

SAS® and Web Publishing -- Wisdom for the Web Challenged

Sally Muller, UNC, Chapel Hill, NC

Caroline Bahler, Biopop, Radford, Va

Arturo Barrios, UNC, Chapel Hill, NC

David Doolittle, Biopop, Radford, Va

Abstract

If you've attended a SUGI conference or visited the SAS® Institute Web site during the last several years, you know that the use of the SAS World-Wide Web publishing tools has exploded in popularity. This hands-on class is meant for the people who feel "web challenged" and who have NO or VERY LITTLE knowledge of either HTML or the SAS Web publishing tools. By the end of the class every participant will have a working knowledge of HTML and the SAS Web publishing tools and feel comfortable enough to write their own Web page using these tools.

Introduction

We would argue that the Web is one of the best, if not the best, tool available in the information technology toolbox today. Therefore, making efficient use of the web has become a high priority for anyone in research, education, or business. Of the document preparation tools available, SAS's set of web publishing tools are some of the easiest and most powerful available.

Although this paper assumes that you know what the Internet is and why it is popular, we will discuss a few basic concepts:

- *URL* -- Uniform Resource Locator, the address or location of a page on the web. A URL has the following format:

service_type://host-name/path/file.html

Where: service_type is the transfer protocol,

for example: http = hypertext transfer protocol

ftp = file transfer protocol

host-name is the web server

path is the path to web server

file.html is the optional actual file name

An example is <http://www.sas.com/rnd/web>, which accesses SAS's macros that speed HTML document publishing.

- *Browser* -- the computer software, such as Netscape® or Microsoft Internet Explorer®, that allows you to view the web pages.
- *Thin client* -- computer with small resources and a browser on it to access web pages.

- *Web Master* -- person who maintains a web server. This person is an extremely valuable resource on how your web server works. The web server is where your document will reside.
- *HyperText Markup Language (HTML)* – The language interpreted by the browser. A HTML document is only text, what it does contain are references to graphic images, video clips or sound clips that allow the browser to access these when a document is displayed². HTML is constantly evolving under the auspices of the W3C. The W3C is an international industry consortium charged with the development of a common set of standards. These standards are then used to enhance HTML and browser functionality. For more information go to <http://www.w3.org>, which is the official W3C web page.

Note: Another common term for a HTML document is a web page.

- *Hypertext link* – is the facility that allows you to move within a HTML document, between web pages residing at the same web site or between web pages residing at different web sites.

The whole concept of a hypertext document system was first proposed in 1989 by Tim Berners-Lee and Robert Carilliau as a way for physicists working within the High Energy Physics project to collaborate. The first working version of HTML was developed in 1990, by Tim Berners-Lee (he coined the term World Wide Web), in conjunction with the first internet browser. The addition of graphics to browsers arrived in 1993 and the commercial aspects of the web started in 1994¹.

Components of an HTML Document

The first step in creating a web page is to learn how a HTML document is published. The set of instructions used to tell internet browsers how to present the information in your document file is HTML. HTML's document formatting mechanism is comparable to the earliest word processors where commands or tags were embedded in the text document to tell a printer to bold or italicize a word or superscript the next set of letters. The major advantage of browser technology is HTML's ability to create hypertext links. This allows a user to access other HTML documents or different sections within a HTML document.

HTML consists of a set of elements and their attributes. Each element is defined within a HTML document by a set of tags in brackets, e.g. <> and </>. A <> tag indicates the start of an element and </> tag indicates that the element is finished. Elements have been defined so that they can be nested within one another and any HTML reference will list not only the attributes for each element but how that element can be nested. HTML is not case sensitive, any internet browser will recognize HTML tags no matter what case is used³.

Types of Elements Available:

- List - dictionary, ordered and unordered lists
- Text appearance - bold, italics, paragraphs, colors, pre-formatted, etc.
- Graphic Images - GIF or JPEG files that are displayed by a browser.
- Tables - formatting tool that allows the document developer to place text, anchors or graphic images in a specific cell.
- Anchors - hypertext links that utilize URL's to allow access to another HTML document. An anchor can be either a graphic image or text.
- Forms - a set of elements that allow document developers to collect information from document user. Example - SUGI registration on the web.
- Frames - this element divides the browser window into two (2) or more sections that can contain their own scroll bars. Each section of a frame can contain a different HTML document (see Advanced Tutorial paper - *SAS and HTML - HTML Publishing Using SAS*⁴).

Composition of an HTML Document

All HTML documents must start with <HTML> and end with </HTML> tags. Inside of the <HTML> tags a HTML document can be divided into following sections: the head and the body. The head section is defined by the <HEAD>.. </HEAD> tags and contains general document information. For example, the document title and whether or not the document is searchable. The body section is defined by the <BODY>..</BODY> tags and contains the text and graphic references that will be formatted by the browser. The <BODY> tag also can define the overall color of the page or specify a graphic image to use as the background. The only section required is the body section³. For example, a basic HTML document is represented below:

```
<HTML>
<BODY>
<P> HELLO WORLD </P>
</BODY>
</HTML>
```

This document contains all of the required elements and tells the browser to represent the text in paragraph form,

<p>. Any SAS data set or output can be converted into an HTML document by either utilizing a DATA _NULL_ step or the SAS web publishing tools. For example, to use the DATA _NULL_ step to create the HTML document above:

```
filename doc
'f:\home\cbahler\myfile.html';

data _null_;
file doc;
put @01 '<HTML>'
    /@01 '<BODY>'
    /@01 '<P> HELLO WORLD </P>'
    /@01 '</BODY>'
    /@01 '</HTML>';
run;
```

The other option is to use the SAS Web Publishing Tools.

SAS Web Publishing Tools

Since December of 1996, the SAS Institute has made available a portfolio of "Web Tools", at the URL - <http://www.sas.com/rnd/web/formatter>. These tools are simple and easy to use and create STATIC web pages. The documentation that accompanies these tools is superb and the section below is meant to acquaint the reader with what is available.

Overview:

The SAS Web Publishing tools are most beneficial if you fall within any of these categories: have data within a SAS data set; have output from a SAS procedure; or have graphics from SAS procedures that need to be presented within an Internet/Intranet application. To accomplish any of these tasks, the SAS Web Publishing tools offer the following advantages:

- allow creation of HTML documents in either batch or interactive mode without any knowledge of HTML;
- allows customization of HTML document output;
- allows paperless distribution of reports and other information on a regular basis;
- SAS Web publishing tools can be integrated into a full fledged Internet/Intranet application that allows the user to create their own reports (a web version of SAS/EIS[®] or SAS/AF[®])

Tools:

The SAS Web Publishing tools consist of a set of macros available from the web site listed above. They are:

1. **Output Formatter** – This tool takes all information contained within either the SAS output or log windows and allows you to save (capture) that information as a HTML document. All HTML formatting tags are added to the log or output

information at time of capture. The information from the log or output windows will be presented within the browser in exactly the same layout as it was presented in the above windows. (See web reference SAS 3.)

2. **Data Set Formatter** – This tool takes all information or a subset of that information from either a SAS data set or view and allows you to save that information as a HTML document. In comparison to the Output Formatter, the information selected will be presented as a HTML 3.x table when displayed by a browser. (See web reference SAS 4.)
3. **Tabulate Formatter** – This tool converts the output from only PROC TABULATE and saves that output as a HTML document. In comparison to the Output Formatter, the PROC TABULATE output will be presented as a HTML 3.x table when displayed by a browser. (See web reference SAS 5.)

Utilization:

All of the SAS Web Publishing Tools can be used in either a batch mode or interactively. The following are the typical steps used to create an HTML document:

1. Create the procedure output or SAS data set that you would like to publish on the web. This includes titles, footnotes and other peripheral information needed.
2. Decide whether or not the creation process can be interactive or needs to be batch.
3. Determine how the document should look. For example, what colors to use for the text and background or if a graphic image will be the background.
4. Determine where the document should be stored (if you have a web master it would be an excellent idea to consult that person).
5. Create the HTML document using the appropriate macro.

Examples - All examples were created utilizing the batch option within the macros. However, all of the HTML documents created below could have been created utilizing the interactive facility included in the SAS Web Publishing tool macros.

Output Formatter - Batch Example 1

```
data daylily;
input @01 ploid $10.
      @12 cultivar $50.;
cards;
Diploid    BITSY
Diploid    CHESTER CYCLONE
Diploid    OPEN HEARTH
Diploid    ORIENTAL RUBY
Diploid    PRAIRIE BLUE EYES
Diploid    PRAIRIE MOONLIGHT
Diploid    STELLA D'ORO
```

```
Tetraploid AMBER BALLERINA
Tetraploid BELOVED BALLERINA
Tetraploid CHICAGO RUBY
Tetraploid CHICAGO SUNRISE
Tetraploid HUDSON VALLEY
Tetraploid OLIVE BAILEY LANGDON
;
run;

proc sort;
  by ploid;
run;

*Create HTML Document - Output Formatter*;
%out2htm(capture=on, window=output,
         runmode=b);

options nocenter ls=80;
proc print noobs;
  title 'Daylily Cultivars';
  by ploid;
  id ploid;
run;

%out2htm(htmlfile=c:\sugi\output.html,
         capture=off, window=output,
         openmode=replace, runmode=b,
         ctext=blue, septype=none);
```

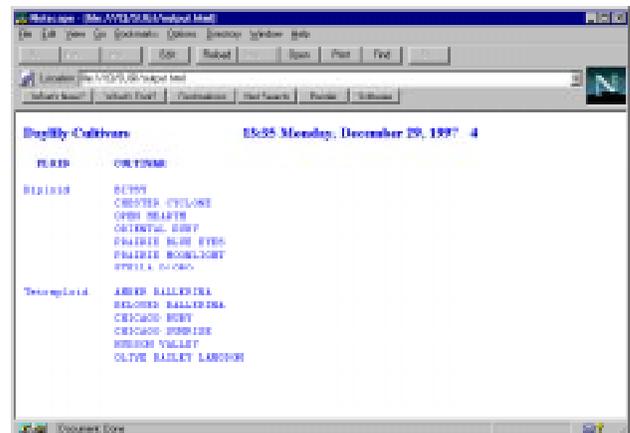


Fig 1. Output Formatter

Tabulate Formatter - Batch Example 2

```
%tab2htm(capture=on,
         runmode=b);

options nocenter ls=80 pageno=1;
proc tabulate data=sasuser.daylilys
FORMCHAR='82838485868788898a8b8c'x ;*
title 'Daylilys - Number of Cultivars
Produced by Hybridizers';
class hybridzr season;
var size;
table hybridzr*season, size*n /
      rtspace=30;
label hybridzr='Hybridizer'
      season='Season'
      size='Bloom Size';
run;

%tab2htm(htmlfile=c:\sugi\tabfmter.html,
         capture=off, openmode=replace,
         septype=none, runmode=b);
```

* Note - the formchar option is required see online documentation.

Daylilies - Number of Cultivars Produced by Hybridizers 1
15:35 Monday, December 28, 1997

		Hybridizer	Count
DELAC	MI	1	100
FRSH	MI	1	100
JARR	MI	1	100
LAMB	MI	1	100
SMAN	MI	4	100
SMIT	MI	3	100
PRFC	MI	1	100
WARR	MI	1	100

Fig 2. Tabulate Formatter

Data Set Formatter - Batch Example 3

```
%ds2htm(runmode=b,
        htmlfile=c:\sugi\dsfmrter.html,
        data=sasuser.daylilys,
        septype=none,
        caption=Daylily Information);
```

Daylily Information

CULTIVAR	HYBRIDER	YEAR	FLOOD	COLOR	SEED	SEASON	HT	TYPE	REBLOOM	FRAGRANT
ANDER BELLEFEL	MIH	1955	T	ANDER TAN BLEED, PEACH GOLD THEOCAT	4	EM	24	SEP	Y	Y
DELAC BELLEFEL	MIH	1955	T	PACRY PINK BLEED	4	EM	28	SEP	Y	Y
MTT	WAR	1963	D	LIMON YELLOW SELF	1.3	EE	58	SEP		
CHESTER CYCLONE	BLAC	1973	D	LIMON YELLOW SELF-ROSE THEOCAT	2.5	M	68	D	Y	Y
CHICAGO BOY	MAK	1977	T	RED SELF-CHERRY THEOCAT	4	M	27	SEP		

Fig 3. Data Set Formatter

Graphic Drivers

In addition, SAS graphs can be exported via graphic drivers to create web usable images from SAS/GRAPH[®] output and all are available with SAS 6.12 (see web reference SAS 6):

- GIF and IMGJPG drivers -- generates GIF (Graphics Interchange Format) files, and enables you to produce single GIF images. (See web reference SAS 7).
- GIF animation driver -- allows you to combine GIF images. (See web reference SAS 8).

- IMGJPEG driver -- enables you to create JPEG images. (See web reference SAS 9).

A good discussion of how to create and use graphic images can be found in these proceeding in the Advanced Tutorial section - *SAS and HTML - HTML Publishing Using SAS*⁴.

Synthesis - Linking it All Together

The main strength of HTML is the hypertext link, i.e. the ability to link documents together to a form whole. In web terminology this is known as a “web site”. A “web site” can contain one HTML document or multiple HTML documents linked together. In addition, different “web sites” can be accessed by placing links from one “web site” to another. Another term used is the “HOME” page of a web site, this is the first HTML document accessed when a web site URL is specified³.

How a set of documents are organized can be as important as what is contained within each document. Therefore, once the SAS Web Publishing Tools are used to create a set of documents the next step is to determine how the documents should be linked together. The same strategies used to create information systems utilizing SAS/AF or SAS/EIS can be utilized to solve this organizational problem. The problems are the same since in essence you are creating an information system using the Web.

The simplest organizational form is a table of contents where a single document is created to act as a menu and all of the other documents can be accessed from that document. In addition, linkages to other documents can be incorporated into those accessed documents when appropriate⁴.

To add a hypertext link to a document created using the SAS Web Publishing tools can be accomplished with the program below:

```
* Creates main menu page - more items can*
* be added by simply adding additional *
* anchor lines (<A> </A>). See Fig 4 *;
data _null_;
file 'c:\sugi\mainmenu.html';
put @01 '<HTML>'
    /@01 '<BODY>'
    /@01 '<H3> Main Menu </H3>'
    /@01 '<A href="myfile1.html"> Daylily
Cultivars </A>'
    /@01 '</BODY>'
    /@01 '</HTML>';
run;
*Create information page - in this case *
* utilizing the Output Formatter. *;
proc sort data=daylily;
by ploid;
run;
```

```

%out2htm(capture=on,
        window=output,
        runmode=b);

options nocenter ls=80;
proc print noobs data=daylily label
split='*';
  title 'Daylily Cultivars';
  by ploid;
  id ploid;
  var cultivar hybridzr season;
  label hybridzr='Hybridizer*-----'
        cultivar='Cultivar*-----'
        ploidy='Ploidy*-----'
        season='Season*-----';
run;

%out2htm(htmlfile=c:\sugi\output.html,
        capture=off,
        window=output,
        openmode=replace,
        runmode=b,
        ctext=red,
        septype=none);

```

```

filename doc 'c:\sugi\output.html';
filename doc1 'c:\sugi\myfile1.html';

* Modify document to contain link by *
* adding an anchor. See Fig.5        *;
data _null_;
  infile doc missover pad;
  input @01 a $char200.;

  file doc1;
  x =compress(a);

* Step 1 - Write out all but second to *
* last line of the original file.      *;
  if x ne '</BODY>' then
    put @01 a $char200.;

* Step 2 - To place anchor at end of the *
* document add a line of text containing *
* the anchor element before the second *
* to last line of the original file.    *;

  if x = '</BODY>' then
    put @01 '<a
href="mainmenu.html">Home</a>'
        /@01 a $char200.;
run;

```

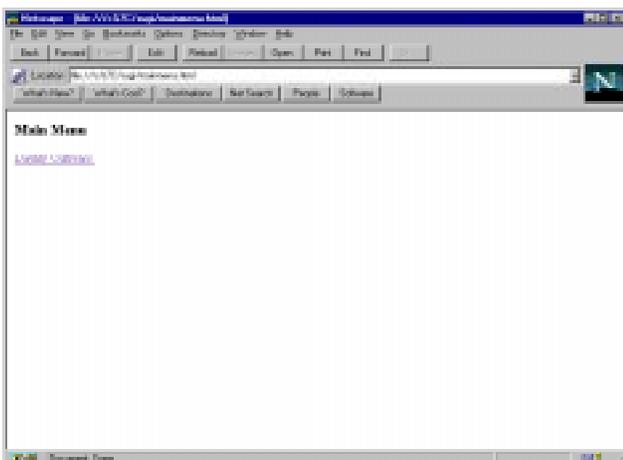


Fig. 4. Main Menu.

 A screenshot of a web browser window displaying a table titled "Daylily Cultivars" with a date "Monday, December 28, 1997". The table has four columns: "Ploidy", "Cultivar", "Hybridizer", and "Season". The data is as follows:

Ploidy	Cultivar	Hybridizer	Season
8	BETSY	BLB	8
	CRESTED CYCLONE	BLB	8
	ORCH BEASTS	LAB	8
	ORIENTAL SUN	PIB	8
	FRANCE BLUE EYES	BLB	8
	FRANCE ROUSLET	BLB	8
	STELLA D'ORO	BLB	8
8	JAMES BALLERINA	BLB	8
	BEAUCON BALLERINA	BLB	8
	CHICAGO PORT	BLB	8
	CHICAGO PRINCE	BLB	8
	EDISON VALLEY	POB	8
	COUNTY VALLEY LADYBIRD	BLB	8

Fig 5. HTML formatted output with addition of the link.

The source created for the main menu is:

```

<HTML>
<BODY>
<H3> Main Menu </H3>
<A href="myfile1.html">Daylily Cultivars
</A>
</BODY>
</HTML>

```

The main menu document could have been created using a simple word processor and saving the file as a text file. It was included in the program to show how a HTML can be created using the DATA _NULL_ step. The important point is that ALL documents that need to have tags interpreted by a browser MUST have the HTM or HTML extension as part of their name.

The above program creates a HTML document using the Output Formatter and then modifies that document by reading the document in using a Data _NULL_ step and writing out a new document with a link to the main menu or "HOME" page of the web site. The original HTML file created by the Output Formatter (example4) and the file after modification (example5) are below.

Note: that the SAS macro uses the <PRE> pre-formatted text format element which does not change the spacing of the PROC PRINT output. In addition, the SAS macro utilizes the head section of an HTML document to provided the following information:

1. A name for the document, in this case GENERATOR
2. A comment - that the page was created using a SAS HTML formatting tool.

Example 4. HTML Document created by the Output Formatter.

```

<HTML>
<HEAD>
  <META    NAME="GENERATOR"
          CONTENT="SAS Institute Inc. HTML Formatting Tools, http://www.sas.com/">
</HEAD>

<BODY TEXT=red>

<PRE><H3>Daylily Cultivars                                13:35 Monday, December 29, 1997   8
</H3></PRE>
<PRE><STRONG>Ploidy      Cultivar      Hybridizer      Season
-----      -----      -----      -----</STRONG></PRE>
<PRE>
D              BITSY              WAR              EE
              CHESTER CYCLONE     BLAK             M
              OPEN HEARTH         LAMB             M
              ORIENTAL RUBY       FISH             ML
              PRAIRIE BLUE EYES   MAR              M
              PRAIRIE MOONLIGHT   MAR              M
              STELLA D'ORO        JAB              EM

T              AMBER BALLERINA     MUN              EM
              BELOVED BALLERINA   MUN              EM
              CHICAGO RUBY        MAR              M
              CHICAGO SUNRISE     MAR              M
              HUDSON VALLEY       PECK             M
              OLIVE BAILEY LANGDON MUN              EM</PRE>

</BODY>
</HTML>

```

Example 5. HTML Document created by the Output Formatter with the addition of a link at the end.

```

<HTML>
<HEAD>
  <META    NAME="GENERATOR"
          CONTENT="SAS Institute Inc. HTML Formatting Tools, http://www.sas.com/">
</HEAD>

<BODY TEXT=red>

<PRE><H3>Daylily Cultivars                                13:35 Monday, December 29, 1997   8
</H3></PRE>
<PRE><STRONG>Ploidy      Cultivar      Hybridizer      Season
-----      -----      -----      -----</STRONG></PRE>
<PRE>
D              BITSY              WAR              EE
              CHESTER CYCLONE     BLAK             M
              OPEN HEARTH         LAMB             M
              ORIENTAL RUBY       FISH             ML
              PRAIRIE BLUE EYES   MAR              M
              PRAIRIE MOONLIGHT   MAR              M
              STELLA D'ORO        JAB              EM

T              AMBER BALLERINA     MUN              EM
              BELOVED BALLERINA   MUN              EM
              CHICAGO RUBY        MAR              M
              CHICAGO SUNRISE     MAR              M
              HUDSON VALLEY       PECK             M
              OLIVE BAILEY LANGDON MUN              EM</PRE>

<a href="mainmenu.html"> Home </a>
</BODY>
</HTML>

```

Summary

Overview of Web Conversion Process

Once you have determined that you need to publish information from SAS data sets, views, or procedure outputs on the web the following steps should make the conversion process easier:

1. Isolate the information that needs to be accessed on the web.
2. Determine where and how the information needs to be accessed on the web. (Is this going to be a stand-alone page or part of an information system?)
3. Determine which web publishing tool would be appropriate to use and create the HTML document.
4. Make what modifications are necessary.

Using the Web to Your Full Advantage

1. "Surf" the web to see how other document developers are handling similar problems. A good place to start is SAS's Technical Support Page of Web examples.
2. All browsers have under the VIEW button at the top of the screen a DOCUMENT SOURCE choice. If you click on this choice then a copy of the HTML source document that created the page will appear in a separate window. Therefore, if you like what you see the how is readily available.
3. The number of HTML reference pages available on the Web is staggering and increasing daily. A short list of some of the pages available is at the end of this paper.

Is this really necessary:

The reason that the web is exploding onto the information dissemination scene so quickly is the fact that a user friendly GUI application can be created easily by people with little or no programming experience. In addition, where you audience is located no longer matters as long as they can access the web the information is immediately available.

References

1. Wilson, Brian (1997), HTML Overview. <http://hepax6.rl.ac.uk/delphi/aye/blooberry/html/history/html.htm>.
2. Ray, Deborah S. and Ray, Eric J. (1997), HTML for Dummies, Quick Reference, Second Edition. Foster City, CA, IDG Books World Wide, Inc. pp. 10-11.
3. Graham, Ian S. (1996), HTML Sourcebook. A Complete Guide to HTML 3.0, Second Edition. New York, New York, John Wiley and Sons, Inc.
4. Bahler, Caroline; Muller, Sally; Doolittle, David; Barrios, Arturo (1998), SAS and HTML - HTML Publishing Using SAS. Proceedings of the Twenty

Third Annual SAS Users Group International Conference, 23.

SAS, SAS/GRAPH, SAS/AF, and SAS/EIS are registered trademarks or trademarks of SAS Institute, Inc. in the USA and other countries. Netscape is a registered trademarks or trademarks of Netscape, Inc. in the USA and other countries. Microsoft Internet Explorer is a registered trademarks or trademarks of Microsoft, Inc. in the USA and other countries. ® indicates USA registration.

Web References

Below is a listing of web sites with information on HTML, Java Script Language and other general information about publishing on the web. All of these references have some interesting and useful material and point to other web sites of interest.

SAS

1. SAS Home Page
<http://www.sas.com>
2. SAS Research and Development Web Tools (Going Places with SAS Tools)
<http://www.sas.com/rnd/web/sitemap.html>
3. Output Formatter
<http://www.sas.com/rnd/web/format/out/index.html>
4. Data Set Formatter
<http://www.sas.com/rnd/web/format/ds/index.html>
5. Tabulate Formatter
<http://www.sas.com/rnd/web/format/tab/index.html>
6. Creating Graphics for the Web
<http://www.sas.com/rnd/web/publish.html#graph>
7. GIF
<http://www.sas.com/rnd/web/driver/GIF/GIF.html>
8. Animated GIF
<http://www.sas.com/rnd/web/driver/GIFANIM/GifAnim.html>
9. IMGJPEG
<http://www.sas.com/rnd/web/driver/JPEG/Jpeg.html>

HTML

1. W3C Home Page
<http://www.w3.org>
2. *HTML Reference Guide*
<http://developer.netscape.com/library/documentation/htmlguid/index.htm>
3. *Advanced HTML*
<http://www.geocities.com/SiliconValley/Park/7476/advancedhtml.htm>
4. *Index Dot HTML: THE Advanced HTML Reference*
<http://hepax6.rl.ac.uk/delphi/adye/blooberry/html/index.html>
5. *Wilbur - HTML 3.2*
<http://www.flirble.org/alister/docs/html3/index.html>
6. *Alphabetic sorted HTML Tag Listing*
<http://hepax6.rl.ac.uk/delphi/adye/blooberry/html/alphabet.htm>
7. *How do they do that with HTML?*
<http://www.nashville.net/~carl/htmlguide/index.html>
8. *THE Netscape Frames Tutorial*
<http://www.newbie.net/frames/3ed/>
9. *The HTML Goodies Home Page - Thanks for stopping by...*
<http://www.htmlgoodies.com/>
10. *Quad's Ultimate HTML Site - Your one stop Web Resource, period.*
<http://www.quadzilla.com/>
11. *Welcome to The HTML Station*
<http://www.december.com/html/>
12. *Yale C/AIM Web Style Guide*
<http://info.med.yale.edu/caim/manual/>
13. *The Ten Commandments of HTML FAQ*
<http://www.visdesigns.com/design/commandments.html>
14. *Willcam's Comprehensive HTML Cross Reference - "The Tag List"*
<http://www.willcam.com/cmat/html/tags.html>
15. *Dynamic HTML*

<http://www.stars.com/Authoring/HTML/Dynamic/>

General Reference

16. *Netscape Gold Rush Tool Chest*
http://home.netscape.com/assist/net_sites/starter/samples/index.html
17. *Webreference.com (sm) - The Webmaster's Reference Library*
<http://www.webreference.com/>
18. *Web Review -- The Legion of DOM: Exploring the Document Object Model*
<http://webreview.com/>
19. *The Web Developer's Virtual Library*
<http://www.stars.com/>
20. *World Wide Web Tools and Guides for Authors*
<http://manuel.brad.ac.uk/WWW/authtool.html>