Journal of Universal Computer Science, vol. 1, no. 4 (1995), 221-231
submitted: 22/7/94, accepted: 14/10/94, appeared: 28/4/95, ©Springer Pub. Co.

About WWW

R. Cailliau,
(CERN, cailliau@www.cern.ch)

Abstract: The World-Wide Web is the most talked-about distributed information system today. This paper does not touch on its workings; it tries to give a brief history and outlines the feelings provoked by the explosive adoption in all circles of WWW as the first vehicle on the Global Information Infrastructure.

Keywords: WWW, World-Wide Web, History, SGML, Cultural Aspects, Society.
Category: H.5.1

1. Introduction

Before the World-Wide Web, networked information was difficult to access. With the Web, browsing through distant data bases has become almost a recreational pleasure.

It is fair to say that the Web is now driving the Internet. In fact many recent articles in newspapers and magazines simply make no distinction between the Internet and the World-Wide Web: it is as if the roads had been lying there for some time, waiting for someone to invent the Volkswagen.

I have followed the development of the Web from the days before it had a name. A brief history is therefore in order.

2. Brief History

In 1989, Tim Berners-Lee and I proposed independently a project for studying hypertexts and their possible uses at CERN. We joined efforts quickly, and Tim had already a prototype and a set of ideas to use the hypertext paradigm over the network. In this, he was no doubt influenced by the earlier work of Ted Nelson (Xanadu, [Nelson (88)]).

In 1990, Tim implemented the first browser/editor under the NeXTStep operating system. This was easily possible, since the NeXTStep system came with an object-oriented development kit which included not only a graphical interface builder (itself wysiwyg!) but also a programmable text editing object with paragraph styles. We were off the ground at least as far as we ourselves were concerned: the browser/–editor as a single tool for both navigating and correcting, editing, composing texts and hypertexts was a real dream object. To this date, the easy of use of that program has not been surpassed in the WWW world.

In 1991, Nicola Pellow, a technical student at CERN, wrote the Line Mode Browser. This was a simple, character-grid oriented client which was written in C (not even ANSI, but just flat C!). It could be compiled on just about anything but the kitchen sink. Through its availability, the Web began to spread outside CERN.

In 1992, the first steps were taken to implement format "negotiation", i.e. a mechanism whereby a server and a client could agree on the format of the document to be transmitted. This was also the year of the integration of all the other useful and existing protocols on the Internet: Gopher, ftp, telnet etc.

In 1993, Marc Andreessen, then a graduate student at NCSA, produced an X-window browser. Though I would term it primitive compared to the elegance and functionality of the NeXTStep browser, it had the marketing advantage of permitting colour images to be included. This sudden availability of colour pictures and proportional type fonts to the grey world of Unix gave the Web a boost it had not derived from anything else. Pictures were clearly the means to capture the imagination of the manager of your manager. The Internet programming community went wild. Mosaic became the synonym of WWW.

At the end of 1993, I decided that it was time to have all the early contributors meet each other in a great brainstorming session: I planned the First WWW Conference, which was held at CERN in May 1994.

1994 can truly be called the "Year of the Web". It became clear that CERN could no longer continue core development without external help, and a project was submitted to the European Community to fund a transitional phase of Web development in Europe. This project has a partner in the US. Its major aim is to ensure that there is a single, open standard in the Web mark-up and the Web communications protocol, based on a working reference implementation which is freely available.

The Web is now mentioned in any magazine at any time. Newspapers tell you how to connect to the Internet, URLs are routinely used in scientific journals to refer to information, they proliferate on the teletext pages of MTV, in short, you can no longer get away from it.

3. A successful system

To be a global success, a system must have two basic properties:

- it must have a small learning threshold so many people will join in,
- it must be sufficiently scaleable to stand up after large numbers of users and publishers have in fact joined.

The Web satisfies these conditions because:

- it defines an easy to understand name space for documents which is open-ended and addresses synthetic documents thus allowing interfacing to data bases and systems generating documents.

- it works over the Internet, making it global and accessible to a large community of programmers,
- early HTML was easy to generate, so populating the Web with information from existing data bases could be done through simple server interfaces.
- separation of form and content allows documents to be shipped without worrying about the capabilities of the client, making the web easily portable to all platforms.
- it is also easy to populate the Web because servers can be set up without prior consultation with previous publishers.
- the name space is based on the Internet naming scheme, therefore the Web scales like the Net.
- there are only fleeting connections, so servers can handle requests serially.

But being a success does not mean you are better than others or even merely good. The Web has also well-known disadvantages:

- it is not easy to write browser/editors, which are more difficult to make than word processors.
- it is not easy to find information because indexes do not scale.
- it is easy to get lost.
- it is not possible to control the quality or authenticity of the information, leading to a social problem.
- being open and easy to add to, the danger of divergence into incompatible systems is great.

In any case, to ensure the future, we must:

- maintain interoperability,
- have open standards,
- keep systems mutatable,
- produce interfaces that "can be understanded by them people"

Some of these points will become clear later.

4. Providing Information

4.1 SGML and layout

For the context of the points following, it is necessary to understand something of the SGML philosophy and to remove a number of popular misconceptions.

HTML is not a subset of SGML.

What is today called HTML should in fact have been named “the Web DTD”. For the sake of those who do not know all about SGML, I’ll briefly describe its most important features.

SGML is not a document format, it is a system for describing structures. It starts from the idea that there are sets of documents that look alike (or should look alike) in structure. For example, all novels are divided into elements called chapters, and each chapter is a sequence of paragraphs.

Elements of a document are marked-up by putting tags at the appropriate places in the text.

The structure of each set of documents can be described by a Document Type Definition (or DTD) which is a formal grammar about its elements.

A DTD defines only the structure, but not the presentation. The presentation of each element is given in a separate object, a so-called style sheet.

SGML then allows for complete separation of structure and presentation.

Now, if a person wants to communicate a document to you, you need in total three objects:

- the document itself, with the mark-up in it (tags),
- the DTD, which tells you what tags mean and how they relate to each other,
- the style sheet, which tells you how to present each element.

The way in which WWW uses SGML is (slightly simplified):

- the Web DTD is agreed beforehand by everyone (and called HTML),
- the presentation is left to the browser (client) so each individual can set it to his/her liking.
- the document only is shipped from server to client.

We are now limited to HTML, since only the document is transferred. However, because there is no objection to ship a complete MIME message from the server, including a DTD and a style sheet, there is room for a complete implementation of distributed SGML systems using the Web.

Keeping presentation separate from contents has enormous advantages:

- automated treatment of information is easy.
- no knowledge is needed of the presentation capabilities of the client.

It is far easier to look for author names when these are tagged properly in simple text documents than when they are embedded in proprietary formats using just text styles.

Thus, HTML tags like <b> for bold and <i> for italic are really meaningless and should be avoided.

The world is alas divided into those who want to supply information and those who want to supply advertising. Information may be given to humans or to their computerised agents. Advertising is normally done by visual impact directly to a receiving human, which requires complete control over layout. The image is omnipresent on the Web, showing that the majority of servers are actually providers of advertising.

4.2 The Paper Metaphor

We are learning how to provide information in the new medium of distributed hypertext. This early phase is dangerous: if we are not bold enough, we may never escape from the paper and book metaphor.

So far, a lot of existing stuff has been made available. This information was destined to be presented on paper. Existing methods and habits for preparing it are all more or less related to text processing for the printed page. The microcomputers in use today are practically all employed for such purposes: printed reports.

The Web does not need printing. It is useless to print a well-conceived hypertext that is kept up-to-date. A paper copy leaves me uneasy: is it still valid? Yet the most frequent question that novice users of the Web ask is "can I print this?". Yes, you can, but why?

For the same reason of attachment to the printed page, we have seen floods of converters from proprietary formats into HTML, but not a serious browser/editor and (to my knowledge) only a single rudimentary tool to assemble a number of Web documents into a larger one for printing.

Everyone seems to see the printed, word-processor document as the original, and the Web document as the derivative. This is bad news.

4.3 Existing Information

Existing information of course has to be published somehow, so converters are not completely undesirable. There are four courses of action to get existing documents published on the Web:

- do nothing at all. This is the easiest method and obviously consumes no resources but also achieves nothing.
- publish references. Here the Web user will find at least that the document exists, and perhaps a way of getting a copy somehow.
- publish as-is. Now some Web users can see the document (if they have an application that is capable of handling the proprietary format after transferring the document file). Others will see only the reference but will not be able to follow the link.
- convert to Web structures. This is best, but consumes most resources, since links have to be put in, redundancies have to be removed, and chunking (dividing the original up into reasonable hypertext portions) is not always obvious.

4.4 New Stuff

For the new stuff, we should think of the far future. The Web is not the ultimate repository of human knowledge. But we want to keep the documents somehow, what they mean and what they have to say. Thus we should think: "how are we going to read this a hundred years from now?" And I'll bet it will not be on an Intel-based PC. So we had better make sure that the semantics are preserved, and that their format is easy to convert from current systems to the next ones. We must make the contents mutatable. For me this is one of the more important reasons to use SGML-like encoding and to encourage authors to use it correctly.

5. Tools

5.1 Collaborative Tools

Currently we have a Read-Only Web. It is not possible to change a displayed document (unless you happen to use a copy of the NeXTStep browser), even if it belongs to you and you have all access rights to the file.

There are projects under way to change this situation, notably the GrifW3 which is based on wysiwyg SGML technology, but we are not there yet. In this respect, the Hyper-G system is far superior to the Web.

Once there is no longer a difference between a browser and a Web editor, people will be able to use the Web as a collaborative tool, developing information and documents together, independently of geographic location or time zones.

Note that an browser/editor gives you these advantages:

- no more problems with starting points, since the construction of my home page (see note at bottom of list) is now easy: it contains the pointers of the places I'm most interested in and perhaps some comments.
- hot lists are a thing of the past, since they are just edited local HTML files, and I can impose any structure on any number of them.
- personal annotations and group annotations and annotation servers and the like are just all unified into sets of HTML files again.
- a lot of converter use goes away.
- pagination, the nightmare of traditional text processing is gone.
- bad HTML is never produced by a proper editor.
- some linking problems disappear (the editor can know about relative links).
- printing diminishes.
- you can organise your own notes in the same hypertext way as anything that you publish, and with drag-and-drop ease on your laptop.

(Note: a "Home page" is where I as an idividual start from when I launch a browser on my computer. Its contents are nobody's business. A "Welcome page" is what I get from a server as the generic page when I specify only the server name in a URL. Currently, the term “home page” is used for what should be called a “welcome page”. The confusion comes from the fact that almost everyone gets a local welcome page when they start up a browser: the absence of editors makes it very difficult for non-expert users to build a home page.)

5.2 Rhetoric

Of course, a new rhetoric has to be learned. As today, some people will write badly for the Web, others will be brilliant authors.

5.3 Navigation

Anyone who has seriously tried to work the Web knows that it takes no effort to get lost. Systems like Hyper-G provide you with excellent navigation tools, Gopher is so structured that it is easy to know where you are.

The distributed nature of the Web makes it difficult to show the user a map, unless he/she is prepared to incur a lot of network traffic overhead.

5.4 Construction

Even with collaborative tools and browser/editors we will still need a variety of tools to help us construct the information. In word processors there are items like outliners, style checkers, even tools to help you organise your thoughts. For a distributed system, we need tools like that, but they must concentrate on other aspects: finding existing materials, generic referencing, suggesting which phrases should be linked to explanations, helping cut long parts up and rearranging and merging shorter ones. Plenty of room for research.

6. Social and Cultural aspects

We have in computing seen the negative effects of too many inventors in too many areas. Short-term commercial interests often force us to adopt computing solutions that are frustratingly complicated and that direct us into dead-end streets, holding back real change for decades. Networking has not been an exception.

Too much attention is devoted to backwards compatibility (a term invented by software developers?). Technology should not be worried about the past. I prefer the attitude expressed in this maxim [De Bono (91)]:

> *Instead of being pushed by our history,*
>
> *we should be pulled by our vision.*
>
> *E. De Bono*

6.1 Change

Inside the network, we are rebuilding the world we know: we use the book metaphor, we want total layout control (which is a help for visual navigation, but definitely not for computer aided navigation!), we talk about adding a worn-out look to frequently consulted objects (do I want my welcome page to look worn out by complete strangers?).

Is it possible to think about what the networked society should look like or is it only possible to let a thousand weeds grow?

6.2 Isolation

Communication is different from the Gigabit/second number. There is a lot of "communicating" being done, yet I know nobody who is happy receiving 100 e-mail messages a day. You feel obliged to answer (a vestige of the days of personal contact?) but what you get is the stress of having done so badly, hurriedly and tiredly.

Increased isolation of individuals has been the result of increased exteriorisation of information. Books made it possible to know something without having to contact the author. Radio and TV spread news without personal contact. The Walkman effectively shields a person from investing in contact with others on public transport trips and certainly is used in this way. A recent article [? (94)] proclaimed as an advantage of the global network that

> *"Sex, location of a partner, video marriage, …*
>
> *You can have any kind of interaction without the inconvenience of having someone in your house"*

The key word is "inconvenience". The lack of social contact is perhaps the most negative side of our Western culture, and is so perceived in most other cultures.

6.3 Life, the Universe and Everything

In an article of 1977, describing his graphical user interface, Alan Kay wrote [Kay (77)]:

> *There are three reactions to the introduction of a new medium:*
>
> *illiteracy,*
>
> *literacy*
>
> *and artistic creation*

He goes on to say:

> *After reading material became available the illiterate were those who were left behind by the new medium. It was inevitable that a few creative individuals would use the written word to express inner thoughts and ideas. The most profound changes were brought about in the literate. They did not necessarily become better people or better members of society, but they came to view the world in a way quite different from the way they had viewed it before, with consequences that were difficult to predict or control.*

How will the networked society influence the daily lives of normal people? Below are a number of questions. In the current social structures, I can think of negative answers only:

- will employment go up?
- will people feel better?
- will general education improve?
- will the world be a safer place?
- what will happen to the service sector?
- will power be in the hands of benevolent people?
- is this what we want?

Is it therefore not more urgent to work on structural reforms in our culture rather than on laying down fibre cables or shooting 700 satellites into the sky?

6.4 Work

The argument against Malthus' predictions of overpopulation was that food production could be made to expand faster than the population growth. However, there really is a limit to the number of people that can live on earth, and only the very obstinate now hold that we have no population problem.

Likewise, an old economic argument has been that the introduction of machines will not take away work, just displace it to other activities. So we have seen a massive movement from agriculture to industry and then to services. But the computer is not a machine like any used in the industrial revolution and after: it displaces people.

Consider the distribution of work. Before 1900, there was lots of work, despite mechanisation. Jobs of low and high intellectual content were plenty. There was even a healthy overlap in the middle.

During this century, we have witnessed a constantly growing separation: jobs are either menial or intellectual, and there are fewer of them. They are less gratifying. Menial jobs are done by immigrant workers, intellectual jobs need high qualification which not many can attain. The young generation is acutely aware of the "No Future" syndrome.

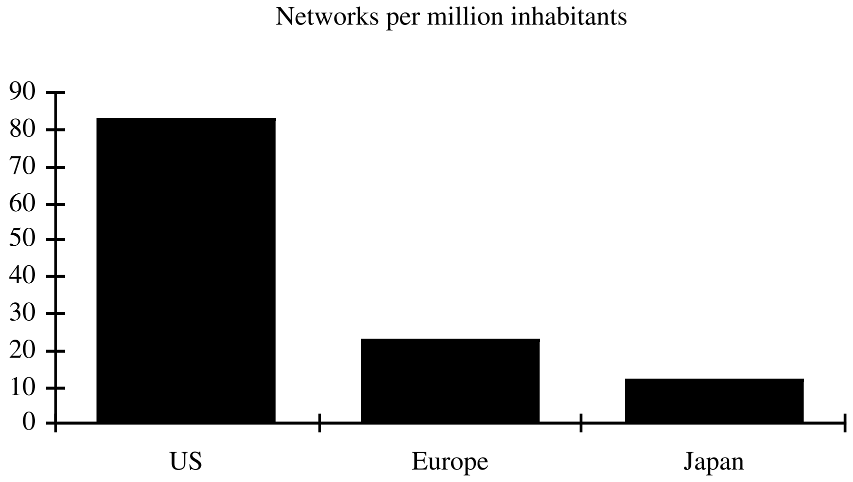
With the massive introduction of computing, fewer and fewer real jobs remain. The service sector which absorbed people during the boom years no longer does: those who are unemployed remain so. The recovery of the economy does not result in higher employment.

Networking makes the problem more acute: when you can learn immediately from the best in the world, why go anywhere local?

Will the networked society be able to create more jobs or will it just lead to much more efficient service companies, leading to higher unemployment? Maybe the time has finally come to start working less and less hard.

7. Europe

I was once told that "Europe has a cultural deficit in networking". Brilliantly and concisely expressed. I set out to calculate the value of the deficit. as a simple expression: the number of networks in use per million inhabitants. The Internet statistics give numbers of networks, an encyclopaedia numbers of people. The graph below needs no comment. One could argue whether I now like this situation or not, given what I wrote before.



Europe's cultural deficit in networking

Europe has many assets, and I am attached to this strange assemblage of peninsulas. But we have one big problem: reacting speedily and nimbly to situations which need mobilisation of large resources. We seem to be poor at exploiting ideas that we generate here, especially in high-tech. Will Europe play a real role in the global networked society or will we have to buy everything from a US software company, even though we keep generating the important ideas?

Three axes are important for any society that wants to partake of the coming network culture [Abramatic (95)]:

- core technology, (you need to understand the options),
- tools (their use determines your competitive efficiency),
- content published (your visibility in the marketplace).

The content of a network server in WWW is what makes people look at it. We can make a big impact there and remain at level with other parts of the world.

The making of the content is dependent on the tools you use. If Europe makes no good tools to support its diverse cultures, then we will have to wait until someone else supplies us with them. Or doesn't.

Tool construction in turn depends on intimate knowledge of the core technology. It is not only important that the WWW standards remain open. It is even more important for us here that Europe maintains a strong core development effort so that our computer scientists can work locally on projects that will give them the necessary expertise. Our continued presence in the tools and contents areas rests on this expertise.

8. Brief Future & Conclusions

Networks will keep up with the traffic: speed will go up, costs will come down. There will, like on normal motorways, be traffic jams here and there. But on the whole, things will keep pace.

Collaborative work will take off especially in research, where it traditionally has been, but also inside big companies.

A new rhetoric will develop, as compelling as that of advertising. This may distort our perception of reality.

The networked society will probably fall apart not just into rich and poor, but into rich informed and poor information-illiterates.

The Web is fast becoming an entertainment medium. Perhaps Andreessen's ideas of a browser already contained the seeds of the Web entertainment business. But for me, a member of the minority who want to stay in real reality, the questions for humanity are:

- do we want entertainment or collaborative tools?
- do we want to drown in multimedia sense overload or do we want text searches?
- do we want proprietary encryption or trusted carriers?
- … or … ?

Maybe I'm just old-fashioned…

9. References

[Nelson (88)] Theodore H. Nelson: "Literary Machines", The Mindful Press, 1988

[De Bono (91)] E. De Bono: "I am right, You are wrong", Penguin, 1991.

[? (94)] Airline in-flight journal, November 1994 (?)

[Abramatic (95)] J.F. Abramatic, INRIA, private communication.

[Kay (77)] Alan C. Kay: Microelectronics and the Personal Computer, Scientific American, September 1977, p.231