Spinning the Web:

A Hands-On Introduction to Building Mosaic and WWW Documents

by James Powell
Scholarly Communications Project
(http://scholar.lib.vt.edu/)
University Libraries
(http://vatech.lib.vt.edu/)
Virginia Polytechnic Institute and State University
(http://www.vt.edu/)

Table of Contents

Introduction to HTML

The **HyperText Markup Language** (HTML) is an SGML application for marking up documents for inclusion in the World Wide Web. It, along with the World Wide Web were invented by Tim Berners-Lee of the CERN High Energy Particle Physics Laboratory in Geneva, Switzerland. HTML allows you to:

Publish documents to the Internet in a platform independent format

Create links to related works from your document

Include graphics and multimedia data with your document

Link to non-World Wide Web information resources on the Internet

What is Markup?

To use HTML, you must understand the concept of markup:

• **Markup** is the act of inserting additional text into a document which is not usually visible to the reader, and is not part of the content, but enhances the document in some way, such as capturing document structure or adding hypertext capability.
• **Markup** also refers to the additional text, also known as tags, which are inserted in the document.

An example of markup:

```
Grocery list
<UL>
<LI>Apples
<LI>Oranges
</UL>
```

HTML Markup

HTML tags are used to markup the structure of a document, as well as some formatting information. HTML has two types of markup: **tags** and **character entities**

Tags are constructed of brackets between which the tag is placed. Tags are placed around segments of text, so there is usually a companion end tag which is identical to the start tag except it includes a forward slash. Here are start and end tags for a title:

```
<TITLE>Introduction to HTML</TITLE>
```

HTML also includes markup called **character entities** . These are used to include international characters as well as characters usually included in tags as markup. Here is a character entity for an ampersand:

```
&amp;
```

Components of HTML Markup

Here is a second example of HTML markup:

```
<A HREF="Virginia.gif">More about Virginia</A>
```

- **<A** is the anchor tag (tags are also referred to as **elements**).
- **HREF="Virginia"** is an attribute of the anchor tag
- **>** closes the anchor tag
- The phrase **More about Virginia** is the tag's contents
- **</A>** is an end tag for the anchor tag.

Tags can have elements, which are only allowed between them. For example, all HTML tags are elements of the <HTML> tag.

Some tags lack contents, attributes or end tags.

HTML Levels

There are currently three levels of HTML conformance. Each encompasses a set of tags and higher levels include tags from all those below it.

Level 0

The minimum tags which constitute an HTML document (most tags currently in use). Level 0 tags are usually rendered consistently from browser to browser.

Level 1

Level 0 tags plus tags for highlighting (also called Logical Tags) and images

Level 2

Level 0 and Level 1 tags plus form tags

Most browsers support all Level 0 tags covered in this presentation. Some level 1 tags such as <EM> for emphasis are not widely supported. Only very recent releases of browsers such as 2.0 Mosaic support Level 2 form tags.

Document-wide tags

HTML, HEAD, BODY

Each HTML document is contained within the **<HTML>** tag. This tag is not required by all browsers but using it is good practice and may be required in the future.

Each HTML document also includes a header section indicated by the **<HEAD>** tag which contains information about the document. It should always be present and at least contain the **<TITLE>** with the document title.

The remainder of the HTML document should be enclosed by the **<BODY>** tag.

The minimal HTML document:

```
<HTML>
<HEAD> <TITLE> Minimal </TITLE> </HEAD>
<BODY>
</BODY> </HTML>
```

HEAD Tags

The following tags are elements of the <HEAD> tag and cannot appear outside the header section of the document:

<TITLE>

This should be a unique and descriptive title for the document. While it is not displayed with the document, many browsers display the title elsewhere and use it when constructing hotlist entries.

```
<TITLE>Introduction to HTML</TITLE>
```
<BASE>

Sets the URL for relative links within the current document. This tag is not required but can be useful when a document includes many local hypertext links.

```
<BASE HREF="/reports/annual-1994/">
```
<ISINDEX>

Indicates a document is keyword searchable - this is usually an unnecessary tag as documents are considered searchable by default.

```
<ISINDEX>
```

HEAD Tags

More tags used within the document header

<NEXTID>

This is usually generated by automated markup systems as a portion of anchors within this particular document. There is no need to insert this tag manually.
```
<NEXTID X=A10>
```
<LINK>

A Level 1 HEAD element that can be used to indicate relationships between documents such as next and previous documents, glossaries, related indexes, document author. Multiple <LINK> tags can be used to include multiple relationships:

```
<LINK HREF="mailto:jpowell@vt.edu">
<LINK HREF="glossary.html">
```

<!-- --> (comment tags)

Comment tags may actually occur anywhere in an HTML document but are most commonly used in the document header to encode miscellaneous information about a document, such as version, creation date, last update, etc.
```
<!-- Version 2.1, last updated August 22, 1994 -->
```

Document Body tags
Headings

HTML provides six levels of **heading tags (<H1>-<H6>)** . All levels are rendered larger than surrounding text, with the lowest level rendered at the largest size. Heading tags also insert vertical space between the text marked and the surrounding text.

You should avoid skipping levels to achieve a certain visual effect as there is no guarantee how the headings will be rendered on any given system. They are intended to indicate **document structure.**

Here is a markup example along with tagged examples for all six levels:
<H1>Level 1</H1>

Level 1

Level 2

Level 3

Level 4

Level 5

Level 6

Information Type Elements

Information type elements are used to markup information content in a document. There are tags for highlighting sections of text, for definitions and citations, and computer oriented data such as program source code and output. Since these are level 1 conformant tags, appearance is not guaranteed to be the same from browser to browser.

Highlighting Elements

• <EM> indicates this portion of text should be emphasized (usually italicized)

8

• <STRONG> indicates stronger emphasis than <EM> (usually bold)

General Text Elements
• <ADDRESS> is used to record information that can be used to contact the document   author.
• <DFN> is used to markup a definition **(proposed tag - not yet supported)**
• <CITATION> is used to markup a citation from another document
• <STRIKE> indicates that the selected text should be struck out of the document **(proposed - not widely supported)**

The following tags are intended for computer related information:
• <CODE> is used to markup sections of program code
• <SAMPLE> indicates the data is output from a computer application
• <KBD> represents the key a user should enter when responding to a computer prompt
• <VAR> marks a variable

Information Type Examples

Each example includes segments of text marked up by the tag listed:

```
    <EM>Emphasized text</EM> :
```
Emphasized text

```
    <STRONG>Strong emphasis</STRONG> :
```
Strong emphasis

```
    <CITATION> HTML 2.0 Specification</CITE> :
```
HTML 2.0 Specification

```
    <DFN> HTML stands for HyperText Markup Language</DFN> :
```
HTML stands for HyperText Markup Language

```
    <STRIKE>completely unacceptable</STRIKE> :
```
completely unacceptable

```
    <CODE>for (i=0; i<10; i++) {printf("Count = %d",i); } </CODE> :
```
for (i=0; i<10; i++) { printf("Count = %d",i); }

```
    <SAMPLE>Count = 1 Count = 2</SAMPLE> :
```
Count = 1 Count = 2

```
    Enter <KBD>Q</KBD> to quit :
```
Enter Q to quit

```
    <VAR>Percent</VAR> :
```
Percent

Most browsers do not render all of these tags differently than surrounding text, although the specification indicates most should do so. It is a good idea to use these tags as needed regardless of whether they are displayed differently, since they enhance the accessibility of the document content.

Physical Style Elements

HTML has a few tags for controlling the appearance of text. There are three text markup elements for controlling the font of a text segment (a fourth is proposed), and three tags for controlling the flow of text.

Font Tags:
These tags are placed around segments of text for the desired effect. A few browsers do not support the <TT> tag and none currently support the proposed <U> tag for underlining text.

<B> will display text in **BOLD**
<I> will display text in *Italics*
<TT> will display text in `a fixed typewriter-like font`
<U> is proposed for displaying text underlined

Text Flow Tags:
Text flow tags should be inserted between sections of text to be combined or divided.
Text flow tags do not have end tags.

 forces a line break
this text is on a new line
<NOBSP> will force a long line to stay together rather than wrap
<HR> will display a rule line between segments of text , like this one:

Text Elements

Text elements, like Information Type Elements are inserted around segments of text as structural elements. There are three HTML tags for text elements:

<P> is intended to mark paragraphs. The paragraph end tag is optional, although it is good practice to insert it.

<PRE> is the preformatted text tag. There are a variety of uses for the preformatted text tag. It can be used to retain formatted ASCII text such as newsletters, calendars, spreadsheets or statistical data in columns (there is not yet an HTML tag for tables). Here is a brief example:

```
            World Wide Web Stats
                  January    February
Document 1             752         697
Document 2             134         232
```

<BLOCKQUOTE> has as its content sections of text included from other sources:

The World Wide Web Initiative (W3) links information throughout the world.

Lists - Unordered, Ordered

HTML supports five types of lists. A list is first marked with the start and end list tag and then each list item is indicated with a list item tag <LI> (unless it is a definition list). List item tags have end tags, but they are optional since a new list item tag implies the end of the previous item. Lists may be nested and if they are nested lists are indented farther than their parent list when displayed.

Unordered List: <UL>
A list of items which may appear in any particular order. It is usually displayed as a bulleted list of items.
<UL>
<LI>Apples   • Apples
<LI>Oranges   • Oranges
</UL>

Ordered List: <OL>
A list of items to be displayed in a particular order. These are usually numbered when displayed.
<OL>

```
<LI>Apples                              1 Apples
<LI>Oranges                             2 Oranges
</OL>
```

Lists - Definition List

A **definition list <DL>** is a list of terms <DT> and their definitions <DD>. Each definition is usually displayed indented slightly in relation to the term. Each term should have a corresponding definition.
```
<DL>
<DT>HTML
<DD>HyperText Markup Language
<DT>SGML
<DD>Standard Generalized Markup Language
</DL>
```

HTML

   HyperText Markup Language

SGML

   Standard Generalized Markup Language

Lists - Menu, Directory

A **directory list <DIR>** is a list of brief items, less than twenty characters long each.

```
<DIR>
<LI>jte-v1n1<LI>jte-v1n2
<LI>jte-v2n1<LI>jte-v2n2
</DIR>
```

jte-v1n1
jte-v1n2
jte-v2n1
jte-v2n2

A **menu list <MENU>** is a compact list of items, usually one per line.
```
<MENU>
<LI>Press 1 for help
```

```
<LI>Press 2 for menu options
<LI>Press 3 to quit
</MENU>
```

Press 1 for help
Press 2 for menu options
Press 3 to quit

Review: Structural Elements

Document-wide

```
HTML – all tags are elements of the <HTML> tag:
Level 0          HEAD (TITLE, BASE, ISINDEX, NEXTID)
                 <!-- -->
                 BODY
Level 1          HEAD (LINK)
```

Document-body

```
BODY – all tags not elements of <HEAD> are elements of the <BODY> tag:
Level 0          H1, H2, H3, H4, H5, H6
                 B, I, TT, U, BR, NOBSP, HR
                 P, PRE, BLOCKQUOTE
                 UL, OL, MENU, DIR (LI)
                 DL (DT, DD)
                 ADDRESS

Level 1          EM, STRONG, CITATION, STRIKE, CODE,
                 SAMPLE, KBD, VAR
```

Adding Hypertext...
HTML Anchor Tag

Anchors are what make HTML a **hypertext language** . The anchor tag consists of a start tag **<A** plus one or more attributes naming or describing the anchor plus **>** then content which becomes the link, followed by an end tag **</A>** :

<A HREF="slide16.html">HTML Anchor Tag</A>
HTML Anchor Tag

There are basically two types of anchors: **start** and **destination** . Start anchors are selectable segments of text such as *HTML Anchor Tag* above, while destination anchors are portions of text that mark an available destination. Here is an example of a start and a destination in which the start example is pointed to by the destination example:

```
Start:          <A NAME="SGML">Standard Generalized Markup Language</A>
Destination:    <A HREF="#SGML">SGML</A>
```

Building Links to HTML Files

Links are built using the **HREF** attribute with the anchor tag. HREF must be assigned some value, a target value for a destination. The target can be within the same document, another document on the same server (filename), a document on a different server (URL), or a portion of text in another document.

```
<A HREF="#SGML">More information about SGML</A>
```
provides a link to a target named SGML in the current document.

```
<A HREF="http://scholar.lib.vt.edu/html-intro.html">More information about
HTML</A>
```
links to a document called html-intro.html on the World Wide Web.

```
<A HREF="http://scholar.lib.vt.edu/html-intro.html#SGML">More information
about SGML</A>
```
links to a target named SGML in a file called html-intro.html on the World Wide Web.

Finally, an anchor can be both a link and a target:

```
<A NAME="SGML2" HREF="#SGML">if you are still lost</A>
```
can be pointed to with the name **SGML2** and points to a target called **SGML** .

Any value assigned to an attribute must be enclosed by double quotes.

Using Uniform Resource Locators

Anchors can link to remote data when a **Uniform Resource Locator (URL)** is used with the **HREF** attribute. Any type of data on almost any type of information server on the Internet can be accessed using a URL. The URL has three main components:

• **Server/Resource Type**
> • file (File Transfer Protocol)
> • gopher
> • http (World Wide Web)
> • news (Usenet News)
> • telnet

• **Internet Name** (and port if required)
• **Filename and path**

Most URLs include the characters :// to divide the server type from the internet address, except for Usenet news URLs. Filename can be truncated to a forward slash / which tells the server to send a default document or directory listing. Telnet does not require a filename or path.
Here are example URLs for several common resource types :

```
file://scholar.lib.vt.edu/pub/next/HTML-Editor.FAT.compressed
gopher://vatech.lib.vt.edu/
http://scholar.lib.vt.edu/library.html
news:comp.infosystems.www.providers
telnet://vtls.vt.edu
```

Summary of Anchor Attributes

HREF

HREF makes an anchor a hypertext link. It is used with an anchor name, filename, or Uniform Resource Locator.

NAME

NAME assigns an identifier to a piece of text which can then be linked to by other anchors.

TITLE

TITLE is an optional attribute which is supported by some browsers. It specifies that the contents of the <TITLE> tag in the HEAD of a target document can be displayed before a user loads the entire document.

Other attributes less commonly used include **URN** (Uniform Resource Number), **METHODS** , and the proposed **REL** and **REV** . Uniform Resource Numbers will become more common as the URN standard is finalized, and may replace URL in the future.

Images

Images can be included with HTML documents using the **<IMG>** tag. Images can be icons, small images of characters HTML cannot support, or photographs. The linked image must be in one of two graphics formats:

• Xbitmap (**XBM**)
• Compuserve's Graphics Interchange Format (**GIF**)

Many tools are available for Macintosh and PC platforms for converting from other graphics formats to those supported by WWW browsers.

IMG has four attributes:

SRC is a required attribute that is assigned the filename or URL of the image to be linked.

```
<IMG SRC="icon.gif">
```

ALIGN can be set to top, middle or bottom and indicates how text following a graphic should be aligned with the image.

```
<IMG SRC="newman.gif" ALIGN="middle">Newman Library
```

ALT specifies text data to be displayed instead of the graphic if the image cannot be displayed.

```
<IMG SRC="warning.gif" ALT="Warning!">
```
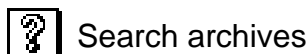
ISMAP is used to make an image a graphical navigation tool. More about this soon.

Image Examples

Here is an icon:
```
<IMG SRC="index.xbm">Search archives
```

 Search archives

Here is an image of non-supported characters:

```
Mathematical symbols <IMG SRC="math.gif"> are not supported by HTML
```

Mathematical symbols $\sum_{i=1}^{n} x_i = \int_{0}^{1} $ are not supported by HTML

Here is a photograph:

```
<IMG SRC="newman.gif" ALT="Newman Library">
```


Images can also be hypertext links:

```
<A HREF="http://vatech.lib.vt.edu/"><IMG SRC="newman.gif"></A>
```


Special Characters

HTML supports the **ISO-Latin-1** character set, a larger character set than ASCII. ISO-Latin 1 characters are represented as character entities in HTML. These are prefixed by an ampersand and followed by a semicolon. Here is an example for the less than sign:
Capital A with acute accent (â) is represented by Á

Characters which are used to construct tags and entities must also be represented as entities:

```
Character                       Entity
----------------------------------------
Less-than sign (<)              &lt;
Greater-than sign (>)           &gt;
Ampersand (&)                   &amp;
Double quote (")                &quot
```

A complete list of ISO-Latin-1 characters and entities is included in your manual.

Review: Anchors, Images Character Entities

Document-body

```
BODY - all tags not elements of <HEAD> are elements of the <BODY> tag:
Level 0          A [HREF, NAME, TITLE, URN, METHODS, REL, REV]
                 &amp; &lt; &gt; &quot; ISO-Latin-1 character entities

Level 1          IMG [SRC, ALIGN, ALT, ISMAP]
```

Special Applications of HTML: Forms

Form support is a recent addition to HTML. Forms can be constructed from five level 2 HTML tags:

• FORM
• INPUT
• OPTION
• SELECT
• TEXTAREA

They provide a user with the ability to enter information which can then be processed on the server as survey information, search information for a database, information request, etc.

Forms by themselves only allow data entry. They require software commonly referred to as gateways, which receive the data, process it, and return a response to the WWW client. Gateways are custom applications written in an available programming language on the server (such as Perl or C). Some gateways are available on the Internet for Z39.50, WAIS, and mail forwarding.

Form Tags - FORM

The **<FORM>** tag is placed around a section of an HTML document which includes FORM elements. Other BODY tags can occur in a form, and multiple forms can occur in a document, but forms cannot be nested.

There are two attributes essential to forms:
ACTION indicates the URL of the processing gateway. This URL will point to a program rather than a document. This program will receive the contents of the form in one of two ways depending on what value is specified for the **METHOD** attribute.

METHOD can be assigned one of two values: GET or POST. Gateways can accept data directly when METHOD is GET or look for it in a special variable if POST is used. If you are using an existing gateway, refer to its documentation for the correct METHOD.

Example:
<FORM METHOD=GET ACTION="http://nebula.lib.vt.edu:8001/cgi-bin/marian-gate">
Sends the contents of a form directly to a gateway called marian-gate.

Form elements should not occur outside <FORM> start and end tags.

Form Tags - INPUT

The **<INPUT>** tag, as its name implies, provides some type of data entry point in the form depending on the value of its **TYPE** attribute:

```
CHECKBOX  and RADIO  specify selectable options:

<INPUT TYPE="CHECKBOX">    <INPUT TYPE="RADIO">

RESET  and SUBMIT  make the INPUT field a button to clear the form or
send its contents to the gateway:

<INPUT TYPE="RESET">  RESET  .<INPUT TYPE="SUBMIT">  . SUBMIT

HIDDEN  is used to conceal a field that has a preset value that will
never change but is unknown to the gateway.

TEXT  specifies that a data input field be displayed:

<INPUT TYPE="TEXT">

IMAGE  displays an image and when the user selects a spot, the
coordinates are passed to the gateway.
```

INPUT is an empty tag, like IMG, with no end tag.

Form Tags - More INPUT Attributes

Each INPUT field must have a **NAME** attribute which is associated with the input value. For example, if you were building an OPAC searching form and had two fields:

```
Enter a Title:

Enter an Author:

<INPUT TYPE=TEXT NAME="Title">
<INPUT TYPE=TEXT NAME="Author">
```
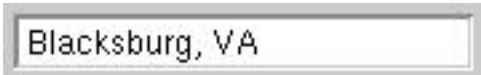
would pair whatever the user enters in the text field with the name Author, and the Title data with the name title, the gateway could construct an author-title search from the

input.

The **VALUE** attribute will set a default value:

```
<INPUT TYPE="TEXT" VALUE="Blacksburg, VA">
```

Blacksburg, VA

Form Tags - Still More INPUT Attributes...

Other INPUT attributes include:

ALIGN and **SRC** used in conjunction with the IMAGE attribute.

MAXLENGTH for limiting the number of characters that can be entered in a TEXT field and **SIZE** for controlling the displayed length of a TEXT field:

```
<INPUT TYPE="TEXT" MAXLENGTH=5 SIZE=5>
```

CHECKED to select a default for a checkbox or set of radio buttons:

```
<INPUT TYPE="CHECKBOX" CHECKED>
```

```
<INPUT TYPE="RADIO"> <INPUT TYPE="RADIO" CHECKED>
```

Despite its complexity, INPUT's role is strictly data acquisition

Form Tags - SELECT, OPTION

SELECT is the form equivalent of a list tag. It is used in conjunction with the **OPTION** tag to build a pull down list of options:

```
<SELECT NAME="Catalog">
<OPTION>Virginia Tech
<OPTION>University of Virginia
<OPTION>Radford University
</SELECT>
```

Virginia Tech

SELECT has a **MUTIPLE** attribute when users are allowed to choose more than one item.

OPTION has two attributes: **SELECTED** to configure a default item which is initially selected, and **VALUE** to specify a value to be returned when a certain option is chosen (otherwise the item label is returned).

Form Tags - TEXTAREA

TEXTAREA allows the user to enter multiple lines of text. It has three attributes:

NAME to associate the field data with some identifier, and **ROWS** and **COLS** to specify the size of the text area:

```
<TEXTAREA NAME="Address" ROWS=4 COLS=60>
University Libraries
Blacksburg, VA
                 24060
</TEXTAREA>
```

The data between the start and end TEXTAREA tag is optional, but the end tag is required even if no default value is specified.

HTML Form Example

This form from our library home page serves as a Z39.50 client for searching several locally mounted databases:

```
<HTML><HEAD>
<TITLE>INFOSHARE Page </TITLE>
</HEAD>
<BODY>
<H1>INFOSHARE Database Query Page</H1>
<IMG SRC="index.xbm"> Search University Libraries INFOSHARE databases using
Z39.50.
<P>
<HR>
<FORM METHOD=POST
ACTION="http://nebula.lib.vt.edu:8001/cgi-bin/infogate">
<H3>     Search       </H3>
NOTE:  Words <B>must</B> be filled in sequential order.<P>
Database: <SELECT NAME="dbname">
```

```
<OPTION SELECTed>WHUM Humanities Index
<OPTION>WGSI General Science
<OPTION>WSSI Social Science
<OPTION>WBAI Biology and Agriculture
<OPTION>WAST Applied Science and Technology
<OPTION>WRGA Reader's Guide Abstracts
<OPTION>WWBA Wilson Business Abstracts
<OPTION>XWIL Test Database
<OPTION>CCON Current Contents
</SELECT><P>
Records to View: <INPUT NAME="maxrecords" SIZE=4 VALUE=100>
Full or Brief displays:
<SELECT NAME="esn">
<OPTION SELECTED>B
<OPTION>F
</SELECT>
<P>
Word 1: <INPUT NAME="term_term_1">
<SELECT NAME="term_use_1">
<OPTION SELECTED>Author
<OPTION>Title
<OPTION>Subject
<OPTION>Word
</SELECT>
<P>
Word 2: <INPUT NAME="term_term_2">
<SELECT NAME="term_use_2">
<OPTION SELECTED>Subject
<OPTION>Word
<OPTION>Title
<OPTION>Author
</SELECT>
<P>
Word 3: <INPUT NAME="term_term_3">
<SELECT NAME="term_use_3">
<OPTION SELECTED>Title
<OPTION>Subject
<OPTION>Word
<OPTION>Author
</SELECT>
<P>
<INPUT TYPE="radio" NAME="operator" VALUE="and"> AND
<INPUT TYPE="radio" NAME="operator" VALUE="or"> OR
<INPUT TYPE="radio" NAME="operator" VALUE="not"> NOT
<P>
<HR>
<INPUT TYPE="reset" VALUE="Clear">
<INPUT TYPE="submit" VALUE= "Do Search">
</FORM>
</BODY>
</HTML>
```

Special Applications of HTML: Image Mappings

The **ISMAP** attribute of the **IMG** tag is a highly specialized anchor for graphical navigation. Any GIF image can serve as a graphical navigator. Coordinates for each selectable portion of the image are recorded using a graphics application. The two pair represent the upper left and lower right section of the rectangle (other shapes are possible). A map file is then created with a URL associated with each set of coordinates. You can also specify a default URL for any click detected within the image but outside the coordinate pairs.

Here is an ISMAP example:

```
<A HREF="http://vatech.lib.vt.edu/cgi-bin/imagemap/library-2"><IMG
SRC="floor2.gif" ISMAP></A>
```

Example uses include selectable floor maps for libraries, weather maps allowing the user to click on their region, shelf of manuals with selectable spines, and learning exercises such as an image map for anatomy.

ISMAP Map File Example

Here is an image map file and an example HTML document which uses it:

```
default gopher://scholar.lib.vt.edu/

# Link to Spectrum search
rect gopher://scholar.lib.vt.edu:70/0/about 7,23,512,109

# Link to VPI home page
rect http://www.vt.edu/ 517,47,598,88

# Link to first article - page 1
rect gopher://scholar.lib.vt.edu:70/0/spectrum/sp940512/spt1a.051294.txt
11,125,597,351

# Link to second article - lower left page 1
rect gopher://scholar.lib.vt.edu:70/0/spectrum/sp940512/spt1b.051294.txt
13,354,149,776

# Link to third article - middle bottom page 1
rect gopher://scholar.lib.vt.edu:70/0/spectrum/sp940512/spt1c.051294.txt
157,547,457,775
```

```
# Link to fourth article - lower right page 1
rect gopher://scholar.lib.vt.edu:70/0/spectrum/sp940512/spt1d.051294.txt
455,352,600,687
```

```
<HTML><HEAD><TITLE>Spectrum Page Images</TITLE><HEAD>
<BODY>
<H1>Spectrum Page Images</H1>
<I>Click on a portion of the page to see the text</I><P>
<A HREF=http://scholar.lib.vt.edu:80/cgi-bin/imagemap/spectrum><IMG
SRC="spectrum1.gif" ISMAP></A>
</BODY>
</HTML>
```

HTML Tag Summary

Document-wide

```
HTML - all tags are elements of the <HTML> tag:
Level 0            HEAD (TITLE, BASE, ISINDEX, NEXTID)
                  <!-- -->
                  BODY
Level 1           HEAD (LINK)
```

Document-body

```
BODY - all tags not elements of <HEAD> are elements of the <BODY> tag:
Level 0           H1, H2, H3, H4, H5, H6
                  B, I, TT, U, BR, NOBSP, HR
                  P, PRE, BLOCKQUOTE
                  UL, OL, MENU, DIR (LI)
                  DL (DT, DD)
                  ADDRESS
                  A [HREF, NAME, TITLE, URN, METHODS, REL, REV]
                  &amp; &lt; &gt; &quot; ISO-Latin-1 character entities


Level 1           EM, STRONG, CITATION, STRIKE, CODE,
                  SAMPLE, KBD, VAR
                  IMG [SRC, ALIGN, ALT, ISMAP]

Level 2           FORM [ACTION, METHOD]
                  INPUT [TYPE, NAME, VALUE, ALIGN, SRC, MAXLENGTH
                        SIZE, CHECKED]
                  SELECT [NAME, MULTIPLE] (OPTION [SELECTED, VALUE])
                  TEXTAREA [NAME, ROWS, COLS]
```

Netscape Extensions Overview

New tags and attributes

Netscape extensions focus primarily on making documents look better. They are an *excellent* way to improve the appearance and impact of a home page in an increasingly Netscape-dominated Web. Netscape extensions **should be avoided** when publishing reports, articles, and other items where the textual content is more important. They could cause problems when HTML 3 is finalized.

Summary
• New tags include **NOBR, WBR, FONT, BASEFONT, CENTER, BLINK**
• Many other tags have new attributes to provide more control over the appearance of text and graphics.
• Tables are supported.
• Documents can include tiled image backgrounds and you can change the color of text on a document.

Netscape HEAD and BODY Changes

The *HEAD* of a document remains essentially unchanged. The only element affected is the ISINDEX tag. It now has a **PROMPT** attribute which can be assigned a string of text that will be displayed next to the text input field normally displayed by ISINDEX:
`<HEAD><ISINDEX PROMPT="Enter a year and month ">`
The PROMPT attribute for ISINDEX is part of HTML 3.0 as well.

-
The *BODY* tag has several new attributes:

BACKGROUND causes Netscape to retrieve an image and display it tiled as a background for the current HTML document. It should be assigned a URL for a GIF image e.g.
`<BODY BACKGROUND="bricks.gif">`
or the background can be set to a different color (see Color Guide):
`<BODY BGCOLOR="#ffee00">`

Changing the background to an image can make text difficult to read so the other Netscape BODY attributes control text color:

BODY text color attributes

TEXT
Set the default text color: `<BODY TEXT="#ccff00">`

LINK
Set the default hypertext anchor color: `<BODY LINK="#00eeff">`

VLINK, ALINK
Set visited or active (selected) link color

Only BACKGROUND is currently supported in the latest HTML 3 specification.

Netscape Physical Style Elements

Netscape supports several tags and attributes that enhance author control over the appearance of the text in a document.

BLINK tag:
The <BLINK> tag can be used around text just like the bold or italics tag. It causes the marked text to flash continually.

CENTER tag:
The <CENTER> tag causes blocks of text to be centered. It can be used to center headings, paragraphs, tables, even entire documents. The following line uses the center tag:
This text is centered in the Netscape window

FONT tag:
The <FONT> tag sets can be used to change the font size used to display a portion of a document. It is always paired with a SIZE attribute which can be assigned a value from 1-7 where 3 is the default. The value can also have a plus or minus to indicate a relative size change from the default:
```
This text is <FONT SIZE=+3>large</FONT> while this text is <FONT
SIZE=-2>small</FONT>
```
This text is large while this text is small

BASEFONT tag:
The <BASEFONT> tag changes the base font used to determine the font size for all text in the document. It is always paired with a SIZE attribute of the range 1-7 where 3 is the default: `<BASEFONT SIZE=4>`

None of these tags are part of the current HTML 3 Specification.

More Physical Style Elements

HR tag:
Netscape supports several attributes that control how the horizontal rule is displayed. **SIZE** sets the thickness of the line. **WIDTH** sets the width of the line in characters or as a percentage of the current Netscape window width. **ALIGN** can be set to left, right or center to control justification of the rule. **NOSHADE** turns off the default shading. Here

are a couple of examples:

```
<HR SIZE=5>
```

-

```
<HR WIDTH=50% ALIGN=right>
```

-

```
<HR NOSHADE WIDTH=20 ALIGN=center>
```

-

BR tag:
The break tag has a **CLEAR** attribute which ensures that the next line will be completely left justified (necessary to compensate for Netscape floating images, around which text normally wraps).

NOBR tag:
The no break tag ensures that the marked text will not be wrapped.

WBR tag: The word break tag indicates that a word can be broken if needed, but does not cause a break like the **BR**eak tag.

None of these tags/attributes are part of the current HTML 3 Specification.

Netscape Textual Elements and Character Entities

UL, OL, LI tags:
Unordered and Ordered lists now have a **TYPE** attribute that controls the labeling of list items. Unordered lists types are *disc, circle* or *square*. The following example uses `<UL TYPE=square>`:

• Apples
• Oranges

Ordered list types are A for capital letters, a for small letters, I for large roman numerals, i for small roman numerals or 1 for numbers (the default). This example uses `<OL TYPE=I>` for roman numerals:

1. Markup
2. HTML Tags

LI tag:
The list item tag has a TYPE attribute as well. This allows you to change the type within the list. It also has a value attribute that allows you to change the numbering of ordered lists within the list:
`<LI TYPE=circle>` changes the bullet for this item and all items that follow to a circle.

`<LI TYPE=a VALUE=4>` changes the numbering type for this ordered list item to lower case letters and starts a new numbering with the letter d.

New Character Entities:
© © Copyright
® ® Registered Trademark

Netscape and Images

IMG tag extensions:
If the IMG **ALIGN** attribute is set to *top, texttop, middle, absmiddle, baseline, bottom,* or *absbottom* then one line of text following the image will be aligned with the image according to the value selected. If ALIGN is set to *left* or *right* then the image is aligned with the left or right margin and the document text flows around the image.

WIDTH, HEIGHT
With Netscape, you can use these IMG attributes if you know the width and height of an image in pixels. It will speed up image loading and document display.

BORDER
The border attribute places a border around an image. Assign it a value to control the thickness of the border.

VSPACE, HSPACE
These attributes control the amount of clear space around a floating (ALIGN=left, right) image.

Here is one example that incorporates several Netscape attributes:
<IMG SRC="newman.gif" ALIGN=left BORDER=2>

Here is some text that inserted merely to demonstrate text flow around an image in Netscape. You may need to resize your Netscape window to make it narrow enough to cause the text to wrap.

HTML+

What is HTML+?

HTML+ is a proposed extension to HTML that has been discussed for well over a year on the Internet as a Request For Comment (RFC) document. It is a whole-sale rewrite of HTML which replaces many tags and adds some new tags to focus more on structure.

In reality, some HTML+ features such as forms have, by necessity, crept into HTML to become level 2.

HTML+ will become **HTML level 3**. Netscape supports some HTML 3 tags.

New BODY Elements in HTML+

A new document prologue is required:

```
<DOCTYPE html public "-//IETF//DTD W3 HTML 3.0//EN">
```

HTML documents currently have a prologue that looks like this:

```
<DOCTYPE html public "-//W3O//DTD W3 HTML 2.0//EN">
```

which has been considered optional.

<SP>
Non-breaking space tag

<ICON>
Inlined graphics or icons

<Q>, <ABBREV>...
Many new logical emphasis tags

<EMBED>
Embedded data in an external format

<INPUT>
Input field exactly as implemented in HTML level 2

<FIGD>
Figure tag, similar to IMG

HTML+ Logical Emphasis Tags

- **<Q>** marks a short quotation
- **<CITE>** is a citation
- **<PERSON>** for a person
- **<ACRONYM>** marks acronyms
- **<ABBREV>** indicates an abbreviation
- **<CMD>** for a computer command

added to KBD, VAR, DFN, CODE and SAMP from HTML level 1.

HTML+ Paragraph Styles

Several new paragraph styles are introduced with HTML+. These can contain multiple paragraphs <P>:

<QUOTE>
Longer quotations which may span several paragraphs

<ABSTRACT>
Document overview

<BYLINE>
Author information

<NOTE> with attribute **ROLE** (TIP, NOTE, WARNING, ERROR)
Special information, suggestions, warnings

HTML+ Tables

HTML+ provides much-needed support for tables with the <TABLE> tag and its elements:

<CAPTION>
Table title tag

<TH>
Table header tag

<TD>
Table data cells

<TR>

Row divider

Table headers and cells can have **ROWSPAN** and **COLSPAN** attributes which can expand a cell or header to any number of cell rows or columns. The **ALIGN** attribute can be set to LEFT, CENTER or RIGHT to position cell data.

Here is an example from an HTML 3 standards document:

```
<TABLE BORDER>
<CAPTION> An Example of a Table</CAPTION>
<TH ROWSPAN=2 TH COLSPAN="2"> average <TH> <TH> <TH>other <TR>
<TH> height <TH> weight <TH> category <TR>
<TH ALIGN=LEFT colspan=2> males <TD> 1.9 <TD> .003 <TD> yyy <TR>
<TH ALIGN=LEFT colspan=2> females <TD> 1.7 <TD> .002 <TD> xxx
</TABLE>
```

HTML+ Mathematical Equation Support

A whole collection of SGML tags for mathematical equations are also proposed for HTML+. The **MATH** tag is used to markup a mathematical formula. Tags such as **SUP, SUB, BOX, ARRAY** and a set of character entities are defined. HTML math support is similar to TeX.

Here is an example:

```
<math>
     H(s) = &int;<sub>0</sub><sup>&infin;</sup> e<sup>-st</sup>
h(t) dt
</math>
```

Here is how this might be rendered:

$$\sum_{i=1}^{n} x_i = \int_0^1 \rfloor$$

References

A Beginner's Guide to HTML

by NCSA (pubs@ncsa.uiuc.edu

http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html

A Beginner's Guide to URLs

by Marc Andraesson

http://www.ncsa.uiuc.edu/demoweb/url-primer.html

Crash Course on writing documents for the Web

by Eamonn Sullivan

http://www.ziff.com/~eamonn/crash_course.html

Composing Good HTML

by James "Eric" Tilton

http://www.willamette.edu/html-composition/strict-html.html

HTML Quick Reference

by Michael Grobe

http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLQuickRef.html

HTML Specification

by Daniel W. Connolly

http://www.hal.com/users/connolly/html-spec/

HTML+ Discussion Document

by David Raggett

http://www.w3.org/hypertext/WWW/MarkUp/HTMLPlus/htmlplus_1.html

ISO Latin 1 character entities

derived from ISO 8879:1986//ENTITIES Added Latin 1//EN

http://www.w3.org/hypertext/WWW/MarkUp/ISOlat1.html

NCSA HTML Style Sheet

by NCSA (pubs@ncsa.uiuc.edu)

http://www.ncsa.uiuc.edu/Pubs/StyleSheet/NCSAStyleSheet.html

Appendix A:
ISO Latin 1 character entities

```
&AElig;              capital AE diphthong (ligature)
&Aacute;             capital A, acute accent
&Acirc;              capital A, circumflex accent
&Agrave;             capital A, grave accent
&Aring;              capital A, ring
&Atilde;             capital A, tilde
&Auml;               capital A, dieresis or umlaut mark
&Ccedil;             capital C, cedilla
&ETH;                capital Eth, Icelandic
&Eacute;             capital E, acute accent
&Ecirc;              capital E, circumflex accent
&Egrave;             capital E, grave accent
&Euml;               capital E, dieresis or umlaut mark
&Iacute;             capital I, acute accent
&Icirc;              capital I, circumflex accent
&Igrave;             capital I, grave accent
&Iuml;               capital I, dieresis or umlaut mark
&Ntilde;             capital N, tilde
&Oacute;             capital O, acute accent
&Ocirc;              capital O, circumflex accent
&Ograve;             capital O, grave accent
&Oslash;             capital O, slash
&Otilde;             capital O, tilde
&Ouml;               capital O, dieresis or umlaut mark
&THORN;              capital THORN, Icelandic
&Uacute;             capital U, acute accent
&Ucirc;              capital U, circumflex accent
&Ugrave;             capital U, grave accent
&Uuml;               capital U, dieresis or umlaut mark
&Yacute;             capital Y, acute accent
&aacute;             small a, acute accent
&acirc;              small a, circumflex accent
&aelig;              small ae diphthong (ligature)
&agrave;             small a, grave accent
&aring;              small a, ring
&atilde;             small a, tilde
&auml;               small a, dieresis or umlaut mark
&ccedil;             small c, cedilla
&eacute;             small e, acute accent
&ecirc;              small e, circumflex accent
&egrave;             small e, grave accent
&eth;                small eth, Icelandic
&euml;               small e, dieresis or umlaut mark
&iacute;             small i, acute accent
&icirc;              small i, circumflex accent
&igrave;             small i, grave accent
&iuml;               small i, dieresis or umlaut mark
&ntilde;             small n, tilde
&oacute;             small o, acute accent
&ocirc;              small o, circumflex accent
&ograve;             small o, grave accent
&oslash;             small o, slash
&otilde;             small o, tilde
&ouml;               small o, dieresis or umlaut mark
&szlig;              small sharp s, German (sz ligature)
&thorn;              small thorn, Icelandic
&uacute;             small u, acute accent
&ucirc;              small u, circumflex accent
&ugrave;             small u, grave accent
&uuml;               small u, dieresis or umlaut mark
&yacute;             small y, acute accent
&yuml;               small y, dieresis or umlaut mark
```

Appendix B:

Netscape/RGB Color Guide

Netscape extensions allow you to change the color of the document background, text and anchors. Colors are assigned with a pound sign followed by three sets of two digit hexadecimal numbers (00 to ff) to specify shades of red, green and blue which are combined to create the color you actually see e.g.

```
<BODY BACKGROUND="#C0C0C0">
```
sets the background color to gray, the default for most web browsers.

Common colors and their RGB codes:

```
Red        #FF0000
Orange     #FF7F00
Yellow     #FFFF00
Green      #00FF00
Cyan       #00FFFF
Blue       #0000FF
Purple     #DB70DB
Black      #000000
White      #FFFFFF
```

Try other combinations to get the hang of blending red, green and blue for new colors: