The Overview and Comparation of HTML and XHTML

Su Wei
860424-T086
candle0601@gmail.com
HTML Example

<html>
<body>
<h1>My First Heading</h1>
<p>My first paragraph.</p>
</body>
</html>
What is HTML?

HTML is a language for describing web pages.

- HTML stands for Hyper Text Markup Language
- HTML is not a programming language, it is a markup language
- A markup language is a set of markup tags
- HTML uses markup tags to describe web pages
What is XHTML

• XHTML stands for EXtensible HyperText Markup Language
• 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
• XHTML is compatible with HTML 4.01.
• All new browsers have support for XHTML.
Why XHTML?

- XHTML is a combination of HTML and XML (EXtensible Markup Language).
- XHTML consists of all the elements in HTML 4.01 combined with the syntax of XML.
- We have reached a point where many pages on the WWW contain "bad" HTML.
- The following HTML code will work fine if you view it in a browser, even if it does not follow the HTML rules:
Continue...

<html>
<head>
<title>This is bad HTML</title>
<body>
<h1>Bad HTML</h1>
</body>
</html>

Bad html  b-xhtml  f-xhtml
Continue...

- XML is a markup language where everything has to be marked up correctly, which results in "well-formed" documents.
- XML was designed to describe data and HTML was designed to display data.
- Today’s market consists of different browser technologies, some browsers run Internet on computers, and some browsers run Internet 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 HTML and XML, and their strengths, we got a markup language that is useful now and in the future — XHTML.
XHTML Syntax

Some 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**
XHTML Syntax

Attribute Names Must Be In Lower Case

This is wrong:
<table WIDTH="100%">

This is correct:
<table width="100%"/>
XHTML Syntax

Attribute Values Must Be Quoted

This is wrong:
<table width=100%>

This is correct:
<table width="100%"
**XHTML Syntax**

**Attribute Minimization Is Forbidden**

This is wrong:

```
<input checked>
<input disabled>
<option selected>
```

This is correct:

```
<input checked="checked" />
<input disabled="disabled" />
<option selected="selected" />
```
XHTML Syntax

The id Attribute 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. Use id instead.
Continue...

This is wrong:
&lt;img src="picture.gif" name="picture1" /&gt

This is correct:
&lt;img src="picture.gif" id="picture1" /&gt

**Note:** To interoperate with older browsers for a while, you should use both name and id, with identical attribute values, like this:

&lt;img src="picture.gif" id="picture1" name="picture1" /&gt

**IMPORTANT Compatibility Note:**
To make your XHTML compatible with today’s browsers, you should add an extra space before the ""/"" symbol.
**XHTML Syntax**

**Mandatory XHTML Elements**

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

**Note:** The DOCTYPE declaration is not a part of the XHTML document itself. It is not an XHTML element, and it should not have a closing tag.
Continue...

This is a minimum XHTML document template:
<!DOCTYPE Doctype goes here>
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<title>Title goes here</title>
</head>
<body>
</body>
</html>
XHTML DTD

<!DOCTYPE> Is Mandatory

An XHTML document consists of three main parts:

* the DOCTYPE declaration
* the <head> section
* the <body> section
Continue...

The basic document structure is:

```html
<!DOCTYPE ...>
<html>
<head>
<title>...</title>
</head>
<body>...</body>
</html>
```
Continue...

This is a simple (minimal) XHTML document:

```xml
<!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>
```
Continue...

Document Type Definitions (DTD)

* A DTD specifies the syntax of a web page in SGML

* DTDs are used by SGML applications, such as HTML, to specify rules for documents of a particular type, including a set of elements and entity declarations

* An XHTML DTD describes in precise, computer-readable language, the allowed syntax of XHTML markup

There are three XHTML DTDs:

* STRICT
* TRANSITIONAL
* FRAMESET
Continue...

- XHTML 1.0 Strict
- <!DOCTYPE html
- PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
- "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
- Use the strict DOCTYPE when you want really clean markup, free of presentational clutter. Use together with CSS.
Continue...

- **XHTML 1.0 Transitional**
- `<!DOCTYPE html`
- `PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
- "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
- **Use the transitional DOCTYPE when you want to still use HTML’s presentational features.**
Continue...

- XHTML 1.0 Frameset
- <!DOCTYPE html
- PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN"
- "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd"
- Use the frameset DOCTYPE when you want to use HTML Frames to split the web page into two or more frames.
XHTML vs HTML

- You can prepare yourself for XHTML by starting to write strict HTML.
- HTML is the antecedent technology to XHTML
XHTML vs HTML

The Most Important Differences:

• 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
XHTML vs HTML

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</i></b>

In XHTML, all elements must be properly nested within each other, like this:
<b><i>This text is bold and italic</i></b>
XHTML vs HTML

This is wrong:

```html
<ul>
  <li>Coffee</li>
  <li>Tea</li>
  <ul>
    <li>Black tea</li>
    <li>Green tea</li>
  </ul>
  <li>Milk</li>
</ul>
```
XHTML vs HTML

This is correct:

```html
<ul>
  <li>Coffee</li>
  <li>Tea</li>
  <ul>
    <li>Black tea</li>
    <li>Green tea</li>
  </ul>
  <li>Milk</li>
</ul>
```
XHTML vs HTML

XHTML Elements Must Always Be Closed
Non-empty elements must have an end tag.

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

This is correct:
<p>This is a paragraph</p>
<p>This is another paragraph</p>
XHTML vs HTML

- Empty Elements Must Also Be Closed
- Empty elements must either have an end tag or the start tag must end with `/>`.

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"/>
/>
XHTML vs HTML

**XHTML Elements Must Be In Lower Case**

The XHTML specification defines that the tag names and attributes need to be lower case.

This is wrong:

```html
<BODY>
<P>This is a paragraph</P>
</BODY>
```

This is correct:

```html
<body>
<p>This is a paragraph</p>
</body>
```
**XHTML vs HTML**

**XHTML Documents Must Have One Root Element**

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

```
<html>
<head> ... </head>
<body> ... </body>
</html>
```
Thank You! 😊