An XHTML document consists of three main parts:

This is a simple (minimal) XHTML document:

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

<html>
<head>
<title>simple document</title>
</head>
<body>
<p>a simple paragraph</p>
</body>
</html>
```

Example 1.2-3

1.2.7 XHTML Validation

An XHTML document can be validated with W3C's validator.

1.2.7.1 Validate XHTML With A DTD

Before an XHTML file can be validated, a correct DTD must be added as the first line of the file.

The Strict DTD includes elements and attributes that have not been deprecated or do not appear in framesets:

**!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd"**

The Transitional DTD includes everything in the strict DTD plus deprecated elements and attributes:

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

The Frameset DTD includes everything in the transitional DTD plus frames as well:

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

You can validate a XHTML document here. http://validator.w3.org/