THE WORLD WIDE WEB

Brian Hayes

This is the year of Internet fever. Newspaper reporters are prowling "cyberspace" and bringing back tales of wonder and woe. Patrons of your local espresso bar are reading *Wired* and *Internet World* and *Mondo 2000* and chatting knowledgeably of SLIP accounts, kill files, flame wars and emoticons. The espresso bar itself may well have an Internet connection. Any day now, the storefront that once sold CB radios and then became a Computerland and now deals in home satellite dishes will be reincorporating as an Internet access provider. That's when we'll know for sure that the wave has passed, and it's time to move on to the next fad.

I am chagrined to be contributing further to this publicity frenzy. It's as if I were writing a "Computing Science" column about Tonya Harding or the O. J. Simpson case. And yet there is something happening on the Internet that is too important to pass over in silence. It is called the World Wide Web—a name that proclaims its ambitions. I believe the Web offers the clearest vision yet of what computer networking, and communication more generally, is going to be like in the future.

The Universe in a Box

What is the Web? Is it a place? A program? A protocol? One of the documents in which the Web describes itself offers this assessment: "The World Wide Web (W3) is the universe of network-accessible information, an embodiment of human knowledge." That about covers it.

For something as vast as a universe, the Web is surprisingly easy to find your way around in. It works like this. On a computer connected to the Internet you start up a program called a browser; the browser goes out over the network and retrieves a document, which we can assume for the moment is simply a page of text. Within the text are some highlighted phrases, displayed in color or underlined. When you select one of the highlighted words by clicking on it with a mouse, a new document appears, with new highlighted "links." Clicking on one of these links takes you to still another document. Each time you follow a link, you may be visiting another network site, perhaps quite distant from your original destination as well as from your own location.

If that's all there is to it, what's the big deal? Well, there *is* something more to it, which I shall discuss momentarily, but even this one mechanism has a remarkable effect on the way the network presents itself. There have long been protocols for transferring various kinds of information over the Internet, but the Web offers the first seamless interface to the entire network. You no longer need to think much about where things are, either physically ("The file I need is in Geneva, Switzerland") or in terms of the syntax of domain names and addressing conventions ("That file is stored in the pub/ directory at info.cern.ch"). Geography disappears, and so does network topology. The Web promotes the illusion that all resources are at your fingertips; the universe of information is inside the little box that sits on your desk.

The highlighted words that serve as links to other documents make the Web a *hypertext* system. The "something more" alluded to above is that the Web is actually a *hypermedia* system, because the linked documents need not be text alone. Most Web documents include graphic elements—photographs, drawings, diagrams, and various kinds of ornamental flourishes. Small images are displayed automatically within the document; larger ones can be acquired by clicking on an icon. Other links lead to digitized sounds or video sequences. Web browsers also know how to deal with links to older Internet services and protocols, such as Usenet news (a collection of bulletin boards), ftp (for file transfers) and Gopher (a hierarchically organized distributed database).

We're Off to See the Web...

Explaining the Web would be a lot easier if I could typographically highlight a word and thereby convert it into a hypertext link. Then you might point to one of the magazine titles mentioned a few paragraphs above and see your copy of *American Scientist* transformed into *Mondo 2000*. Unfortunately, the hypertext printing press is not yet working, and so I can offer only a travelogue, not a real guided tour.

A good place to begin exploring the Web is a document called the World Wide Web Home, maintained at CERN, the European Laboratory for Particle Physics in Geneva, which was also the

Brian Hayes is a former editor of American Scientist. *Address: 211 Dacian Avenue, Durham, NC 27701. Internet:bhayes@mercury.interpath.net.*

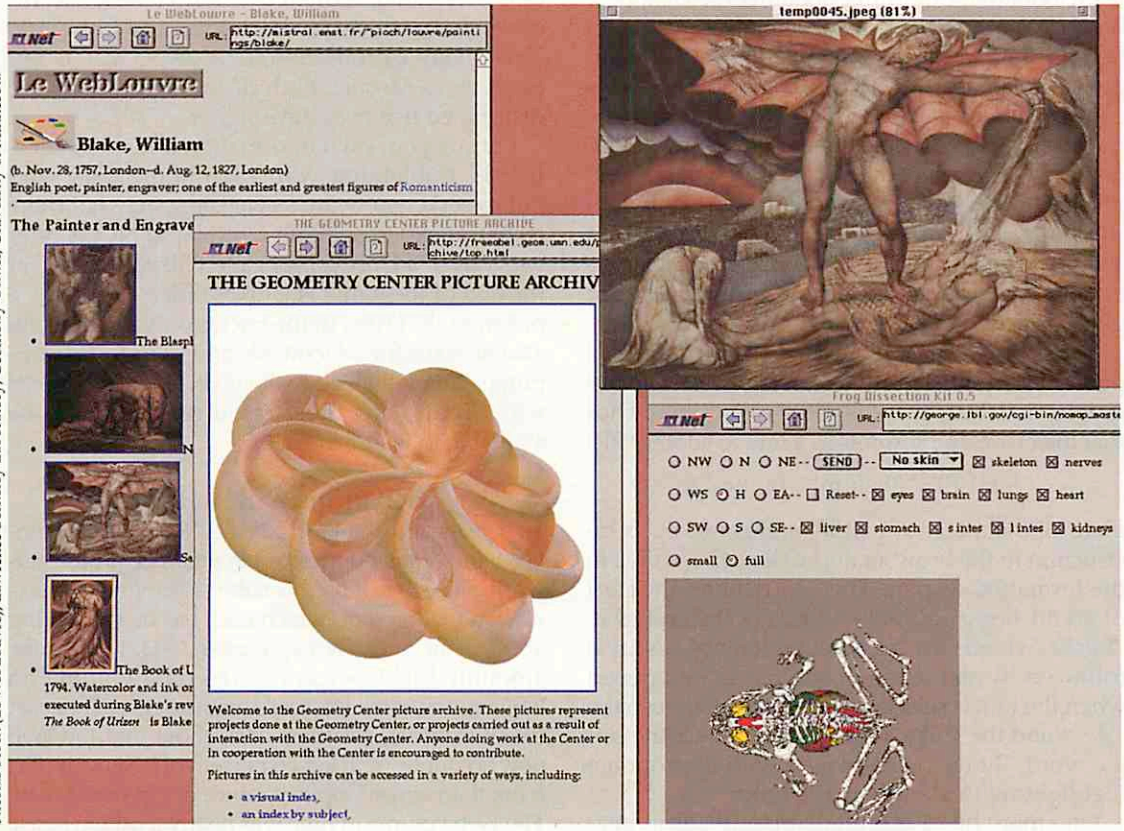
birthplace of the Web. Most such "home pages" include links both to locally stored information and to other sites. In this case the local information has to do with the Web itself—there are manuals, tutorials, histories, software, etc. ("Everything there is to know about W3 is linked directly or indirectly to this document," it says.)

The navigational tools provided at the CERN Web Home include a catalogue of Web nodes organized alphabetically by subject. *Astronomy* is near the beginning of the alphabet, and so off we go to the home page of the American Astronomical Society. There we can read abstracts of papers to be presented at the society's upcoming meeting, or check a register of job openings. A link from the AAS page sends us to a voluminous directory of astronomical and astrophysical web sites compiled at the Space Telescope Science Institute in Baltimore. From this list we might turn in many directions—to the Kitt Peak National Observatory, the Jet Propulsion Laboratory, the European Southern Observatory (want to know the weather in Chile?), or the archive of astrophysics e-prints at the Lawrence Livermore National Laboratory. Back at the Space Telescope Science Institute we can pick up images of the Shoemaker-Levy 9 cometary impacts on Jupiter. (In July, the images were being posted on the Web hours after they were made.)

Astronomers have taken to the Web with particular eagerness, but other disciplines are also well represented. A mathematician might stop first at the home page of the American Mathematical Society, which has a preprint archive, or look over the latest issue of the *Electronic Journal of Combinatorics*. The Geometry Center at the University of Minnesota has one of the most carefully tended Web pages. It includes a gallery of geometric images and a facility for running software installed on computers at the center; for example, you could experiment with Eugenio Durand's program for generating quasiperiodic tilings.

For the biologist on the Web a likely starting point is the biosciences list maintained at Harvard University as part of the World Wide Web Virtual Library project. Essentially all of the genomic and protein databases that have become so central to the practice of molecular biology are now accessible on the Web. A site at the National Institutes of Health provides direct access to GenBank and the Brookhaven Protein Data Bank. The Genome Database at the Johns Hopkins University has a hypertext WWW interface, and Hopkins also offers Web access to sequence-alignment software available at the Oak Ridge National Laboratory. Running a DNA query sequence through the Oak Ridge programs yields a list of possible matches, with each item in the list a hypertext link to the database entry on that sequence. For another kind of biology, the Lawrence Berkeley Laboratory has a "frog dissection kit"; unlike the real thing, it's as easy to put back together as it is to take apart.

Perhaps the most important nodes on the Web are those that index and organize resources. Wandering aimlessly from node to node is always engrossing but not always very productive;



Nicolas Pioch (Le WebLouvre); Lawrence Berkeley Laboratory; Geometry Center, University of Minnesota

Pages from the Web: Catalogue and painting from Le WebLouvre, a dissected frog, pictures from the Geometry Center.

it is like wandering through the stacks of a large library and stopping whenever a title catches your eye: You will probably find something interesting but not necessarily what you want. The various lists of links mentioned above are analogous to a library catalogue, but the Web also has more flexible search-and-retrieval services. For example, the World Wide Web Worm (at the University of Colorado) and the Web Crawler (at the University of Washington) are programs that periodically explore every node they can reach and compile a database of what they find; when you query the database, the results are presented as a list of hypertext links, so that you need only click on an item to be taken there.

How the Web Works

The inventor and chief architect of the World Wide Web is Tim Berners-Lee, a text-processing expert on the staff at CERN. He first proposed the system in 1989 as a tool for creating and reading structured documents, such as software manuals. The first components of the system were working by 1991, but the Web did not begin to spread outside the high-energy physics community until 1993.

The most fundamental technology underlying the Web is the Hypertext Transport Protocol, or HTTP, which is the set of rules governing communication between a browser, or client, and a Web server. The ordinary user of the Web needn't know anything of this protocol.

At a somewhat higher level of abstraction is HTML, the Hypertext Markup Language, which is the notation for writing documents that appear on the Web. HTML is a simplified derivative of SGML, the Standard Generalized Markup Language. A key feature of HTML is that it describes content rather than appearance. Tags inserted into the text identify its component parts—headings, paragraphs, references, etc.—without saying explicitly how those parts should be presented on the page. How a subheading or a quotation should be formatted is a decision left to the browser software. In most cases the browser in turn gives stylistic control to the reader.

The mechanism for creating hypertext links is part of HTML. Just as there are tags to label paragraphs and lists, there is a tag called an anchor that the browser interprets as a link to another file:

```
<A HREF="file.html">Link</A>
```

Here all the material within angle brackets is instruction to the browser and will not be visible in the formatted display. The "A" denotes the start of an anchor, and the "/A" marks the end of it. "HREF" stands for hypertext reference, and it is followed by the name of the file to be opened when the link is selected. Everything between the <A...> and the </A>, which in this case is the single word "Link," is displayed with appropriate highlighting to identify it as a link.

One more bit of essential apparatus is a notation for specifying files and other resources stored on remote computers. The notation is called a uniform resource locator, or URL, and it looks like this:

```
http://machine.domain/directory/file.html
```

The first part of the URL, in front of the colon, identifies the kind of resource; "http," naturally, specifies a document to be accessed with the Web's own hypertext transport protocol. Other possible designators include "ftp," "gopher" and "news." The rest of the URL tells where to find the resource, using a syntax that amalgamates the domain-name system of the Internet with the hierarchical directory structure of Unix and other operating systems.

It bears emphasizing that you can navigate the Web without learning anything about HTTP or HTML or even URLs. They are part of the infrastructure of the Web and are normally kept out of sight. The component of the Web that *is* visible—indeed, conspicuous—is the client software. And a major reason for the Web's popularity is that it has been blessed with spectacular client software. The premier browser is a program called Mosaic, originally written by Marc Andreesen at the National Center for Supercomputing Applications in Illinois and developed further by a team of programmers at NCSA. It is software that works so well and so naturally that you can't help smiling the first time you lay hands on it. Mosaic is so closely associated with the Web that the two things are sometimes confused, but there are several other browsers as well. A new one for the Macintosh, called MacWeb, written by John Hardin and his colleagues at the Microelectronics and Computer Technology Corporation in Texas, is just as impressive as Mosaic. Both of these programs are distributed free over the Internet.

Putting your own information onto the Web—that is, publishing your own home page—requires server software as well as somewhat more technical expertise. Both CERN and NCSA offer versions of a Unix server called httpd (the "d" at the end of the name is Unixese for a "demon," a program that runs in the background). There are also servers for Macintosh and Windows computers. You will need to create HTML documents with links to any other resources that are to be available to visitors.

Why the Web Works

Two years ago, the World Wide Web was an unnoticed and almost undetectable presence on the network. Since then it has been the fastest-growing service on the Internet (which itself has been growing at a fearful rate). By September 1993, Web traffic amounted to 1 percent of the data volume on NSFnet, one of the Internet's main arteries. The most recent statistics, for June 1994, show that the Web now accounts for at least 6 percent of NSFnet traffic, more than either Gopher service or electronic mail. The Web volume in June was nearly a trillion bytes, and it has been doubling every two or three months.

The Web has some distinctive properties that set it apart from other Internet services:

The Web is transparent. You can begin to forget where you are and think of the entire Web as one big file system. Whether a node is in Geneva or New Zealand, it's just a click away.

The Web is democratic. It is a web, not a tree-like hierarchy. There is no official top node. You can start anywhere and follow any path from site to site.

The Web is sessionless. Each request issued to a server is a separate transaction, which lasts only as long as it takes to transfer the data. Between requests, while you are reading, the browser consumes no network resources.

The Web encourages sharing and discourages hoarding. Protocols such as Gopher and ftp help you find a document and then bring home a copy. The Web can be used in the same way, but most Web documents work best when they are left in place. People generally retain only a reference to a document (typically as an item on a "hotlist" that the browser maintains).

The Web is paperless. Yes, you can print a copy of a document, but all the life goes out of it as soon as it leaves the screen, because the hypertext links become inactive. Sounds and videos, of course, cannot be printed at all. There is a positive feedback loop at work here. If a document makes use of devices that do not work on paper, then people will read it on the computer; if people read documents on the computer, then authors will be liberated to exploit devices that do not work on paper.

The Web has a degree of formality. Whereas e-mail and Usenet news are media for hasty jottings and rantings, material published on the Web is generally prepared with greater thought and care. Home pages are meticulously designed (although not always *well*-designed).

The Web is modular and extensible. Most browsers do not attempt to understand all possible media and file formats. Instead they rely on "helper" programs that are automatically launched when needed, such as when there is a sound or a video to be played. With this mechanism it is fairly easy to accommodate new media and formats.

Other aspects of the Web give rise to misgivings.

Availability is one problem. There are text-only browsers that can be used with almost any Internet account, but to get the full effect of the Web, you need a computer that is directly present on the Internet. Many potential readers and authors are therefore shut out.

Solving the availability problem will probably create more serious worries about bandwidth—that is, the information-carrying capacity of the network. The Web protocol is reasonably efficient, but because of the graphics and other hypermedia elements, someone browsing the Web generates far more network traffic than someone reading electronic mail or Usenet news.

I have concerns of my own about the effect of the Web on the nature of writing. Hypertext offers an intoxicating opportunity to incorporate ideas by reference rather than by quoting or paraphrase. This ought to be a good thing. Based on the evidence on the Web today, however, hypertext has an alarming tendency to degenerate into a list, or at best an outline, without narrative continuity. Among Web authors there is too much compiling going on and too little writing.

The visual presentation of information on the Web is also in a primitive state. Developing the art and technology of printing a magazine like this one has taken a couple of centuries; the necessary skills cannot be transferred overnight to a

The Weekend Webster

There is more to the Web than physics and mathematics and other earnest pursuits. Here are some lighter links. If this were a page on the Web itself, you could merely click on the boldface headings to be taken to each site. Since this is printed on unclickable paper, you will have to type the URLs into a Web browser.

Information on the Superhighway. A color-coded map shows average speeds at several dozen points along the San Diego freeway system, updated minute by minute. Is your car computer on the Web?

http://www.scubed.com:8001/caltrans/transnet.html

Le WebLouvre. Unlike other museums, this one lets you take the paintings home with you—pixel by pixel. The galleries opened in March, and when I last visited in July, I was issued ticket number 73,310.

http://mistral.enst.fr/~pioch/louvre/

The Juggling Information Service. See Anthony Gatto juggle eight rings. Get software that teaches the juggler's art.

http://www.hal.com/services/juggle/

Internet-accessible Coke Machines. Check the stock of your favorite vending machines at Carnegie-Mellon, Columbia and other caffeine capitals. At Berkeley you can also see who's been eating the HoHos and the Death Muffins. If you have an account, you can buy a Coke from a machine in Perth while sitting at a terminal in Peoria.

http://www.cs.cmu.edu:8001/afs/cs.cmu.edu/user/bsy/www/coke.html

LEGO. The history of LEGO bricks; the LEGOland theme park in Denmark; a tour of the LEGO factory in Enfield, Connecticut; LEGO robots.

http://legowww.itek.norut.no/

The Weather. Several weather Gophers can be accessed through the Web, but there is also a native HTTP page at Michigan State. Superimposed on a national weather-radar map are the current temperatures in a few hundred cities; clicking on a temperature calls up a detailed weather report.

http://rs560.cl.msu.edu/weather/interactive.html

Recipes. There are several large recipe archives in the U.S., but when I went looking for Cajun cooking, I found it in New Zealand.

http://www.vuw.ac.nz/who/Amy.Gale/recipe-index.html

Lunch. See what Sho Kuwamoto, graduate student at Purdue University, has in his brown bag today.

http://physics.purdue.edu/~sho/lunch_main.html

Bluedog Can Count!! It's too stupid for words. Woof, woof.

http://hp8.ini.cmu.edu:5550/bdf.html

new medium. The fact that the Web puts typographic controls in the hands of the reader complicates the situation further.

Watching myself browse the Web brings up still another worry. With the excuse of preparing to write this column, I have spent many late nights wandering the Web from one node to the next, glassy-eyed, weary, waiting to be entertained. It is easy to fall into an almost hypnotic state, clicking the mouse on new links in much the same way that you click the button on the television remote control. Thus it appears we have another addictive, passive medium, ready to kidnap America's children—those few who have not already succumbed to television itself or to video games. If the Web succeeds as I expect it to, the next generational decline in SAT scores may well be blamed on WWW rather than TV. In my view the hazard of immersion in alternative realities is a real one, but our response to it betrays some curious cultural biases. Look at a child reading a book, turning page after page as if in a trance, oblivious to the "real" world. This behavior too looks obsessive or addictive, but no one seems to be upset about the menace of all those libraries in the heart of America's cities.

As I was writing this essay, and trying to find ways of conveying my enthusiasm for the Web, I kept toying with a half-remembered quotation from an unremembered source: "I have seen the future, and it works." Yes, I thought, that's just how I feel, coming back to the world of paper and faxes and telephones after a few hours on the Web. Finally I resorted to *Bartlett's* to find out who said it first, and about what. The actual words are: "I have been over into the future, and it works." They were spoken by Lincoln Steffens in 1918. He was just back from a visit to the revolutionary new regime in Russia.

Bibliography

Berners-Lee, T. J., and R. Cailliau. 1990. WorldWideWeb: Proposal for a hypertext project. Available as: http://info.cern.ch/hypertext/WWW/Proposal.html.

Berners-Lee, T. J., R. Cailliau, J. F. Groff and B. Pollermann. 1992. World-Wide Web: The information universe. *Electronic Networking: Research, Applications and Policy* 2:52–58.

Boutell, Thomas. 1994. World Wide Web frequently asked questions. Available as: http://siva.cshl.org/~boutell/www_faq.html.

Hughes, Kevin. 1994. Entering the World-Wide Web: A guide to cyberspace. Available as: http://www.hcc.hawaii.edu/guide/www.guide.html.

National Center for Supercomputing Applications. 1994. A beginner's guide to HTML. Available as: http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html

Rugier, Mario (compiler). 1994. *Preliminary Proceedings of the First International Conference on the World Wide Web*. Available as: http://www1.cern.ch/WWW94/PrlimProcs.html.

Torkington, Nathan. 1994. World Wide Web primer. Available as: http://www.vuw.ac.nz/who/Nathan.Torkington/ideas/www-primer.html.