The World-Wide Web Initiative

T. Berners-Lee[1], R. Cailliau[2], N. Pellow[3], A. Secret[4]

Abstract

The World-Wide Web (W3) is a way of viewing all the online information available on the Internet as a seamless, browsable continuum. Using hypertext jumps and searches, the user navigates through an information world partly hand-authored, partly computer-generated from existing databases and information systems. The web today incorporates all information from more basic information systems such as Gopher as WAIS, as well as sophisticated multimedia and hypertext information from many organizations. As a user interface tool, W3 clients provides a comprehensive point-and-click network access tools, while W3 servers provide an efficient, friendly method of providing data to real users. This paper answers some commonly asked questions about W3, such as those comparing it with other systems, and about recent developments and future directions global hypermedia will take.

I. What is the World-Wide Web?

If you haven't come across the web before, the best way to find out about it is to try it. At the end of this paper are some recipes for getting hold of W3 clients. Given one of these, you will quickly find out all you need to know, and much more besides. For something to read on the plane, or you don't have Internet access from your desk top machine, you could read our paper in "Electronic Networking" [W3] for an overview of the project, material which we will not repeat here.

To summarize, a W3 "client" program runs on your computer. It displays a "document" fetched from another computer, the "server". The reader can either request a search, typing in plain text (or complex commands) to send to the server, or can follow a link from a highlighted phrase to another document. In either case, the client sends a request off to the server, often a completely different machine in some other part of the world, and within (typically) a second, the related information, hypertext again, plain text or multimedia, is presented. This is done repeatedly, and by a sequence of selections and searches one can find anything which is "out there". Some important things to note are:

- Whatever type of server, the user interface is the same, so users do not need to understand the differences between the many protocols in common use;
- Links can point to anything which can be displayed (including search result lists);
- While menus and directories can be used, the extra option of hypertext gives a more powerful communication tool, but you do not have to write hypertext to run a W3 server.
- There is a very extendable system for introducing new formats for multimedia data;
- There are many clients for different platforms, each displaying formatted hypertext as well as it can on that given platform.

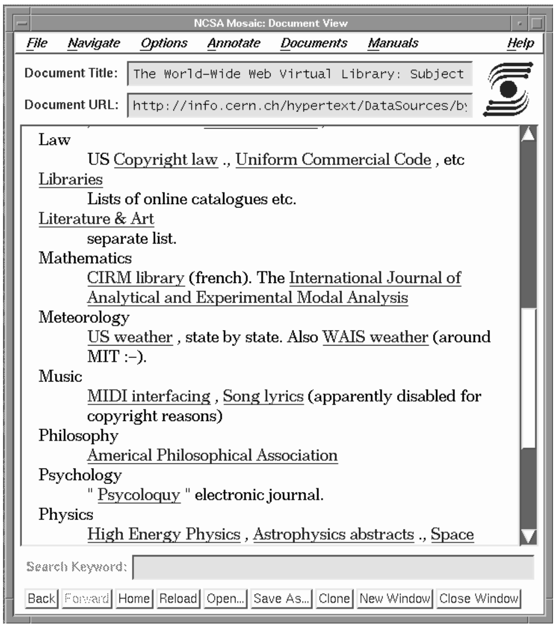

Figure 1. A view of NCSA's Mosaic W3 browser for X. Here we are looking at part of a list of Internet resources by subject. The document title and address ("URL") are displayed at the top. Linked text is underlined here, although color is normally used when available. The "Search keyword" box at the bottom is enabled when the server allows for a search

1 Tim Berners-Lee is with CERN, the European Laboratory for Particle Physics in Geneva and may be reached as timbl@info.cern.ch.

2 Robert Cailliau is at CERN and may be reached as cailliau@cernnext.cern.ch.

3 Nicola Pellow is freelance at CERN for a short while, future e-mail uncertain.

4, Arthur Secret is Technical Student at CERN, and may be reached as secret@dxcern.cern.ch

at this point. Buttons provide the usual navigation aids, and a "hot list" of pointers to book-marked documents is saved between sessions.

II. What does W3 define?

W3 has come to stand for a number of things, which it is as well to distinguish. These include

- The idea of a world in which all information items have a reference by which they can be retrieved;
- The addressing system [URL] which the project put in to make this possible, with many different servers;
- A network protocol [HTTP] used by native W3 servers giving performance and features not otherwise available;
- A markup language [HTML] which every W3 client is required to understand, and is used for the transmission of basic things such as text, menus and simple online help information across the net;
- The body of data available on the Internet using all or some of the above.

It is worth clarifying the relationships here:

The URL syntax allows objects (menus, documents, etc.) to be addressed not only using HTTP, but also using the other common networked information protocols in use today ([FTP], [NNTP], [Gopher], and [WAIS]). The URL syntax is defined by an Internet Draft document which, when a standard, will allow different applications to exchange pointers to information on the network.

Rather than being a protocol for transferring hypertext, HTTP is a protocol for retrieving information with the efficiency necessary for making hypertext jumps. HTTP does not only transfer HTML documents. Although HTML comprehension is required of W3 clients, so that menus can be presented, HTTP is used for retrieving documents in an unbounded and extensible set of formats. This is done by the client sending a list of the formats it can handle, and the server replying with data in any format which it also can produce. This allows proprietary formats to be used between consenting programs in private, without the need for standardization of those formats. This is important both for high-end users who share data in sophisticated forms, and also as a hook for formats which have yet to be invented.

HTML is an Internet protocol. It is similar in style to the FTP and NNTP protocols use to transfer files and news on the internet for many years.

HTML is a markup language which does not have to be used with HTTP. It can be used in hypertext mail (it is proposed as a format for Multipurpose Internet Mail Extensions, [MIME]), news, and anywhere where basic hypertext is needed. It includes simple structure elements, such as several levels of headings, bulleted lists, menus and compact lists, all of which are useful when presenting choices, and in online documents. Whilst being in fact a valid SGML format (the DTD is an appendix to the specification), it is very simple, and many servers actually work by running small programs to generate HTML on the fly in response to client requests.

W3 and Other Systems

Two other systems, WAIS and Gopher, share W3's client-server architecture and a certain amount of its functionality. Table 1 indicates some of the differences. (Note: registered server figures taken 27 April 1993. WAIS: from TMC directory, number of

	WAIS	**Gopher**	**W3**
Original target application:	Information retrieval (IR)	Campus-wide information (CWIS)	Collaborative work (CSCW)
Protocol Style:	ISO	Internet	Internet
Formats:	yes	yes	yes
Plain text	with keywords	yes	yes
Graphics	-	-	yes
Hypertext	-	-	yes
Functions			
Search	yes	yes	yes
Relevance feedback	yes	-	-
Reference to other servers	-	yes	yes
Registered servers (approx.)	113	455	62

Table 1: Features of networked information systems

distinct hosts. Gopher: from "All the Gophers in the world" register at UMN; W3: from Geographical list kept at CERN. In all cases more servers exist which are not directly registered, so these are a very rough guide only with no indication of quantity or quality of information at each host)

The WAIS protocol is influenced largely by the z39.50 protocol designed for networking library catalogues. It allows a text-based search, and retrieval following a search. Indexes to be searched are found by searching in a master index. This two-stage search has been demonstrated to be sufficiently powerful to cover the current world of WAIS data. There are however no navigational tools to allow the reader to be shown the available resources, or guided through the data: the reader is parachuted in to a hopefully relevant spot in the information world, but left without context.

	WAIS	Gopher	W3
WAIS	yes	gateway	yes (v. 2.04)
Gopher	-	yes	yes
HTTP	-	-	yes
FTP	-	gateway	yes
Net news	-	gateway	yes

Table 2. protocols supported by clients.

Gopher provides a WAIS-like search mechanism, but principally uses menus. A menu is a list of titles, from which the user may pick one. While gopher space is in fact a web containing many loops, the menu system gives the user the impression of a tree. The Veronica server provides a master index for gopher space.

The W3 data model is similar to the gopher model, except that menus are generalized to hypertext documents. In both cases, simple file servers generate the menus or hypertext directly from the file structure of a server. The W3 hypertext model gives the program more power to communicate the options available to the reader, as it can include headings, and various forms of list structure for example within the hypertext. W3 uses the conventions of the MIME project for multimedia mail, so opening the door to integration of hypermedia mail, news, and information retrieval.

All three systems allow for the provision of graphics, sound and video, although as the WAIS system only has access by text search, text has to be associated with graphics files to allow them to be found.

The uptake of the systems has been largely affected by the software available. A public domain version of WAIS provided an indexing system which in fact was used as the search engine of several Gopher and W3 servers. The existence of free clients on small platforms (Mac and PC) for Gopher was a major feature in its popularity, as was the simplicity of setting up the servers, which were at the time available free of charge. W3 enthusiasm was initially greater in the high end, among workstation users for whom the best clients existed, and for those providers who demanded the quality of service provided by a hypertext interface. The recent availability of clients for PC and Mac for W3, and the existence of "drop-in" servers for unix systems is now starting to change this picture.

W3 clients provide access to servers of all types, as a single simple interface to the whole web is considered very important. Unknown to the user, several protocols are in use behind the scenes. Whereas one would not wish to see greater proliferation of protocols, the existence of more than one protocol probably allows for the most rapid progress during this phase in the development of the field.

III. Recent W3 developments

When initially the push was for a good X11 interface for W3, there are now no less than five X-based browsers. Of these, NCSA's "Mosaic" for X, released for free use by the academic public, has achieved great popularity. At the opposite end of the scale, five terminal-based browsers cater for different needs of vt100 and dumb terminal based users, the "Lynx" full-screen vt100-based browser from the University of Kansas being the latest addition. Whilst NCSA is currently developing Mosaic ports to Macintosh and PC/Windows platforms, CERN has a simple Mac browser, and the Legal Information Institute at Cornell Law School says its existing but currently unreleased Windows-based "Cello" browser will be made available in July. One measure of the uptake rate of clients is the load on the "info" W3 server at CERN, which has more than doubled every 4 months over the last two years.

Information providers have also blossomed. Some of these provide simple overviews of what is available at particular institutes or in particular fields. Others use the power of the W3 model to provide a virtual world of great richness. Examples of servers which use hypertext in interesting ways are the RAL-Durham Particle Database, the Legal Information Institute's hypertexts of several great tomes of American law.

The kernel W3 code (a common code library, and basic server and clients) from CERN is in the public domain. (All protocols and specifications are public domain). The latest release of the library now includes, when used in conjunction with the free-WAIS product from CNIDR, direct access to WAIS servers from W3 clients without the necessity of passing through gateways.

IV. The future

The W3 initiative sits at the meeting point of many fields of technology. Users put pressure and effort into bringing about the adoption of W3 in new areas. Some of the developments which we look forward to in the next few years include

- The implementation of a name service which will allow documents to be referenced by name, independent of their location;
- Hypertext editors will allow non-expert users to make hypertext links to organize published information. This will bring the goal of computer-supported collaboration closer, with front-end update, and annotation.
- The development of a common syntax for forms to be filled in by users, allowing users to perform data input to servers for numerous applications;
- More sophisticated document type definitions providing for the needs of commercial publishers of online material;
- The development of a common format for hypertext links from two-dimensional images will give more exciting interface possibilities;
- Easy-to-use servers for low-end machines will increase the rate of publication of information by small groups and individuals;
- Conventions on the Internet for charging and commercial use will allow direct access to for-profit services.

V. Getting Started

If you have an X-windows based workstation, then pick up the xmosaic product by anonymous FTP from ftp.ncsa.uiuc.edu, directory Web/xmosaic. Just get the binary (sources are available if the binary is not there for your machine), uncompress it, set it executable and run it.

If you have a vt100 terminal, you can try out a full screen interface by telnetting to ukanaix.cc.ukans.edu and logging in as www. With any terminal, you can telnet to info.cern.ch for the simplest interface. These browsers are also available in source and in some cases binary form. Details of status and coordinates of all browsers are available on the web - just follow a link to World-Wide Web, and select "software available". Line mode and Mac browsers are both available from info.cern.ch in directories /pub/www/bin and /pub/www/src.

Hopefully after reading this article you will have an idea of what W3 is, where it fits in with other systems in the field, and where it is going. There is much more to be said, in particular about the providing of data, but this is described on the web.

VI. Glossary with References

FTP File Transfer Protocol. J. Postel and J. Reynolds, File Transfer Protocol, Internet RFC 959, October 1985.

Gopher The Internet Gopher. F. Anklesaria et.al., The Internet Gopher Protocol, Internet RFC 1436, March 1993.

HTML Hypertext Markup language. T. Berners-Lee and D. Connolly, Hypertext Markup Language, ftp://info.cern.ch/pub/www/doc/html-spec.ps, .txt

HTTP Hypertext Transfer Protocol. T. Berners-Lee, Hypertext Transfer Protocol, ftp://info.cern.ch/pub/www/doc/http-spec.ps, .txt

MIME Mulipurpose Internet Mail Extensions. N. Borenstein and N. Freed, MIME (Multipurpose Internet Mail Extensions): Mechanisms for Specifying and Describing the Format of Internet Message Bodies", Internet RFC 1341, June 1992

NNTP Network News Transfer Protocol. Kantor, B & Lapsley, P (1986). A proposed standard for the transmission of news. Internet RFC 977

URL Uniform Resource Locator. T. Berners-Lee, Uniform Resource Locators, ftp://info.cern.ch/pub/ietf/url4.ps, .txt, tempoarily in discussion as Internet Draft draft-ietf-uri-url-00

WAIS Wide Area Information Servers. See T. Addyman, WAIS: Strengths, Weaknesses and Opportunities, proceedings of "Information Networking 93", London, May 1993, Meckler, London.

W3 T.J. Berners-Lee, R. Cailliau, J-F Groff, B. Pollermann, CERN, "World-Wide Web: The Information Universe", published in "Electronic Networking: Research, Applications and Policy", Vol. 2 No 1, pp. 52-58 Spring 1992, Meckler Publishing, Westport, CT, USA. See also documents in ftp://info.cern.ch/pub/www/doc and information referenced by http://info.cern.ch/hypertext/WWW/TheProject.html

Overviews of WAIS and Gopher are also in the same issue as [W3]

Registered servers figures culled from documents with URLs as follows:

wais://quake.think.com/directory of servers

gopher://gopher.tc.umn.edu:70/11/Other%20Gopher%20and%20Information%20Servers/all

http://info.cern.ch/hypertext/DataSources/WWW/Geographical.html

Author Information

Tim Berners-Lee originated the World-Wide Web in 1990, to enable the sharing of knowledge in complex distributed groups of teams. At CERN, the European Particle Physics Laboratory in Geneva, Switzerland, he now coordinates W3 development by collaborating institutes around the world. Originally acquiring his BA in physics from Oxford University in 1976, he immediately moved into commercial software. His experience since has been in text processing, graphics, and communications software, and system design with a number of companies in the United Kingdom. In 1984 Tim moved to CERN where, prior to initiating the W3 project, he developed tools for distributed real-time software. Tim may be reached at timbl@info.cern.ch

Formerly in programming language design and compiler construction, Robert Cailliau has been interested in document production since 1975, when he designed and implemented a widely used document markup and formatting system. He ran CERN's Office Computing Systems group from 87 to 89. He is a long-time user of Hypercard, and has been working on W3 since 1991, contributing many ideas, and software for the Macintosh. He now coordinates the use of W3 by CERN experiments and other physics institutes. Robert's mail address is cailliau@cernnext.cern.ch.

Nicola Pellow first worked on the W3 project from November 1990 to August 1991, when she wrote the original line mode W3 browser. A graduate of Leicester Polytechnic, UK, Nicola is currently visiting CERN again and is working on the Macintosh browser.

Arthur Secret wrote the first Gateway giving W3 access to a relational database in 1992. He is studying Computer Science at Ecole Internationale des Sciences du Traitement de l'Information in Paris, France, but currently in CERN’s Technical Student Programme, where he is working on many aspects of W3 software.