Lecture 2

Introducing HTML

Objective

This lecture provides an insight into the following:

- The web browsers in use
- The use of hypertext
- How to design a web page using web documents
- What Hyper Text Markup Language is and is not
- The HTML Elements
- How do you format your HTML text
- Rules for the HTM Language

Lecture - 2

2.1 Snap Shot

The World Wide Web is the set of all Web sites and the documents they can provide to clients (users). This lecture introduces the Internet, in which the WWW is a subset and HTML (Hypertext Markup Language), which lies foundation and builds the WWW. A Web page is basically a text file that contains the text to be displayed and references to elements such as images, sounds, and, of course, hyperlinks to other documents. These concepts including basic components of HTML are discussed in this lecture.

2.2 Overview of Web Browsers

Web browsers provide a unique tool for accessing information on any network, whether it is an internal Intranet or the Internet. They remove the mystery of the Internet and eliminate the need for users to understand arcane commands. Most people begin accessing resources the first time they use a browser. Little training is necessary, and most browser software is free. Browsers do most of the work in accessing and displaying the documents, making the process almost transparent to the user. The traditional role of a Web browser has been to contact a Web site and obtain information from the site in the form of an HTML page. Let us take a look at the various Web browsers available to our users.

NCSA Mosaic:

NCSA (National Center of Supercomputing Applications), University of Illinois, in late 1993 introduced the Web browser NCSA Mosaic. Mosaic first introduced millions of Internet users to linking graphics and text in a browser window to documents located throughout the Internet. Mosaic's popularity was aided by its cross-platform presence; versions were available for the Windows and MacOS environments in addition to several flavors of Unix. Mosaic is still maintained by the NCSA and is free to users. Many browsers are based on the original Mosaic code.

Netscape:

The original developers of Mosaic left the NCSA in 1994 to form what eventually became Wall Street and Media darling Netscape Communications. The company's flagship product, the Netscape Navigator Web browser, is estimated to be used by up to 70 percent of all users of the Web. Netscape Navigator was the first browser to allow inline display of graphics written in JPEG format. Netscape Navigator also supports encrypted HTTP connections thereby facilitating growth of a booming online transaction market. With this browser, users are able to send encrypted information for use in online purchase of goods and services

Internet Explorer:

With Netscape dominating the Web browser and server market, Microsoft, in the early 1996 introduced a new Web browser and server to the Internet community. Microsoft Internet Explorer is destined to be one of the most popular Web browser for several reasons. Not only is the browser easy to use and is supported by the world's largest software company, it also supports several other specific HTML extensions. These include the following

- Background sounds played automatically when the Web page is loaded
- Inline animations of AVI files instead of graphic images
- Marquees that scroll across the browser window

2.3 Hypertext

Traditional text in the form of a book is typically defined as sequential or linear because there is an order in which the text must be read-page two follows page one, and so on. There are many advantages to this method of presenting information. It provides a logical sense of order. It can, however, be an inefficient way to access large bodies of information. (Imagine reading an entire 20-volume encyclopedia page by page to find a single relevant bit of information.)

A variety of mechanisms can speed a user's search for information within documents. For example, a book such as this one uses an index, table of contents, and section headings to speed access to various bits of information. The index provides a mapping from an idea to a particular page in the document containing these related pieces of information. Non-sequential ways to access information such as footnotes, references, and indexes are useful way to deal with navigating and organizing large bodies of related information. With the amount of information available for consumption, exploring an alternative to sequential access seems appropriate. This is where the idea of hypertext comes in.

A hypertext document is an electronic document that contains links to related pieces of information. It could be characterized as providing generalized footnotes. For example, a hypertext document about cows may feature a link from the word milk, as shown in Figure 2.1 Hypertext is a nonlinear way to access information. Many people find it similar to the way they think about problems.

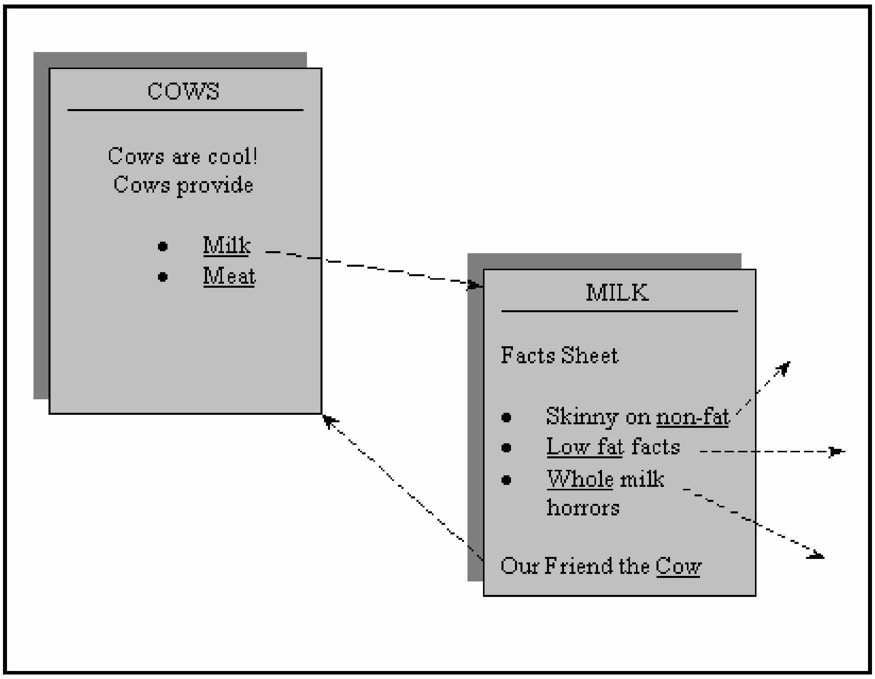


Figure 2.1 *Hypertext documents with underlined links*

2.4 Hypertext Markup Language (HTML)

HTML (Hypertext Markup Language) is the language that puts the face on the Web by helping to prepare documents for online publications. HTML documents are also called *Web documents,* and each HTML document is known as *Web page*. A page is what is seen in the browser at any time. Each *Web site*, whether on the Internet or Intranet, is composed of multiple pages. And it is possible to switch among them by following hyperlinks. The collection of HTML pages makes up the *World Wide Web*.

HTML pages can be created using simple text editor such as Notepad or WYSIWYG (What–You-See-Is-What-You-Get) Web page editors such as Microsoft FrontPage. In either case the result is a plain text file that computers can easily exchange. The *browser* interprets this text file and renders on the client computer.

Without knowing the HTML even a bit, one can create Web pages by using the Web Page editors. But this method has a drawback. These editors often introduce their own little nuances or quirks to the way code is produced. Moreover even the best editors do not support all the tags that are part of HTML at any given time.

This brings the first reason to learn the HTML. Sometimes it is necessary to directly modify the source of the page to add or change tags and attributes. Learning HTML will clarify how the tags relate to each other.

Second, HTML is as much an organizational tool, as it is design tool. Even with Web page editors, the rationale behind the tags is to give a structure and purpose to a page. Learning how HTML organizes a page, leads to better planning and designing of Web pages for the readers.

Finally, HTML can be downright fun. It gives certain satisfaction from building a Web page from the ground up. It's like building an own house, in which every brick, every board, ever nail etc. are known. This facilitates the easy modification to acquire the desired result. It also makes it much easier to take a look at someone else's page and know how they achieved their effect.

What HTML Is Not

HTML is a powerful technology, but there are many misconceptions about it. Many people have a basic notion of what HTML is, but few seriously consider what it isn't.

HTML is not a Programming Language

Many people think that making pages is like programming. However, HTML is unlike programming in that it does not specify logic. It specifies only the layout and structure of a document. HTML is a structured language; but unlike a programming language, in which the rules are enforced by the compiler, HTML rules are not strictly enforced by browsers. The language does not provide any sort of execution order like programming languages do.

HMTL is Not a WYSIWYG Design Language

HTML is not a specific, screen or printer-precise formatting language. Many people struggle against the technology on a daily basis trying to create perfect layouts by using HTML elements inappropriately, or using images to make up for HTML's lack of screen and font-handling features.

HTML Is Not Complete

HTML is not finished. The language does not provide all the facilities it should. If HTML is meant to be a physical design language, it still does not provide font or pixel-level control, which is very important.

HTML Is Not All, Needed to Create Good Web Pages

While HTML is the basis for Web pages, it is required to know a lot more in order to build useful Web pages unless the page is very simple. Document design, graphic design, and even programming, are often necessary to create sophisticated Web pages. HTML serves as the foundation environment for all these tasks, and complete understanding of HTML technology can only aid document authors.

2.5 Basic components of HTML

HTML documents are created by combining special markup code called *tags*. The tags define the structure of the document and provide the framework for holding the actual content, which can be text, images, or other special content. When a browser reads a document that has HTML markup in it, it determines how to render it onscreen by considering the HTML elements (tags) embedded within the document (Figure 2.2)

An HTML document is simply a text file that contains the information, which is to be published. It also contains embedded instructions, called elements (tags) that indicate how a Web browser should structure or present the document. HTML elements generally consist of a pair of angle-bracketed tags surrounding some text. The *end-tag* (**</TAG>**) is just like the *start tag* (**<TAG>**), except that it has a slash (/) in it as shown:

```
<TAG>        ← Start Tag
        -----------
        Text that the tags affect
        -----------
</TAG>       ←End Tag
```

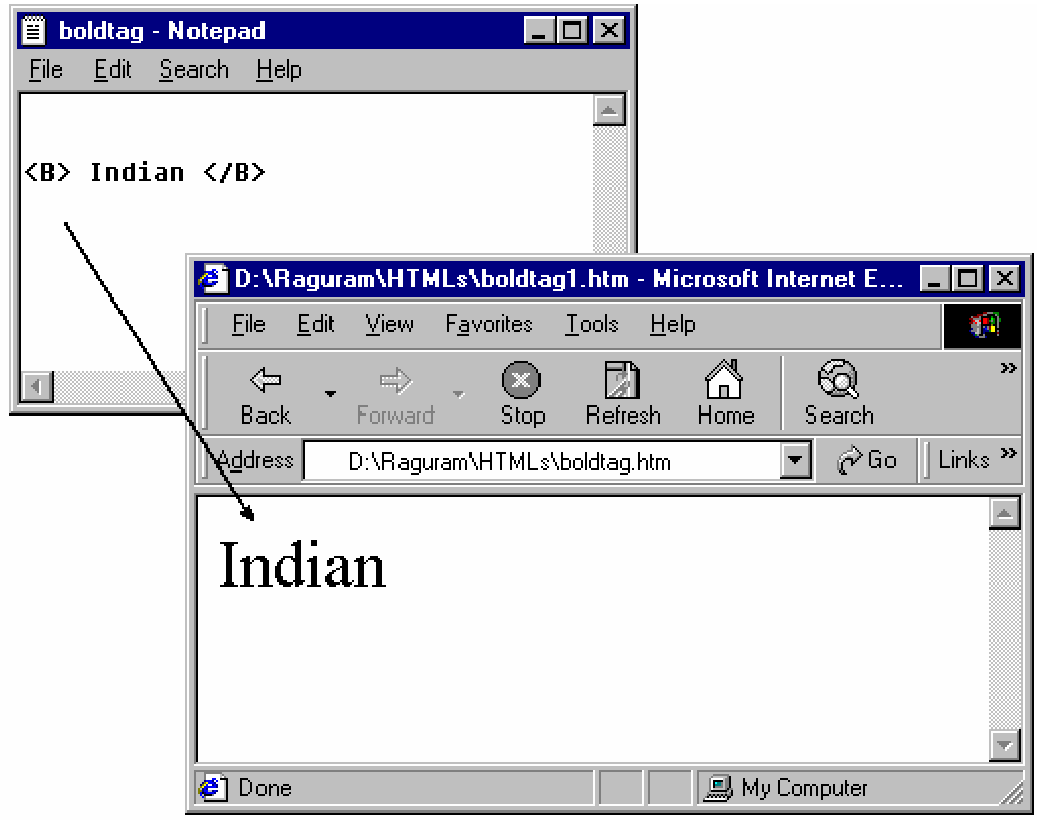
HTML elements indicate the "markup" on the surrounded text. They may indicate the meaning of the enclosed information (for example, a citation) or how it should be rendered (for example, in bold). HTML elements normally consist of a pair of tags called *container* tags, because content goes between them. However, some elements, such as the horizontal rule tag **<HR>**, do not have a corresponding end tag. These are termed empty elements. For some other elements, like the paragraph element <P>, the end tag may be optional. It is always good idea to use the end tag if one is available. Given the following HTML code,

```
<B> Indian </B>
```

a Web browser should render the phrase "Indian" in a bold typeface.

An HTML file usually begins with the <HTML> element, which indicates that the contents of the file include markup. The file should end with that element's end tag, </HTML>. The rest of a typical HTML file is composed of the head and the body.

The head, which is enclosed with the <HEAD> element (consisting of the <HEAD> and </HEAD> tags), includes supplementary information about the document, such as the title of the document, which most browsers display in a title bar at the top of the browser window. The title is indicated between the <TITLE> and </TITLE> tags. The document tittle is required under the current HTML specification. While some browser may not require the inclusion of the <TITLE> element, it should always be included for correctness, book marking, and the sake of good HTML style.



boldtag - Notepad
File Edit Search Help
D:\Raguram\HTMLs\boldtag1.htm - Microsoft Internet E...
File Edit View Favorites Tools Help
Back
Forward
Stop
Refresh
Home
Search
Address
D:\Raguram\HTMLs\boldtag.htm
Go
Links
Done
My Computer

```
<B> Indian </B>
```

Indian

Figure 2.2 *Interpretation of a Web page with HTML markup.*

The body, which is enclosed between <BODY> and </BODY> tags, contains the actual content and appropriate markup tags needed to render the page.

A basic HTML template is as follows:

```
<HTML>
<HEAD>
        <TITLE>
                Document Title
        </TITLE>
</HEAD>
<BODY>
        The Document's body: text, images, sounds, and HTML commands
</BODY>
</HTML>
```

The HTML Elements

The following sections describe the basic elements used in formatting an HTML document and also some other most commonly used elements.

Setting the Boundaries with <HTML>

The first tag to learn about is <HTML>. It's paired with </HTML> to encase all the other tags in an HTML page; the two mark the absolute beginning and end of the file.

Syntax

```
<HTML>
---document and tags---
</HTML>
```

The <HTML> tag is an optional element for building Web pages. If it's not explicitly included most browsers or other user agents will assume its existence, along with the appropriate defaults for text direction. However, some browsers might be confused when handed a document without any clear indication of type with HTML, resulting in a page that's displayed unpredictably, displayed as plain text or that refuses to display at all.

There are a couple of good reasons to use the <HTML> tag. First, HTML isn't the only markup language on the Web. There are cousins to HTML, such as extensible markup language (XML) that are interpreted in slightly different ways and are gaining in acceptance and use. Second, using the <HTML> tag is good style and shows that whoever builds the document had some idea what to do.

The HEAD Element

Theoretically, every document has header and a body. The header of the document is where global settings are defined; it's contained between the <HEAD> and </HEAD> tags.

Syntax

```
<HEAD>
        header content
</HEAD>
```

It is a good place to place <TITLE>. Besides it is a favorite place to include scripting language function definitions.

Giving to a Page a <TITLE>

Probably the most commonly used HEAD feature in HTML, and the only required element is the <TITLE> tag.

<TITLE>*Text*</TITLE>

In this tag, *text,* is a short, one-line name for the document that's displayed in the browser's title bar. Without title, most browsers default to the HTML filename. Because file-names aren't always terrible descriptive and are sometimes long and clunky, it's good practice to supply a title for all HTML documents.

Only one title is allowed per document and its size is limited by the size of the user's browser window. Of course, a title can be of any length, but it is truncated in the browser's title bar if it stretches beyond the limits.

The BODY Element

Like the <HEAD> tag, the <BODY> tag's primary purpose is to delineate the main portion of the document - the part seen by the user.

Syntax

```
<BODY
        [ ALINK = color name | #RRGGBB ]
        [ BACKGROUND = path of background image ]
        [ BGCOLOR = color name | #RRGGBB ]
        [ BOTTOMMARGIN = pixels ]
        [ LEFTMARGIN = pixels ]
        [ LINK = color name | #RRGGBB ]
        [ RIGHTMARGIN = pixels ]
        [ SCROLL = YES | NO ]
        [ TEXT = color name | #RRGGBB ]
        [ TITLE = advisory text ]
        [ TOPMARGIN = pixels ]
        [ VLINK = color name | #RRGGBB ] >

        --The Document's body: text, images, sounds, and HTML commands
</BODY>
```

The attributes of the <BODY> element is described below:

ALINK

This attribute sets the color for active links within the document. Active links represent the sate of a link as it is being pressed. The value of the attribute can either be a browser-dependent named color or a color specified in the hexadecimal #RRGGBB format.

BACKGROUND

This attribute contains a path for an image file, which will be titled to provide the document background.

BGCOLOR

This attribute sets the background color for the document. Its value can be either a browser-dependent named color or color specified using the hexadecimal #RRGGBB format.

BOTTOM MARGIN

This attribute specifies the bottom margin for the entire body of the page and overrides the default margin. When set to 0 or “”, the bottom margin is the bottom edge of the window or frame the content is displayed in.

LEFTMARGIN

This Internet Explorer-specific attribute sets the left margin for page in pixels, overriding the default margin. When set to 0 or “”, the left margin is the left edge of the window or the frame.

LINK

This attribute sets the color for hyperlinks within the document that have not yet been visited. Its value can be either a browser-dependent named color or color specified using the hexadecimal #RRGGBB format.

RIGHTMARGIN

This Internet Explorer-specific attribute sets the right margin for page in pixels, overriding the default margin. When set to 0 or “”, the right margin is the right edge of the window or the frame.

SCROLL

This attribute turns the scroll bars on or off. The default value in YES.

TEXT

This attribute sets the text color for the document. Its value can either be a browser-dependent named color or a color specified in the hexadecimal #RRGGBB format.

TOPMARGIN

This Internet Explorer-specific attribute sets the top margin for page in pixels, overriding the default margin. When set to 0 or “”, the top margin will be exactly on the top edge of the window or frame.

VLINK

This attribute sets the color for links within the document that have already been visited. Its value can be either a browser-dependent named color or color specified using the hexadecimal #RRGGBB format.

Headings <H1> Through <H6>

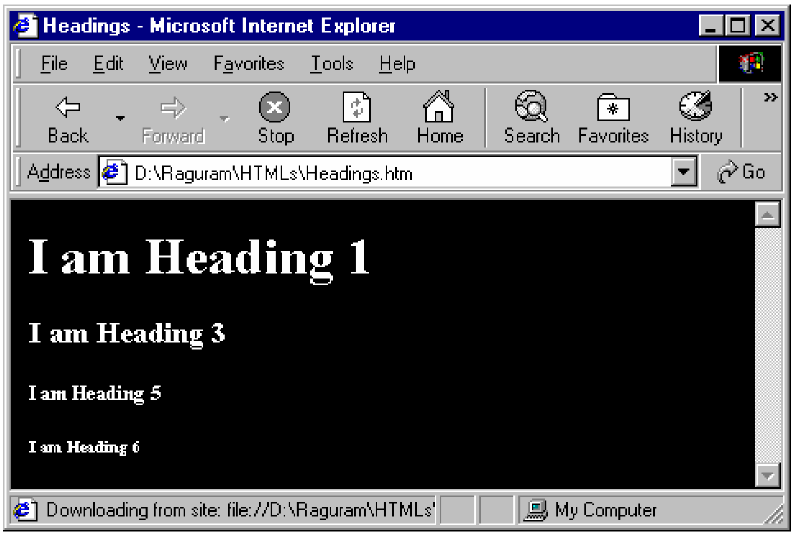
No page begins with a lead paragraph but with a headline. The six HTML heading styles are a way of showing the level of importance among different parts a page.

Syntax

<H_n> *Heading Text* </H_n>

Here, n is an integer from 1 to 6 indicating the level of heading used, with 1 being the most important and 6 being the least. Although browsers vary in the size and typestyle given to the six headings levels, every browsers follows the basic rule of giving the biggest and boldest style to <H1> and the smallest and most unobtrusive style to <H6>. The following lines produce the document as in Figure 2.3.

```
<HTML>
<HEAD>
<TITLE>Headers</TITLE>
</HEAD>
<BODY BGCOLOR="Black" Text="White">

<H1> I am Heading 1 </H1>
<H3> I am Heading 3 </H3>
<H5> I am Heading 5 </H5>
<H6> I am Heading 6 </H6>

</BODY>
</HTML>
```


Headings - Microsoft Internet Explorer
File
Edit
View
Favorites
Tools
Help
Back
Forward
Stop
Refresh
Home
Search
Favorites
History
Address
D:\Raguram\HTMLs\Headings.htm
Go
Downloading from site: file://D:\Raguram\HTMLs'
My Computer

I am Heading 1

I am Heading 3

I am Heading 5

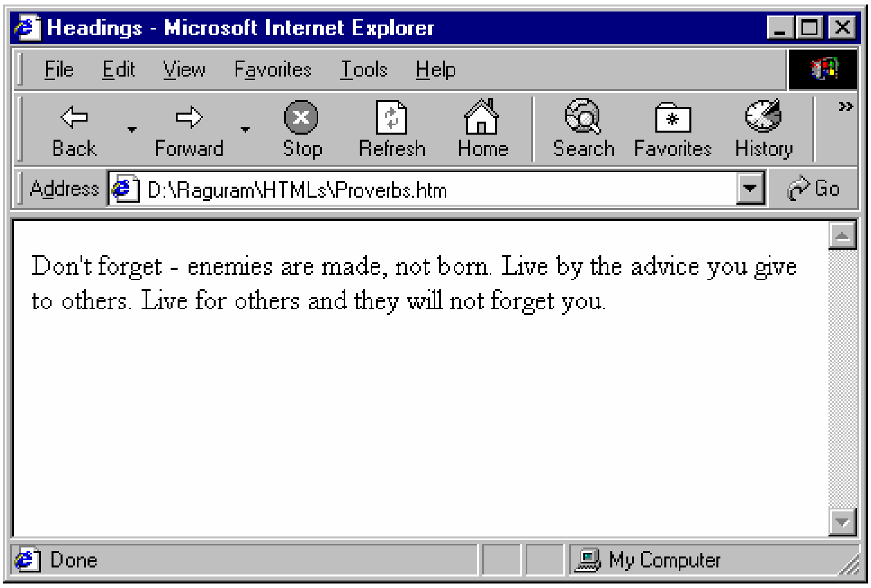
I am Heading 6

Figure 2.3 *Different Styles of Headings.*

A New Paragraph: <P>

Now that the headings are in place, it's time to turn the attention to the rest of the document. The next step is to start breaking the text into paragraphs. The following listing appears to be a Web page broken into logical paragraphs, but look at Figure 2.4 to see how a browser interprets it.

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

Don't forget  - enemies are made, not born.

Live by the advice you give to others.

Live for others and they will not forget you.
</BODY>
</HTML>
```


Headings - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back
Forward
Stop
Refresh
Home
Search
Favorites
History
Address
D:\Raguram\HTMLs\Proverbs.htm
Go
Done
My Computer

Don't forget - enemies are made, not born. Live by the advice you give to others. Live for others and they will not forget you.

Figure 2.4 *All lines appearing together without any line break.*

This result is obviously isn't what the designer had in mind. The browser has ignored the extra line breaks and runs the lines together into one long paragraph.

The first way to break text into paragraphs is to use the paragraph tag: <P>.

Syntax

<P [STYLE = Style Information] > ….*Text*….</P>

Text is a line or paragraph text that should begin on a new line and remain together. The closing tag is optional and can be omitted. The STYLE attribute has Property:Value pair to specify appearance. Each Property:Value pair is delimited by semi-colon (;) and, Property and Value itself by colon (:). For example, the following line constructs a paragraph with green text color, yellow background, font-weight normal and font-size 20.

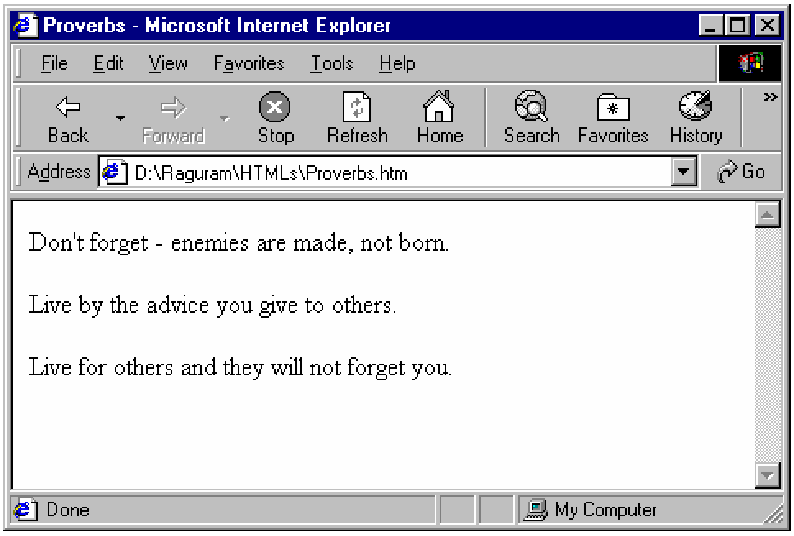
<P STYLE= "Color:Green; Background-color:Yellow; Font-weight:Normal;Font-size:20">

Don't forget - enemies are made, not born. </P>

The preceding listing is rewritten with this paragraph tag as given below:

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

<P>Don't forget  - enemies are made, not born. </P>

<P>Live by the advice you give to others.</P>

<P>Live for others and they will not forget you.</P>

</BODY>
</HTML>
```

This listing produces the document as in Figure 2.5


Proverbs - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History
Address D:\Raguram\HTMLs\Proverbs.htm
Go
Done
My Computer

Don't forget - enemies are made, not born.

Live by the advice you give to others.

Live for others and they will not forget you.

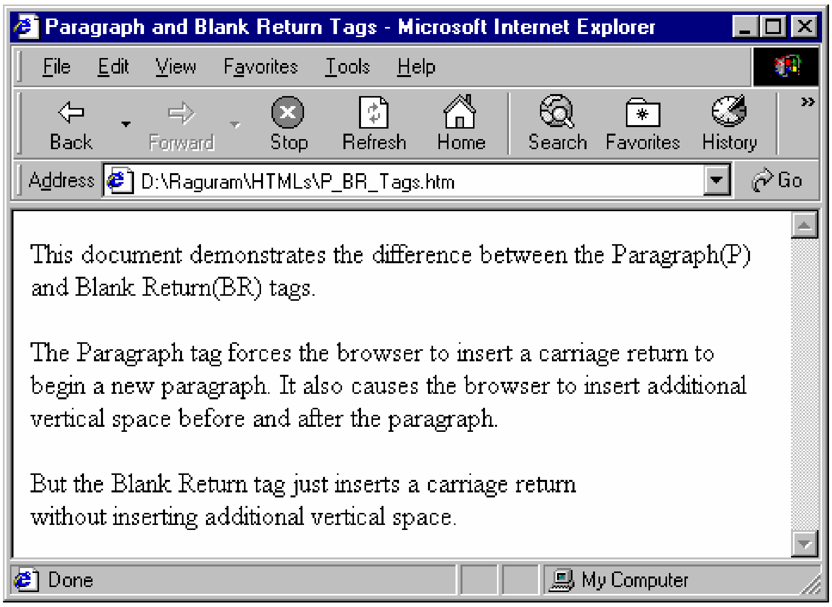
Figure2.5 *Sentences as new paragraphs due to <P> tag.*

A New Line: <BR>

A blank return tag, <BR>, which is an empty tag, is similar to a paragraph tag. But it behaves in a slightly different way. It starts a new line within the current paragraph, but it doesn't start a new paragraph.

The following lines produce a document in Figure 2.6 demonstrating the difference between the <P> tag and <BR> tag.

```
<HTML>
<HEAD>
<TITLE>
Paragraph and Blank Return Tags
</TITLE>
</HEAD>

<BODY>
This document demonstrates the difference between
the Paragraph(P) and Blank Return(BR) tags.

<P> The Paragraph tag forces the browser to insert a carriage return to begin a
new paragraph. It also causes the browser to insert additional vertical space before
and after the paragraph.
</P>
<P>
But the Blank Return tag just inserts a carriage return <BR>
without inserting additional vertical space.
</P>
</BODY>
</HTML>
```



D:\Raguram\HTMLs\P_BR_Tags.htm

This document demonstrates the difference between the Paragraph(P) and Blank Return(BR) tags.

The Paragraph tag forces the browser to insert a carriage return to begin a new paragraph. It also causes the browser to insert additional vertical space before and after the paragraph.

But the Blank Return tag just inserts a carriage return
without inserting additional vertical space.

Figure 2.6 *An HTML document demonstrating the difference between the <P> and <BR> tags.*