The ethos now seems to be that patents are a matter of whatever you can get away with. Engineers, asked by company lawyers to provide patentable ideas every few months, resignedly hand over "ideas" that make the engineers themselves cringe.

It is time for a change, to an ethos in which companies use patents to defend their own valid products, rather than serendipitously suing based on claims even they themselves would have thought applied. The threshold of "innovation" is too low. Corporate lawyers are locked into a habit of arguing whatever advantage they can, and probably only determined corporate leadership can set the industry back on a sane track. The consortium members have, at the time of writing, been delivering on what to do, but it is not clear what the result will be.

The Semantic Web, like the Web already, will make many things previously impossible just obvious. As I write about the new technology, I do wonder whether it will be a technical dream or a legal nightmare.

CHAPTER 14

Weaving the Web

Can the future Web change the way people work together and advance knowledge in a small company, a large organization, a country? If it works for a small group and can scale up, can it be used to change the world? We know the Web lets us do things more quickly, but can it make a phase change in society, a move to a new way of working—and will that be for better or for worse?

In a company with six employees, everybody can sit around a table, share their visions of where they're going, and reach a common understanding of all the terms they're using. In a large company, somebody defines the common terms and behavior that make the company work as an entity. Those who have been through the transition know it only to well: It typically kills diversity. It's too rigid a structure. And it doesn't scale, because as the company gets bigger, the bureaucratic boundaries cut off more and more of its internal communications, its lifeblood. At the other extreme is the utopian commune with no structure, which doesn't work either because nobody actually takes out the garbage.

Whether a group can advance comes down to creating the right connectivity between people—in a family, a company, a country, or the world. We've been trying to figure out how to create this for years. In many ways, we haven't had to decide, as geography has decided for us. Companies, and nations, have always been defined by a physical grouping of people. The military stability of a nation was based on troop placements and marching distances. The diversity of culture we've had also has stemmed from two-dimensional space. The only reason the people in a little village in Switzerland would arise speaking a unique dialect was that they were surrounded by mountains. Geography gave the world its military stability and cultural boxes. People didn't have to decide how large their groups would be or where to draw the boundaries. Now that the metric is not physical distance, not even time zones, but clicks, we do have to make these decisions. The Internet and the Web have pulled us out of two-dimensional space. They've also moved us away from the idea that we won't be interrupted by anybody who's more than a day's march away.

At first, this violation of our long-held rules can be unsettling, destroying a geographical sense of identity. The Web breaks the boundaries we have relied on to define us and protect us, but it can build new ones, too.

The thing that does not scale when a company grows is intuition—the ability to solve problems without using a well-defined logical method. A person, or a small group brainstorming out loud, ruminates about problems until possible solutions emerge. Answers arrive not necessarily from following a logical path, but rather by seeing where connections may lead. A larger company fails to be intuitive when the person with the answer isn't talking with the person who has the question.

It's important that the Web help people be intuitive as well as analytical, because our society needs both functions. Human beings have a natural balance in using the creative and analytical parts of their brains. We will solve large analytical problems by turning computer power loose on the hard data of the Semantic Web.

Scaling intuition is difficult because our minds hold thousands of ephemeral tentative associations at the same time. To allow group intuition, the Web would have to capture these threads—half thoughts that arise, without evident rational thought or inference, as we work. It would have to present them to another reader as a natural complement to a half-formed idea. The intuitive step occurs when someone following links by a number of independent people notices a relevant relationship, and creates a shortcut link to record it.

This all works only if each person makes links as he or she browses, so writing, link creation, and browsing must be totally integrated. If someone discovers a relationship but doesn't make the link, he or she is wiser but the group is not.

To make such a shortcut, one person has to have two pieces of inference in his or her head at the same time. The new Web will make it much more likely that somebody somewhere is browsing one source that has half of the key idea, and happens to have just recently browsed the other. For this to be likely, the Web must be well connected—have few "degrees of separation." This is the sort of thing researchers are always trying to do—get as much in their heads as possible, then go to sleep and hope to wake up in the middle of the night with a brilliant idea and rush to write it down. But as the problems get bigger, we want to be able to work this brainstorming approach on a much larger scale. We have to be sure to design the Web to allow feedback from the people who've made new intuitive links.

If we succeed, creativity will arise across larger and more diverse groups. These high-level activities, which have occurred just within one human's brain, will occur among ever-larger, more interconnected groups of people acting as if they shared a larger intuitive brain. It is an intriguing analogy. Perhaps that late-night surfing is not such a waste of time after all: It is just the Web dreaming.

Atoms each have a *valence*—an ability to connect with just so many other atoms. As an individual, each of us picks a few channels to be involved in, and we can cope with only so much. The advantage of getting things done faster on the Web is an advantage only to the extent that we can accept the information faster, and there are definite limits. By just pushing the amount we have to read and write, the number of e-mails we have to cope with, the number of Web sites we have to surf, we may scrape together a few more bytes of knowledge, but exhaust ourselves in the process and miss the point.

As a group works together, the members begin to reach common understandings that involve new concepts, which only they share. Sometimes these concepts can become so strong that the group finds it has to battle the rest of the world to explain its decisions. At this point, the members may realize for the first time that they have started using words in special ways. They may not realize how they have formed a little subculture until they begin explaining their decisions to colleagues outside the group. They have built a new understanding, and at the same time built a barrier around themselves. Boundaries of understanding have been broken, but new ones have formed around those who share the new concept.

A choice has been made, and there is a gain and a loss in terms of shared understanding.

What should guide us when we make these choices? What kind of a structure are we aiming for, and what principles will help us achieve it? The Web as a medium is so flexible that it leaves the choice to us. As well as the choice of links we make

individually, we have a choice in the social machines we create, the variously shaped parts in our construction game. We know that we want a well-connected structure for group intuition to work. We know it should be decentralized, to be resilient and fair.

The human brain outperforms computers by its incredible level of parallel processing. Society, similarly, solves its problems in parallel. For the society to work efficiently on the Web, massive parallelism is required. Everybody must be able to publish, and to control who has access to their published work. There should not be a structure (like a highway system or mandatory Dewey decimal system) or limitation that precludes any kind of idea or solution purely because the Web won't allow it to be explained.

The Internet before the Web thrived on a decentralized technical architecture and a decentralized social architecture. These were incrementally created by the design of technical and social machinery. The community had just enough rules of behavior to function using the simple social machines it invented. Starting from a flat world in which every computer had just one Internet address and everyone was considered equal, over time the sea of chattering people imposed some order on itself. Newsgroups gave structure to information and people. The Web started with a similar lack of preset structure, but soon all sorts of lists of "best" sites created a competition-based structure even before advertising was introduced. While the Internet itself seemed to represent a flight from hierarchy, without hierarchy there were too many degrees of separation to prevent things from being reinvented. There seemed to be a quest for something that was not a tree, but not a flat space, either.

We certainly need a structure that will avoid those two catastrophes: the global uniform McDonald's monoculture, and the isolated Heaven's Gate cults that understand only themselves. By each of us spreading our attention evenly between groups of different size, from personal to global, we help avoid these extremes.

Link by link we build paths of understanding across the web of humanity. We are the threads holding the world together. As we do this, we naturally end up with a few Web sites in very high demand, and a continuum down to the huge number of Web sites with only rare visitors. In other words, appealing though equality between peers seems, such a structure by its uniformity is not optimal. It does not pay sufficient attention to global coordination, and it can require too many clicks to get from problem to solution.

If instead everyone divides their time more or less evenly between the top ten Web sites, the rest of the top one hundred, the rest of the top one thousand, and so on, the load on various servers would have a distribution of sizes characteristic of "fractal" patterns so common in nature (from coastlines to ferns) and of the famous "Mandelbrot set" mathematical patterns. It turns out that some measurements of all the Web traffic by Digital Equipment employees on the West Coast revealed very closely this $1/n$ law: The Web exhibits fractal properties even though we can't individually see the patterns, and even though there is no hierarchical system to enforce such a distribution.

This doesn't answer the question, but it is intriguing because it suggests that there are large-scale dynamics operating to produce such results. A fascinating result was found by Jon Kleinberg, a computer scientist at Cornell University who discovered that, when the matrix of the Web is analyzed like a quantum mechanical system, stable energy states correspond to concepts under discussion. The Web is starting to develop large-scale structure in its own way. Maybe we will be able to produce new metrics for checking the progress of society toward what we consider acceptable.

The analogy of a global brain is tempting, because Web and brain both involve huge numbers of elements—neurons and Web pages—and a mixture of structure and apparent randomness. However, a brain has an intelligence that emerges on quite a different level

from anything that a neuron could be aware of. From Arthur C. Clarke to Douglas Hofstader, writers have contemplated an "emergent property" arising from the mass of humanity and computers. But remember that such a phenomenon would have its own agenda. We would not as individuals be aware of it, let alone control it, any more than the neuron controls the brain.

I expect that there will be emergent properties with the Semantic Web, but at a lesser level than emergent intelligence. There could be spontaneous order or instability: Society could crash, much as the stock market crashed in October 1987 because of automatic trading by computer. The agenda of trading—to make money on each trade—didn't change, but the dynamics did; so many huge blocks of shares were traded so fast that the whole system became unstable.

To ensure stability, any complex electronic system needs a damping mechanism to introduce delay, to prevent it from oscillating too wildly. Damping mechanisms have since been built into the stock-trading system. We may be able to build them into the Semantic Web of cooperating computers—but will we be able to build them into the web of cooperating people? Already the attention of people, the following of links, and the flow of money are interlaced inextricably.

I do not, therefore, pin my hopes on an overpowering order emerging spontaneously from the chaos. I feel that to deliberately build a society, incrementally, using the best ideas we have, is our duty and will also be the most fun. We are slowly learning the value of decentralized, diverse systems, and of mutual respect and tolerance. Whether you put it down to evolution or your favorite spirit, the neat thing is that we seem as humans to be tuned so that we do in the end get the most fun out of doing the "right" thing.

My hope and faith that we are headed somewhere stem in part from the repeatedly proven observation that people seem to be naturally built to interact with others as part of a greater system. A person who's completely turned inward, who spends all his or her time alone, is someone who has trouble making balanced decisions and is very unhappy. Someone who is completely turned outward, who's worried about the environment and international diplomacy and spends no time sitting at home or in his local community, also has trouble making balanced decisions and is also very unhappy. It seems a person's happiness depends on having a balance of connections at different levels. We seem to have built into us what it takes in a person to be part of a fractal society.

If we end up producing a structure in hyperspace that allows us to work together harmoniously, that would be a metamorphosis. Though it would, I hope, happen incrementally, it would result in a huge restructuring of society. A society that could advance with intercreativity and group intuition rather than conflict as the basic mechanism would be a major change.

If we lay the groundwork right and try novel ways of interacting on the new Web, we may find a whole new set of financial, ethical, cultural, and governing structures to which we can choose to belong, rather than having to pick the ones we happen to physically live in. Bit by bit those structures that work best would become more important in the world, and democratic systems might take on different shapes.

Working together is the business of finding shared understandings but being careful not to label them as absolute. They may be shared, but often arbitrary in the larger picture.

We spend a lot of time trying to tie down meanings and fighting to have our own framework adopted by others. It is, after all, a lifelong process to set ourselves up with connections to all the concepts we use. Having to work with someone else's definitions is difficult. An awe-inspiring talent of my physics tutor, Professor John Moffat, was that when I brought him a problem I had worked out incorrectly, using a strange technique and symbols different from the well-established ones, he not only would follow my weird reasoning to find out where it went wrong, but would then use my own strange notation to explain the right answer. This great feat involved looking at the world using my definitions, comparing them with his, and translating his knowledge and experience into my language. It was a mathematical version of the art of listening. This sort of effort is needed whenever groups meet. It is also the hard work of the consortium's working groups. Though it often seems to be no fun, it is the thing that deserves the glory.

We have to be prepared to find that the "absolute" truth we had been so comfortable with within one group is suddenly challenged when we meet another. Human communication scales up only if we can be tolerant of the differences while we work with partial understanding.

The new Web must allow me to learn by crossing boundaries. It has to help me reorganize the links in my own brain so I can understand those in another person's. It has to enable me to keep the frameworks I already have, and relate them to new ones. Meanwhile, we as people will have to get used to viewing as communication rather than argument the discussions and challenges that are a necessary part of this process. When we fail, we will have to figure out whether one framework or another is broken, or whether we just aren't smart enough yet to relate them.

The parallels between technical design and social principles have recurred throughout the Web's history. About a year after I arrived to start the consortium, my wife and I came across Unitarian Universalism. Walking into a Unitarian Universalist church more or less by chance felt like a breath of fresh air. Some of the association's basic philosophies very much match what I had been brought up to believe, and the objective I had in creating the Web. People now sometimes even ask whether I designed the Web based on these principles. Clearly, Unitarian Universalism had no influence on the Web. But I can see how it could have, because I did indeed design the Web around universalist (with a lowercase *u*) principles.

One of the things I like about Unitarianism is its lack of religious trappings, miracles, and pomp and circumstance. It is minimalist, in a way. Unitarians accepted the useful parts of philosophy from all religions, including Christianity and Judaism, but also Hinduism, Buddhism, and any other good philosophies, and wrapped them not into one consistent religion, but into an environment in which people think and discuss, argue, and always try to be accepting of differences of opinion and ideas.

I suppose many people would not classify "U-Uism" as a religion at all, in that it doesn't have the dogma, and is very tolerant of different forms of belief. It passes the Test of Independent Invention that I apply to technical designs: If someone else had invented the same thing independently, the two systems should work together without anyone having to decide which one was "central." For me, who enjoyed the acceptance and the diverse community of the Internet, the Unitarian church was a great fit. Peer-to-peer relationships are encouraged wherever they are appropriate, very much as the World Wide Web encourages a hypertext link to be made wherever it is appropriate. Both are philosophies that allow decentralized systems to develop, whether they are systems of computers, knowledge, or people. The people who built the Internet and Web have a real appreciation of the value of individuals and the value of systems in which individuals play their role, with both a firm sense of their own identity and a firm sense of some common good.

There's a freedom about the Internet: As long as we accept the rules of sending packets around, we can send packets containing anything to anywhere. In Unitarian Universalism, if one accepts the basic tenet of mutual respect in working together toward some greater vision, then one finds a huge freedom in choosing one's own words that capture that vision, one's own rituals to help focus the mind, one's own metaphors for faith and hope.

I was very lucky, in working at CERN, to be in an environment that Unitarian Universalists and physicists would equally appreciate: one of mutual respect, and of building something very great through collective effort that was well beyond the means of any one person—without a huge bureaucratic regime. The environment was complex and rich; any two people could get together and exchange views, and even end up working together somehow. This system produced a weird and wonderful machine, which needed care to maintain, but could take advantage of the ingenuity, inspiration, and intuition of individuals in a special way. That, from the start, has been my goal for the World Wide Web.

Hope in life comes from the interconnections among all the people in the world. We believe that if we all work for what we think individually is good, then we as a whole will achieve more power, more understanding, more harmony as we continue the journey. We don't find the individual being subjugated by the whole. We don't find the needs of the whole being subjugated by the increasing power of an individual. But we might see more understanding in the struggles between these extremes. We don't expect the system to eventually become perfect. But we feel better and better about it. We find the journey more and more exciting, but we don't expect it to end.

Should we then feel that we are getting smarter and smarter, more and more in control of nature, as we evolve? Not really. Just better connected—connected into a better shape. The experience of seeing the Web take off by the grassroots effort of thousands gives me tremendous hope that if we have the individual will, we can collectively make of our world what we want.

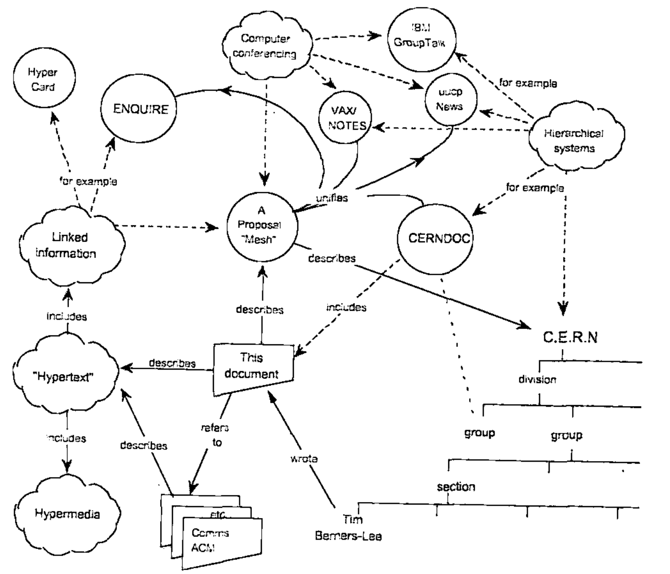
Publisher's Note: *This appendix contains the original proposal for the World Wide Web. At the author's request, it is presented here as a historical document in its original state, with all of its original errors intact—including typographical and style elements—in order to preserve the integrity of the document.*

Information Management: A Proposal

TIM BERNERS-LEE, CERN

MARCH 1989, MAY 1990

This proposal concerns the management of general information about accelerators and experiments at CERN. It discusses the problems of loss of information about complex evolving systems and derives a solution based on a distributed hypertext system.



OVERVIEW

Many of the discussions of the future at CERN and the LHC era end with the question - "Yes, but how will we ever keep track of such a large project?" This proposal provides an answer to such questions. Firstly, it discusses the problem of information access at CERN. Then, it introduces the idea of linked information systems, and compares them with less flexible ways of finding information.

It then summarises my short experience with non-linear text systems known as "hypertext", describes what CERN needs from such a system, and what industry may provide. Finally, it suggests steps we should take to involve ourselves with hypertext now, so that individually and collectively we may understand what we are creating.

LOSING INFORMATION AT CERN

CERN is a wonderful organisation. It involves several thousand people, many of them very creative, all working toward common goals. Although they are nominally organised into a hierarchical management structure, this does not constrain the way people will communicate, and share information, equipment and software across groups.

The actual observed working structure of the organisation is a multiply connected "web" whose interconnections evolve with time. In this environment, a new person arriving, or someone taking on a new task, is normally given a few hints as to who would be useful people to talk to. Information about what facilities exist and how to find out about them travels in the corridor gossip and occasional newsletters, and the details about what is required to be done spread in a similar way. All things considered, the result is remarkably successful, despite occasional misunderstandings and duplicated effort.

A problem, however, is the high turnover of people. When two years is a typical length of stay, information is constantly being lost. The introduction of the new people demands a fair amount of their time and that of others before they have any idea of what goes on. The technical details of past projects are sometimes lost forever, or only recovered after a detective investigation in an emergency. Often, the information has been recorded, it just cannot be found.

If a CERN experiment were a static once-only development, all the information could be written in a big book. As it is, CERN is constantly changing as new ideas are produced, as new technology

becomes available, and in order to get around unforeseen technical problems. When a change is necessary, it normally affects only a small part of the organisation. A local reason arises for changing a part of the experiment or detector. At this point, one has to dig around to find out what other parts and people will be affected. Keeping a book up to date becomes impractical, and the structure of the book needs to be constantly revised.

The sort of information we are discussing answers, for example, questions like

- Where is this module used?
- Who wrote this code? Where does he work?
- What documents exist about that concept?
- Which laboratories are included in that project?
- Which systems depend on this device?
- What documents refer to this one?

The problems of information loss may be particularly acute at CERN, but in this case (as in certain others), CERN is a model in miniature of the rest of world in a few years time. CERN meets now some problems which the rest of the world will have to face soon. In 10 years, there may be many commercial solutions to the problems above, while today we need something to allow us to continue[1].

LINKED INFORMATION SYSTEMS

In providing a system for manipulating this sort of information, the hope would be to allow a pool of information to develop which could grow and evolve with the organisation and the projects it describes. For this to be possible, **the method of storage must not place its own restraints on the information.**

This is why a "web" of notes with links (like references) between them is far more useful than a fixed hierarchical system. When describing a complex system, many people resort to diagrams with circles and arrows. Circles and arrows leave one free to describe the interrelationships between things in a way that tables, for example, do not. The system we need is like a diagram of cir-

[1] The same has been true, for example, of electronic mail gateways, document preparation, and heterogeneous distributed programming systems.

cles and arrows, where circles and arrows can stand for anything.

We can call the circles nodes, and the arrows links. Suppose each node is like a small note, summary article, or comment. I'm not over concerned here with whether it has text or graphics or both. Ideally, it represents or describes one particular person or object. Examples of nodes can be

- People
- Software modules
- Groups of people
- Projects
- Concepts
- Documents
- Types of hardware
- Specific hardware objects

The arrows which links circle A to circle B can mean, for example, that A...

- depends on B
- is part of B
- made B
- refers to B
- uses B
- is an example of B

These circles and arrows, nodes and links[2], have different significance in various sorts of conventional diagrams:

Diagram	Nodes are	Arrows mean
Family tree	People	"Is parent of"
Dataflow diagram	Software modules	"Passes data to"
Dependency	Module	"Depends on"
PERT chart	Tasks	"Must be done before"
Organisational chart	People	"Reports to"

2 Linked information systems have entities and relationships. There are, however, many differences between such a system and an "Entity Relationship" database system. For one thing, the information stored in a linked system is largely comment for human readers. For another, nodes do not have strict types which define exactly what relationships they may have. Nodes of simialr type do not all have to be stored in the same place.

The system must allow any sort of information to be entered. Another person must be able to find the information, sometimes without knowing what he is looking for.

In practice, it is useful for the system to be aware of the generic types of the links between items (dependences, for example), and the types of nodes (people, things, documents..) without imposing any limitations.

THE PROBLEM WITH TREES

Many systems are organised hierarchically. The CERNDOC documentation system is an example, as is the Unix file system, and the VMS/HELP system. A tree has the practical advantage of giving every node a unique name. However, it does not allow the system to model the real world. For example, in a hierarchical HELP system such as VMS/HELP, one often gets to a leaf on a tree such as

HELP COMPILER SOURCE_FORMAT PRAGMAS DEFAULTS

only to find a reference to another leaf: "Please see

HELP COMPILER COMMAND OPTIONS DEFAULTS PRAGMAS"

and it is necessary to leave the system and re-enter it. What was needed was a link from one node to another, because in this case *the information was not naturally organised into a tree*.

Another example of a tree-structured system is the uucp News system (try 'rn' under Unix). This is a hierarchical system of discussions ("newsgroups") each containing articles contributed by many people. It is a very useful method of pooling expertise, but suffers from the inflexibility of a tree. Typically, a discussion under one newsgroup will develop into a different topic, at which point it ought to be in a different part of the tree. (See Fig 1).

From mcvax!uunet!pyrdc!pyrnj!rutgers!bellcore!geppetto!duncan Thu Mar...
Article 93 of alt.hypertext:
Path: cernvax!mcvax!uunet!pyrdc!pyrnj!rutgers!bellcore!geppetto!duncan

>From: duncan@geppetto.ctt.bellcore.com (Scott Duncan)
Newsgroups: alt.hypertext
Subject: Re: Threat to free information networks
Message-ID: <14646@bellcore.bellcore.com>
Date: 10 Mar 89 21:00:44 GMT
References: <1784.2416BB47@isishq.FIDONET.ORG>
<3437@uhccux.uhcc...
Sender: news@bellcore.bellcore.com
Reply-To: duncan@ctt.bellcore.com (Scott Duncan)
Organization: Computer Technology Transfer, Bellcore
Lines: 18

Doug Thompson has written what I felt was a thoughtful article on censorship
-- my acceptance or rejection of its points is not
particularly germane to this posting, however.

In reply Greg Lee has somewhat tersely objected.

My question (and reason for this posting) is to ask where we might logically
take this subject for more discussion. Somehow alt.hypertext does not seem
to be the proper place.

Would people feel it appropriate to move to alt.individualism or even one of
the soc groups. I am not so much concerned with the specific issue of censor-
ship of rec.humor.funny, but the views presented in Greg's article.

Speaking only for myself, of course, I am...
Scott P. Duncan (duncan@ctt.bellcore.com OR ...!bellcore!ctt!duncan)
(Bellcore, 444 Hoes Lane RRC 1H-210, Piscataway, NJ...)
(201-699-3910 (w) 201-463-3683 (h))

Fig 1. An article in the UUCP News scheme.

The Subject field allows notes on the same topic to be linked together within a "newsgroup". The name of the newsgroup (alt.hypertext) is a

hierarchical name. This particular note is expresses a problem with the strict tree structure of the scheme: this discussion is related to several areas. Note that the "References", "From" and "Subject" fields can all be used to generate links.

THE PROBLEM WITH KEYWORDS

Keywords are a common method of accessing data for which one does not have the exact coordinates. The usual problem with keywords, however, is that two people never chose the same keywords. The keywords then become useful only to people who already know the application well.

Practical keyword systems (such as that of VAX/NOTES for example) require keywords to be registered. This is already a step in the right direction.

A linked system takes this to the next logical step. Keywords can be nodes which stand for a concept. A keyword node is then no different from any other node. One can link documents, etc., to keywords. One can then find keywords by finding any node to which they are related. In this way, documents on similar topics are indirectly linked, through their key concepts.

A keyword search then becomes a search starting from a small number of named nodes, and finding nodes which are close to all of them.

It was for these reasons that I first made a small linked information system, not realising that a term had already been coined for the idea: "hypertext".

A SOLUTION: HYPERTEXT

PERSONAL EXPERIENCE WITH HYPERTEXT

In 1980, I wrote a program for keeping track of software with which I was involved in the PS control system. Called *Enquire*, it allowed one to store snippets of information, and to link related pieces together in any way. To find information, one progressed via the links from one sheet to another, rather like in the old computer game "adventure". I used this for my personal record of people and modules. It was similar to the application Hypercard produced more recently by Apple for the Macintosh. A difference was that Enquire, although lacking the fancy graphics, ran on a multiuser system, and allowed many people to access the same data.

```
Documentation of the RPC project                              (concept)

  Most of the documentation is available on VMS, with the two
  principle manuals being stored in the CERNDOC system.

   1) includes: The VAX/NOTES conference VXCERN::RPC
   2) includes: Test and Example suite
   3) includes: RPC BUG LISTS
   4) includes: RPC System: Implementation Guide
                Information for maintenance, porting, etc.
   5) includes: Suggested Development Strategy for RPC Applications
   6) includes: "Notes on RPC", Draft 1, 20 feb 86
   7) includes: "Notes on Proposed RPC Development" 18 Feb 86
   8) includes: RPC User Manual
                How to build and run a distributed system.
   9) includes: Draft Specifications and Implementation Notes
  10) includes: The RPC HELP facility
  11) describes: THE REMOTE PROCEDURE CALL PROJECT in DD/OC

 Help Display Select Back Quit Mark Goto_mark Link Add Edit
```

Fig 2. A screen in an Enquire scheme.

This example is basically a list, so the list of links is more important than the text on the node itself. Note that each link has a type ("includes" for example) and may also have comment associated with it. (The bottom line is a menu bar.)

Soon after my re-arrival at CERN in the DD division, I found that the environment was similar to that in PS, and I missed

Enquire. I therefore produced a version for the VMS, and have used it to keep track of projects, people, groups, experiments, software modules and hardware devices with which I have worked. I have found it personally very useful. I have made no effort to make it suitable for general consumption, but have found that a few people have successfully used it to browse through the projects and find out all sorts of things of their own accord.

HOT SPOTS

Meanwhile, several programs have been made exploring these ideas, both commercially and academically. Most of them use "hot spots" in documents, like icons, or highlighted phrases, as sensitive areas. touching a hot spot with a mouse brings up the relevant information, or expands the text on the screen to include it. Imagine, then, the references in this document, all being associated with the network address of the thing to which they referred, so that while reading this document you could skip to them with a click of the mouse.

"Hypertext" is a term coined in the 1950s by Ted Nelson [...], which has become popular for these systems, although it is used to embrace two different ideas. One idea (which is relevant to this problem) is the concept:

> "Hypertext": Human-readable information linked together in an unconstrained way.

The other idea, which is independent and largely a question of technology and time, is of multimedia documents which include graphics, speech and video. I will not discuss this latter aspect further here, although I will use the word "Hypermedia" to indicate that one is not bound to text.

It has been difficult to assess the effect of a large hypermedia system on an organisation, often because these systems never had seriously large-scale use. For this reason, we require large amounts of existing information should be accessible using any new information management system.

CERN REQUIREMENTS

To be a practical system in the CERN environment, there are a number of clear practical requirements.

REMOTE ACCESS ACROSS NETWORKS.

CERN is distributed, and access from remote machines is essential.

HETEROGENEITY

Access is required to the same data from different types of system (VM/CMS, Macintosh, VAX/VMS, Unix)

NON-CENTRALISATION

Information systems start small and grow. They also start isolated and then merge. A new system must allow existing systems to be linked together without requiring any central control or coordination.

ACCESS TO EXISTING DATA

If we provide access to existing databases as though they were in hypertext form, the system will get off the ground quicker. This is discussed further below.

PRIVATE LINKS

One must be able to add one's own private links to and from public information. One must also be able to annotate links, as well as nodes, privately.

BELLS AND WHISTLES

Storage of ASCII text, and display on 24x80 screens, is in the short term sufficient, and essential. Addition of graphics would be an optional extra with very much less penetration for the moment.

DATA ANALYSIS

An intriguing possibility, given a large hypertext database with typed links, is that it allows some degree of automatic analysis. It is possible to search, for example, for anomalies such as undocumented software or divisions which contain no people. It is possible to generate lists of people or devices for other purposes, such as mailing lists of people to be informed of changes.

It is also possible to look at the topology of an organisation or a project, and draw conclusions about how it should be managed, and how it could evolve. This is particularly useful when the database becomes very large, and groups of projects, for example, so interwoven as to make it difficult to see the wood for the trees.

In a complex place like CERN, it's not always obvious how to divide people into groups. Imagine making a large three-dimensional model, with people represented by little spheres, and strings between people who have something in common at work.

Now imagine picking up the structure and shaking it, until you make some sense of the tangle: perhaps, you see tightly knit groups in some places, and in some places weak areas of communication spanned by only a few people. Perhaps a linked information system will allow us to see the real structure of the organisation in which we work.

LIVE LINKS

The data to which a link (or a hot spot) refers may be very static, or it may be temporary. In many cases at CERN information about the state of systems is changing all the time. Hypertext allows documents to be linked into "live" data so that every time the link is followed, the information is retrieved. If one sacrifices portability, it is possible so make following a link fire up a special application, so that diagnostic programs, for example, could be linked directly into the maintenance guide.

NON REQUIREMENTS

Discussions on Hypertext have sometimes tackled the problem of copyright enforcement and data security. These are of secondary importance at CERN, where information exchange is still more important than secrecy. Authorisation and accounting systems for hypertext could conceivably be designed which are very sophisticated, but they are not proposed here.

In cases where reference must be made to data which is in fact protected, existing file protection systems should be sufficient.

SPECIFIC APPLICATIONS

The following are three examples of specific places in which the proposed system would be immediately useful. There are many others.

DEVELOPMENT PROJECT DOCUMENTATION.

The Remote procedure Call project has a skeleton description using *Enquire*. Although limited, it is very useful for recording who did what, where they are, what documents exist, etc. Also, one can keep track of users, and can easily append any extra little bits of information which come to hand and have nowhere else to be put. Cross-links to other projects, and to databases which contain information on people and documents would be very useful, and save duplication of information.

DOCUMENT RETRIEVAL.

The CERNDOC system provides the mechanics of storing and printing documents. A linked system would allow one to browse through concepts, documents, systems and authors, also allowing references between documents to be stored. (Once a document had been found, the existing machinery could be invoked to print it or display it).

THE "PERSONAL SKILLS INVENTORY".

Personal skills and experience are just the sort of thing which need hypertext flexibility. People can be linked to projects they have worked on, which in turn can be linked to particular machines, programming languages, etc.

THE STATE OF THE ART IN HYPERMEDIA

An increasing amount of work is being done into hypermedia research at universities and commercial research labs, and some commercial systems have resulted. There have been two conferences, Hypertext '87 and '88, and in Washington DC, the National Institute of Standards and Technology (NST) hosted a workshop on standardisation in hypertext, a followup of which will occur during 1990.

The Communications of the ACM special issue on Hypertext contains many references to hypertext papers. A bibliography on hypertext is given in [NIST90], and a uucp newsgroup alt.hypertext exists. I do not, therefore, give a list here.

BROWSING TECHNIQUES

Much of the academic research is into the human interface side of browsing through a complex information space. Problems addressed are those of making navigation easy, and avoiding a feeling of being "lost in hyperspace". Whilst the results of the research are interesting, many users at CERN will be accessing the system using primitive terminals, and so advanced window styles are not so important for us now.

INTERCONNECTION OR PUBLICATION?

Most systems available today use a single database. This is accessed by many users by using a distributed file system. There are few products which take Ted Nelson's idea of a wide "docuverse" literally by allowing links between nodes in different databases. In order to do this, some standardisation would be necessary. However, at the standardisation workshop, the emphasis was on standardisation of the format for exchangeable media, nor for networking. This is prompted by the strong push toward publishing of hypermedia information, for example on optical disk. There seems to be a general consensus about the abstract data model which a hypertext system should use.

Many systems have been put together with little or no regard for portability, unfortunately. Some others, although published, are proprietary software which is not for external release. However, there are several interesting projects and more are appearing all the time. Digital's "Compound Document Architecture" (CDA) , for example, is a data model which may be extendible into a hypermedia model, and there are rumours that this is a way Digital would like to go.

INCENTIVES AND CALS

The US Department of Defence has given a big incentive to hypermedia research by, in effect, specifying hypermedia documentation for future procurement. This means that all manuals for parts for defence equipment must be provided in hypermedia form. The acronym CALS stands for "Computer-aided Acquisition and Logistic Support).

There is also much support from the publishing industry, and from librarians whose job it is to organise information.

WHAT WILL THE SYSTEM LOOK LIKE?

Let us see what components a hypertext system at CERN must have.

The only way in which sufficient flexibility can be incorporated is to separate the information storage software from the information display software, with a well defined interface between them. Given the requirement for network access, it is natural to let this clean interface coincide with the physical division between the user and the remote database machine[3].

3 A client/server split at his level also makes multi-access more easy, in that a single server process can service many clients, avoiding the problems of simultaneous access to one database by many different users.

This division also is important in order to allow the heterogeneity which is required at CERN (and would be a boon for the world in general).

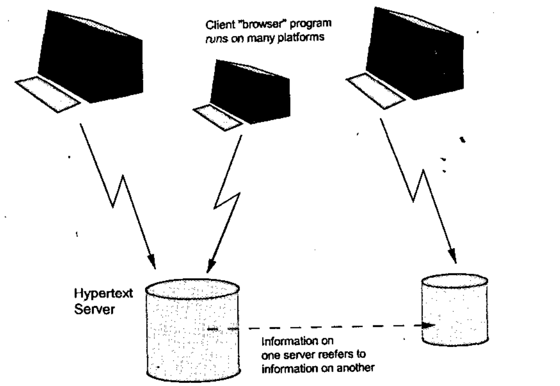


Fig 2. A client/server model for a distributed hypertext system.

Therefore, **an important phase in the design of the system is to define this interface.** After that, the development of various forms of display program and of database server can proceed in parallel. This will have been done well if many different information sources, past, present and future, can be mapped onto the definition, and if many different human interface programs can be written over the years to take advantage of new technology and standards.

ACCESSING EXISTING DATA

The system must achieve a critical usefulness early on. Existing hypertext systems have had to justify themselves solely on new data. If, however, there was an existing base of data of personnel, for example, to which new data could be linked, the value of each new piece of data would be greater.

What is required is a gateway program which will map an existing structure onto the hypertext model, and allow limited (perhaps read-only) access to it. This takes the form of a hypertext server written to provide existing information in a form matching the standard interface. One would not imagine the server actually generating a hypertext database from and existing one: rather, it would generate a hypertext view of an existing database.

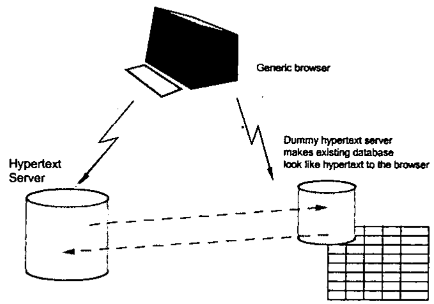


Fig 3. A hypertext gateway allows existing data to be seen in hypertext form by a hypertext browser.

Some examples of systems which could be connected in this way are

uucp News This is a Unix electronic conferencing system. A server for uucp news could makes links between notes on the same subject, as well as showing the structure of the conferences.

VAX/Notes This is Digital's electronic conferencing system. It has a fairly wide following in FermiLab, but much less in CERN. The topology of a conference is quite restricting.

CERNDOC This is a document registration and distribution system running on CERN's VM machine. As well as documents, categories and projects, keywords and authors lend themselves to representation as hypertext nodes.

File systems This would allow any file to be linked to from other hypertext documents.

The Telephone Book Even this could even be viewed as hypertext, with links between people and sections, sections and groups, people and floors of buildings, etc.

The unix manual This is a large body of computer-readable text, currently organised in a flat way, but which also contains link information in a standard format ("See also..").

Databases A generic tool could perhaps be made to allow any database which uses a commercial DBMS to be displayed as a hypertext view.

In some cases, writing these servers would mean unscrambling or obtaining details of the existing protocols and/or file formats. It may not be practical to provide the full functionality of the original system through hypertext. In general, it will be more important to allow read access to the general public: it may be that there is a limited number of people who are providing the information, and that they are content to use the existing facilities.

It is sometimes possible to enhance an existing storage system by coding hypertext information in, if one knows that a server will be generating a hypertext representation. In 'news' articles, for example, one could use (in the text) a standard format for a reference to another article. This would be picked out by the hypertext gateway and used to generate a link to that note. This sort of enhancement will allow greater integration between old and new systems.

There will always be a large number of information management systems - we get a lot of added usefulness from being able to cross-link them. However, we will lose out if we try to constrain them, as we will exclude systems and hamper the evolution of hypertext in general.

CONCLUSION

We should work toward a universal linked information system, in which generality and portability are more important than fancy graphics techniques and complex extra facilities.

The aim would be to allow a place to be found for any information or reference which one felt was important, and a way of finding it afterwards. The result should be sufficiently attractive to use that it the information contained would grow past a critical threshold, so that the usefulness the scheme would in turn encourage its increased use.

The passing of this threshold accelerated by allowing large existing databases to be linked together and with new ones.

A PRACTICAL PROJECT

Here I suggest the practical steps to go to in order to find a real solution at CERN. After a preliminary discussion of the requirements listed above, a survey of what is available from industry is obviously required. At this stage, we will be looking for a systems which are future-proof:

- portable, or supported on many platforms,
- Extendible to new data formats.

We may find that with a little adaptation, pars of the system we need can be combined from various sources: for example, a browser from one source with a database from another.

I imagine that two people for 6 to 12 months would be sufficient for this phase of the project.

A second phase would almost certainly involve some programming in order to set up a real system at CERN on many machines. An important part of this, discussed below, is the integration of a hypertext system with existing data, so as to provide a universal system, and to achieve critical usefulness at an early stage.

(... and yes, this would provide an excellent project with which to try our new object oriented programming techniques!)

TBL March 1989, May 1990

REFERENCES

[NEL67] Nelson, T.H. "Getting it out of our system" in Information Retrieval: A Critical Review", G. Schechter, ed. Thomson Books, Washington D.C., 1967, 191-210

[SMISH88] Smish, J.B and Weiss, S.F,"An Overview of Hypertext",in Communications of the ACM, July 1988 Vol 31, No. 7,and other articles in the same special "Hypertext" issue.

[CAMP88] Campbell, B and Goodman, J,"HAM: a general purpose Hypertext Abstract Machine",in Communications of the ACM July 1988 Vol 31, No. 7

[ASKCYN88] Akscyn, R.M, McCracken, D and Yoder E.A,"KMS: A distributed hypermedia system for managing knowledge in originations", in Communications of the ACM , July 1988 Vol 31, No. 7

[HYP88] Hypertext on Hypertext, a hypertext version of the special Comms of the ACM edition, is avialble from the ACM for the Macintosh or PC.

[RN] Under unix, type man rn to find out about the rn command which is used for reading uucp news.

[NOTES] Under VMS, type HELP NOTES to find out about the VAX/NOTES system

[CERNDOC] On CERNVM, type FIND DOCFIND for infrmation about how to access the CERNDOC programs.

[NIST90] J. Moline et. al. (ed.) Proceedings of the Hypertext Standardisation Workshop January 16-18, 1990, National Institute of Standards and Technology, pub. U.S. Dept. of Commerce