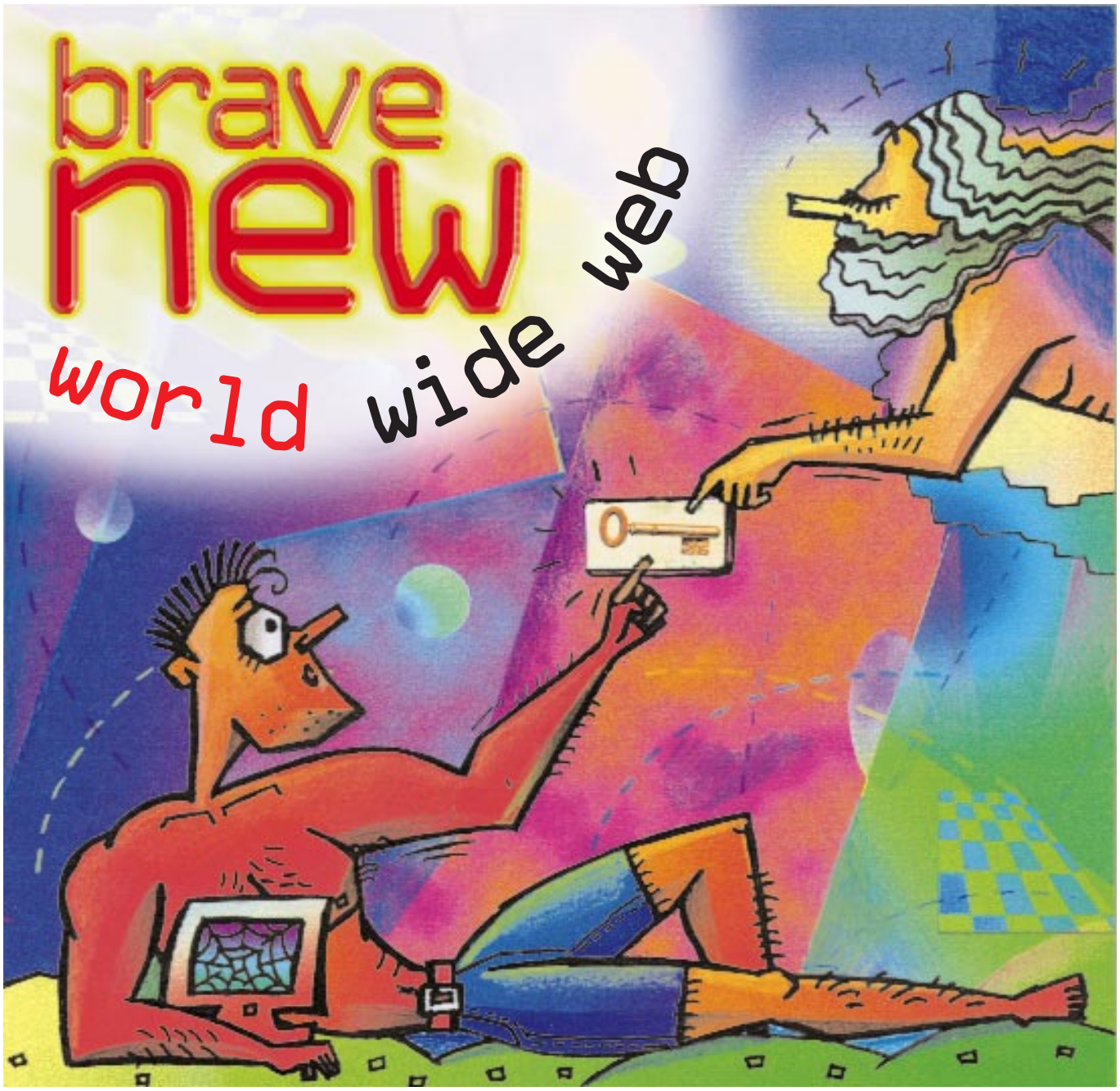
Illustration: FARZANA COOP-




**Ever clicked on a promising hyperlink only to be greeted with a cheery 'HTTP 404 Not Found'? Slow browsing and dead links will soon be extinct if current developments in Internet technology are anything to go by**

Let us assume that you want to access the homepage of CHIP and you don't know the address. The straightforward way of finding the URL would be to go to a search engine such as AltaVista or HotBot and type 'chip' in the Find box. Unfortunately, such a search is likely to throw up millions of hits, ranging from pages devoted to microchip manufacture to Aunt Edna's recipe for chocolate chip cookies!

Blame it on HTML? You would not be entirely wrong. HTML, or HyperText Markup Language, is only a presentation language that is more concerned about how the data on a page will be displayed rather than throwing any light on the content itself. It is true that HTML is easy to learn; its syntax is simple and compact. But the rigid formatting, the inability to layout text and graphics precisely how we want them,

the links that are easily broken... clearly, HTML could do with an overhaul.

A complete makeover is just what the Web is readying for, as the Internet matures from adolescence to young adulthood. The hottest thing to hit the Web lately—eXtensible Markup Language (XML)—is only a taste of exciting things to come.

### eXstatic About XML

What is it that makes XML so radically different from HTML? For one thing, XML defines the data structure rather than describing how the data will be presented. Unlike HTML, XML also lets you define your own tags; thereby giving you control over the document structure (*See box 'XML–the Components'*).

Big deal, did you say? What good are data structure descriptions if each person describes data in his/her own way? XML enables authors to devise new element names and publish them for others to use, and thus help establish standard terms for common data elements. These published 'namespaces' will prevent name conflicts and make search engines more efficient.

XML is fully internationalised for all major European and Asian languages. Though more complex than HTML, it is far easier to use than its parent, SGML (*See box 'In the beginning...'*). To quote from Richard Light's *Presenting XML*, "XML offers 80% of the benefits of SGML for 20% of its complexity."

A host of applications are expected to boost the relevance of XML in the immediate future. The most important application will probably be the Resource Description Framework (RDF), which lets applications describe to each other not only data in each document but also new data fields and classes. Microsoft's Channel Definition Format (CDF) is a fairly new XML application that allows Web publishers to control Push technology. Open Financial Exchange (OFX), a framework for exchanging financial data that Microsoft says will soon be XML-compliant, will enable companies to connect directly to their customers. The Open Software Description (OSD) format, which is ready for ratification by the World Wide Web Consortium is a set of XML tags that describes software packages and their dependencies, and for software running on multiple platforms. With MathML or Mathematical Markup Language (an XML application) one can accurately depict mathematical expressions and content—the first time this is possible on the Web.

### This ASP has bite!

Not all developments are radically new. Consider Active Server Pages, for example. You might have occasionally come across a Web page with an address that ends in .asp instead of .html. You would probably have failed to notice any major differences between ASP sites and standard HTML sites. As far as the end user (who is accessing a Web site using Active Server Pages) is concerned, there is no difference. The differences are all behind the scene, at the server end. ASP is a major step in bridging the gap between applications and Web sites, enabling developers to make their sites as functional as possible. ASP pages, such as Allaire ColdFusion's CFM pages and Java's JHTML pages, are handled at the server end itself.

How does ASP work? ASP breaks the static nature of normal HTML by using a new technology known as Active Server Components, a broad category of Internet technologies. This category includes scripting languages such as Jscript, VBscript or even programming languages that previously were not actively used on the Internet (say, FORTRAN or C++). Prior to ASP, the limiting factor for the content of a Web site was the ability of the end user's system to compile the code for that Web site—the HTML code as well as any scripting code used to enhance the site. Active Server Components are not compiled on the end user's system, but on the server itself.

This implies that the Web developers no longer have to worry about the end user's ability to compile the code of a

## XML—the Components

Three components of XML deserve special mention: Document Type Definition (DTD), Extensible Style Language (XSL) and Extensible Link Language (XLL).

Though not strictly a mandatory component of XML, DTDs specify the elements of a page and its logical structure. This enables the validation of the tags used by a page. The definitions should be oriented towards describing the data structure rather than how the data will be displayed; the latter should ideally be handled by scripts and style sheets.

XSL is the language used to specify the style sheets for XML documents. It enables browsers to change the presentation of a document (the order in which data is presented, for instance) without further interaction with the server. It can handle an unlimited number of tags in an unlimited number of ways, precisely because of its extensibility. It also enhances the layout of a Web page by enabling such features as rotated text, multiple columns, and support for international alphabets.

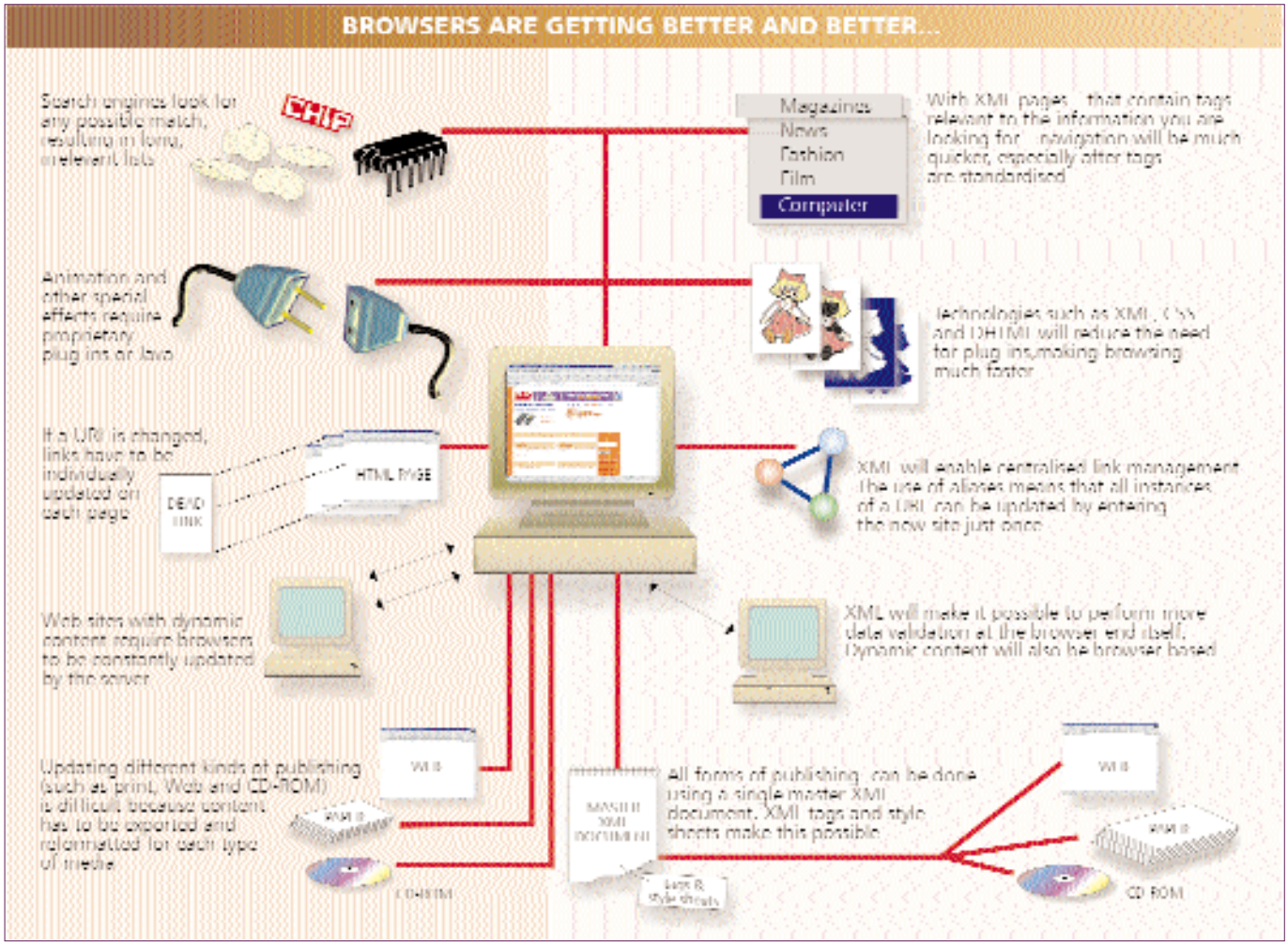
XLL is XML's eXtensible Link Language, and is currently under development. It will implement extended links, including indirect links. Of course, it will continue to support simple links as they exist today. Hopefully, dead links will soon be a thing of the past

## In the beginning...

there was SGML (Standard Generalised Markup Language), a meta-language that actually predates the Web. SGML describes how marked-up electronic text should be formatted it is a universal code for generating text, a guideline for how other languages should be developed. The foundation for the Web sites we see today comes from SGML. Since its inception, several sub-languages evolved out of SGML, the most notable of which is Hypertext Markup Language (HTML).

As the Internet began to take shape in the early nineties, it was essential that there be a common format for all Web sites to make sure that any operating system could read any site. Using SGML as a template, HTML was developed, with the first specifications being released in 1991.

Unlike SGML, HTML is a fixed programming language, which means that it has a clearly defined instruction set in the form of unique Tags that all HTML documents must follow. Subsequently, there have been several revisions to this language (the most current being HTML 4.0). XML can be considered to be an updated, abridged version of SGML; midway between SGML and HTML that tries to combine the best of both worlds.




certain language. They have to be only concerned about getting the proper compiler on their server in order to compile the code for the end user. Whenever a user accesses a site on this particular server, the server generates the code for the site, converts the code to HTML format and sends it to the user's system where the user's browser compiles the HTML and displays the Web site. Unfortunately, many sites still use standard HTML (and will probably continue to use it for a long time). As there are limits as to what can be accomplished using standard HTML code, client side scripts, Java classes or ActiveX controls may still be required to obtain the desired effects.

The server can be programmed to detect if a particular user has been to a site before, and therefore compile the site in a different way so that users who access a site for the first time will see a different output than those who have visited the site before. Another feature of ASP is its integration with the Active Database Object, a programming interface that allows ASP-enabled servers to access most major database architectures. ASP-enabled servers also have the ability to check whether the user's browser is capable of viewing any client-side active content such as ActiveX controls or Java classes. If the browser is incapable of compiling the active content, if its ability to compile them has been disabled, or if it

## The World Wide Web Consortium (W3C)

Founded in 1994 and hosted by the Massachusetts Institute of Technology s Laboratory for Computer Science (MITLCS) in the United States, the Institut National de Recherche en Informatique et en Automatique (INRIA) in Europe and the Keio University in Asia, the W3C is concerned with developing and publishing standards for the World W ide Web. These standards encompass everything from programming languages used on the Internet to connection protocols like TCP/IP or modem command strings.

In other words, if a technology is associated in any way with the Internet, the W3C is likely to already have published a standard for it, or is probably working on one. These standards are not meant to be enforced on every company developing technology for the Internet. Instead, they are guidelines intended to make it easier for a developer to ensure cross-compatibility. In other words, a W3C ratification is a good guarantee that the standard or technology will meet with global acceptance.

has been limited to static content, the server can generate an HTML document to inform the user that it may not be possible to view the special content of the site. It can also inform the user how to enable the special content of the site.

### HTML on Steroids

Dynamic HTML (DHTML), Microsoft's latest advance in HTML, is intended to make the language more functional by adding abilities normally found in application programming languages like C++ or Java. Now a site developer can code an HTML document that uses animated text, can be set for timed events, scrolling text as well as the ability to access databases—all without the need for additional components like ActiveX Controls or Java class files.

What enables DHTML to access databases is a new technology known as Data Binding. Data Binding allows the developer to link the data gathered in forms within an HTML document with either a text file or a database. This is further enhanced by another new feature of DHTML known as the Document Object Model (DOM). DOM is a platform- and language-neutral interface that allows programs and scripts such as Java and ActiveX to dynamically access and update the content, structure and style of documents. The document can be further processed and the results of that processing can be incorporated into the presented page. DOM does not place programs and scripts outside the normal HTML tags and treat them as separate entities. Rather, it enables HTML to directly access scripting languages and other technologies by the use of Object tags.

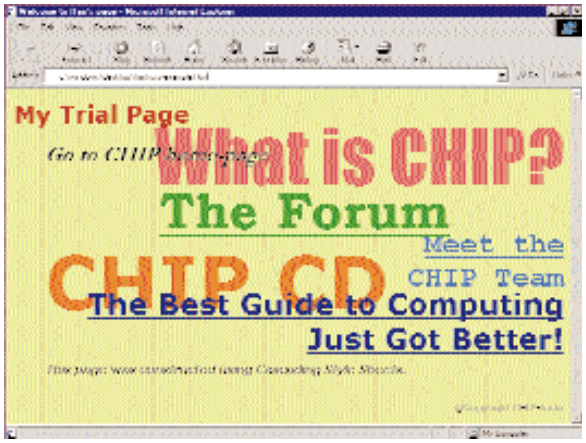
The latest versions of Internet Explorer and Netscape Communicator support DHTML to varying degrees. Even if you use an older version, there is little reason to worry. The technology for DHTML has been developed so that if a page with DHTML tags is accessed by a browser that does not support DHTML, the DHTML tags will 'degrade gracefully'. Unsupported browsers will still be able to view the page without the active DHTML content. Thus, although DHTML is still considered a proprietary technology, it virtually eliminates the concern for support for browsers that cannot read its special tags.

### A Matter of Style

Cascading Style Sheets (CSS) is not a variant of HTML but a new technology integrated into HTML. HTML versions 1.0 through 3.2 had major limitations when it came to the formatting of text. If a developer wanted to position a chunk of text or an image on a specific spot on a Web page, he or she would have to use tables or tab tags—a rather tedious process.

CSS was developed specifically to add to the flexibility of HTML by providing about 70 new attributes that can be assigned to objects within an HTML document. These additional attributes include more colour possibilities, spacing, positioning and borders. Objects can even be positioned on top of each other; a feature commonly referred to as Layering.

There have been several published versions of a standard for the instruction set for HTML. Both Netscape and Microsoft have followed the main points of these standards, but both have added support for their own features in their browsers. These features range from the tag for adding an image to an HTML document to support for other technologies like ActiveX controls or Javascript. One hitch is that Netscape and Microsoft have their own definitions for Cascading Style Sheets and both support different aspects of HTML 4.0. Recent versions of Netscape Communicator and Microsoft Internet Explorer support CSS far more comprehensively, though Netscape lags behind Microsoft.




CSS enables you to play with fonts you can select typestyles and colours of your choice, and text can even be superimposed

If a user encounters a Web site that uses HTML tags that are not supported by his or her browser, there will not be any error message. Instead, the Web site will lack a component. Moreover, if the developer has considered this situation and made allowances for it, the user will see the page in a format tailored to the capabilities of the browser.

Microsoft's Internet Explorer 4.0 supports CSS-1, the first revision of Cascading Style Sheets. IE 4.0 also supports some attributes of CSS-2, which has recently been ratified by the W3C. Unique to CSS-2 is its aural style sheets—this will help the sight-impaired to browse effectively.

The rapid development of Internet-related technology in recent months, together with the realisation of the need for platform-independent standards, will ensure that things can only get better for Internet users in the near future.

HARIKRISHNAN MENON

## Call of the Wired

What does the future hold for Web surfers? Online addicts who suffer from withdrawal symptoms when they are away from their computers, cricket buffs who absolutely need to know the latest scores, moviegoers who want to order tickets to the latest Hollywood flick all could soon meet their needs through voice-enabled applications that connect them to the Internet via telephones.

As a decisive step towards achieving this, a Voice eXtensible Markup Language (VXML) forum was recently formed by companies such as AT&T, Lucent, and Motorola. These companies had been working separately on different versions of VXML, but the forum expects to post a standard on its Web site, and a proposed specification will be submitted to the W3C shortly.

Besides fulfilling consumer demands, the VXML standard will lead to business applications such as call centres, banking transactions and electronic commerce. Other applications will let Web users check e-mail from telephones, access weather reports, stock quotes and other data found online. Eventually, people will be able to write VXML applications to fit their needs.