A Beginner's Guide to HTML

This is a primer for producing documents in HTML, the markup language used by the World Wide Web.

Introduction

Acronym expansion

WWW
: World Wide Web

SGML
: Standard Generalized Markup Language – This is perhaps best be thought of as a programming language for style sheets.

DTD
: Document Type Definition – This is a specific implementation of document description using SGML. One way to think about this is: Fortran is to a computer program as SGML is to a DTD.

HTML
: HyperText Markup Language – HTML is a SGML DTD. In

practical terms, HTML is a collection of styles used to define the various components of a World Wide Web document.

What this primer doesn't cover

This primer assumes that you have:

- at least a passing knowledge of how to use NCSA Mosaic or other WWW browser
- a general understanding of how World Wide Web servers and client browsers work and
- access to a World Wide Web server for which you would now like to produce HTML documents

Creating HTML documents

HTML documents are in plain text format and can be created using any text editor (e.g., Emacs or vi on Unix machines). A couple of WWW browsers (tkWWW for X Window System machines and CERN's WWW browser for the NeXT) do include rudimentary HTML editors in a WYSIWYG environment, and you may want to try one of them first before delving into the details of HTML.

You can preview documents in progress with NCSA Mosaic (and some other WWW browers). Open the document using the Open Local option under the **File** *menu. Use the Filters, Directories, and Files fields to locate the document or enter the path and name of the document in the Name of Local Document to Open field. Press* **OK**.

If you see edits you want to make, enter them in the source file. Save the changes. Return to NCSA Mosaic and press the **Reload** *button on the bottom menu. The edits are reflected in the on−screen display.*

The minimal HTML document

Here is a barebones example of HTML:

```
<TITLE>The simplest HTML example</TITLE>

<H1>This is a level one heading</H1>

Welcome to the world of HTML.
This is one paragraph.<P>

And this is a second.<P>
```

Click here to see the formatted version of the example.

HTML uses tags to tell the World Web viewer how to display the text. The above example uses

- the `<TITLE>` tag (which has a corresponding `</TITLE>` tag), which specifies the title of the document,
- the `<H1>` header tag (with corresponding `</H1>`), and
- the `<P>` end–of–paragraph tag.

HTML tags consist of a left angular bracket (<), known as a ‘‘less than’’ symbol to mathematicians, followed by some text (called the directive) and closed by a right angular bracket (>). Tags are usually paired, e.g. `<H1>` and `</H1>`. The ending tag looks just like the starting tag except a slash (/) precedes the text within the brackets. In the example, `<H1>` tells the viewer to start formatting a top level heading; `</H1>` tells the viewer that the heading is complete.

The primary exception to the pairing rule is the `<P>` end–of–paragraph tag. There is no such thing as `</P>`.

Note: HTML is *not* case sensitive. `<title>` is completely equivalent to `<TITLE>` or `<TiTlE>`.

Not all tags are supported by all World Wide Web browsers. If a browser does not support a tag, it should just ignore it, though.

Titles

Every HTML document should have a title. A title is generally displayed separately from the document and is used primarily for document identification in other contexts (e.g., a WAIS search). Choose about half a dozen words that describe the document’s purpose.

> *In NCSA Mosaic, the Document Title field is at the top of the screen just below the pulldown menus.*

The directive for the title tag is `<title>`. The title generally goes on the first line of the document.

Headings

HTML has six levels of headings (numbered 1 through 6), with 1 being the most prominent. Headings are displayed in larger and/or bolder fonts than the normal body text. The first heading in each document should be tagged

<H1>. The syntax of the heading tag is:

```
<Hy>Text of heading</Hy>
```

where *y* is a number between 1 and 6 specifying the level of the heading.

For example, the coding for the ‘‘Headings’’ section heading above is

```
<H3>Headings</H3>
```

Title versus first heading: In many documents (including this one), the first heading is identical to the title. For multi−part documents, the text of the first heading should be suitable for a reader who is already browsing related information (e.g., a chapter title), while the title tag should identify the node in a wider context (e.g., include both the book title and the chapter title).

Paragraphs

Unlike documents in most word processors, carriage returns and white space in HTML files aren't significant. Word wrapping can occur at any point in your source file, and multiple spaces are collapsed into a single space (except in the <TITLE> field). Notice that in the barebones example, the first paragraph is coded as

```
Welcome to HTML.
This is the first paragraph. <P>
```

In the source file, there is a line break between the sentences. A Web browser ignores this line break and starts a new paragraph only when it reaches a <P> tag.

Important: You must end each paragraph with <P>. The viewer ignores any indentations or blank lines in the source text. Without the <P> tags, the document becomes one large paragraph. HTML relies almost entirely on the tags for formatting instructions. (The exception is text tagged as ‘‘preformatted,’’ explained below.) For instance, the following would produce identical output as the first barebones HTML example:

```
<TITLE>The simplest HTML example</TITLE><H1>This is a level
one heading</H1>Welcome to the world of HTML. This is one
paragraph.<P>And this is a second.<P>
```

However, to preserve readability in HTML files, headings should be on separate lines, and paragraphs should be separated by blank lines.

Linking to other documents

The chief power of HTML comes from its ability to link regions of text (and also images) to another document (or an image). These regions are typically highlighted by the browser to indicate that they are hypertext links.

> *In NCSA Mosaic, hypertext links are in color and underlined by default. It is possible to modify this in the Options menu as well as in your .Xdefaults file.*

HTML's single hypertext–related directive is `A`, which stands for anchor. To include anchors in your document:

1. Start by opening the anchor with the leading angle bracket and the anchor directive followed by a space: `<a`
2. Specify the document that's being pointed to by giving the parameter `href="filename.html"` followed by a closing angle bracket: >
3. Enter the text that will serve as the hypertext link in the current document (i.e., the text that will be in a different color and/or underlined)
4. Enter the ending anchor tag: `</A>`

Below is an sample hypertext reference:

```
<a href="MaineStats.html">Maine</a>
```

This entry makes ''Maine'' the hyperlink to the document `MaineStats.html`.

Uniform Resource Locator

A Uniform Resource Locator (URL) refers to the format used by WWW documents to locate other files. A URL gives the type of resource being accessed (e.g., gopher, WAIS) and the path of the file. The format used is:

scheme://*host.domain*[:*port*]/*path*/*filename*

where scheme is one of:

`file`
: a file on your local system, or a file on an anonymous ftp server

`http`
: a file on a World Wide Web server

`gopher`
: a file on a Gopher server

`WAIS`
: a file on a WAIS server

The scheme can also be `news` or `telnet`, but these are used much less often than the above. The port number can generally be omitted from the URL.

For example if you wanted to insert a link to this primer, you would insert

```
<A HREF="http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html">
NCSA's HTML Primer</A>
```

into your document. This would make the text ''NCSA's HTML Primer'' a hyperlink leading to this document.

Refer to the Addressing document prepared by CERN for additional information about URLs. *A Beginner's Guide to URLs* is located on the NCSA Mosaic Help menu.

Anchors to Specific Sections in Other Documents

Anchors can also be used to move to a particular section in a document. Suppose you wish to set a link from document A and a specific section in document B. First you need to set up what is called a named anchor in document B. For example, to add an anchor named ''Jabberwocky" to document B, you would insert

```
Here's <A NAME="Jabberwocky">some text</a>.
```

Now when you create the link in document A, you include not only the filename, but also the named anchor, separated by a hash mark(''#''):

```
  This is my <A HREF="documentB.html#Jabberwocky">link</a>.
```

Now clicking on the word ''link'' in document A would send the reader directly to the words ''some text'' in document B.

Anchors to Specific Sections within the Current Document

The technique is exactly the same except the file name is now omitted.

Note: The NCSA Mosaic **Back** button does not work for an anchor within a document because the **Back** button is designed to move to a previous document. Move back manually within the document using the scroll bar. (The **Back** button will return to the start of a hyperlink effective with Version 2.0 of NCSA Mosaic.)

Additional markup tags

The above is sufficient to produce simple HTML documents. For more complex documents, HTML also has tags for several types of lists, extended quotes, character formatting and other items, all described below.

Lists

HTML supports unnumbered, numbered, and descriptive lists. For list items, no paragraph separator is required. The tags for the items in the list terminate each list item.

Unnumbered Lists

1. Start with an opening list `<ul>` tag.
2. Enter the `<li>` tag followed by the individual item. (Remember that no closing tag is needed.)
3. End with a closing list `</ul>` tag.

Below an example two–item list:

```
<UL>
<LI> apples
<LI> bananas
</UL>
```

The output is:

- apples
- bananas

Note that different viewers display an unordered list differently. A viewer might use bullets, filled circles, or dashes to show the items.

Numbered Lists

A numbered list (also called an ordered list, from where the abbreviation comes) uses the `<ol>` directive to start a list rather than the `<ul>` directive. The items are tagged using the same `<li>` tag as for a bulleted list. For example:

```
<OL>
<LI> oranges
<LI> peaches
<LI> grapes
</OL>
```

The list looks like this online:

1. oranges
2. peaches

3. grapes

Descriptive Lists

A description list usually consists of alternating a description title (abbreviated as dt) and a description description (abbreviated as dd). The description generally starts on a new line, because the viewer allows the full line width for the contents of the dt field.

Below is an example description list as included in your source file:

```
<DL>
<DT> National Center for Supercomputing Applications
<DD> NCSA is located on the campus of the University
of Illinois at Urbana-Champaign. NCSA is a one of
four member institutions in the National Metacenter for
Computational Science and Engineering.
<DT> Cornell Theory Center
<DD> CTC is located on the campus of Cornell
University in Ithaca, New York. CTC is another member
of the National Metacenter for  Computational Science
and Engineering.
</DL>
```

The output looks like this:

National Center for Supercomputing Applications
: NCSA is located on the campus of the University of Illinois at Urbana–Champaign. NCSA is a one of four member institutions in the National Metacenter for Computational Science and Engineering.

Cornell Theory Center
: CTC is located on the campus of Cornell University in Ithaca, New York. CTC is another member of the National Metacenter for Computational Science and Engineering.

The `<DT>` and `<DD>` entries can contain multiple paragraphs (separated by paragraph tags), lists, or other description information.

Nested Lists

Lists can be arbitrarily nested. A list item can itself contain lists. You can also have a number of paragraphs, each themselves containing nested lists, in a single list item, and so on.

Remember that the display of an unordered list varies with the viewer. A browser may not provide successive levels of indentation or modify the bullets used at each level.

NCSA Mosaic indents the second level in the following list and *changes the ''bullet'' from a bullet to a small box.*

An example nested list:

```
<UL>
<LI> A few New England states:
  <UL>
  <LI> Vermont
  <LI> New Hampshire
  </UL>
<li> One Midwestern state:
  <UL>
  <LI> Michigan
  </UL>
</UL>
```

The nested list is displayed as

- A few New England states:
  - Vermont
  - New Hampshire
- One Midwestern state:
  - Michigan

Preformatted Text

Use the `pre` tag (which stands for ''preformatted'') to include text in a fixed−width font and to cause spaces, new lines, and tabs to be significant. This is useful for program listings. For example, the following lines in your source file:

```
<PRE>
  #!/bin/csh
  cd $SCR
  cfs get mysrc.f:mycfsdir/mysrc.f
  cfs get myinfile:mycfsdir/myinfile
  fc -02 -o mya.out mysrc.f
  mya.out
  cfs save myoutfile:mycfsdir/myoutfile
  rm *
</PRE>
```

display as:

```
  #!/bin/csh
  cd $SCR
  cfs get mysrc.f:mycfsdir/mysrc.f
  cfs get myinfile:mycfsdir/myinfile
  fc -02 -o mya.out mysrc.f
  mya.out
  cfs save myoutfile:mycfsdir/myoutfile
  rm *
```

Hypertext references (and other HTML tags) can be used within <pre>

sections.

Extended quotes

Use the `<blockquote>` and `</blockquote>` tags to include quotations in a separate block on the screen.

For example

```
<blockquote>
Let us not wallow in the valley of despair. I say to you, my
friends, we have the difficulties of today and tomorrow. <P>

I still have a dream. It is a dream deeply rooted in the
American dream. <P>

I have a dream that one day this nation will rise up and
live out the true meaning of its creed. We hold these truths
to be self-evident that all men are created equal. <P>
</blockquote>
```

The result is

> Let us not wallow in the valley of despair. I say to you, my friends, we have the difficulties of today and tomorrow.
>
> I still have a dream. It is a dream deeply rooted in the American dream.
>
> I have a dream that one day this nation will rise up and live out the true meaning of its creed. We hold these truths to be self–evident that all men are created equal.

Addresses

The `<ADDRESS>` tag is generally used within HTML documents to specify the author of a document and provides a means of contacting the author (e.g., an email address). This is usually the last item in a file and generally starts on a new, left–justified line.

For example, the last part of the HTML file for this primer is

```
<ADDRESS>
A Beginner's Guide to HTML / NCSA / pubs@ncsa.uiuc.edu
</ADDRESS>
```

The result is:

A Beginner's Guide to HTML / NCSA / pubs@ncsa.uiuc.edu

Character formatting

Individual words or sentences can be put in special styles. Logical styles are those that are configured by your viewer. For example, `<CITE>` may be defined as italic by your viewer. Each time you enter `<CITE>` tags, the viewer automatically displays the text in italics.

A physical style is one that you determine, and the viewer displays what you have coded. For example `<I>` tells the viewer to display your text in italics.

For HTML–coded documents, you should use logical styles whenever possible. Future implementations of HTML may not implement physical styles at all.

- Italic
  - `<I>text</i>` puts text in italics (*HTML Primer*)
  - `<em>text</em>` also italicizes text (only *one* viewer)
  - `<cite>text</cite>` is used for citations of names of manuals, sections, or books (*HTML Primer*)
  - `<var>text</var>` indicates a variable (*filename*)
- Bold
  - `<b>text</b>` puts text in bold (**Important**)
  - `<strong>text</strong>` also emphasizes text (**Note:**)
- Fixed width font
  - `<tt>text</tt>` puts text in a fixed–width font (`1 SU = 1 CPU hour`)
  - `<code>text</code>` also puts text in a fixed–width font (`1 SU = 1 CPU hour`)
  - `<samp>text</samp>` formats text for samples (`–la`)
  - `<kbd>text</kbd>` displays the names of keys on the keyboard (`HELP`)
- Other (the following special tag currently does not display in NCSA Mosaic)
  - `<dfn>text</dfn>` displays a definition in italics

Special Characters

Three characters out of the entire ASCII (or ISO 8859) character set are *special* and cannot be used ‘‘as–is’’ within an HTML document. These characters are left angle bracket (<), right angle bracket (>), and ampersand (`&`).

The angle brackets are used to specify HTML tags (as shown above), while the ampersand is used as the escape mechanism for these and other characters:

- `<` is the *escape sequence* for <
- `>` is the escape sequence for >
- `&` is the escape sequence for &

Note that ‘‘escape sequence’’ means that the given sequence of characters represents the single character in an HTML document and that the semicolon is required. The conversion to the single character itself takes place when the document is formatted for display by a reader.

There are additional escape sequences, such as a whole set of sequences to support 8−bit character sets (ISO 8859−1). For example:

- `ö` is the escape sequence for a lowercase `o` with an umlaut: ö
- `ñ` is the escape sequence for a lowercase `n` with an tilde: ñ
- `È` is the escape sequence for an uppercase `E` with a grave mark: È

Many such escapes exist and are available in a listing from CERN.

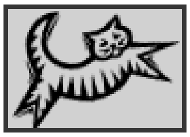
Inline Images

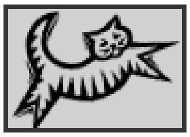
NCSA Mosaic is can display X Bitmap (XBM) or GIF format images inside documents. *Each image takes time to process and slows down the initial display of the document*. Using a particular image multiple times in a document causes very little performance degradation compared to using the image only once.

NOTE: The `<img>` tag is an HTML extension first implemented in NCSA Mosaic. Currently it is not understood by most other World Wide Web browsers.

To include an inline image in your document, enter:

```
<IMG SRC="filename.GIF">
```

 By default the bottom of an image is aligned with the text as shown in this paragraph.

 Include the `align=top` parameter if you want the viewer to align adjacent text with the top of the image as shown in this paragraph. The full inline image tag with the top alignment is:

```
<IMG ALIGN=top SRC="filename.GIF">
```

If you have a larger image (i.e., one that fills most of your screen), you should insert an end of paragraph tag (<p>) before inserting the image parameter. End with another paragraph tag. (Or you might want to have the image open a new window, which is explained below.)

External Images

You may want to have an image open as a separate document when a user activates a link on either a word or a smaller version of the image that you have inlined into your document. This is considered an external image and is particularly useful because (assuming you use a word for your hypertext link) you do not have any processing time degradation in the main document. Even if you include a small image in your document as the hyperlink to the larger image, the processing time for the ‘‘postage stamp’’ image is less than the full image.

To include a reference to a graphic in an external document, use

```
<A HREF = "filename.gif">link anchor</A>
```

Make certain the image is in GIF, TIFF, JPEG, RGB, or HDF format.

Troubleshooting

- While certain HTML constructs can be nested (for example, you can have an anchor within a header), they cannot be overlapped. For example, the following is *invalid* HTML:

  ```
  <h1>This is <a name="foo">invalid HTML.</h1></a>
  ```

  Because many current HTML parsers aren’t very good at handling invalid HTML, avoid overlapping constructs.

- In NCSA Mosaic, when an <img> tag points at an image that does not exist or cannot be otherwise obtained from whatever server is supposed to be serving it, the NCSA logo is substituted. For example, entering `<img href="DoesNotExist.gif">` (where "DoesNotExist.gif" is a nonexistent file) causes the following to be displayed:

  If this happens to you, first make sure that the referenced image does in fact exist and that the hyperlink has the correct information in the link entry. Next verify that the file permission is set appropriately (world–readable).

A Longer Example

Here is a longer example of a HTML document:

```
<TITLE>A Longer Example</TITLE>
<H1>A Longer Example</H1>

This is a simple HTML document. This is the first
paragraph. <P>

This is the second paragraph, which shows special effects.  This is a
word in <I>italics</I>.  This is a word in <B>bold</B>.
Here is an inlined GIF image: <IMG SRC="myimage.gif">.
<p>

This is the third paragraph, which demonstrates links.  Here is
a hypertext link from the word <A HREF="subdir/myfile.html">foo</A>
to a document called "subdir/myfile.html". (If you
try to follow this link, you will get an error screen.) <P>

<H2>A second-level header</H2>

Here is a section of text that should display as a
fixed-width font: <P>

<PRE>
    On the stiff twig up there
    Hunches a wet black rook
    Arranging and rearranging its feathers in the rain ...
</PRE>

This is a unordered list with two items: <P>

<UL>
<LI> cranberries
<LI> blueberries
</UL>

This is the end of my example document. <P>

<address>Me (me@mycomputer.univ.edu)</address>
```

Click here to see the formatted version.

For More Information

More information on HTML is available through the following hyperlinks.

- HTML Quick Reference Guide, which gives a comprehensive listing of HTML codes
- the official HTML specs
- the in−development HTML RFC (Request for Comments)
- a description of SGML, the Standard Generalized Markup Language
- the URL (Uniform Resource Locator) specification
- the style guide for online hypertext document structures