

CCP5 information on the Daresbury World Wide Web server

Maurice Leslie

1 Introduction

The CCP5 community is widely distributed throughout the world and serves readers with computing facilities from the most sophisticated to very basic. This article has two sections. The first describes what World Wide Web is and gives sufficient information for readers to set up their own servers if they wish. This is followed by a description of information which may be obtained about CCP5 from the Daresbury server.

1.1 What is World Wide Web?

By far the easiest way to find out about World Wide Web is to use it. The following is therefore primarily aimed at those of our readers who do not have access to World Wide Web, or who's sites are not connected to Internet. In order not to interrupt the flow of the document too much I have included a glossary of the large number of terms and acronyms at the end.

World Wide Web (W³ or WWW) aims to give access in an uncomplicated way to information stored on a computer system anywhere in the world. The project was initiated at CERN with a view to coordinating the high-energy physics community. Consider as an analogy reading a scientific paper in your university library. Two of the activities which you will find yourself needing to do most frequently are to look up a cross-reference in the document and to search an index for a particular keyword. World Wide Web is designed to mimic both of these human activities.

To carry out the cross-referencing activity, World Wide Web documents are written in **hypertext**. This is just the same as ordinary text, except that it contains references within the text which are designed to be interpreted by the computer. These references are called **hypertext links**. They may refer to a separate part of the same document, another document on the same file system, or a document on another file system anywhere in the world. The hypertext links could refer to the author, giving background information, or to the institution. One consequence of this is that documents cease to be static; in the CCP5 library page described below the link from a program name will always point to the most up to date version. The cross-referenced documents may themselves have links; in this way the user is rapidly able to traverse the globe through the Web. Links in documents need not refer to items which are text. It is possible to refer to an item which is a graphical image, an audio sample or a movie, for which the term "hypermedia links"

has been coined. Links can also refer to “virtual documents”, created by running some program as a result of activating the link.

The activities of providing information and accessing information are kept distinct in WWW. Information providing programs, called servers, provide data requested by information gathering programs called clients. Accessing information is done by means of a ‘browser’ program, of which there are a number available. The most popular is “Mosaic” written by NCSA. This will run on most UNIX platforms, IBM-compatibles running Microsoft Windows and Macintoshes. There are also browsers for vt100 terminals and line-mode terminals. The fundamental difference between the browsers lies in how the hypermedia links are presented to the user and activated. In the case of mosaic, which is the only browser for which I have personal experience, links are highlighted using different colours on a colour screen and by underlining on a monochrome screen, and are activated using the mouse. When a browser is started, the users terminal will display a home page containing links to other documents. In the case of mosaic, by default the NCSA home page is displayed. In order to view other documents, the user either follows a link or provides a URL from the file menu.

It was a design aim of WWW that it should run on a wide variety of platforms. In order to achieve this aim, a number of new protocols and conventions needed to be established. WWW uses a hypertext data format called HTML, which is defined in terms of SGML. HTML is a simple text formatter with capabilities for document title, headings, lists, verbatim text, embedded images and character fonts. HTML also contains anchors, which mark the beginning (or occasionally the end) of a hypertext link. An anchor which is the beginning of a link will contain a document reference and sensitive text which will be highlighted by the browser that the user is running. The document reference needs to refer in an unambiguous way to the document to which it is linked. This is done by supplying a URL. A URL has a number of components, not all of which need to be given, within an anchor in a HTML documents. These components are

- A protocol which the client will use to obtain the document.
- The Internet address of the computer where the document is stored.
- The location of the document within the filesystem.
- A fragment-id, pointing to a location within the document.

A new network protocol HTTP was defined by the WWW project to give features not otherwise available. This is a protocol not only for the transfer of hypertext, but also it permits the client and server to communicate via the Internet, prior to sending the document, and agree on the most appropriate form of data transfer. Documents transferred need not be written in HTML, they can for example be in postscript. Examples in the next section should clarify this. WWW is able to access a wide number of existing services, including Gopher, anonymous ftp and WAIS.

In order to obtain further information, for readers without a WWW service, use anonymous ftp to

Site	Directory	Information supplied
info.cern.ch	/pub/www	Transfer and read README.txt
ftp.ncsa.uiuc.edu	Web	Transfer and read README.Mosaic

1.2 CCP5 information on WWW

This section is written to describe verbally the contents of the CCP5 WWW page, although it will be much simpler for readers with WWW browsers to try it out. The URL for the CCP5 home page is

<http://www.dl.ac.uk/CCP/CCP5/main.html>

I will now describe some of the contents using mosaic; readers should be aware that different browsers may behave differently. There is a brief introductory paragraph, followed by a menu of items. (history of the project, organisation, research interests, library, meetings and workshops, visitors, index of newsletter articles, copy of recent newsletters and registration form). To get this newsletter, select the *newsletters* option. Follow the directory path 41/ps and select the file `newsletter_part1.ps.gz`. This is a compressed postscript file. The client program at Daresbury transfers the compressed file, opens an external viewer and displays the newsletter. (Other browsers may not perform all of these operations). The newsletter can then be read on the screen. Suppose the user selects *library*. A new document is displayed with information about the program library. Further down this page is an index of programs in the library. Selecting *directory* in the introductory paragraph will display the entire directory in which the program library is stored. If you move *back* to the library page and select the entry MDMEGA, the code for the program is displayed on the terminal. To transfer a copy to the local host, open the file menu and use the save as option. You will be prompted for the name of the local file. Look at the entry for the program MOLDY. If the link BE is followed, the user finds that the Beeman algorithm is used. Move *back* to the MOLDY entry and follow the *manual* link. This link executes an anonymous ftp transfer to the program author's host computer, where the most up-to-date version will always be stored. Now move back to the library page and look at the NEMD entry. This is a binhexed Mackintosh file (binary and manual) (.hqx) The Daresbury browser in this case just requests a filename to download the file for subsequent processing as there is no sensible way to display the information.

The other capability of WWW is its ability to carry out searches of data. This is carried out using either its own search protocols, or by acting as an interface to one of the existing information services. (Gopher, xarchie, WAIS). A WWW browser can send a request to a client to search a document for a keyword. However, the whole question of how to create an index of information on WWW is currently under active debate. A number of projects have been set up in which a *robot*

automatically scans documents, following links in them, extracting titles and header information, or information contained within URL's within the documents. However, such programs can impose a heavy load on the Internet and it is not at present clear that this will be the best way to create indices. The information returned by the robot is often of limited use to the subsequent search of the database, being geographical in nature or covering too broad an area. Readers with WWW browsers are referred to the two documents below; the first is a general description of robots and the second is a useful searchable index.

URL	Document Title
http://web.nexor.co.uk/mak/doc/robots/robots.html	World Wide Web Wanderers, Spiders and Robots
http://www.stir.ac.uk/jsbin/js	JumpStation Front Page

I have constructed a list of links I have found which are relevant to CCP5. This is not complete, but I hope it will quickly give new people an idea of the scope of WWW. Readers are welcome to contribute to this list if they wish. I would also welcome comments from users of WWW on this article, either for information or for publication in this newsletter, by Email to m.leslie@dl.ac.uk.

Glossary

anchor	A point in an HTML document where there is a hypertext link.
browser	Interface between the user and the systems WWW client program.
client	Program run to access WWW information.
Gopher	A menu-driven information service.
HTML	Acronym for HyperText Markup Language, the data format used for hypertext on all WWW systems.
HTTP	HyperText Transfer Protocol.
hypermedia link	A link from a document to some form of information, stored elsewhere. (synonym hyperlink)
NCSA	National Centre for Supercomputing Applications
robot	A program which recursively follows links in documents in order to build up a database of information about the whole or part of WWW.
server	Program which supplies information.
SGML	Standard Generalised Markup Language, an ISO standard for defining structured document types and markup languages.
URL	A Uniform Resource Locator, a syntax for expressing the names and addresses of objects on the network.
WAIS	Wide-Area Information Service