



## Hypertext Markup Language

### Why HTML?

Documents on the WWW must be able to be displayed by a variety of different computer platforms (such as Mac, IBM-PC, UNIX etc.), and a variety of different clients (these are browsers such as Netscape Navigator, Mosaic, or Internet Explorer). This is made possible by storing all the documents in the Hypertext Markup Language (HTML). The HTML documents themselves consist only of ASCII text, allowing them to be transferred across the Internet to any platform. The language HTML provides a way to specify how the document is to be displayed (or presented) by the client.

### Structural and Surface Formatting

The wide range of platforms upon which the documents are to be displayed causes a design problem for HTML. Some of the clients are only capable of displaying plain text, others cannot display colours while still others only have a few fonts. It would be pointless to try to specify the exact look of a page which is desired, because many of the clients may not have the capability to display the document in that way. For this reason, HTML provides a structural formatting methodology where the language specifies general properties of text, and the client decides how to display those properties.

For example, HTML specifies text is a “heading” or a “list” or is to be “emphasised”, and the client decides how to display the text. Different clients choose to display the same property in different ways (eg; emphasis may be italic on one client, and underlined on another). Because properties about the text are indicated by reference to their *function* (eg; specifying text as a heading), the commands are known as structural markup. Structural markup is used in word processors when the user defines a style to use for their own purposes (eg. Defining a new style called Character\_Name for use when laying out a script).

You should be aware that HTML also supports a small number of surface formatting commands which specify text is “bold” or “italic”. These commands are known as surface markup because they indicate features about what the text looks like, without reference to how the text is *used* in the document. In a word processor, changing the font size, typeface, or altering margins without using a defined style is surface markup. Surface HTML commands should be avoided where possible.

### Tags

The commands used by HTML to specify how to markup the document are known as “tags”. There are two kinds of tags, container tags and separator tags. A container tag has a beginning and an end. The command specified by the tag applies to all the text which appears between the beginning and the end of the tag. Separator tags usually divide up the document into different sections. All tags are enclosed in angle brackets  $\langle \rangle$ .

```
<EM> This text is within an emphasis container tag </EM>

This text is divided
<HR>
by using a separator tag
```

### Essential parts of HTML

All HTML documents consist of the information contained within an HTML container tag. Within the HTML container, there are two further containers, a HEAD and a BODY. The HEAD container holds information for the client, whereas the BODY contains information which will be displayed by the client.

```
<HTML>
<HEAD>
    <TITLE> This is the title of the document </TITLE>
</HEAD>
<BODY>
    This is a simple HTML document.
</BODY>
</HTML>
```

### Head tags

The head of the document contains information used by the client and server. The information within the head of an HTML document is not intended to be viewed by a human user. Most Browsers use the title to label the pages and aid navigation (ie; used in the bookmarks, and to label the browser window). The only tag which you will commonly use in the head of a document is the TITLE container.

#### Title

The title of a document is the name used by the Client to refer to the document. The title is the name which will appear your Bookmarks, or in the **Go** menus to aid in navigation. Note that the title does not appear within the page displayed on the browser.

```
<TITLE> ... </TITLE>
```

### Body tags

All the information within the body of the document is intended for the user. If you print a document, you will get all the information in the body printed, and none of the head information. There are many tags used to format the document itself. Experiment with the different tags to get an idea about how they are all used. Some of the common tags will be explained below.

#### Heading

There are 6 levels of headings which correspond to their importance. The most important heading uses the tag  $\langle H1 \rangle \dots \langle /H1 \rangle$ , and the least important is  $\langle H6 \rangle \dots \langle /H6 \rangle$ . Netscape displays the text as shown below.

```
<H1> ... </H1>      -   Heading 1
<H2> ... </H2>      -   Heading 2
<H6> ... </H6>      -   Heading 6
```

#### Strong

Strong is used for important information, and is usually displayed as Bold text  $\langle STRONG \rangle \dots \langle /STRONG \rangle$  - **Strong** - for important information

#### Emphasis

Emphasis is provided by Netscape by making the text italic.  $\langle EM \rangle \dots \langle /EM \rangle$  - *Emphasis*

### Unordered List

The unordered list required two different tags. The unordered list container is used to surround the entire list. The list item tag <LI> is used to distinguish each item in the list. The list will appear on the page with bullets beside each list item.

```
<UL> ... </UL>      -   Unordered List
<LI> ... </LI>      -   Item in a List
```

### Ordered List

The ordered list also required two different tags. The ordered list container is used to surround the entire list. The list item tag <LI> is used to distinguish each item in the list. The list will appear on the page with numbers beside each list item.

```
<OL> ... </OL>      -   Ordered List
<LI> ... </LI>      -   Item in a List
```

### Images

In order to include an image on your page, you need to specify the location of the image. The browser will load the image from the location specified, and display it in the page at the point of your image tag.

```
<IMG SRC=" ... specify a URL or filename here ... "> -Image Tag
```

### Horizontal Rule

This tag will place a horizontal line across the page. Note that it is a separator tag, not a container tag (ie; it is a single tag used to separate parts of a page).

```
<HR>      -   Horizontal Rule
```

### Paragraph

In order to add a blank line between different paragraphs you should use the paragraph tag. If you use more than one of these tags in a row, they are treated as if they were a single paragraph tag by the browser, and you only get a single line. In order to get multiple blank lines in a page, you will have to use the line break tag.

```
<P> ... </P>      -   Paragraph Break
```

### Line Break

To force the browser to place a blank line on your page (just like a carriage return), you can use the line break tag. You can use this tag repeatedly to get multiple blank lines.

```
<BR>      -   Line Break
```

### Hypertext Reference

The foundation of the WWW is the notion of hypertext. It is vitally important to know how hypertext references work in HTML. The tag provides a link to a distant document by specifying some text to act as an anchor (on which you can click), and a URL which is the destination or target document.

```
<A HREF=" ... specify URL here ... "> ...text appearing on page...</A>
```

The URL specified must refer to a particular document which forms part of the WWW. Note that if the location of the document is moved, your page will still contain the original address. To update the page, you would have to manually alter the address.

### Tables

Tables are constructed from rows and columns. The entire table must be enclosed within a TABLE tag, each row must be enclosed within a TR tag, and within each row, each item of data is required to be within a TD tag. So to create a table with one row and three columns the following code would be used:

```
<TABLE>
<TR><TD> .. 1 . </TD><TD> .. 2 .. </TD><TD> .. 3 .. </TD></TR>
</TABLE>
```

### Comments

A comment can be used in HTML to provide information about the author of the page, or information to others who may be looking at the source code. These comments do not appear on the actual page when viewed by a web browser, but are available for people who look at the HTML itself. Note the start and end of the tag must include the minus signs exactly as shown below.

```
<!--      the comment goes here      -->
```

### Example Source Code

```
<HTML>
<HEAD>
  <TITLE>
    Example Page
  </TITLE>
</HEAD>
<BODY>
  <H1>HTML Example</H1>

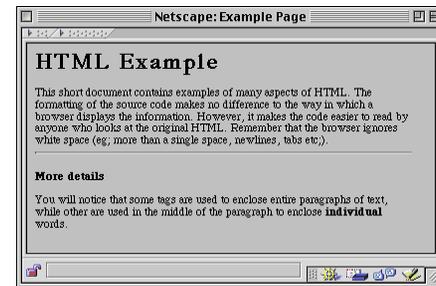
  This short document contains examples of many aspects
  of HTML. The formatting of the source code makes no
  difference to the way in which a browser displays the
  information. However, it makes the code easier to read
  by anyone who looks at the original HTML. Remember that
  the browser ignores white space (eg; more than a single
  space, newlines, tabs etc;).

  <HR>

  <H4>More details</H4>

  You will notice that some tags are used to enclose entire
  paragraphs of text, while other are used in the middle of
  the paragraph to enclose <STRONG> individual </STRONG> words.

</BODY>
</HTML>
```



### Useful References

<http://archive.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html>  
<http://www.davesite.com/webstation/html/>  
<http://werbach.com/barebones/>  
<http://werbach.com/barebones/wwwhelp.html>