

HTML Basics

What is HTML?

HTML stands for HyperText Markup Language. HTML is based on SGML (the Standard Generalized Markup Language), a more encompassing document-processing system. HTML, by virtue of its SGML heritage, is a standardized language for describing structured documents. A HTML document can be read on any computer as long as you have a browser written for that computer that can interpret HTML.

Most documents have common elements, such as, titles, paragraphs, and lists. These elements are identified in HTML through the use of tags. Everything that is not a HTML tag is part of the document content. With a few exceptions, HTML does not allow you to describe the exact appearance or layout of a document. The designers of HTML did this on purpose. Why? Because you don't know the capabilities of the computer that is going to display the document, i.e. the size of the screen, the fonts that are installed, etc. By separating the structure of a document and its appearance, a browser program that reads and interprets HTML can make formatting decisions based on the capabilities of the end user's computer.

When you retrieve a HTML document with a browser such as Netscape Navigator, Mosaic or Lynx, the browser reads, or parses, the HTML information and formats the text and images on the screen. If you use different browsers, you may notice that the same document may appear differently in each browser (yes, on the same computer). The headings may be centered in one browser, or in a larger font. Therefore, it is important that you do not design your documents based on what they look like in one browser. Focus instead on providing clear, well-structured content that is easy to read and understand. The more complicated (with all the bells and whistles) a document is, the less likely users will be able to view it correctly, because of the limitations of various browsers.

A Note on HTML Extensions

In addition to using standard HTML, companies that make browsers, such as Netscape, develop extensions to the language. The extensions are tags to allow web page authors to get around some of HTML's current limitations, including centered text, aligning text next to images, and varying font sizes.

The problem with extensions, of course, is that they are not part of the HTML standard. This makes pages that include extensions strange looking or even unreadable in browsers that can not interpret the extensions.

Netscape browsers are readily downloadable free for students and educators and are widely used at Illinois State University. For these reasons, Netscape extensions will be included throughout this tutorial.

What Does HTML Look Like?

The following is an example HTML document that displays the first paragraph of this document:

```
<HTML>
<HEAD>
<TITLE> ISU ACS Dept HTML Tutorial-Basics </TITLE>
</HEAD>
<BODY BGCOLOR="#FFFFFF" TEXT="#000000">
<H3>What is HTML?</H3>
<P>
HTML stands for <B>H</B>yper<B>T</B>ext <B>M</B>arkup
<B>L</B>anguage. HTML is based on SGML (the Standard Generalized
Markup Language), a more encompassing document-processing system. HTML, by
virtue
of its SGML heritage, is a standardized language for describing <EM>structured
documents</EM>. A HTML document can be read on any computer
as long as you have a browser written for that computer that can interpret HTML.
</P>
</BODY>
</HTML>
```

As you can see, the paragraph is preceded and followed by HTML tags enclosed in brackets (< >), e.g. <P> and </P>. The paragraph itself also includes several tags, i.e. , , and .

HTML Tags

Most HTML tags are used in pairs, e.g. <TagName>the affected text</TagName>

The tag name itself (here, TagName) is enclosed in brackets (< >).

HTML tags generally have a beginning and an ending tag, surrounding the text that they affect. The beginning tag "turns on" a feature (such as headings, bold, etc.), and the ending tag turns it off. Closing tags have the same name as the beginning tag except that the closing tag name is preceded by a slash (/). For example, the following tag will make the word basic in the following sentence bold when displayed by a browser.

This is a basic tag tutorial.

All HTML tags are case-insensitive; that is, you can specify them in upper or lower case, or in a mixture. is the same as or . It is common practice, however, to specify all tags in uppercase because they are easier to spot when editing an HTML document.

Viewing HTML Code in a Web Document

You can view the HTML source code for just about any document you view with your browser. Most browsers have a View Source or View HTML button or menu item which allows you to view the unformatted (raw) HTML of a document. The file contents, with all the HTML tags, are displayed in a new window. This is an excellent way to see how HTML is used and to learn tips and constructs. You can also save an HTML document on your disk and use it as a template for one of your Web pages or modify the format to suit your purposes. I suggest you view several such documents to acquaint yourself with the code. Don't worry if you don't understand it; that will come later.

Try viewing the HTML code for this document. If you are using Netscape, click on the View button and then on then select Page Source. Close the view window to return to this document.

Creating a Simple HTML Document

To create a HTML document you will need a text editor. A text editor is a program that saves documents in the ASCII format, i.e. as plain text, with no special formatting such as tab characters and page breaks. On UNIX systems, vi, emacs, and pico are all text editors. On Windows systems, Notepad and DOS Edit are good basic text editors. On the Macintosh you can use Simple Text.

Open up your text editor, and type the following code. You don't have to understand what any of it means at this point. You will learn about it later in the Tags Tutorial. This is a simple example to get you started.

```
<HTML>
<HEAD>
<TITLE> My Sample HTML Document</TITLE>
</HEAD>
<BODY>
<H1>This is my HTML Document.</H1>
</BODY>
</HTML>
```

Every HTML document begins with a <HTML> tag followed by a Head Section, then a Body Section, and finally terminates with a </HTML> tag. The Head Section contains the document's title, and the Body Section contains the actual text of the document. Browsers are designed to expect this specific format so be sure to follow this format when developing HTML documents.

The elements required in all HTML documents are the <HTML>, <HEAD>, <TITLE>, and <BODY> tags and their corresponding end tags. Because you should include these tags in each HTML document, you can create a "template" document that contains only these elements. Then use the template whenever you create a new HTML document so that you do not have to type these required elements each time. (Some browsers will

format your HTML file correctly even if all of these tags are not included. But some browsers won't! So make sure to include them.)

File Naming Conventions

After you create your HTML document, save it to a disk as a file. When you pick a name for the file there are three rules:

The filename should be at most eight characters long and have an extension of .htm (.html for UNIX systems)

If you are creating a homepage that you will keep on an Illinois State University web server name your HTML file welcome.htm. Then when you move the document to the web server (which uses UNIX) change the name to welcome.html.

Also, make the eight character filename descriptive enough so that you can recognize it. Use simple names. Avoid using spaces or special characters (&, ", <, etc.) in the filename and be aware that upper and lowercase are considered to be different characters by most Web software.

Viewing Your HTML Document with a Web Browser

Startup your Web browser. You don't have to be connected to the Internet to use it. If your browser tries to connect to the Internet, click the Stop button. Once your browser is running, if you are using Netscape, click on the menu item File and then the item Open Page. Other browsers may have the Open Page under another menu item; just keep looking for it. Once in the Open Page menu, Browse for the HTML file you just made (welcome.htm) and highlight the filename. Click Open. Your HTML document should appear. If not, go back into your text editor and compare your file to the example above. Make sure all the tags have corresponding closing tags, and that all the < characters are matched with > characters. If you find mistakes, fix them and save the file again under the same name. Return to your browser and click on the Reload button. The browser will read the new version of your file. You can edit and preview and edit and preview until you get your document just as you want it.

Text and HTML Editors

To create a HTML document all you need is a text editor. A text editor is a program that saves documents in the ASCII format, i.e. as plain text, with no special formatting or special characters. On UNIX systems, vi, emacs, and pico are all text editors. On Windows systems, Notepad and DOS Edit are good basic text editors. On the Macintosh you can use Simple Text. You can also use word-processing software if you remember to save your document as "ASCII", "DOS Text", or "text only with line breaks." But be careful, special characters, such as the apostrophe, are not saved correctly by all word processors. To be safe, use a text editor.

Another type of editor, a HTML editor, helps you write HTML by providing buttons than may be clicked to enter a tag (a different button is provided for each type of tag). If you have one, use it. If not, and you want one, go to Download.com and search on HTML Editor. You will be presented with a list of editors (some free) that you may download. Try Arachnophilia. It's Freeware and lets you add your own buttons.