A Staff Developers' Briefing on the World Wide Web

Sue Cunningham, University of Manchester

1. Introduction

This guide is not a HTML tuto guide to setting up your own web server. Rather it has rial or a been written to give the new user of the World Wide Web (WWW) an overview of the issues and technologies surrounding it.

The concept of hypertext has been around a long time, and was first described by Vannevar Bush in 1945. Ted Nelson coined the actual term hypertext in the 60's to describe his Xanadu project. A hypertext document allows the user to navigate through it in a non-linear or non-sequential fashion, by selecting parts of the text which are linked to other parts of the same document or other documents. Hypermedia is a hypertext system that is not restricted to text documents, but includes other media, effectively multimedia hypertext.

The WWW is a platform independent, distributed hypermedia system, allowing the user to access hypermedia documents stored on remote servers around the world using a range of different computing platforms.

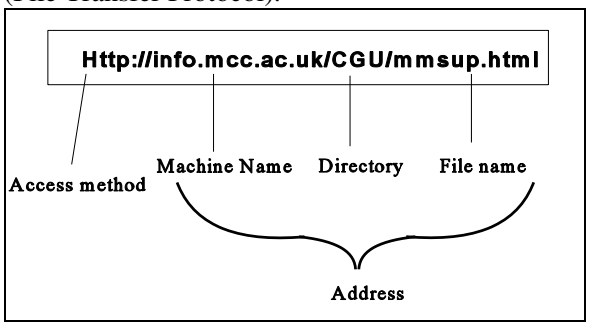
2. WWW

2.1 History

The World-Wide Web was initially developed by Tim Berners-Lee and Robert Cailliau of CERN Laboratories, Geneva, to allow particle physicists throughout Europe to share information. Initially mainly a text-based system, its cross platform capability and ease of use ensured its uptake into the wider community and its continuing development. Although initially accessed with a text only browser, the majority of people now use one of the many graphical browsers now available, such as Mosaic.

2.2 URLs

URL stands for Uniform Resource Locator, and is used for specifying an object, such as a file, and how to access it. It is in effect the address of the file. It consists of two main parts, and takes the form access-method://address, where the access method is the protocol for retrieving the file. These include HTTP (Hypertext Transport Protocol) and FTP (File Transfer Protocol).



2.3 HTML

HTML stands for Hypertext Markup Language, and is a simple markup language recognised by all WWW browsers. A standard HTML document is still the most effective way to widely disseminate information on the WWW. A HTML document consists of the normal text of the document that you see and read, and markup tags that define elements of the document, such as title and paragraph.

```
<h1> Multimedia Support for Higher Education</h1>
<A HREF=
"http://info.mcc.ac.uk/CGU/staff/cunningham.html">
Sue Cunningham
<p>Manchester is host to the JISC funded Multimedia
Support Officer . This post was setup to provide general
support for multimedia in the academic community via a
phone and email helpdesk, courses etc. </p>
```

The code fragment above shows a number of tags, including a level one heading <H1> and paragraph <p>. All tags are enclosed within angle brackets, and many have a start and end tag, then end tag starting with </. The other tag shown in this fragment (<A HREF...) creates a link to another page.

HTML files can be created directly using any text editor, using a HTML editor, using a word processor add-on, such as Internet Assistant for Word, or by using a program to convert an existing file into HTML.

Writing HTML
There are a number of HTML primers and beginners guides available, both on the WWW itself and as books. Some examples are given in the bibliography. Since web pages are just text files, they can be created with even the most basic text editor. However, a number of programs have been written to help with markup and automatically convert between HTML and other formats.

HTML Editors
These are mostly simply text-based editors, such as HTML Notepad for the PC, which is an extension of Windows Notepad. Text is created or imported as in a standard text editor, then marked-up using the additional pull down menus. The display shows the text and the markup tags and a separate browser is necessary to view the finished HTML document. Increasingly however, WYSIWYG HTML editors are becoming available, such as Microsoft's Front Page and GNN Press. For a list of HTML editors see:*http://www.bendigo.net.au/tucows/softhtm.html*

Conversion programs
There are many programs around to convert existing documents from various formats to HTML, including Word, Word Perfect, RTF (Rich Text Format) and Framemaker. The WWW Consortium maintains a list of converters at:
http://www.w3.org/pub/WWW/Tools/
The success with which documents can be converted often depends on how they were written originally. For example, a Word document in which headings were created using the heading styles feature is more likely to convert correctly than one in which the headings were generated by manually increasing the font size. The more 'structural' information of this nature in the original document, the better the conversion will be.

3 Clients and Servers

3.1 Introduction

The WWW is based on a client server architecture. The client, often called a browser, is the software which runs on the local machine allowing the user to view documents. The server is the software which delivers the information to the client. Server is also used to mean the actual machine the server software sits on. For example, a document is displayed in the browser which has several hypertext links to other documents. When the user clicks on, or otherwise selects, a link, the browser requests the new file from server (which may be a different server than the original document was located on), the server retrieves the file and delivers it to the browser.

3.2 Browsers

Text browsers
Text-based browsers provide WWW access when a graphical interface is not available, or impractical, for example for blind users. However, they lack many of the functions of graphical browsers, and increasingly Web designers are making use of graphics as an integral part of their pages. One of the best known text-based browsers is Lynx, developed at the University of Kansas. Versions for Unix systems and DOS are available.

Graphical Browsers
These comprise the bulk of the clients currently in use. They can display not only text, but also handle images, and increasing other file formats also. The most popular browser at the current time is Netscape, available on a range of platforms.

Non-standard tags
HTML was not designed to define the final appearance of a document. The markup tags are interpreted by the browser, so that the page's appearance will depend on the particular browser and how the reader has configured the browser. For example, the author specifies a piece of text is in style heading one, but the reader can decide which font headings should be in and how large they are. However, both Netscape and Microsoft's Internet Explorer support a range of browser specific markup tags in addition to standard HTML. These tags are designed to give the author more control over the final appearance of the document, but may cause unpredictable results if the viewer is not using that particular browser.

Helper Applications
Helper applications are external programs which display file types that the browser cannot handle, for example, a Postscript or Acrobat viewer. Browsers can be configured to automatically launch such applications when particular file types are encountered. Setting up helper applications is browser specific, and you will need to consult the documentation. Increasingly browsers are supporting a much wider range of file types internally, and the need for helper applications for common file types is decreasing. For example, Microsoft's Internet Explorer will support several image and sound formats, MPEG, AVI and Quicktime movie files and VRML (Virtual Reality Markup Language).

4 Accessing Information

The WWW continues to grow rapidly, and so, as the amount of information available increases, finding it becomes harder, as there is no central catalogue or cataloguing system. There are a number of general directories, such as Yahoo (*http://www.yahoo.com*) which provide page listings by topic, but perhaps more useful are the search engines.

4.1 Search engines

Search engines are databases indexing the contents of large numbers of web pages, which are accessed over the WWW using a query form. The usefulness of the engine depends on how many pages it has indexed, what information it has indexed from those pages and how intelligent a query you can submit, e.g. does it support 'AND', 'OR' and 'NOT'.
Some of the best search engines include Lycos (*http://www.lycos.com/*) and AltaVista (*http://www.altavista.digital.com/*), but it is always worth trying more than one, as their databases differ. Meta search engines, such as Metacrawler (*http://metacrawler.cs.washington.edu:8080/index.html*) allow you to search more than one engine with a single query. The query is submitted to the meta engine, and it automatically passes it on to several engines and collates the results.

4.2 Domain Names

Domain names are unique addresses on the Internet. Usually a company or organisation will have its own domain name, e.g. mcc.ac.uk (University of Manchester) and addresses of specific machines within an organisation will end in the domain name. It is often possible to guess the correct URL given a limited amount of information. The first part of the domain name will relate to the company or organization name, and the last part refers to the type of company/organization. For example:
.ac.uk - UK Higher Education
.edu - US educational site
.co.uk - UK business
.com - US business
.org - organizations, usually non-profit.

5 Providing Information

5.1 Home Pages

Many departments now have their own web servers, and the departmental home page should name the web master, who you will need to contact to set up your own pages on that server. Once you have setup your own home page there are a number of points to bear in mind:

- Follow any local codes of practice (see below).
- The WWW is not static, users expect information to be up to date. Consider updating your pages on a regular basis to include new information and links.
- Periodically check that external links are still valid.
- To ensure a wide audience, ask authors of related web sites to include a link to your pages, announce it in relevant newsgroups and submit the URL to some various search engines.

5.2 Good practice for the WWW

The following list is based on the SIMA report by Margaret Isaacs - Guide to good practices for WWW authors (see Bibliography).

- For maximum portability and usability, use standard HTML.
- Validate your HTML markup. Programs are available which will check your HTML file to ensure it is correct and conforms to the standard.
- Make sure your documents are usable on a range of browsers.
- Give a consistent appearance to a collection of documents.
- Use a template as a starting point for composing documents.

- Provide navigational aids at all times.
- Always give a concise, meaningful title in the HEAD part of the document.
- Limit pages to a manageable size.
- Don't use "Click here" for links.
- Make regular checks of the links in your documents.
- Minimise the size of graphics files and do not include large graphics files as inline images. Large files take much longer to download, slowing the network and irritating the viewer.
- Always provide an alternative to images, not everyone will be using graphical browsers.
- Provide a text alternative to hyperlinked graphics.

5.3 Codes of Practice

It is important to remember that intellectual property rights, defamation and data protection legislation all apply to the WWW as well as other forms of media. On top of this, most institutions will have their own codes of practice laying out acceptable practice, which you should read before setting up any web pages. Institutions are also bound by the JANET acceptable use policy, which governs the use of the links between the institutions.

Usually, such codes of practice prohibit inclusion of (or direct links to):

- official logos or similar material on personal, unofficial pages
- sexist, racist, pornographic or other similarly offensive material
- material to which a third party holds an intellectual property right, without the express written permission of the rightholder.
- defamatory material
- personal data about third parties, without their permission or registration with the Data Protection Registrar

For a sample code of practice and a full discussion of the legal aspects of the WWW, see Andrew Charlesworth's SIMA report, A Comprehensive Survey of the Legal Issues Relating to the Development and use of WWW Technology at Educational Sites.

6 Bibliography

SIMA Reports

http://info.mcc.ac.uk/CGU/SIMA/simapj.html
Running a World Wide Web Server, Brian Kelly

A Comprehensive Survey of the Legal Issues relating to the development and use of WWW Technology at Educational Sites, Andrew Charlesworth

Guide to good practices for WWW authors, Margaret Isaacs

Online Resources

Procedures for Creating a Home Page, University of Maine *http://maine.maine.edu/hpproc.html*

A Resource on WWW Search Engines, North Suburban Library System
http://www.nslsilus.org/youth/search/nslssrch.html

Creating Web Pages: A Tutorial Guide for Windows Users, Peter Christian, Computer Services Guide, Goldsmiths College
http://www.gold.ac.uk/guides/c606.htm

Introduction to HTML Version 2.0, Trevor McMullan, University of Ulster
http://www.edsm.ulst.ac.uk/trainhtml/

Entering the WWW - a Guide to Cyberspace, Kevin Hughes, Enterprise Integration Technologies, 1994
http://www.eit.com/web/www.guide/

Guide to Publishing on the Web, collection of online tutorials and resources
http://www.gn.apc.org/gn/webtools/howtohtml.html

A Beginner's Guide to HTML, Marc Andreessen
http://www.ncsa.uiuc.edu/demoweb/html-primer.html

HTML Validation Service
http://www.hensa.ac.uk/html-val-svc/

JANET Acceptable Use Policy, Editor: R A Day
http://www.ja.net/documents/use.html