

INTRODUCTION TO COMPUTER PROGRAMMING

Richard Pierse

Class 8: Writing Web Pages in *HTML*

Introduction

One of the important uses of the *Java* language is to write *applets*: small programs that execute within web pages. This is the subject of next week's class. To prepare for this, we need to understand web pages and the language in which they are written: **H**yper-**T**ext **M**ark-up **L**anguage, or **HTML** for short. While many programs exist to help you write web pages without learning *HTML* (examples are *DreamWeaver*, *HoTMetaL*, *FrontPage*), none of these programs is really necessary. The *HTML* language is actually very simple and web pages can be written using nothing more than a simple text editor such as *Notepad*.

A Brief History of the Internet

The first attempt to link two computers via a telephone line was made in 1965 by Lawrence Roberts and Thomas Merrill. This led to the setting up of the *ARPANET* (Advanced Research Projects Agency Network), a network linking computers of the US Department of Defense and the key University Departments that provided their (military) research. By 1969, four computers were linked together. By its nature, access to the *ARPANET* was highly restricted. However, it led to the development of the basic concepts used in Internet traffic to this day: the *File Transfer Protocol (FTP)* and the *Transmission Control Protocol/Internet Protocol (TCP/IP)* which was adopted in 1983.

The Joint Academic Network (*JANET*) in the UK (1984) and the National Science Foundation Network (*NSFNET*) in the US (1985), were the first attempts to create networks aimed at serving the entire higher education community. These networks gradually expanded to take in commercial users. Access was mainly through *TELNET* (logging in) or *FTP* (downloading or uploading of files). Both would usually require a password on the remote system but, gradually, some open access (*anonymous FTP*) became available.

The *World Wide Web*

As the Internet grew, the amount of available material expanded and, by the early 1990s, attempts were being made to facilitate access by organising and cataloguing the information. Early systems were *Archie* (an archive of *FTP* sites) and *Gopher*, an organisation system with a searching facility called *Veronica*. The *World Wide Web* (*WWW*) and its language, *HTML*, were invented by Tim Berners-Lee at *CERN* (Centre Européenne pour la Recherche Nucléaire: the European particle physics laboratory) in 1990. The *WWW* is a document system where the documents can contain hyper-links to other documents or files, situated anywhere else on the Internet. Initially designed for use within the high-energy physics community, it spread rapidly as users came to appreciate its simplicity and the power of its decentralised design. In 1993, *Mosaic* was released, the first web browser available for PCs. By the end of 1993, there were 500 web servers; one year later there were 10,000.

The *HTML* Language

An *HTML* document is simply a document containing text and special tags, delimited by angled brackets `<>`. The text defines the content of the document and the tags define its structure and appearance. Some tags appear on their own such as:

`<P>`

which causes a paragraph break. Other tags appear in pairs such as:

` `

which causes any text between the `` tag and the `` tag to be displayed in bold font.

Tags are not case sensitive, although for clarity, in this handout I follow the recommendation of the *World Wide Web Consortium* (*W3C*) and always use upper-case. Tags that are not recognised by a web browser are simply be ignored. This makes the *HTML* language tolerant of mistyping as well as allowing for future expansion of the *HTML* language.

The appearance of an *HTML* document depends to some extent upon the particular web browser being used to display it. It is always good practice to test out your web page on more than one browser to check that it looks OK.

HTML Files

HTML documents are stored in text files having the special file extension `.htm` or `.html`.

HTML Document Structure

The structure of an HTML document is as follows:

```
<HTML>
<HEAD>
<TITLE>document title goes here</TITLE>
</HEAD>
<BODY>
document text goes here
</BODY>
</HTML>
```

The whole document is enclosed between the `<HTML></HTML>` tags. It is divided into two sections. The first is the **header** section enclosed between the `<HEAD></HEAD>` tags. This is where the document title is defined. The header section may also contain other tags that identify the document author and specify other parameters, but nothing in the header appears in the visible page. The **body** section, enclosed between the `<BODY></BODY>` tags, is where you put the actual content of the page.

Simple Formatting Tags

The following are commonly used tags for formatting of text.

<code>
</code>	causes break to new line
<code><P></code>	causes paragraph break
<code><HR></code>	inserts a horizontal rule
<code></code>	text between the tags is in bold font
<code><I></I></code>	text between the tags is in italic font
<code><U></U></code>	text between the tags is underlined
<code><CENTER></CENTER></code>	text between the tags is centred

Heading Tags

Headings can be specified using one of the six heading tags in decreasing size from `<H1></H1>` through to `<H6></H6>`.

Hyperlinks

Hyperlinks are tags that allow your document to link with other *HTML* pages or other files elsewhere on the web. The format is

```
<A HREF="URL"></A>
```

where `URL` is any valid web address that defines the web page to which to link, and the text between the `<A>` and `` tags will form the clickable hyperlink itself.

For example, the *HTML* code

```
a <A HREF="http://www.nasa.gov">cool</A> site
```

includes a hyperlink to the main web site of the United States National Aeronautics and Space Administration. The visible text on the page will be the phrase "a cool site" but the word "cool" will be highlighted in some way (depending on the browser) to indicate that it is a hyperlink. Clicking anywhere on the word "cool" will initiate the link and transfer the viewer to the NASA home page.

Specifying Web Addresses

Hyperlinks can be links to web pages on external sites or to other pages within your own web site. In the latter case, addresses may be specified either absolutely or relative to the current page. For example:

```
HREF="/hobbies/climbing/k2.html">
```

is a reference to an *HTML* page called `k2.html` in the sub-directory `hobbies/climbing` of your own site. This is an **absolute** address. This page could also be referred to from a page in the sub-directory `hobbies`, using the **relative** address:

```
HREF="climbing/k2.html">
```

or

```
HREF="./climbing/k2.html">
```

`./` indicates "starting from the current directory".

`../` indicates "going up one directory from the current directory".

Images

Apart from text and hyperlinks, web pages often contain graphical images. The image tag takes the form:

```
<IMG SRC="file">
```

where `file` is the address of a file containing the image. Optional parameters in the `IMG` tag control the appearance of the image; for example, `ALIGN="RIGHT"` positions the image to the right of the page and `BORDER=0` suppresses the border.

Image files can be in one of three standard recognised formats: *GIF*, *PNG* and *JPEG*. Typically, the *JPEG* format is used for photographic images and the *GIF* or *PNG* formats for icons and other small graphics. (Note: the *PNG* format is an alternative to *GIF* that avoids the *UniSys* software patent on the latter format).

An image can be made into a hypertext link as in:

```
<A HREF="http://www.nasa.gov">  
    <IMG SRC="images/spaceman.jpg">  
</A>
```

In this example, the image `spaceman.jpg` is a clickable hyperlink, transferring the viewer to the NASA home page.

Lists

Lists of items can easily be created in *HTML*. Lists may be ordered or unordered:

```
<UL></UL> denotes an unordered (bulleted) list  
<OL></OL> denotes an ordered (numbered) list.
```

List items are indicated by the `` tag. For example:

```
<UL><LI>books<LI>CDs<LI>DVDs</UL>
```

displays an unordered list of three items, each on a new line.

Tables

The `<TABLE></TABLE>` tag provides a way to display tabular data. More generally, however, tables can be used to manage the layout of a document, allowing blocks of text or images to be positioned on the page as cells of a table. By default, tables are automatically formatted so that each cell has the size of the largest element in any cell. Tables can be nested within each other, making it possible to design very complex layouts.

Within a table, rows are denoted by the tag pair `<TR>/<TR>` and within a row, cells are denoted by the tag pair `<TD>/<TD>`.

For example:

```
<TABLE>
<TR><TD>11</TD><TD>12</TD><TD>13</TD></TR>
<TR><TD>21</TD><TD>22</TD><TD>23</TD></TR>
<TR><TD>31</TD><TD>32</TD><TD>33</TD></TR>
<TR><TD>41</TD><TD>42</TD><TD>43</TD></TR>
</TABLE>
```

displays a table of numbers having four rows and three columns.

Frames

Frames allow several different web pages to be displayed as independent windows or frames in the same page. When a page is too big to fit in its frame, scrollbars are automatically created. (Note that frames were not supported by early versions of *HTML*).

The use of frames is best illustrated by a simple example:

```
<FRAMESET ROWS="50%,50%" COLS="30%,70%">
  <FRAME SRC="frame1.html">
  <FRAME SRC="frame2.html">
  <FRAME SRC="frame3.html">
  <FRAME SRC="frame4.html">
</FRAMESET>
```

The `<FRAMESET>` tag defines a set of frames with two rows and 2 columns. Each row is allocated 50% of the total page height. The first column is allocated 30% of the total page width and the second, 70%. The `<FRAME>` tags then specify the web pages to be placed in each frame, the frames being ordered by row.

A Test *HTML* Document

The following complete *HTML* document illustrates many of the features discussed above. This document can be typed into any text editor such as *Notepad* or *Wordpad*. It needs to be saved with the extension `.htm` or `.html`.

```
<HTML>
<HEAD> <TITLE>A simple web page</TITLE> </HEAD>
<BODY>
  <CENTER><H2>A Heading</H2></CENTER>
  Welcome to my web page.
  <P>
  A web page can have text in <B>bold</B> or <I>italic</I>
  fonts and can include
  <A HREF="http://www.nasa.gov">links</A> to other web pages.
  <P>
  It can also include pictures like <IMG SRC="bird.gif">
  or <IMG SRC="ladybug.jpeg">.
  <P>
  <H3>A Table</H3>
  The following displays a table with two rows and three
  elements per row.
  <TABLE>
  <TR><TD>a11</TD><TD>a12</TD><TD>a13</TD></TR>
  <TR><TD>a21</TD><TD>a22</TD><TD>a23</TD></TR>
  </TABLE>
  <H3>An Unnumbered List</H3>
  <UL><LI>Item 1<LI>Item 2</UL>
</BODY>
</HTML>
```

Displayed in a web browser, the page will look something like this:

A Heading

Welcome to my web page.

A web page can have text in **bold** or *italic* fonts and can include [links](#) to other web pages.



It can also include pictures like

A Table

The following displays a table with two rows and three elements per row.

a11	a12	a13
a21	a22	a23

An Unnumbered List

- Item 1
- Item 2

Uploading Your Web Pages

Once a web page has been created, it needs to be uploaded to a web server where you have permission to create web pages (web space). For this class, you have been given web space on the Economics Department web server. To upload your web page, you will use the *FTP* (File Transfer Protocol) program ***WS_FTP*** available on the University server.

Once you have entered these details, click the *OK* button. If the details have been entered correctly, *WS_FTP* will connect you with the web server.

This window has two panes. The left pane shows a sub-directory of the local hard-drive on your computer. The right-pane shows the remote web directory on the Economics server where you will upload your page. The arrow buttons between the panes allow you to upload files from your local machine to the server (→) and download files from the server to your local machine(←).

Transferring files: ASCII and Binary

When you upload or download a file using *FTP*, you need to specify whether the file is *Binary* or *ASCII*. *HTML* files and other text files are *ASCII*. Image files (*GIF*, *JPEG*) and *Acrobat PDF* files are *Binary*. You specify the file type to *WS_FTP* by selecting the appropriate radio button situated immediately below the two panes.

Changing Directory

You will need to change your local directory in the left-pane to the directory where you created your web page, e.g. `c:\tmp`. You can do this either by using the `ChgDir` button or by navigating through the directory tree in the drop-down list box above the left-pane.