WorldWideWeb

Proposal for a HyperText Project

To: P.G. Innocenti/ECP, G. Kellner/ECP, D.O. Williams/CN

Cc: R. Brun/CN, K. Gieselmann/ECP, R. Jones/ECP, T. Osborne/CN, P. Palazzi/ECP, N. Pellow/CN, B. Pollermann/CN, E.M. Rimmer/ECP

From: T. Berners-Lee/CN, R. Cailliau/ECP

Date: 12 November 1990

The attached document describes in more detail a Hypertext project.

HyperText is a way to **link** and **access** information of various kinds as a web of nodes in which the user can **browse** at will. It provides a **single user-interface** to large classes of information (reports, notes, data-bases, computer documentation and on-line help). We propose a simple scheme incorporating **servers** already available at CERN.

The project has **two phases**: firstly we make use of existing software and hardware as well as implementing simple browsers for the user's workstations, based on an analysis of the requirements for information access needs by experiments. Secondly, we extend the application area by also allowing the users to add new material.
Phase one should take 3 months with the full manpower complement, phase two a further 3 months, but this phase is more open-ended, and a review of needs and wishes will be incorporated into it.

The **manpower** required is 4 software engineers and a programmer, (one of which could be a Fellow). Each person works on a specific part (eg. specific platform support).
Each person will require a state-of-the-art **workstation**, but there must be one of each of the supported types. These will cost from 10 to 20k each, totalling 50k. In addition, we would like to use **commercially available software** as much as possible, and foresee an expense of 30k during development for one-user licences, visits to existing installations and consultancy.
We will assume that the project can rely on some **computing support** at no cost: development file space on existing development systems, installation and system manager support for daemon software.

T. Berners-Lee R. Cailliau

WorldWideWeb:

Proposal for a HyperText Project

T. Berners-Lee / CN, R. Cailliau / ECP

Abstract: *HyperText is a way to link and access information of various kinds as a web of nodes in which the user can browse at will. Potentially, HyperText provides a single user-interface to many large classes of stored information such as reports, notes, data-bases, computer documentation and on-line systems help. We propose the implementation of a simple scheme to incorporate several different servers of machine-stored information already available at CERN, including an analysis of the requirements for information access needs by experiments.*

Introduction

The current incompatibilities of the platforms and tools make it impossible to access existing information through a common interface, leading to waste of time, frustration and obsolete answers to simple data lookup. There is a potential large benefit from the integration of a variety of systems in a way which allows a user to follow links pointing from one piece of information to another one. This forming of a web of information nodes rather than a hierarchical tree or an ordered list is the basic concept behind HyperText.

At CERN, a variety of data is already available: reports, experiment data, personnel data, electronic mail address lists, computer documentation, experiment documentation, and many other sets of data are spinning around on computer discs continuously. It is however impossible to "jump" from one set to another in an automatic way: once you found out that the name of Joe Bloggs is listed in an incomplete description of some on-line software, it is not straightforward to find his current electronic mail address. Usually, you will have to use a different lookup-method on a different computer with a different user interface. Once you have located information, it is hard to keep a link to it or to make a private note about it that you will later be able to find quickly.

Hypertext concepts

The principles of hypertext, and their applicability to the CERN environment, are discussed more fully in [1], a glossary of technical terms is given in [2]. Here we give a short presentation of hypertext.

A program which provides access to the hypertext world we call a ***browser***. When starting a hypertext browser on your workstation, you will first be presented with a hypertext page which is personal to ***you***: your personal notes, if you like. A hypertext page has pieces of text which refer to other texts. Such references are highlighted and can be selected with a mouse (on dumb terminals, they would appear in a numbered list and selection would be done by entering a number). When you select a reference, the browser presents you with the text which is referenced: you have made the browser follow a hypertext ***link***:

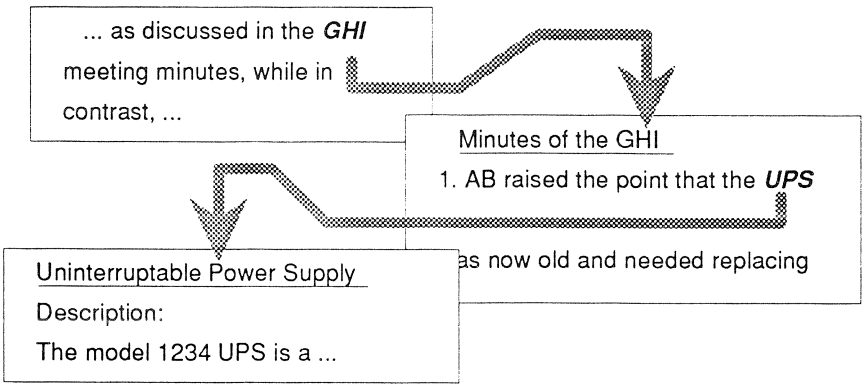

Fig. 1: hypertext links.

That text itself has links to other texts and so on. In fig. 1, clicking on the ***GHI*** would take you to the minutes of that meeting. There you would get interested in the discussion of the UPS, and click on the highlighted word ***UPS*** to find out more about it.

The texts are linked together in a way that one can go from one concept to another to find the information one wants. The network of links is called a ***web***. The web need not be hierarchical, and therefore it is not necessary to "climb up a tree" all the way again before you can go down to a different but related subject. The web is also not complete, since it is hard to imagine that all the possible links would be put in by authors. Yet a small number of links is usually sufficient for getting from anywhere to anywhere else in a small number of hops.

The texts are known as ***nodes***. The process of proceeding from node to node is called ***navigation***. Nodes do not need to be on the same machine: links may point across machine boundaries. Having a world wide web implies some solutions must be found for problems such as different access protocols and different node content formats. These issues are addressed by our proposal.

Nodes can in principle also contain non-text information such as diagrams, pictures, sound, animation etc. The term ***hypermedia*** is simply the expansion of the hypertext idea to these other media. Where facilities already exist, we aim to allow graphics interchange, but in this project, we concentrate on the universal readership for text, rather than on graphics.

Applications

The application of a universal hypertext system, once in place, will cover many areas such as document registration, on-line help, project documentation, news schemes and so on. It would be inappropriate for us (rather than those responsible) to suggest specific areas, but experiment online help, accelerator online help, assistance for computer center operators, and the dissemination of information by central services such as the user office and CN and ECP divisions are obvious candidates. WorldWideWeb (or W^3) intends to cater for these services across the HEP community.

Scope: Objectives and non-Objectives

The project will operate in a certain well-defined subset of the subject area often associated with the "Hypertext" tag. It will aim:

- to provide a common (simple) protocol for requesting human readable information stored at a remote system, using networks;
- to provide a protocol within which information can automatically be exchanged in a format common to the supplier and the consumer;
- to provide some method of reading at least text (if not graphics) using a large proportion of the computer screens in use at CERN;
- to provide and maintain at least one collection of documents, into which users may (but are not bound to) put their documents. This collection will include much existing data. (This is partly to give us first hand experience of use of the system, and partly because members of the project will already have documentation for which they are responsible)
- to provide a keyword search option, in addition to navigation by following references, using any new or existing indexes (such as the CERNVM FIND indexes). The result of a keyword search is simply a hypertext document consisting of a list of references to nodes which match the keywords.
- to allow private individually managed collections of documents to be linked to those in other collections.
- to use public domain software wherever possible, or interface to proprietary systems which already exist.
- to provide the software for the above free of charge to anyone.

The project will ***not*** aim

- to provide conversions where they do not exist between the many document storage formats at CERN, although providing a framework into which such conversion utilities can fit;
- to force users to use any particular word processor, or mark-up format;
- to do research into fancy multimedia facilities such as sound and video;
- to use sophisticated network authorisation systems. data will be either readable by the world (literally), or will be readable only on one file system, in which case the file system's protection system will be used for privacy. All network traffic will be public.

Requirements Analysis

In order to ensure response to real needs, a requirements analysis for the information access needs of a large CERN experiment will be conducted at the very start, in parallel with the first project phase.

This analysis will at first be limited to the activities of the members of the Aleph experiment, and later be extended to at least one other experiment. An overview will be made of the information generation, storage and retrieval, independent of the form (machine, paper) and independent of the finality (experiment, administration).

The result should be:

1. lists of sources, depots and sinks of information,
2. lists of formats,
3. diagrams of flow,
4. statistics on traffic,
5. estimated levels of importance of flows,
6. lists of client desires and / or suggested improvements,
7. estimated levels of satisfaction with platforms,
8. estimated urgency for improvements.

This analysis will itself not propose solutions or improvements, but its results will guide the project.

Architecture

The architecture of the hypertext world is one of data stored on server machines, and client processes on the same or other machines. The machines are linked by some network (fig. 2).

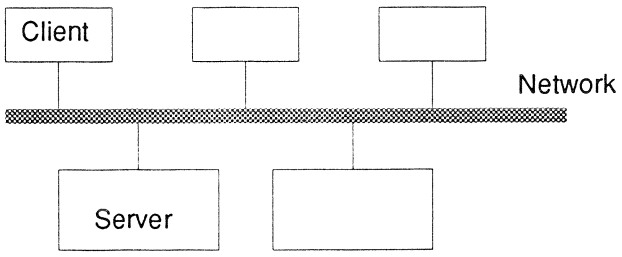

Fig. 2: proposed model for the hypertext world

A workstation is either an independent machine in your office or a terminal connected to a close-by computer, and connected to the same network. The servers are active processes that reply to requests. The hypertext data is explicitly accessible to them. Servers can be many on the same computer system, but then each caters to a specific hypertext base. Clients are browser processes, usually but not necessarily on a different computer system. Information passed is of two kinds: nodes and links.

Building blocks

Browsers and servers are the two building blocks to be provided.

- A *browser* is a native application program running on the client machine:
  - it performs the display of a hypertext node using the client hardware & software environment. For example, a Macintosh browser will use the Macintosh interface look-and-feel.
  - it performs the traversal of links. For example, when using a Macintosh to browse on CERNVM FIND it will be the Macintosh browser which remembers which links were traversed, how to go back etc., whereas the CERNVM server just responds by handing the browser nodes, and has no idea of which nodes the user has visited.
  - it performs the negotiation of formats in dialog with the server. For example, a browser for a VT100 type display will always negotiate ASCII text only, whereas a Macintosh browser might be constructed to accept PostScript or SGML.
- A *server* is a native application program running on the server machine:
  - it manages a web of nodes on that machine.
  - it negotiates the presentation format with the browser, performing on-the-fly (or cached) conversions from its own internal format, if any..

Operation

A link is specified as an ASCII string from which the browser can deduce a suitable method of contacting an appropriate server. When a link is followed, the browser addresses the request for the node to the server. The server therefore has nothing to know about other servers or other webs and can be kept simple.

Once the server has located the requested node, it will know from the node contents what the node's format is (eg. pure ASCII, marked-up, word processor storage and which word processor etc.). The server then begins a negotiation with the browser, in which they decide between them what format is acceptable for display on the user's screen. This negotiation will be based only on existing conversion programs and formats: it is not in the scope of W^3 to write new converters. The last resort in the negotiation is the binary transfer of the node contents to a file in the user's file space. Negotiating the format for presentation is particular to W^3.

Project phases

Provided with resources mentioned below, we foresee the first two phases of the project as achieving the following goals:

Phase 1

Target: 3 months from start

- Browsers on dumb terminal to open readership to anyone with a computer or PC.(?)
- Browsers on vt220 terminals to give cursor-oriented readership to a very large proportion of readers;
- A browser on the Macintosh in the Macintosh style;
- A browser on the NeXT using the NeXTStep tools, as a fast prototype for ideas in human interface design and navigation techniques.
- A server providing access to the world of Usenet/Internet news articles. *
- A server providing access to all the information currently stored on CERNVM and mentioned in the FIND index. This should include CERN program library notes, IBM and CERN CMS help screens, CERN/CN writeups, Computer Newsletter articles, etc.
- A server which may be installed on any machine to allow files on that machine to be accessed as hypertext.
- The ability for users to write, using markup tags, their own hypertext for help files. No other hypertext editing capability will necessarily be implemented in this phase.
- A gateway process to allow access between the Internet and DECnet protocol worlds.
- A set of guidelines on how to manage a hypertext server.
- A requirements analysis of the information access needs for a large experiment.

At this stage, readership is universal, but the creation of new material relies on existing systems. For example, the introduction of new material for the FIND index, or the posting of news articles will use the same procedures as at present. we gain useful experience in the representation of existing data in hypertext form, and in the types of navigational and other aids appreciated by users in high energy physics.

Phase 2

Target: 6 months from start

In this important phase, we aim to allow

- The creation of new links and new material by readers. At this stage, authorship becomes universal.
- A full-screen browser on VM/XA for those using CERNVM, and other HEP VM sites;
- An X-window browser/editor, giving the sophisticated facilities originally prototyped under NeXTStep to the wide X-based community. (We imagine using OSF/Motif subject to availability)
- The automatic notification of a reader when new material of interest to him/her has become available. This is essential for news articles, but is very useful for any other material.

The ability of readers to create links allows annotation by users of existing data, allows to add themselves and their documents to lists (mailing lists, indexes, etc). It should be possible for users to

link public documents to (for example) bug reports, bug fixes, and other documents which the authors themselves might never have realised existed.This phase allows collaborative authorship. It provides a place to put any piece of information such that it can later be found. Making it easy to change the web is thus the key to **avoiding obsolete information**. One should be able to trace the source of information, to circumvent and then to repair flaws in the web.

Resources required

1. People

The following functions are identifiable. They do not necessarily correspond to individuals on a one to one basis. The initials[1] in brackets indicate people who have already expressed an interest in the project and who have the necessary skills but do not indicate any commitment as yet on thier part or the part of their managers. We are of course very open to involvement from others.

- System architect. Coordinate development, protocol definition, etc; ensures integrity of design. (50% TBL?)
- Market research and product planner. Discuss the project and its features with potential and actual users in all divisions. Prepare criteria for feature selection and development priority. (50% RC?)
- Hyper-Librarian. Oversees the web of available data, ensuring its coherency. Interface with users, train users. Manages indexes and keyword systems. Manages data provided by the project itself. (100% KG?)
- Software engineer: NeXTStep. Provide browser/editor interface under the NeXTStep human interface tools. Experiment with navigational aids. Keep a running knowledge of the NeXTStep world. (50%TBL?)
- Software engineer: X-windows and human interface. Provide browser/editor human interface under OSF/Motif. Respond to user suggestion for ease of use improvements and options. Create an aesthetic, practical human interface. Keep a running knowledge of the X world. (75%RJ?)
- Software engineer: IBM mainframe. Provide browser service on CERNVM and other HEP VM sites. Maintain the FIND server software. Keep up a running knowledge of the CMS, Rexx world. (75% BP?)
- Software engineer: Macintosh. Provide browser/editor for the mac, using whatever tools are appropriate (Thnk-C, HyperCard, etc?). (50%RC?)
- Software engineer: C. Help write code for dumb terminal or vt100 browsers, and portable browser code to be shared between browers. This could include a technical student project. (100% NP? + A.N.Other?)

We foresee that a demand may arise for browsers on specific systems, for specific customizations, and for servers to make specific existing data available online as hypertext. We intend to enthusiastically support such widening of the web. Of course, we may have to draw on more manpower and specific expertise in these cases.

2. Other resources

We will require the following support in the way of equipment and services.

---

1 Tim Berners-Lee/CN (TBL), Robert Cailliau/ECP(RC), Karin Gieselmann/ECP(KG) , R. Jones/ECP(RJJ), Nicola Pellow/CN (NP), Bernd Pollermann/CN (BP)

- We feel it is important for those involved in the project to be able to work close to each other and exchange ideas and problems as they work. An **office area** or close group of offices is therefore required.
- Each person working on the project will require a state-of-the-art **workstation**. Experience shows that a workstation has to be upgraded in some way every two years or so as software becomes more cumbersome, and memory/speed requirements increase. This, and the cost of software upgrades, we foresee as reasonable expenses. We imagine using a variety of types of workstation as we provide software on a variety of machines, but otherwise NeXTs. For VMS machines, we would like the support of an existing VAXcluster to minimize our own system management overheads.
- We would like to be able to purchase licenses for **commercial** hypertext **software** where we feel this could be incorporated into the project, and save development and maintenance time, or where we feel we could gain useful experience from its use. (Approximate examples are: *Guide* license: CHF750; *KMS* full author license CHF1500, evaluation kit CHF100. *FrameMaker:* CHF2000)
- We will require computing support. In particular, we will require a reliable backed up NFS (or equivalent) **file server support** for our development environment. We will also need to run daemon software on machines with Internet, DECnet and BITNET connectivity, which will require a certain amount of support from operators and system managers.

Future paths

The two phases above will provide an extremely useful set of tools. Though the results seem ambitious, the individual steps necessary are well within our abilities with available technology. Future developments which would further enhance the project could include:

- Daemon programs which run overnight and build indexes of available information.
- A server automatically providing a hypertext view of a (for example Oracle) database, from a description of the database and a description (for example in SQL) of the view required.
- Work on efficient networking over wide areas, negotiation with other sites to provide compatible online information.
- A serious study of the use and abuse of the system, the sociology of its use at CERN.

References

[1] T. Berners-Lee/CN, *HyperText and CERN.* An explanation of hypertext, and why it is important for CERN. A background document explaining the ideas behind this project.

[2] T. Berners-Lee/CN, *Hypertext Design Issues.* A detailed look at hypertext models and facilities, with a discussion of choices to be made in choosing or implementing a system.

[3] Other documentation on the project is stored in hypertext form. See <file://cernvax/userd/tbl/hypertext/TheProject.html> which leads to further references.