



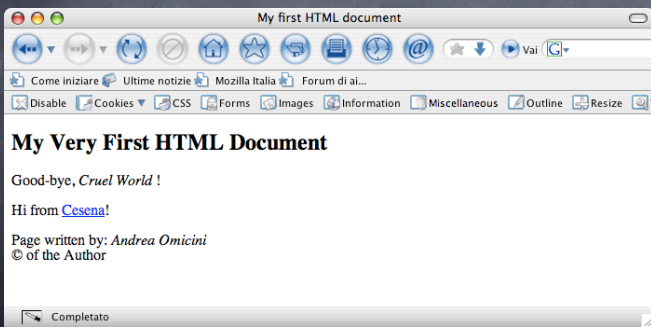
XHTML: An Introduction

Prof. Ing. Andrea Omicini
Ingegneria Due, Università di Bologna a Cesena
andrea.omicini@unibo.it
2006-2007

Good-bye Cruel World!

```
<?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>My first HTML document</title>
</head>
<body>
<h2>My Very First HTML Document</h2>
<p>
  Good-bye, <i>Cruel World </i>!
</p>
<i-- "Hello World" did not seem enough -->
<p>
  Hi from <a href="http://www.ing2.unibo.it/">Cesena</a>!
</p>
<p>
  Page written by: <cite>Andrea Omicini</cite>
  <br />
  &#169; of the Author
</p>
</body>
</html>
```

We Obtain...



What is HTML?

- It is a markup language
 - allows you to annotate text, and to embody annotations in along with text in a document
 - annotations provide text with properties
 - e.g., printing properties as annotations, in order to separate them from content
 - SGML subset
 - Standard Generalized Markup Language
- A family of standards
 - W3C: consortium in charge of Web standards
 - <http://w3c.org/Markup>
 - Develops over time
 - either official or proprietary extensions
 - proposals, drafts, releases and recommendations

Versions

- From 1.0, 2.0, 3.2, 4.0 to 4.01
 - HTML 4.01 is the last recommendation
 - <http://www.w3.org/TR/html4/>
 - XHTML 1.0 current standard
 - defined based on HTML 4.01
 - as more or less its XML-compliant version
 - <http://www.w3.org/TR/xhtml1/>
 - XHTML 2.0 still ongoing work
 - <http://www.w3.org/TR/xhtml2/>
 - 26/7/2006: last public Working Draft of XHTML 2.0 has been published
 - <http://www.w3.org/TR/2006/WD-xhtml2-20060726>



XHTML™ 2.0

W3C Working Draft 26 July 2006

This version: <http://www.w3.org/TR/2006/WD-xhtml2-20060726>

Latest version: <http://www.w3.org/TR/xhtml2>

Previous version: <http://www.w3.org/TR/2005/WD-xhtml2-20050527>

Diff-marked version: [xhtml2-diff.html](http://www.w3.org/TR/2005/WD-xhtml2-20050527)

Editors:

- Jonny Axelsson, Opera Software
- Mark Birbeck, x-port.net
- Micah Dubinko, Invited Expert
- Beth Epperson, Websense
- Masayasu Ishikawa, W3C
- Shane McCarron, Applied Testing and Technology
- Ann Navarro, WebGeek, Inc.
- Steven Pemberton, CWI (HTML Working Group Chair)

This document is also available in these non-normative formats: [Single XHTML file](#), [PostScript version](#), [PDF version](#), [ZIP archive](#), and [ScripsTAR archive](#).

Copyright ©2006 W3C® (MIT, ERCIM, Keio). All Rights Reserved. W3C liability, trademark, and document use rules apply.

From SGML to HTML

- SGML is a very intricate markup language
 - Web needs just a subset of it
- SGML is also a meta-markup language
 - can be used to define other markup languages
 - by defining their own DTD (Document Type Definition)
- HTML is defined via a SGML DTD
 - <http://www.w3.org/TR/html4/sgml/dtd.html>

7

HTML / XHTML Document

- Suffix .html (or .htm, for windows fault)
 - Text file
- How can I create my own HTML file?
 - with any text editor, saving text with .html suffix
 - with any word processors allowing "Save as Text"
 - with any web-page creation tool
 - Composer, Dreamweaver, etc.
 - please no FrontPage
- How does computer see the HTML file?
 - in the same way as we do?
 - different levels of "perception"
 - OS ≠ editors ≠ browsers ≠ ...

8

Elements, Tags & Attributes

- An HTML document contains
 - elements** & sections delimited by **tags**
- Generally, tags delimit start & end of a section / element
 - `<tag>section or element</tag>`
 - it is obviously useful to learn the main HTML tags
- Tags may contain further specifications called **attributes**
 - some of them **required**, some **optional**
 - ``
 - src is mandatory
 - no closing tag, `/>` is used
- Remind: tags and elements are in general different things
 - `<p>` is a tag, `<p>Paragraph</p>` is an element

9

Some Details

- White spaces have no meaning
 - if not within strings
- HTML is not case sensitive
 - `<p>` or `<P>`, it is the same
- Please notice: the same does NOT hold for XHTML!
 - `<p>` is correct, `<P>` is wrong

10

Types of Tags

- Section tag
- Header tag
- Content tag
- Styling tag
- Empty elements (?)
- Anchor / Hyperlink tags

11

Section Tags

- dividing HTML document in sections
- root tag
 - `<html>` starts HTML document
 - may not be the beginning of the HTML file...
 - `</html>` ends it
 - while the file might go on...
- 2 sections: Header & Body

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

12

Header tags

- ⌚ within the header, between `<head>` and `</head>`
- ⌚ not displayed directly by the browser
- ⌚ main header tags
 - ⌚ `<title>` defines page title
 - ⌚ in the title bar of the browser window
 - ⌚ `<meta />` carries meta-information on the document content
 - ⌚ e.g.: `<meta name="author" content="Andrea Omicini" />`
 - ⌚ like a comment, but can be referenced and used

13

Content tags

- ⌚ within the body, between `<body>` and `</body>`
- ⌚ used by browser for display
- ⌚ most of the useful tags
 - ⌚ `<p>` for paragraphs
 - ⌚ `<table>` for tables
 - ⌚ `<h1>` for 1st-level headers
 - ⌚ `<h2>`, `<h3>`, ... next levels
 - ⌚ comments
 - ⌚ `<!-- this is a comment -->`

14

Styling tags

- ⌚ two kinds
 - ⌚ based on the content nature
 - ⌚ based on formatting
- ⌚ content-based: examples
 - ⌚ `<blockquote>` contains a block for a quotation
 - ⌚ `<cite>` contains a reference to a citation
- ⌚ format-based: examples
 - ⌚ `` bold, `<i>` italic
- ⌚ sometimes no differences in display by browsers
 - ⌚ ``, `<cite>`, `<i>`, `<dfn>`
 - ⌚ but the source shows the differences in markup
 - ⌚ that could be used anyway for some reasons

15

Empty tags

- ⌚ In XHTML
 - ⌚ `
` line break
 - ⌚ `<hr />` horizontal rule
 - ⌚ `` inline image
- ⌚ In HTML, `
` & `<hr>` are ok
- ⌚ in XHTML they should be "closed" somehow
- ⌚ Pay attention to attributes!
 - ⌚ required & optional
 - ⌚ e.g., attribute `src` in `img` is required
 - ⌚ take a close look to specifications
 - ⌚ check when needed
 - ⌚ exploit tools!
 - ⌚ along with their embedded knowledge

16

Anchor / Hypertext tags

- ⌚ tag `<a>` for both
 - ⌚ "anchor" denotes portions of a document
 - ⌚ to be directly referred to with `#`
 - ⌚ "hypertextual reference" denotes other docs
 - ⌚ or portions of them
 - ⌚ obviously contains a URL
- ⌚ `...`
- ⌚ `...`
 - ⌚ relative / absolute URL
 - ⌚ `` denotes an anchor within an href
- ⌚ Pay attention to quotes!

17

Limits of HTML (1)

- ⌚ Content intermixed with presentation
 - ⌚ from 1.0 to 4.01 things have improved
 - ⌚ but too many biases from browsers
 - ⌚ to be absolutely AVOIDED
 - ⌚ in general
 - ⌚ here in this course for sure :)
- ⌚ Not "well-formed"
 - ⌚ as XML is instead
 - ⌚ too much forgiving
 - ⌚ elements can be interleaved, tag can be wrong, closing tags or attributes may be missing, etc...

18

Limits of HTML (2)

- ④ It is more a sort of "structural" markup language
 - ④ describes text structure
 - ④ structural markup
 - ④ rather than the nature of content
 - ④ descriptive / semantic markup
 - ④ not easy to adapt to the different natures of media
- ④ That is why HTML moved toward XML
 - ④ through XHTML
- ④ Goal: a language aimed at being
 - ④ disciplined and easy to check
 - ④ powerful but simple
 - ④ descriptive

19

XML in short

- ④ Extensible markup language
 - ④ to define markup languages
- ④ XML application
 - ④ a markup language defined via XML
 - ④ XHTML is an XML application
- ④ Essential remark: XML has no predefined tag / elements
 - ④ anyone can define tags and structures that better fits the chosen contents

20

Fundamental Parts of a XML Document

1. XML Document (properly said)
 - ④ content built according to XML rules
2. Document Type Definition (DTD)
 - ④ which tags and their meaning
3. Style Sheet
 - ④ for presentation

21

Benefits of XML

- ④ Portable
 - ④ text format, so that any platform is ok, and many applications are available to read & write XML
- ④ Configurable / Extensible
 - ④ anybody can define his/her own markup language
- ④ Self-descriptive
 - ④ an XML document is self-contained: presentation, meaning, data & their structure

22

XHTML = HTML + XML

- ④ XHTML defined using XML as a meta-language
 - ④ HTML defined instead in SGML
- ④ vocabulary taken from HTML, syntax from XML
 - ④ backward compatibility
 - ④ in particular, in the "human legacy"
- ④ XML properties
 - ④ well-formed, not error-prone, extensible via XML

23

XHTML - HTML = ?

- ④ XML Prologue
 - ④ first element of the XHTML document
 - ④ `<?xml version="1.0" encoding="UTF-8" ?>`
 - ④ like corresponding `<meta />` for old browser
- ④ Document type declaration
 - ④ what is the document DTD?
 - ④ `<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">`
 - ④ address, or embedded DTD
 - ④ both before `<html>` tag

24

Other differences

- ④ Recommendation: define the namespace
 - ④ to give meaning to tags
 - ④ we could write
 - `<http://www.w3.org/1999/xhtml:p>` for `<p>` tag
 - ④ and then the same all the others...
 - ④ however, it seems easier to write
 - `<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">`
 - ④ also, we could add our namespaces
 - ④ and declare the languages
- ④ Other
 - ④ case sensitivity
 - ④ full nesting
 - ④ required elements: `<head>`, `<body>`, `<html>`, `<title>`

25

What should we learn from the lab activity?

- ④ Structure of the HTML / XHTML document
 - ④ header, body, and their content
- ④ Inline elements, comments, lists, special chars
- ④ Attributes
 - ④ shared by most elements
 - ④ attributes to affect presentation
- ④ Anchors / Hypertextual references
- ④ Images
- ④ Tables
- ④ Forms
- ④ Frames, perhaps

26