World-Wide Web

Tim Berners-Lee, Robert Cailliau
C.E.R.N.
CH - 1211 Genève 23
timbl@info.cern.ch, cailliau@cernnext.cern.ch

Abstract

The W3 project merges networked information retrieval and hypertext to make an easy but powerful global information system. It aims to allow information sharing within internationally dispersed groups of users, and the creation and dissemination of information by support groups. W3's ability to provide implementation-independent access to data and documentation is ideal for a large HEP collaboration. W3 now defines the state of the art in networked information retrieval, for user support, resource discovery and collaborative work. W3 originated at CERN and is in use at CERN, FNAL, NIKHEF, SLAC and other laboratories. This paper gives a brief overview and reports the current status of the project.

Introduction

The World-Wide Web (W3) project allows access to the universe of online information using two simple user interface operations. It operates without regard to where information is, how it is stored, or what system is used to manage it. Previous papers give general [1] and technical [2] overviews which will not be repeated here. This paper reviews the basic operation of the system, and reports the status of W3 software and information.

Operation

The W3 world view is of ***documents*** referring to each other by ***links***. For its likeness to a spider's construction, this world is called the ***Web***. This simple view is known as the ***hypertext*** paradigm. The reader sees on the screen a document with sensitive parts of text representing the links. A link is followed by mere pointing and clicking (or typing reference numbers if a mouse is not available).

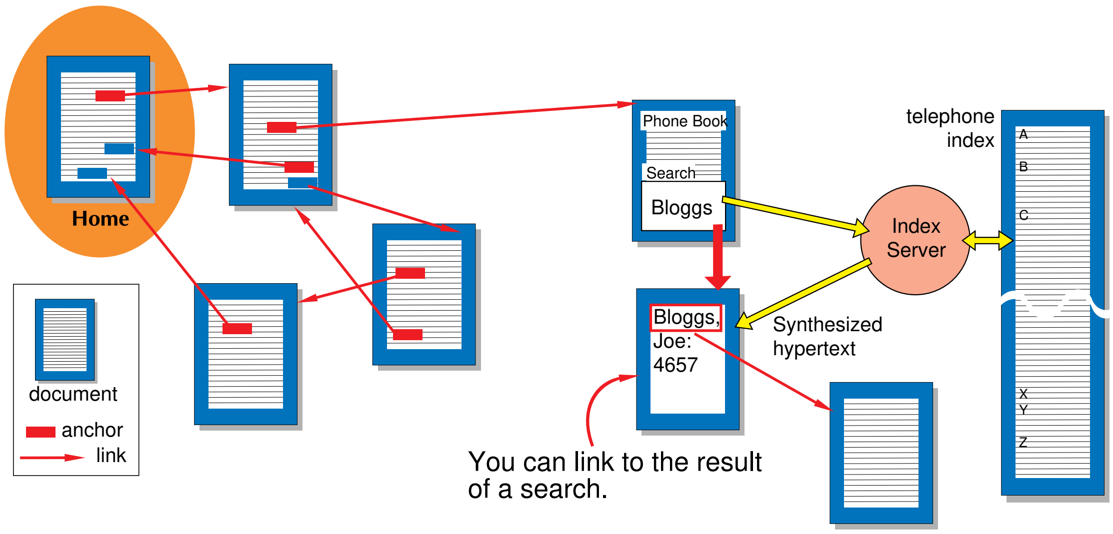

Fig 1. The basic hypertext model is enhanced by searches.

[An invited talk at the conference: *Computing in High Energy Physics 92*, Annecy, France, 23-27 September 1992. This copy is a preprint of a paper to be published in the conference proceedings as a CERN yellow report.]

Hypertext alone is not practical when dealing with large sets of structured information such as are contained in data bases: adding a ***search*** to the hypertext model gives W3 its full power (fig. 1). Indexes are special documents which, rather than being read, may be searched. To search an index, a reader gives keywords (or other search criteria). The result of a search is another document containing links to the documents found.

The architecture of W3 (fig. 2) is one of browsers (clients) which know how to *present* data but not what its origin is, and servers which know how to *extract* data but are ignorant of how they will be presented. Servers and clients are unaware of the details of each other's operating system quirks and exotic data formats.

All the data in the Web is presented with a uniform human interface (Fig. 3). The documents are stored (or generated byalgorithms) throughout the internet by computers with different operating systems and data formats. Following a link from the SLAC home page (the entry into the Web of a SLAC user) to the NIKHEF telephone book is as easy and quick as following the link to a SLAC Working Note.

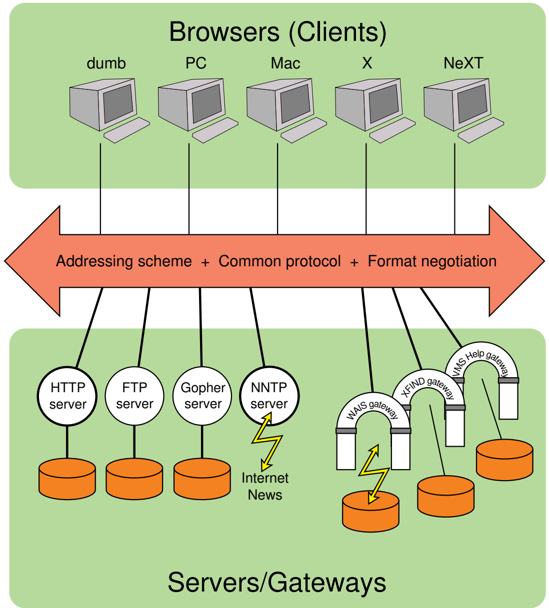

Fig. 2: Architecture of W3

Providing Information

Authors can create documents by simply typing files (in plain text, using hypertext SGML markup or a W3 editor) and linking them into the Web. This is most useful in collaborative work: the latest text is accessible on-line, no copies, drafts or out-of-date printouts. If the data is stored in an existing data-base, a server can be tailored to provide its data to the Web. Hypertext links may be made to any data in non-W3 servers (FTP, Gopher, WAIS or internet news) as W3 clients have the ability to present all such data as hypertext.

In the case of an existing information system containing a large mass of information, one should consider writing a server to provide a hypertext view of the data without touching the data itself or the procedures by which the database is maintained (Fig. 4). An existing server may be taken as an example to be modified and enhanced to provide the functionality required. Typically, it is modified to call a program which already exists to access the data. The server merely reformats the W3 document address (and/or search criteria) into a request to the program, and then reformats the program output as hypertext.

[An invited talk at the conference: *Computing in High Energy Physics 92*, Annecy, France, 23-27 September 1992.
This copy is a preprint of a paper to be published in the conference proceedings as a CERN yellow report.]

Software status

The success of the W3 initiative can be attributed to enthusiasts and collaborators in many institutes. The W3 team at CERN has incorporated some of their work into software releases; other work is distributed and maintained by the original authors. This is a summary: details are available on the Web.

Client software

The initial prototype development for W3 clients was done on two platforms. A "dumb terminal" browser was written at CERN by Nicola Pellow to demonstrate access from lowest common denominator platforms supporting only a C compiler and internet access. This program is now a mature product much in demand both as a simple interactive browser and as a general data access and text formatting tool which can be built into more complex programs.

The prototype window-oriented browser and hypertext editor was developed on a NeXTStep™ platform. It has been frozen in its prototype form until further notice.

For X-Windows, four clients exist, at various levels of development between alpha and beta test. Differing principally in the underlying toolkits used, each has different and interesting possibilities of extension. Sources of all four are available:

> The ***ViolaWWW*** client was written and is maintained by Pei Wei of O'Reilly Associates. It is a fully-fledged hypertext browser with search facility, bookmarks and history recall panel. At beta test level, this browser has to date been ported to SGI, Sun, IBM rs6000 and DECstation™ platforms.
>
> The ***MidasWWW*** client has a Motif look-and-feel. It was written recently by Tony Johnson of SLAC using his Midas toolkit.
>
> The ***tkWWW*** client was written by Joseph Wang at the MIT Athena project, based on the existing "tk" toolkit.
>
> The ***Erwise*** W3 client was written as a student project by four students at the Helsinki Technical University, and is not maintained.

A Macintosh client is being written at CERN with help from FNAL as a stand-alone Macintosh application for any Mac with TCP-IP.

For the IBM-compatible PC, a W3 browser is being written on top of Microsoft's Word for Windows as a CERN supported student project by Alain Favre, with CNAM, France. Neither Mac nor PC browser is available at the time of writing.

The clients share a common library of network information access code which is available separately.

[An invited talk at the conference: *Computing in High Energy Physics 92*, Annecy, France, 23-27 September 1992.
This copy is a preprint of a paper to be published in the conference proceedings as a CERN yellow report.]

Server software

Currently, W3 servers exist for Unix, VMS and VM and must be configured by system managers. When servers for personal computers are available, we expect a great increase in publishing directly by authors, reviewers and documentation managers. Existing servers include those for:

Files File servers run on Unix, VMS or VM to distribute existing files to hypertext browsers. Directories of the file system are represented as hypertext lists of the files they contain. Authors may provide plain text files or marked-up hypertext. Any anonymous FTP server may also be accessed by the W3 clients with some speed penalty compared to a W3 server.

VMS/Help For information in VMS/Help™ format, a server runs under VMS.

Oracle A generic Oracle™ server has been written by Arthur Secret (CERN/EISTI) to allow access to Oracle databases from W3 clients. This currently accepts SQL "select" statements as search terms and runs under Unix.

GNU Info Written by Philippe Defert (CERN), this "perl" script runs under Unix and provides an existing Gnu Info database of online documentation as hypertext.

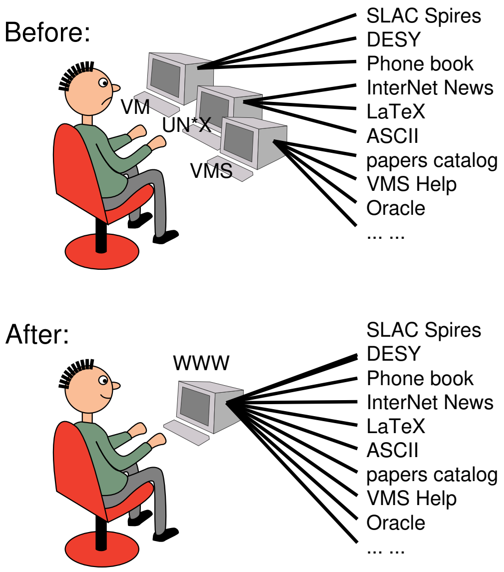

Fig 3: Unification for the user

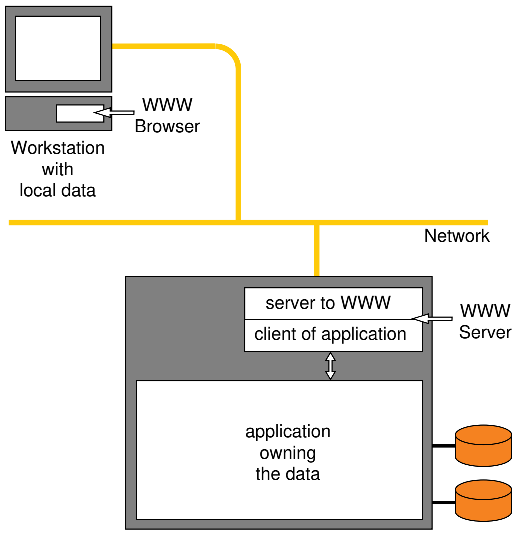

Fig. 4: No operative changes for the provider

[An invited talk at the conference: *Computing in High Energy Physics 92*, Annecy, France, 23-27 September 1992. This copy is a preprint of a paper to be published in the conference proceedings as a CERN yellow report.]

The Spread of the Web

Over the last year, the existence of browsers has prompted several HEP institutes, and several other sites, to put up W3 servers. Thanks to the creativity and vision of those involved, there is a great variety of information available. Whilst the most commonly accessed may be "phone book"-type information from CERN, NIKHEF and SLAC, there is also deeper online documentation. Figure 4 shows locations of some current and prospective server sites (note: Archie, Gopher and WAIS are themselves network information systems, accessible through W3 as a subset of the Web. Only their location of origin is shown).

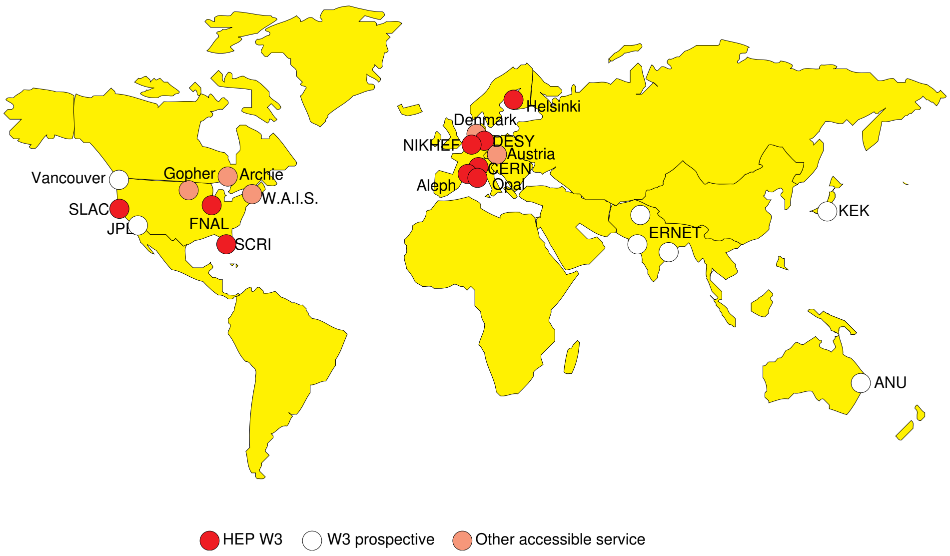

Fig. 4: known servers at September 92

Aleph, Opal, and SLD (and, experimentally, D0) have experiment-specific information. At Fermilab, the existing documentation schemes for online and offline systems have been made available among other things. At SLAC, the "WWWizards" have servers running on VM and Unix, making available the "SPIRES" database information (including the popular preprint index), and a database about the "FreeHEP" software collection.

Future enhancements

The next generation of the W3 protocol is being tested at CERN by Carl Barker of Brunel University. The protocol provides simple password based control over access to sensitive information. It also allows client and server programs to negotiate commonly understood data formats. This will open the door to the transfer of graphics files, as well as raw numeric data for manipulation and analysis, and text in an open set of formats such as word processor formats, TeX, and general SGML document types.

We have W3 servers and clients running over DECnet™, and aim to release a DECnet-Internet W3 gateway in the near future.

[An invited talk at the conference: *Computing in High Energy Physics 92*, Annecy, France, 23-27 September 1992. This copy is a preprint of a paper to be published in the conference proceedings as a CERN yellow report.]

Getting started

To further the spread of W3, we suggest you encourage the adoption of first client programs and then servers at your own site. Information on how to get and install W3 code is available from our line-mode server. On a computer connected to the internet, type:

```
telnet info.cern.ch
```

This will give you the very basic line mode interface. Use it to find out how to install that or more advanced browsers on your local system. The source code, binaries and various articles are available by anonymous FTP from info.cern.ch, under directory /pub/www. (info.cern.ch is currently 128.141.201.74). All information about the W3 project is on the web, so you should look for it there and mail us if you find problems. A plain text version of the installation instructions is included in the tar file for the line mode browser.

Also on the Web is information about how to subscribe to the mailing lists we keep for announcements of new facilities and for technical discussions. The W3 team at CERN will also try to answer questions sent to www-bug@info.cern.ch.

References

[1] 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, Spring 1992, Meckler Publishing, Westport, CT, USA. *See this paper for other technical references*.

[2] T.J. Berners-Lee, R. Cailliau, J-F Groff, B. Pollermann, CERN, "World-Wide Web: An Information Infrastructure for High-Energy Physics", Presented at "Artificial Intelligence and Software Engineering for High Energy Physics" in La Londe, France, January 1992. Proceedings published by World Scientific, Singapore, ed. D Perret-Gallix.

[An invited talk at the conference: *Computing in High Energy Physics 92*, Annecy, France, 23-27 September 1992. This copy is a preprint of a paper to be published in the conference proceedings as a CERN yellow report.]