

**Chapter
2**

**HTML
DOCUMENT
STRUCTURE**

*Get on the
Fast Track!*



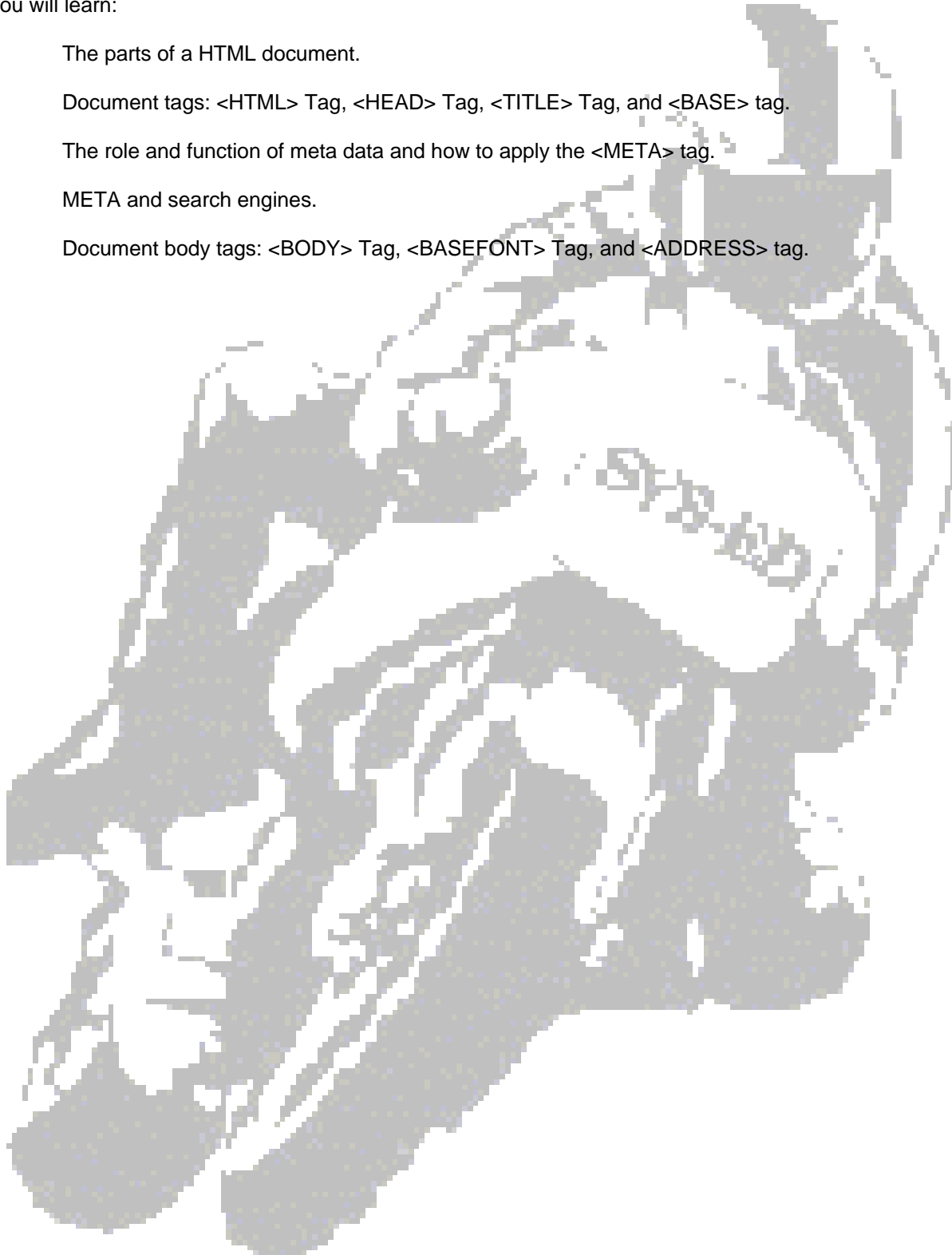
TM

**SYS-ED/
COMPUTER
EDUCATION
TECHNIQUES, INC.**

Objectives

You will learn:

- C The parts of a HTML document.
- C Document tags: <HTML> Tag, <HEAD> Tag, <TITLE> Tag, and <BASE> tag.
- C The role and function of meta data and how to apply the <META> tag.
- C META and search engines.
- C Document body tags: <BODY> Tag, <BASEFONT> Tag, and <ADDRESS> tag.



1 Parts of Document

An HTML document is composed of three parts:

- C A line containing HTML version information.
- C A declarative header section, which delimited by the HEAD tag.
- C A body, which contains the document's actual content. The body can be implemented by the BODY tag or the FRAMESET tag.

Example: A Simple HTML Document

```
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0//EN"  
2 "http://www.w3.org/TR/REC-html40/strict.dtd">  
3 <HTML>  
4 <HEAD>  
5 <TITLE>My first HTML document</TITLE>  
6 </HEAD>  
7 <BODY>  
8 <P>Hello world!  
9 </BODY>  
10 </HTML>
```

1.1 Coding Style

White space--such as spaces, newlines, tabs, and comments--may appear before or after each section. Sections should be delimited by the HTML tag.

Although indentation is not required, HTML authors should use spacing and indentation for clarity.

Comments should be used for inline documentation.

1.2 HTML Version Information

A valid HTML document declares the version of HTML to be used in the document. The document type declaration names the document type definition (DTD) in use for the document.

HTML specifies three DTDs, so authors must include one of the following document type declarations in their documents. The DTDs vary in the tags they support.

1. The HTML 4.0 Strict DTD includes all tags and attributes that have not been deprecated or do not appear in frameset documents.

For documents with this DTD, use this document type declaration:

```
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0//EN"  
2   "http://www.w3.org/TR/REC-html40/strict.dtd">
```

2. The HTML 4.0 Transitional DTD includes everything in the strict DTD plus deprecated tags and attributes, most of which concern visual presentation.

For documents with this DTD, use this document type declaration:

```
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN"  
2   "http://www.w3.org/TR/REC-html40/loose.dtd">
```

3. The HTML 4.0 Frameset DTD includes everything in the transitional DTD plus frames as well.

For documents with this DTD, use this document type declaration:

```
1 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Frameset//EN"  
2   "http://www.w3.org/TR/REC-html40/frameset.dtd">
```

2 <HTML> Tag

The <HTML> tag denotes the file as an HTML document.

Syntax:

```
<HTML>...</HTML>
```

This tag has no attributes.

Example:

```
1     <HTML>
2       <HEAD>
3         <TITLE>
4           .....
5         </TITLE>
6       </HEAD>
7
8       <BODY>
9         <P>This is an HTML document.
10      </BODY>
11     </HTML>
```

3 The Document Head

3.1 <HEAD> Tag

The HEAD tag contains information about the current document, such as its title, keywords that may be useful to search engines, and other data that is not considered document content.

User agents do not generally render tags that appear in the HEAD as content. They may, however, make information in the HEAD available to users through other mechanisms.

Syntax:

```
<HEAD>
```

The end tag is not required.

Example:

```
1 <HEAD>
2 <TITLE>A Simple Document</TITLE>
3 </HEAD>
```

3.2 <TITLE> Tag

Every HTML document should have a TITLE tag in the HEAD section. Authors should use the TITLE tag to identify the contents of a document.

Since users often consult documents out of context, authors should provide context-rich titles. Thus, instead of a title such as "Introduction", which doesn't provide much contextual background, authors should supply a title such as "Introduction to Medieval Bee-Keeping" instead.

This tag is valid only within the HEAD tag. The end tag is required.

Syntax:

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

Example:

```
1      <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0//EN"
2          "http://www.w3.org/TR/REC-html40/strict.dtd">
3      <HTML>
4      <HEAD>
5      <TITLE>A study of population dynamics</TITLE>
6      ... other head tags...
7      </HEAD>
8      <BODY>
9      ... document body...
10     </BODY>
11     </HTML>
```

For reasons of accessibility, user agents must always make the content of the TITLE tag available to users, including TITLE tags that occur in frames.

Titles may contain character entities (for accented characters, special characters, etc.), but no other markups are allowed.

3.3 The title Attribute

The title attribute contains advisory information about the tag for which it is set.

Unlike the TITLE tag, which provides information about an entire document and may only appear once, the title attribute may annotate any number of tags.

Values of the title attribute may be rendered by user agents in a variety of ways.

For instance, visual browsers frequently display the title as a "tool tip". A tool tip is a short message that appears when the pointing device pauses over an object.

```
1      ...some text...
2      Here's a photo of
3      <A href="http://someplace.com/neatstuff.gif" title="Me scuba diving">
4          me scuba diving last summer
5      </A>
6      ...some more text...
```

3.4 <BASE> Tag

The BASE tag must appear in the HEAD section of an HTML document before any tag referencing an external source.

The path information specified by the BASE tag only affects URLs in the document where the tag appears.

Syntax:

```
<BASE  
  HREF=url  
  TARGET=window>
```

Attribute	Purpose								
HREF=url	Specifies the document's full URL in case the document gets read out of context and the reader wants to refer to the original.								
TARGET=window	<p>Specifies to load all of the links on the page into the targeted window.</p> <p>This can be overridden by specifying a different target attribute for a specific link.</p> <p>Specifies to load the link into the targeted window. The window must begin with an alpha-numeric character to be valid, except for the following four target windows.</p> <table><tbody><tr><td>_blank</td><td>Specifies to load the link into a new blank window. This window is not named.</td></tr><tr><td>_parent</td><td>Specifies to load the link into the immediate parent of the document the link is in.</td></tr><tr><td>_self</td><td>Specifies to load the link into the same window the link was clicked in.</td></tr><tr><td>_top</td><td>Specifies to load the link into the full body of the window.</td></tr></tbody></table>	_blank	Specifies to load the link into a new blank window. This window is not named.	_parent	Specifies to load the link into the immediate parent of the document the link is in.	_self	Specifies to load the link into the same window the link was clicked in.	_top	Specifies to load the link into the full body of the window.
_blank	Specifies to load the link into a new blank window. This window is not named.								
_parent	Specifies to load the link into the immediate parent of the document the link is in.								
_self	Specifies to load the link into the same window the link was clicked in.								
_top	Specifies to load the link into the full body of the window.								

Examples:

```
1 <BASE HREF="http:// www.sample.com/hello.htm">  
2  
3 <BASE HREF="http:// www.sample.com/hello.htm" TARGET="viewer">
```

4 Meta Data

Meta data provides a form in which authors can specify information about a document rather than document content - in a variety of ways.

To specify the author of a document, use the META tag as follows:

```
<META name="Location" content="New York">
```

The META tag specifies a property (ie. "Location") and assigns a value to it (ie. "New York").

This specification does not define a set of legal meta data properties. The meaning of a property and the set of legal values for that property should be defined in a reference lexicon called a profile.

For example, a profile designed to help search engines index documents might define properties such as "author", "copyright", "keywords", etc.

4.1 <META> Tag

Meta tag is a general purpose tag typically used to store application-specific information in a HTML document.

Syntax:

```
<META attributes>
```

Attribute	Purpose
HTTP-EQUIV="REFRESH"	Causes a document to be automatically reloaded on a regular basis, specified in seconds.
CONTENT="n; URL=url"	Tells the browser to reload in n seconds. If URL=url is specified, the browser will load the URL after the time specified has elapsed. If no URL is specified, it will reload the current document.

4.2 META and HTTP Headers

The http-equiv attribute can be used in place of the name attribute and has a special significance when documents are retrieved via the Hypertext Transfer Protocol.

HTTP servers can use the property name specified by the http-equiv attribute to create a style header in the HTTP response.

Example: Sample META Declaration

```
1 <META http-equiv="Expires" content="Tue, 20 Aug 1996 14:25:27 GMT">
```

will result in the HTTP header:

```
1 Expires: Tue, 20 Aug 1996 14:25:27 GMT
```

This can be used by caches to determine when to fetch a fresh copy of the associated document.

Some user agents support the use of META to refresh the current page after a specified number of seconds, with the option of replacing it with a different URL.

```
<META http-equiv="refresh" content="3,http://www.acme.com/intro.html">
```

The content is a number specifying the delay in seconds, followed by the URL to load when the time is up.

This mechanism is generally used to show users a fleeting introductory page. However, since some user agents do not support this mechanism, authors should include content on the introductory page to allow users to navigate away from it.

4.3 META and Search Engines

META is typically used to specify keywords that a search engine can use for improving the quality of search results.

When several META tags provide language-dependent information about a document, search engines may filter on the lang attribute to display search results using the language preferences of the user.

Example:

```
1      <-- For speakers of US English -->
2      <META name="keywords" lang="en-us"
3          content="vacation, Greece, sunshine">
4      <-- For speakers of British English -->
5      <META name="keywords" lang="en"
6          content="holiday, Greece, sunshine">
```

The effectiveness of search engines can also be increased by using the LINK tag to specify:

- C Links to translations of the document in other languages.
- C Links to versions of the document in other media (e.g., PDF).
- C When the document is part of a collection, links to an appropriate starting point for browsing the collection.

Examples:

```
1      <HTML>
2      <HEAD>
3      <META HTTP-EQUIV="REFRESH" CONTENT=2>
4      <TITLE>Reload Document</TITLE>
5      </HEAD>
6      <BODY>
7      <P>This document will be reloaded every two seconds.
8      </BODY>
9      </HTML>
10
11     <HTML>
12     <HEAD>
13     <META HTTP-EQUIV="REFRESH" CONTENT="5;
14     URL=http://www.sample.com/next.htm">
15     <TITLE>Load Next Document</TITLE>
16     </HEAD>
17     <BODY>
18     <P>After five seconds have elapsed, the document
19     "http://www.sample.com/next.htm" will be loaded.
20     </BODY>
21     </HTML>
```

5 The Document Body

The body of a document contains the document's content. The content may be presented by a user agent in a variety of ways. For example, for visual browsers, the body can be thought of as a canvas where the content appears: text, images, colors, graphics, etc.

5.1 <BODY> Tag

The <BODY>Tag specifies the beginning and end of the document body. It also allows the background image, the background color, the link colors, and the top and left margins of the page to be set.

Syntax:

```
<BODY
BACKGROUND=url
BGPROPERTIES=FIXED
BGCOLOR=color
TEXT=color
LINK=color
VLINK=color>
```

Attribute	Purpose
BACKGROUND=url	Specifies a background picture. The picture is tiled behind the text and graphics on the page.
BGCOLOR=color	Sets the background color of the page. The color can be either a hexadecimal, red-green-blue color value or a predefined color name.
BGPROPERTIES=FIXED	Specifies a watermark, which is a background picture that does not scroll.
LEFTMARGIN=n	Specifies the left margin for the entire body of the page and overrides the default margin. Set to "0", the left margin will be exactly on the left edge.
LINK=color	Sets the color of hyperlinks that have not yet been visited. The color can be either a hexadecimal, red-green-blue color value or a predefined color name.
TEXT=color	Sets the color of text on the page. The color can be either a hexadecimal, red-green-blue color value or a predefined color name.
TOPMARGIN=n	Specifies the margin for the top of the page and overrides the default margin. Set to "0" the top margin will be on the precise top edge.
VLINK=color or colorname	Sets color of hyperlinks that have already been visited. The color can be either a hexadecimal, red-green-blue color value or a predefined color name.

Examples:

```

1      <BODY BACKGROUND="/ie/images/watermrk.gif" BGPROPERTIES=FIXED
2      BGCOLOR=#FFFFFF TEXT=#000000 LINK=#ff6600 VLINK=#330099>
3
4      <HTML>
5          <HEAD>
6              <TITLE>My First Webpage</TITLE>
7          </HEAD>
8
9          <BODY>
10             Here's a Web page!
11          </BODY>
12     </HTML>

```

Documents that contain framesets replace the BODY tag by the FRAMESET tag.

5.2 <BASEFONT> Tag

The BASEFONT tag sets the base font size using the size attribute.

Font size changes achieved with FONT are relative to the base font size set by BASEFONT. If BASEFONT is not used, the default base font size is 3.

The base font size does not apply to headings, except where these are modified using the FONT tag with a relative font size change.

Syntax:

```
<BASEFONT
SIZE=n>
```

Attribute	Purpose
SIZE=n	<p>Sets the base font size. The n can be between 1 and 7 inclusive; default is 3; 7 is largest.</p> <p>Throughout the document, relative font size settings (e.g.) are set according to this.</p>

Example:

```

1      <BASEFONT SIZE=3> This sets the base font size to 3.
2      <FONT SIZE=+4> Now the font size is 7.
        <FONT SIZE=-1> Now the font size is 2.

```

6 <ADDRESS> Tag

The ADDRESS tag may be used by authors to supply contact information for a document or a major part of a document such as a form.

This tag typically is used at the bottom of a document. Text is displayed in italics.

Syntax:

```
<ADDRESS>...</ADDRESS>
```

Example:

```
1 <ADDRESS>  
2   SYS-ED  
3   19 W 34th Street  
4   New York, NY 10001  
5   www.sysed.com  
6 </ADDRESS>
```