



Overview

- Scope
- Conformance
- Terms
- HTML and XML
- HTML Document Elements
 - HEAD and BODY
 - Headings: H1 ... H6
 - Block Structuring Elements
 - List Elements
 - Hyperlinks
- Page Design

Scope

- The Hypertext Markup Language (HTML) is a simple markup language used to create hypertext documents that are platform independent.
- HTML documents are one type of SGML/XML document. The tags and semantics were designed for representing simple information from a wide range of domains.
 - HTML has been in use since 1990.
 - HTML 2.0 -- features as of June 1994
 - HTML 4.0 -- features as of April 1998
- The HTML Document Type Definition (DTD) is a formal definition of the HTML syntax in terms of SGML.
- HTML as an Internet MediaType[IMEDIA] and MIME Content Type[MIME] called 'text/html'.

September 28, 2001

Introduction to HTML

3

Well-formed and Valid

- Well formed and valid
 - Documents meeting the HTML/XML syntax rules are well formed
 - Valid documents also conform to the DTD (HTML or other)
 - The HTML 4.0 specification is available from the W3C
 - There are a number of syntactic idioms that are not supported or are supported inconsistently in some historical user agent implementations.
- Character sets are independent of the character encoding scheme used to represent a document.
 - For example, the 'ISO-2022-JP' character encoding scheme can be used for HTML documents, since its repertoire is a subset of the [ISO-10646] repertoire.

September 28, 2001

Introduction to HTML

4

Feature Test Entities

- The HTML DTD defines a standard HTML document type and several variations, by way of feature test entities. Feature test entities are declarations in the HTML DTD that control the inclusion or exclusion of portions of the DTD.
 - HTML.Recommended
 - HTML.Deprecated
- Documents generated by translation software or editing software should not contain deprecated idioms.

September 28, 2001

Introduction to HTML

5

User Agents

- An HTML user agent conforms if:
 - It parses the characters of an HTML document into data characters and markup according to XML.
 - It supports the 'ISO-8859-1' character encoding scheme and processes character in the ISO Latin Alphabet No. 1
 - To support non-western writing systems, HTML user agents are encouraged to support 'ISO-10646-UCS-2'
 - It allows the user to traverse (or at least attempt to traverse, resources permitting) all hyperlinks from <A> elements in an HTML document.
- An HTML user agent is level 2 if, additionally:
 - It allows the user to express all form field values specified in an HTML document and to (attempt to) submit the values as requests to information services.

September 28, 2001

Introduction to HTML

6

Terms

- *anchor* one of two ends of a hyperlink
- *hyperlink* a relationship between two anchors, called the head and the tail.
- *URI* A Uniform Resource Identifier is a formatted string that identifies a resource
 - The World Wide Web uses URL's - Uniform Resource Locators, which are one form of URI
 - There is a significant effort to move to Uniform Resource Names - URN's, another form of URI that would be location independent.
- *base URL* an absolute URL used in combination with a relative URL to determine another absolute URL.
- *fragment identifier* the portion of an HREF attribute value following the '#' character which modifies the presentation of the destination of a hyperlink.

September 28, 2001

Introduction to HTML

7

Terms

- *markup* Syntactically delimited characters added to the data of a document to represent its structure.
- *tag* Markup that delimits an element. A tag includes a name which refers to an element declaration in the DTD, and may include attributes.
 - *start-tag* Descriptive markup that identifies the start of an element, generic identifier and attributes.
 - *end-tag* markup that identifies the end of an element.
- *element* A component of the hierarchical structure defined by a document type definition; it is identified by descriptive markup --a start-tag and end-tag.

September 28, 2001

Introduction to HTML

8

Terms

- *character* An atom of information -- letter or a digit.
 - Graphic characters have associated glyphs
 - Control characters have associated processing semantics.
- *data character* Characters other than markup, which make up the content of elements.
- *reference markup* that refers to a character by its code position in the document character set.
- *text entity* A finite sequence of characters. A text entity typically takes the form of a sequence of octets with some associated character encoding scheme.

September 28, 2001

Introduction to HTML

9

HTML, SGML, and XML

- HTML was an application of ISO 8879:1986 -- Standard Generalized Markup Language (SGML).
- SGML is a system for defining structured document types and markup languages to represent instances of those document types.
- SGML syntax grammar separates an SGML document into three parts:
 - SGML declaration
 - a prologue
 - an instance.
- HTML's relationship with SGML is now superseded by the development of XML, a simplified and updated form of SGML.

September 28, 2001

Introduction to HTML

10

HTML Lexical Syntax

- To identify information as an HTML document, each document must start with a document type declaration.

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
```

- Within a document:
 - Elements are nested hierarchically
 - Tags delimit elements
 - A start tag is of the form <name>
 - An end tag is of the form </name>
 - The content of an element is a sequence of data chars
 - In a start-tag, white space and attribute-value pairs are allowed closing delimiter. Values may be:
 - A string literal
 - A name token
- Entities within a document are of the form &name;
- A comment is of the form <!-- comment -->

September 28, 2001

Introduction to HTML

11

Example HTML Document

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<HTML>
<!-- Here's a good place to put a comment. -->
<HEAD>
<TITLE>Structural Example</TITLE>
</HEAD><BODY>
<H1>First Header</H1>
<P>This is a paragraph in the example HTML file. Keep in mind
that the title does not appear in the document text, but that
the header (defined by H1) does.</P>
```

September 28, 2001

Introduction to HTML

12

Example HTML Document (continued)

```
<OL>
<LI>First item in an ordered list.
<LI>Second item in an ordered list.
<UL COMPACT>
<LI> Note that lists can be nested;
<LI> Whitespace may be used to assist in reading the
    HTML source.
</UL>
<LI>Third item in an ordered list.
</OL>
<P>This is an additional paragraph. Technically, end tags are
not required for paragraphs, although they are allowed. You can include
character highlighting in a paragraph. <EM>This sentence of the paragraph
is emphasized.</EM> Note that the <lt;/P> end tag has been omitted.
<P> <IMG SRC ="triangle.xbm" alt="Warning: ">
Be sure to read these <b>bold instructions</b>.
</BODY></HTML>
```

September 28, 2001 Introduction to HTML 13

Document Structure

- An HTML document is a tree of elements, including:
 - a head
 - a body
 - headings
 - paragraphs
 - lists
- Forms
- Imagemaps
- Applets

September 28, 2001

Introduction to HTML

14

The HTML document element

- consists of a head and a body
- The head of an HTML document is an unordered collection of information about the document, e.g.:

```
<HEAD>  
<TITLE>Introduction to HTML</TITLE>  
</HEAD>
```
- The title identifies the document in a global context.
- The <BASE> element provides a base address for interpreting relative URL's
- The <ISINDEX> element indicates that the user agent should allow the user to search an index by keywords
- The <LINK> element represents a hyperlink
- The <META> element is an extensible container for use in identifying specialized document meta-information.

September 28, 2001

Introduction to HTML

15

BODY

- The <BODY> element contains the text flow of the document, including headings, paragraphs, lists, etc. For example:

```
<BODY>  
<h1>Important Stuff</h1>  
<p>Explanation about important stuff...  
</BODY>
```
- The six heading elements, <H1> through <H6> denote section headings.
 - Although the order and occurrence of headings is not constrained by the HTML DTD, documents should not skip levels (for example, from H1 to H3), as converting such documents to other representations is often problematic.

September 28, 2001

Introduction to HTML

16

Typical renderings

- H1 Bold, very-large font, centered. One or two blank lines above and below.
- H2 Bold, large font, flush-left. One or two blank lines above and below.
- H3 Italic, large font, slightly indented from the left margin. One or two blank lines above and below.
- H4 Bold, normal font, indented more than H3. One blank line above and below.
- H5 Italic, normal font, indented as H4. One blank line above.
- H6 Bold, indented same as normal text, more than H5. One blank line above.

September 28, 2001

Introduction to HTML

17

Block Structuring Elements

- Block structuring elements include paragraphs, lists, and block quotes.
 - They must not contain heading elements
 - They may contain phrase markup
 - They may, in some cases, be nested.
- The `<P>` element indicates a paragraph.
 - The exact indentation, leading space, etc. of a paragraph is not specified and may be a function of other tags, style sheets, etc.
 - Typically, paragraphs are surrounded by a vertical space of one line or half a line. The first line in a paragraph is indented in some cases.

September 28, 2001

Introduction to HTML

18

Block Structuring Elements

- The `<PRE>` element represents a character cell block of text and is suitable for text that has been formatted for a monospaced font.
- The `<XMP>` and `<LISTING>` elements are similar to the `<PRE>` element. Their content is declared as CDATA, which means that no markup except the end-tag open delimiter-in-context is recognized
- The `<ADDRESS>` element contains such information as address, signature and authorship, often at the beginning or end of the body of a document.
- The `<BLOCKQUOTE>` element contains text quoted from another source. A typical rendering might be a slight extra left and right indent, and/or italic font.

September 28, 2001

Introduction to HTML

19

List Elements

- HTML includes a number of list elements. They may be used in combination; for example, a `` may be nested in an `` element of a ``
- Lists include ``, ``, `<DL>`, `<DIR>`, `<MENU>`
- The `COMPACT` attribute suggests compact rendering.
- The content of a list is a sequence of `` elements
- The `` is typically rendered as a bulleted list.
- The `` is typically rendered as a numbered list.

September 28, 2001

Introduction to HTML

20

List Elements (continued)

- The `<DIR>` is similar to the `` element.
 - It represents a list of short items, typically up to 20 characters each.
 - Items in a directory list may be arranged in columns, typically 24 characters wide.
- The `<MENU>` element is a list of items with typically one line per item.
 - The menu list style is typically more compact than the style of an unordered list.
- The content of a `<DL>` element is a sequence of `<DT>` elements and/or `<DD>` elements, usually in pairs.
 - Multiple `<DT>` may be paired with a single `<DD>` element.
 - Documents should not contain multiple consecutive `<DD>` elements.

September 28, 2001

Introduction to HTML

21

Phrase Markup

- The `<CITE>` element is used to indicate the title of a book or other citation
- The `<CODE>` element indicates an example of code, typically rendered in a mono-spaced font.
- The `` element indicates an emphasized phrase, typically rendered as italics.
- The `<KBD>` element indicates text typed by a user, typically rendered in a mono-spaced font.
- The `<SAMP>` element indicates a sequence of literal characters, typically rendered in a mono-spaced font.
- The `` element indicates strong emphasis, typically rendered in bold.
- The `<VAR>` element indicates a placeholder variable, typically rendered as italic.

September 28, 2001

Introduction to HTML

22

Typographic and Other Elements

- `` element indicates bold text. Where bold typography is unavailable, an alternative representation may be used.
- `<I>` element indicates italic text. Where italic typography is unavailable, an alternative representation may be used.
- `<TT>` element indicates teletype (monospaced) text. Where a teletype font is unavailable, an alternative representation may be used.
- `
` element specifies a line break between words
- `<HR>` element is a divider between sections of text; typically a full width horizontal rule or equivalent graphic.

September 28, 2001

Introduction to HTML

23

Anchor

- The `<A>` element indicates a hyperlink anchor.
 - `NAME` and/or `HREF` attributes should be present.
 - Attributes of the `<A>` element:
 - `HREF` gives the URI of the head anchor of a hyperlink.
 - `NAME` gives the name of the anchor, and makes it available as a head of a hyperlink.
 - `TITLE` suggests a title for the destination resource -- advisory only. The `TITLE` attribute may be used for display prior to accessing the destination resource or for resources that do not include a title, such as graphics
 - `REL` gives the relationship(s) described by the hyperlink. The value is a whitespace separated list of relationship names. The semantics of link relationships are not specified.
 - `REV` same as the `REL` attribute, but the semantics of the relationship are in the reverse direction.
 - `URN` specifies a preferred, persistent identifier for the head anchor of the hyperlink. Semantics are not yet specified.
 - `METHODS` specifies methods to be used in accessing the destination

September 28, 2001

Introduction to HTML

24

Images

- The element refers to an image or icon via a hyperlink
- Attributes of the element include:
 - ALIGN alignment of the image with respect to the text baseline.
 - TOP specifies that the top of the image aligns with the tallest item on the line containing the image.
 - MIDDLE specifies that the center of the image aligns with the baseline of the line containing the image.
 - BOTTOM specifies that the bottom of the image aligns with the baseline of the line containing the image.
 - ALT text to use in place of the referenced image resource, for example due to processing constraints or user preference.
 - ISMAP indicates an image map
 - SRC specifies the URI of the image resource. In practice, the image resources are limited to gif and jpeg.
 - HEIGHT and WIDTH specify the display height and width.

September 28, 2001

Introduction to HTML

25

Example of an Image

- In line image
 - ` Be sure to read these instructions.`
- An image as an anchor
 - ``
 - ``
 - ``
- An image as a map (server side)
 - ``
 - ``
 - ``
- An image as a map (client side)
 - ``
 - `<MAP NAME="mymap" <AREA SHAPE="rect" COORDS="1,1,x,y" HREF="xyz.html"></MAP>`

September 28, 2001

Introduction to HTML

26

Selected Keys to Good Pages

- Navigation Aids
 - Give each page a descriptive title
 - Use very light – subliminal – pastel solids or bitmaps images shaded light to establish “neighborhoods”
 - Put a text menu as a footer on each page allowing access to key “landmarks” – home, search, sitemap, etc.
- Format
 - Use headings and levels to specify structure
 - If using styles, use a different font family for headings
 - Try to avoid 640x480 white space
 - Use tables to help align “chunks”
 - use width % and not absolute pixel size
 - Don't use colors except for links

September 28, 2001

Introduction to HTML

27

Selected Keys to Good Pages

- Anchors
 - Preference for RED, GREEN, and BLUE for links, alinks and vlinks
 - Make each anchor as explicit as possible – semantically clear
- List Formating
 - Make use of COMPACT
 - Make use of start and type in lists
 - UL types include DISC, SQUARE, CIRCLE
 - OL types include I, 1, A, I, a
 - Start allows a number(letter) to be specified

September 28, 2001

Introduction to HTML

28