We summarize

- [XHTML 1.0 The Extensible HyperText Markup Language], W3C Recommendation. 2002.

1 What is XHTML?

- An XHTML document is an XML document.
- XHTML extends and subsets HTML 4.
- An XHTML document can be written to be understood by older (HTML) browsers.
- XHTML documents can use applications that rely on the HTML DOM or or the XML DOM.

Note:
Because of the lack of a formal HTML specification, the early browsers extended and modified HTML as they pleased. They also modified their parsers so as to display markup that was not valid but common nonetheless in an effort to please the users. That is, users would complain that the browser did not display a certain page even thought it was the page itself that was incorrectly written. As such, the HTML language evolved to be very complex for machines to parse.

XML, on the other hand, had been designed to be easy for machines to parse. The retro-fitting of HTML into an XML framework gave rise to XHTML.

2 Why XHTML?

- Because it is eXtensible. New features can be added with the use of namespaces.
- XHTML is being designed so it can be used by different user agents (mobile phones, text-to-speech, autonomous agents).

3 XHTML Conformance

1. Must obey the XHTML DTDs.
2. Root element must be html.
3. The html must contain an xmlns pointing to the the XHTML namespace.
   `<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">`
4. A DOCTYPE must appear before the html and point to the appropriate XHTML DTD.
5. Cannot override any parameter entitied in the DTD.

Example:
4 XHTML and other Namespaces

- Although it is not strictly conforming, XHTML can include other namespaces.
- For example, in order to get something like
  \[ \int e^x \, dx = e^x \]
  
you would use

```xml
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">
  <head>
    <title>A Math Example</title>
  </head>
  <body>
    <math xmlns="http://www.w3.org/1998/Math/MathML" mode="display">
      \int e^x \, dx = e^x
    </math>
  </body>
</html>
```

5 Differences with HTML

- Must be well-formed
  
  ```xml
  <p>This is <em>bad</em></p>
  ```

- Element and attribute names in lower case.
- Use end tags.
  
  ```xml
  <p>This</p>
  ```
  or
  
  ```xml
  This
  ```
  but not
  
  ```xml
  This
  ```
• Attribute values must always be quoted, like `<td rowspan="1">this</td>`

• Only attribute-value pairs are allowed. `<dl compact>` is not allowed, use `<dl compact="compact">` instead.

**Note:**

You should always use XHTML, especially when the content is generated automatically, for example, by a servlet. XHTML imposes restrictions for the writer of the markup which then serve to significantly lighten the load of the reader (browser). This means that web pages can be rendered faster and processed by automated agents. The hope is that if we have more strictly formatted content then automated agents and other aggregators will be easier to develop which will begin the evolution towards the Semantic Web.

**Notes**

[http://www.w3.org/TR/xhtml1/](http://www.w3.org/TR/xhtml1/)

This talk is available at [http://jmvidal.cse.sc.edu/talks/xhtml](http://jmvidal.cse.sc.edu/talks/xhtml)

Copyright © 2004 Jose M Vidal. All rights reserved.