The Web

Table of contents

1 INTRODUCTION

The **World Wide Web** ("**WWW**" or simply the "**Web**") is a *global information space* which people can read and write via computers connected to the Internet. The term is often mistakenly used as a synonym for the Internet itself, but the Web is actually a service that operates over the Internet, just like e-mail.

The World Wide Web is the combination of four basic ideas:

- *hypertext*, that is the ability, in a computer environment, to move from one part of a document to another or from one document to another through internal connections among these documents (called "*hyperlinks*");
- *resource identifiers*, that is the ability, on a computer network, to locate a particular resource (computer, document or other resource) on the network through a unique identifier;
- the *client-server* model of computing, in which client software or a client computer makes requests of server software or a server computer that provides the client with resources or services, such as data or files; and
- *markup language*, in which characters or codes embedded in text indicate to a computer how to print or display the text, e.g. as in italics or bold type or font.

On the World Wide Web, a client program called a *web browser* retrieves information resources, such as *web pages* and other computer files, from web servers using their network addresses and displays them, typically on a computer monitor, using a markup language that determines the details of the display. One can then follow hyperlinks in each page to other resources on the World Wide Web of information whose location is provided by these hyperlinks. It is also possible, for example by filling in and submitting web forms, to send information back to the server to interact with it. The act of following hyperlinks is often called "*browsing*" or "*surfing*" the Web. Web pages are often arranged in collections of related material called "*websites*."

The phrase "surfing the Internet" was first popularised in print by Jean Armour Polly, a librarian, in an article called Surfing the INTERNET, published in the Wilson Library Bulletin in June, 1992. Although Polly may have developed the phrase independently, slightly earlier uses of similar terms have been found on the Usenet from 1991 and 1992, and some recollections claim it was also used verbally in the hacker community for a couple years before that. Polly is famous as "NetMom" in the history of the Internet.

Although the English word worldwide is normally written as one word (without a space or hyphen), the proper name World Wide Web and abbreviation WWW are now well-established even in formal English.

At its core, the Web is made up of three *standards*:

- the *Uniform Resource Identifier (URI)*, which is a universal system for referencing resources on the Web, such as Web pages;
- the *HyperText Transfer Protocol (HTTP)*, which specifies how the browser and server communicate with each other; and
- the *HyperText Markup Language (HTML)*, used to define the structure and content of

hypertext documents.

Berners-Lee now heads the World Wide Web Consortium (W3C), which develops and maintains these and other standards that enable computers on the Web to effectively store and communicate different forms of information.

2 How the Web works

When a viewer wants to access a web page or other "resource" on the World Wide Web, he or she normally begins either by typing the *URL* of the page into his or her web browser, or by following a hypertext link to that page or resource. The first step, behind the scenes, is for the server-name part of the URL to be resolved into an IP address by the global, distributed Internet database known as the Domain name system or DNS.

The next step is for an HTTP request to be sent to the web server working at that IP address for the page required. In the case of a typical web page, the HTML text, graphics and any other files that form a part of the page will be requested and returned to the client (the web browser) in quick succession.

The web browser's job is then to render the page as described by the HTML, CSS and other files received, incorporating the images, links and other resources as necessary. This produces the on-screen 'page' that the viewer sees.

Most web pages will themselves contain hyperlinks to other relevant and informative pages and perhaps to downloads, source documents, definitions and other web resources.

Such a collection of useful, related resources, interconnected via hypertext links, is what has been dubbed a 'web' of information. Making it available on the Internet produced what Tim Berners-Lee first called the World Wide Web in the early 1990s.

2.1 Publishing web pages

The Web is *available* to individuals outside mass media. In order to "publish" a web page, one does not have to go through a publisher or other media institution, and potential readers could be found in all corners of the globe.

Unlike books and documents, hypertext does not have a linear order from beginning to end. It is not broken down into the hierarchy of chapters, sections, subsections, etc.

Many different kinds of information are now available on the Web, and for those who wish to know other societies, their cultures and peoples, it has become easier. When travelling in a foreign country or a remote town, one might be able to find some information about the place on the Web, especially if the place is in one of the developed countries. Local newspapers, government publications, and other materials are easier to access, and therefore the variety of information obtainable with the same effort may be said to have increased, for the users of the Internet.

Although some websites are available in multiple languages, many are in the local language only. Also, not all software supports all special characters. These factors would challenge the notion that the World Wide Web will bring a unity to the world.

The increased opportunity to publish materials is certainly observable in the countless personal pages, as well as pages by families, small shops, etc., facilitated by the emergence of

free *web hosting* services, which provide individuals, organizations and users with online systems for storing information, images, video, or any content accessible via the Web. Web hosts are companies that provide space on a server they own for use by their clients as well as providing Internet connectivity, typically in a data center. Web hosts can also provide data center space and connectivity to the Internet for servers they do not own to be located in their data center.

2.2 Sociological implications

The Web, as it stands today, has allowed *global interpersonal exchange* on a scale unprecedented in human history. People separated by vast distances, or even large amounts of time, can use the Web to exchange — or even mutually develop — their most intimate and extensive thoughts, or alternately their most casual attitudes and spirits. Emotional experiences, political ideas, cultural customs, musical idioms, business advice, artwork, photographs, literature, can all be shared and disseminated digitally with less individual investment than ever before in human history.

Although the existence and use of the Web relies upon material technology, which comes with its own disadvantages, its information does not use physical resources in the way that libraries or the printing press have. Therefore, propagation of information via the Web (via the Internet, in turn) is not constrained by movement of physical volumes, or by manual or material copying of information. And by virtue of being digital, the information of the Web can be searched more easily and efficiently than any library or physical volume, and vastly more quickly than a person could retrieve information about the world by way of physical travel or by way of mail, telephone, telegraph, or any other communicative medium.

The Web is the most far-reaching and extensive medium of personal exchange to appear on Earth. It has probably allowed many of its users to interact with many more groups of people, dispersed around the planet in time and space, than is possible when limited by physical contact or even when limited by every other existing medium of communication combined.

3 Uniform Resource Identifier (URI)

A **Uniform Resource Identifier (URI)**, is an Internet protocol element consisting of a short string of characters that conform to a certain syntax. The string comprises a name or address that can be used to refer to a resource. It is a fundamental component of the World Wide Web.

A URI can be classified as a locator or a name or both. A **Uniform Resource Locator (URL)** is a URI that, in addition to identifying a resource, provides means of acting upon or obtaining a representation of the resource by describing its primary access mechanism or network "location". For example, the URL http://www.wikipedia.org/ is a URI that identifies a resource (Wikipedia's home page) and implies that a representation of that resource (such as the home page's current HTML code, as encoded characters) is obtainable via HTTP from a network host named www.wikipedia.org.

A **Uniform Resource Name (URN)** is a URI that identifies a resource by name in a particular namespace. A URN can be used to talk about a resource without implying its location or how to dereference it. For example, the URN urn:ISBN:0-395-36341-1 is a URI that, like an International Standard Book Number (ISBN), allows one to talk about a book, but doesn't suggest where and how to obtain an actual copy of it.

The URI *syntax* is essentially a **URI scheme** name like "*http*", "*ftp*", "*mailto*", "*urn*", etc., followed by a colon character, and then a **scheme-specific part**. The syntax and semantics of the scheme-specific part are determined by the specifications that govern the schemes, although the URI syntax does force all schemes to adhere to a certain general syntax that, among other things, reserves certain characters for special purposes, without always saying what those purposes are.

The URI syntax also enforces restrictions on the scheme-specific part, in order to, for example, provide for a degree of consistency when the part has a hierarchical structure. Percent-encoding is an often misunderstood aspect of URI syntax.

A **URI reference** is another type of string that represents a URI, and, in turn, the resource identified by that URI. The distinction between a URI and a URI reference is not often maintained in informal usage, but protocol documents should not allow for ambiguity.

A URI reference may take the form of a full URI, or just the scheme-specific portion of one, or even some trailing component thereof —even the empty string. An optional fragment identifier, preceded by "#", may be present at the end of a URI reference. The part of the reference before the "#" indirectly identifies a resource, and the fragment identifier identifies some portion of that resource.

In web document markup languages, URI references are frequently used in places where there is a need to point to other resources, such as external documents or specific portions of the same logical document.

Examples of absolute URIs

- http://somehost/absolute/URI/with/absolute/path/to/resource.txt
- ftp://somehost/resource.txt
- urn:issn:1535-3613

Examples of URI references

- http://example/resource.txt#frag01
- http://somehost/absolute/URI/with/absolute/path/to/resource.txt
- /relative/URI/with/absolute/path/to/resource.txt
- relative/path/to/resource.txt
- ../../../resource.txt
- resource.txt
- /resource.txt#frag01
- #frag01

4 HyperText Transfer Protocol (HTTP)

HyperText Transfer Protocol (HTTP) is the method used to transfer or convey information on the World Wide Web. The original purpose was to provide a way to publish and receive HTML pages.

HTTP is a *request/response protocol* between clients and servers. The originating client, such

as a web browser, spider, or other end-user tool, is referred to as the user agent. The destination server, which stores or creates resources such as HTML files and images, is called the origin server. In between the user agent and origin server may be several intermediaries, such as proxies, gateways, and tunnels.

A HTTP client initiates a request by establishing a Transmission Control Protocol (TCP) connection to a particular port on a remote host (port 80 by default). A HTTP server listening on that port waits for the client to send a Request Message. Upon receiving the request, the server sends back a status line, such as "HTTP/1.1 200 OK", and a message of its own, the body of which is perhaps the requested file, an error message, or some other information.

5 HyperText Markup Language (HTML)

The **HyperText Markup Language (HTML)** is a markup language designed for the creation of web pages with hypertext and other information to be displayed in a web browser. HTML is used to structure information — denoting certain text as headings, paragraphs, lists and so on — and can be used to describe, to some degree, the appearance and semantics of a document.

Early versions of HTML were defined with looser syntactic rules which helped its adoption by those unfamiliar with web publishing. Web browsers commonly made assumptions about intent and proceeded with rendering of the page. Over time, the trend in the official standards has been to create an increasingly strict language syntax; however, browsers still continue to render pages that are far from valid HTML.

6 Web Server

The term Web server can mean one of two things:

1. A computer that is responsible for accepting HTTP requests from clients, which are known as Web browsers, and serving them Web pages, which are usually HTML documents.
2. A computer program that provides the functionality described in the first sense of the term.

Although Web server programs differ in detail, they all share some basic common features. Every Web server program operates by accepting HTTP requests from the network, and providing an HTTP response to the requester. The HTTP response typically consists of an HTML document, but can also be a raw text file, an image, or some other type of document.

Usually Web servers have also the capability of logging some detailed information, about client requests and server responses, to log files; this allows the Webmaster to collect statistics by running log analyzers on log files.

The origin of the content sent by server is called static if it comes from an existing file or dynamic if it is dynamically generated by some other program or script called by Web server. Serving static content is usually much faster than serving dynamic content.

Web servers usually translate the path component of a Uniform Resource Locator (URL) into a local file system resource. The URL path specified by the client is relative to the Web server's root directory. Consider the following URL as it would be requested by a client:

http://www.example.com/path/file.html

The client's Web browser will translate it into a connection to www.example.com with the following HTTP 1.1 request:

GET /path/file.html HTTP/1.1

Host: www.example.com

The Web server on www.example.com will append the given path to the path of its root directory. On Unix machines, this is commonly /var/www/htdocs. The result is the local file system resource:

/var/www/htdocs/path/file.html

The Web server will then read the file, if it exists, and send a response to the client's Web browser. The response will describe the content of the file and contain the file itself.

The four top most common Web or HTTP server programs are:

- Apache HTTP Server from the Apache Software Foundation.
- Internet Information Services (IIS) from Microsoft.
- Sun Java System Web Server from Sun Microsystems, formerly Sun ONE Web Server, iPlanet Web Server, and Netscape Enterprise Server.
- Zeus Web Server from Zeus Technology.

There are thousands of different Web server programs available, many of them are specialized for some uses and can be tailored to satisfy specific needs.

7 WEB SITES

A website (or web site) is a collection of Web pages, typically common to a particular domain name or sub-domain on the World Wide Web on the Internet. To date, there are nearly 80 million websites in the world with registered domains.

A web page is an HTML/XHTML document accessible generally via HTTP. All publicly accessible websites are seen as constituting a "World Wide Web" of information.

The pages of a website will be accessed from a common root URL called the homepage, and usually reside on the same physical server. The URLs of the pages organise them into a hierarchy, although the hyperlinks between them control how the reader perceives the overall structure and how the traffic flows between the different parts of the sites.

A website may be the work of an individual, a business or other organization and is typically dedicated to some particular topic or purpose. Any website can contain a hyperlink to any other website, so the distinction between individual sites, as perceived by the user, may sometimes be blurred.

Websites are written in, or dynamically converted to, *HTML* (Hyper Text Markup Language) and are accessed using a software program called a *web browser*, also known as a *HTTP client*. Web pages can be viewed or otherwise accessed from a range of computer based and Internet enabled devices of various sizes, examples of which include desktop computers, laptop computers, PDAs and cell phones.

A website is *hosted* on a computer system known as a *web server*, also called an *HTTP*

server, and these terms can also refer to the software that runs on these system and that retrieves and delivers the web pages in response to requests from the website users.

A *static website* is one that has content that is not expected to change frequently and is manually maintained by some person or persons using some type of editor software. There are two broad categories of editor software used for this purpose which are

- *Text editors* such as Notepad, where the HTML is manipulated directly within the editor program
- *WYSIWYG editors* such as Microsoft FrontPage and Macromedia Dreamweaver, where the site is edited using a GUI interface and the underlying HTML is generated automatically by the editor software.

A *dynamic website* is one that may have frequently changing information. When the web server receives a request for a given page, the page is automatically generated by the software in direct response to the page request; thus opening up many possibilities including for example: a site can display the current state of a dialogue between users, monitor a changing situation, or provide information in some way personalised to the requirements of the individual user.

There are a large range of software systems, such as Active Server Pages (ASP), Java Server Pages (JSP) and the PHP programming language that are available to generate dynamic web systems and dynamic sites also often include content that is retrieved from one or more databases or by using XML-based technologies such as RSS.

Static content may also be dynamically generated periodically or if certain conditions for regeneration occur (cached) to avoid the performance loss of initiating the dynamic engine on a per-user or per-connection basis.

Plugins are available for browsers, which use them to show active content, such as Flash, Shockwave or applets written in Java. Dynamic HTML also provides for user interactivity and realtime element updating within Web pages (i.e., pages don't have to be loaded or reloaded to effect any changes), mainly using the DOM and JavaScript, support for which is built-in to most modern browsers.

7.1 Types of websites

There are many varieties of websites, each specialising in a particular type of content or use, and they may be arbitrarily classified in any number of ways. A few such classifications might include:

- *Blog* (or weblog) site: site used to log online readings or to post online diaries; may include discussion forums. Examples: blogger, Xanga.
- *Business site*: used for promoting a business or service.
- *Commerce site or eCommerce site*: for purchasing goods, such as Amazon.com.
- *Community site*: a site where persons with similar interests communicate with each other, usually by chat or message boards.
- *Database site*: a site whose main use is the search and display of a specific database's content such as the Internet Movie Database or the Political graveyard.
- *Development site*: a site whose purpose is to provide information and resources related

to software development, Web design and the like.

- *Directory site*: a site that contains varied contents which are divided into categories and subcategories, such as Yahoo! directory, Google directory and Open Directory Project.
- *Download site*: strictly used for downloading electronic content, such as software, game demos or computer wallpaper.
- *Game site*: a site that is itself a game or "playground" where many people come to play, such as MSN Games.
- *Information site*: contains content that is intended merely to inform visitors, but not necessarily for commercial purposes; such as: RateMyProfessors.com, Free Internet Lexicon and Encyclopedia.
- *News site*: similar to an information site, but dedicated to dispensing news and commentary.
- *Search engine site*: a site that provides general information and is intended as a gateway or lookup for other sites
- *Vanity site (or "personal site")*: run by an individual or a small group (such as a family) that contains information or any content that the individual wishes to include.
- *Web portal site*: a website that provides a starting point, a gateway, or portal, to other resources on the Internet or an intranet.
- *Wiki site*: a site which users collaboratively edit (such as Wikipedia).
- *Political site*: A website on which people may voice political views. Example: New Confederacy.

Some sites may be included in one or more of these categories. For example, a business website may promote the business's products, but may also host informative documents, such as white papers. There are also numerous sub-categories to the ones listed above

Many business Websites have the appearance of brochures—that is, an advertisement that can be strolled around. Some websites act as vehicles for users to communicate with other people via webchat.

7.2 Web Portal

Web portals are sites on the World Wide Web that typically provide personalized capabilities to their visitors. They are designed to use distributed applications, different numbers and types of middleware and hardware to provide services from a number of different sources.

In addition, business portals are designed to share collaboration in workplaces. A further business-driven requirement of portals is that the content be able to work on multiple platforms such as personal computers, personal digital assistants (PDAs), and cell phones.

In the late 1990s, the Web portal was a hot commodity. After the proliferation of Web browsers in the mid-1990s, many companies tried to build or acquire a portal, to have a piece of the Internet market. The Web portal gained special attention because it was, for many users, the *starting point* of their Web browser.

Many of the portals started initially as either Internet directories (notably Yahoo!) and/or search engines (Excite, Lycos, AltaVista, infoseek, and Hotbot among the old ones). The

expansion of service provision occurred as a strategy to secure the user-base and lengthen the time a user stays on the portal. Services which require user registration such as free email, customization features, and chatrooms were considered to enhance repeat use of the portal. Game, chat, email, news, and other services also tend to make users stay longer, thereby increasing the advertisement revenue.

7.2.1 Regional Web portals

Along with the development and success of international Web portals such as Yahoo!, regional variants have also sprung up. Some regional portals contain local information such as weather forecasts, street maps and local business information. Another notable expansion over the past couple of years is the move into formerly unthinkable markets such as the People's Republic of China.

"Local content - global reach" Web portals have emerged from countries like India (Rediff) and China (Sina). Such portals reach out to the widespread diaspora spread across the world.

7.2.2 Government Web portals

At the end of the dot-com boom in the 1990s, many governments had already committed to creating portal sites for their citizens. In the United States the main portal is FirstGov.gov; in the United Kingdom the main portals are directgov (for citizens) and businesslink.gov.uk (for businesses).

Many U.S. states have their own portals which provide direct access to eCommerce applications (i.e Hawaii Business Express and myIndianaLicense), agency and department web sites, and more specific information about living in, doing business in and getting around the state.

Many states have chosen to out-source the operation of their portals to third-party vendors. The most successful company to date for this is NICUSA which runs 18 state portals. NICUSA focuses on the self-funded model, and does not charge the state for work. Instead it is supported by transaction fees for its applications.

7.2.3 Enterprise Web portals

In the early 2000s, a major industry shift in Web portal focus has been the corporate intranet portal, or "enterprise Web". Where expecting millions of unaffiliated users to return to a public Web portal has been something of a mediocre financial success, using a private Web portal to unite the Web communications and thinking inside a large corporation has begun to be seen by many as both a labor-saving and a money-saving technology. Some corporate analysts have predicted that corporate intranet Web portal spending will be one of the top five areas for growth in the Internet technologies sector during the first decade of the 21st century.

Some features of enterprise portals are:

- Single touch point — the portal becomes the delivery mechanism for all business information services.
- Collaboration — portal members can communicate synchronously (through chat, or messaging) or asynchronously through threaded discussion and email digests (forums) and blogs.

- Content and document management — services that support the full life cycle of document creation and provide mechanisms for authoring, approval, version control, scheduled publishing, indexing and searching.
- Personalization — the ability for portal members to subscribe to specific types of content and services. Users can customize the look and feel of their environment.
- Integration — the connection of functions and data from multiple systems into new components/portlets.

Most enterprise portals provide single sign-on capabilities to their users. This requires a user to authenticate only once. Access control lists manage the mapping between portal content and services over the portal user base.

Enterprise portals may be referred to by the community they serve. For instance, an employee-facing portal may be described as a "Business-to-employee" portal, or B2E portal. Other enterprise portal classifications are "B2C" (business-to-customer/consumer), "B2D" (business-to-dealer/distributor), "B2B" (business-to-business/supplier), and "B2G" (business-to-government). Enterprises may develop multiple "B2x" portals based on business structure and strategic focus, but leverage a common architectural framework, reusable component libraries, and standardized project methodologies.

7.3 Wiki

A **wiki** is a type of website that allows users to add and edit content easily and is especially suited for collaborative writing.

The name is based on the *Hawaiian term* wiki, meaning "quick", "fast", or "to hasten" (Hawaiian dictionary). Sometimes the reduplication wikiwiki (or Wikiwiki) is used instead of wiki (Hawaiian dictionary). The term Wiki also sometimes refers to the collaborative software itself (wiki engine) that facilitates the operation of such a website.

In essence, wiki is a simplification of the process of creating HTML web pages combined with a system that records each individual change that occurs over time, so that at any time, a page can be reverted to any of its previous states. A wiki system may also provide various tools that easily allow the user community to monitor the constantly changing state of the wiki and discuss the issues that emerge in trying to achieve a general consensus about wiki content. Wiki content can also be misleading as users may add incorrect information to the Wiki page.

Some wikis will allow completely unrestricted access so that people are able to contribute to the site without necessarily having to undergo a process of 'registration', as had usually been required by various other types of interactive websites such as Internet forums or chat sites.

Wiki is sometimes interpreted as the backronym for "*What I know is*", which describes the knowledge contribution, storage and exchange function.

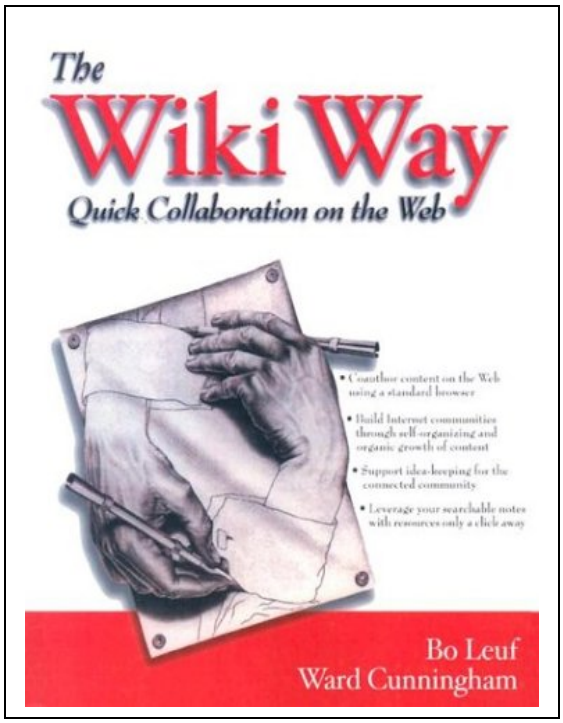

Fig. 1 - A book on Wiki: *The Wiki Way. Quick collaboration on the Web, Addison-Wesley (April 2001).*

A wiki enables documents to be written collectively in a very simple markup language using a web browser. A single page in a wiki is referred to as a "wiki page", whilst the entire body of pages, which are usually highly interconnected via hyperlinks, is "the wiki"; in effect, a very simple, easy to use database.

A defining characteristic of wiki technology is the ease with which pages can be created and updated. Generally, there is no review before modifications are accepted. Most wikis are open to the general public without the need to register any user account. Sometimes session log-in is requested to acquire a "wiki-signature" cookie for autosigning edits. More private wiki servers require user authentication. However, many edits can be made in real-time, and appear almost instantaneously online. This can often lead to abuse of the system.

7.3.1 Wiki pages

In a traditional wiki, there are three representations for each page:

- The *user-editable "source code"*, which is also the format stored locally on the server. It is usually plain text, made visible to the user only when the edit operation shows it in a browser form.
- A *template* (possibly internally generated) that defines layout and elements common to all pages.
- The *rendered HTML code* produced by the server on the fly from the source text when a particular page is requested.

The source format, sometimes known as "*wikitext*", is augmented with a simplified markup language to indicate various structural and visual conventions. An often used example of one such convention is to start a line of text with an asterisk ("*") in order to mark it as an item in a bulleted list. Style and syntax can vary a great deal among implementations, some of which also allow HTML tags.

The reasoning behind this design is that HTML, with its many cryptic tags, is not especially human-readable. Making typical HTML source visible makes the actual text content very hard to read and edit for most users. It is therefore better to promote plain-text editing with a few simple conventions for structure and style.

Wiki syntax (MediaWiki)	HTML	Rendered output
"'''Doctor''? No other title? A ''scholar''? And he rates above the civil authority?" "Why, certainly," replied Hardin, amiably. "We're all scholars more or less. After all, we're not so much a world as a scientific foundation—under the direct control of the Emperor."	 "Doctor? No other title? A scholar? And he rates above the civil authority?" "Why, certainly," replied Hardin, amiably. "We're all scholars more or less. After all, we're not so much a world as a scientific foundation—under the direct control of the Emperor." 	"*Doctor*? No other title? A *scholar*? And he rates above the civil authority?" "Why, certainly," replied Hardin, amiably. "We're all scholars more or less. After all, we're not so much a world as a scientific foundation—under the direct control of the Emperor."

Fig. 2 - The three representations of each Wiki page.

Some recent wiki engines use a different method: they allow *"WYSIWYG" editing*, usually by means of JavaScript or an ActiveX control that translates graphically entered formatting instructions, such as "bold" and "italics", into the corresponding HTML tags. In those implementations, the markup of a newly-edited HTML version of the page is generated and submitted to the server transparently, and the user is shielded from this technical detail. Users who do not have the necessary plugin can generally edit the page, usually by directly editing the raw HTML code. More recently, wiki engines are generating wiki syntax instead of HTML. This way, users who are comfortable editing in wiki syntax can carry on.

While for years the de facto standard was the syntax of the original WikiWikiWeb, currently the formatting instructions vary considerably depending on the wiki engine. Simple wikis allow only basic text formatting, whereas more complex ones have support for tables, images, formulas, or even interactive elements such as polls and games. Many people switch between wiki engines. Because of the difficulty in using several syntaxes, many people are putting considerable effort into defining a wiki markup standard.

Wikis are a true hypertext medium, with non-linear navigational structures. Each page typically contains a large number of links to other pages. Hierarchical navigation pages often exist in larger wikis, often a consequence of the original page creation process, but they do not have to be used. Links are created using a specific syntax, the so-called "link pattern".

Most wikis offer at least a title search, and sometimes a full-text search. The scalability of the search depends on whether the wiki engine uses a database; indexed database access is necessary for high speed searches on large wikis. On Wikipedia, the so-called "Go button" allows readers to view a page that matches the entered search criteria as closely as possible. The MetaWiki search engine was created to enable searches across multiple wikis. Search is keyword-based.

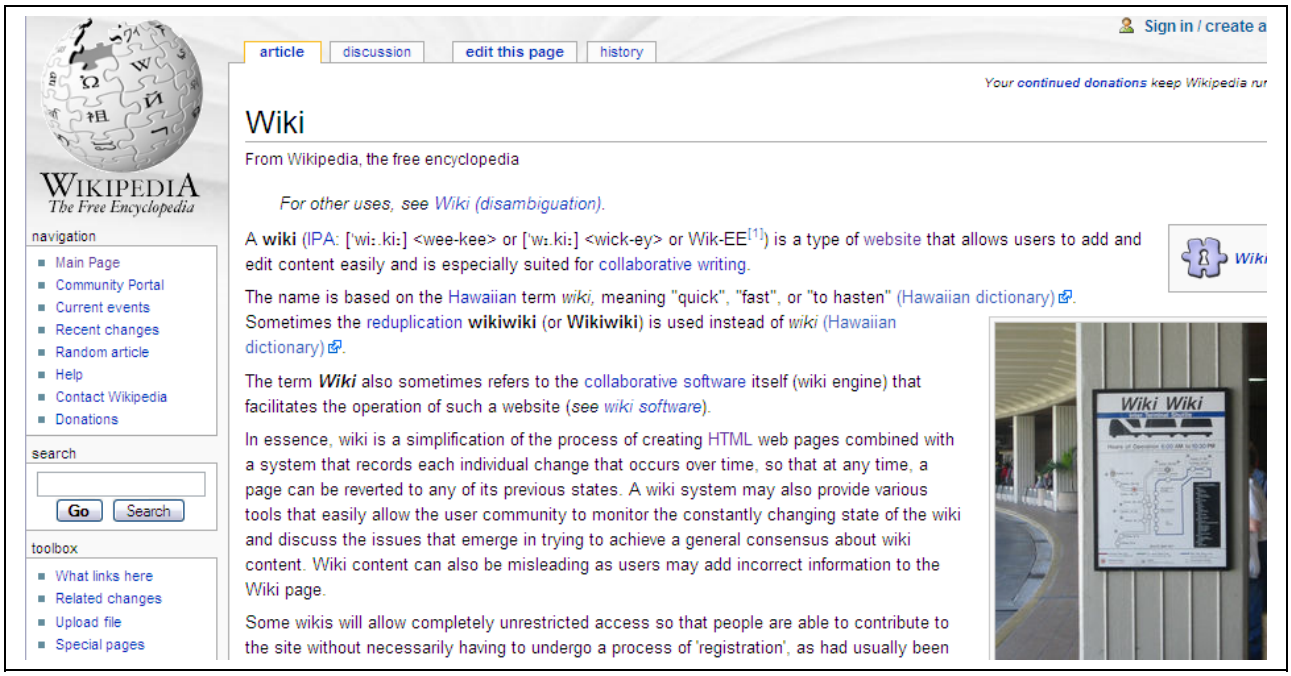

Fig. 3 - The Wiki description in the Wikipedia.

8 Java and JavaScript

Another significant advance in the technology was Sun Microsystems' Java programming language. It initially enabled Web servers to embed small programs (called applets) directly into the information being served, and these applets would run on the end-user's computer, allowing faster and richer user interaction. Eventually, it came to be more widely used as a tool for generating complex server-side content as it is requested. Java never gained as much acceptance as Sun had hoped as a platform for client-side applets for a variety of reasons, including lack of integration with other content (applets were confined to small boxes within the rendered page) and poor perfomance (particularly start up delays) of Java VMs on PC hardware of that time.

JavaScript, however, is a scripting language that was developed for Web pages. The standardised version is ECMAScript. While its name is similar to Java, it was developed by Netscape and not Sun Microsystems, and it has almost nothing to do with Java, with the only exception being that like Java its syntax is derived from the C programming language. Like Java, Javascript is also object oriented but like C++ and unlike Java, it allows mixed code - both object oriented as well as procedural. In conjunction with the Document Object Model, JavaScript has become a much more powerful language than its creators originally envisioned. Sometimes its usage is expressed under the term Dynamic HTML (DHTML), to emphasise a shift away from static HTML pages.

9 W3C

The World Wide Web Consortium (W3C) is an international consortium where member organizations, a full-time staff, and the public, work together to develop standards for the World Wide Web.

W3C's mission is: "*To lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web*". W3C also engages in education and outreach, develops software, and serves as an open forum for discussion about the Web.

The Consortium is headed by **Tim Berners-Lee**, the original creator of the World Wide Web and primary author of the URL (Uniform Resource Locator), HTTP (HyperText Transfer Protocol) and HTML (HyperText Markup Language) specifications, the principal technologies that form the basis of the Web.

9.1 History

In October 1994, Tim Berners-Lee, inventor of the World Wide Web, moved from the European Organization for Nuclear Research (CERN), where the Web originated, and founded the World Wide Web Consortium (W3C) at the Massachusetts Institute of Technology, Laboratory for Computer Science (MIT/LCS) with support from Defense Advanced Research Project Agency (DARPA), and the European Commission.

The consortium was created to ensure compatibility and agreement among industry members in the adoption of new standards. Prior to its creation, incompatible versions of HTML were offered by different vendors, increasing the potential for incompatibilities between web pages. The consortium was created to get all those vendors to agree on a set of core principles and components which would be supported by everyone.

It was originally intended that CERN host the European part of W3C. However, CERN wished to focus on particle physics, not information technology. In April 1995 the Institut National de Recherche en Informatique et Automatique (INRIA) became the European host of W3C, joined by Keio University, Japan in September 1996. Starting in 1997, W3C created regional Offices around the world; as of October 2005 it has fifteen World Offices covering Australia, Benelux, Finland, Germany and Austria, Greece, Hong Kong, Hungary, India, Israel, Italy, Korea, Morocco, Spain, Sweden, the United Kingdom, and Ireland.

In January 2003, the European host was transferred from INRIA, a French computer science laboratory, to ERCIM, which represents European national computer science laboratories.

9.2 Recommendations and Certifications

In accord with the W3C Process Document, a Recommendation progresses through the maturity levels of *Working Draft (WD)*, *Last Call Working Draft*, *Candidate Recommendation (CR)*, and *Proposed Recommendation (PR)*, culminating ultimately as a *W3C Recommendation (REC)*. A Recommendation may be updated by separately-published Errata until enough substantial edits accumulate, at which time a new edition of the Recommendation may be produced (e.g., XML is now in its third edition). W3C also publishes various kinds of informative Notes which are not intended to be treated as standards.

The Consortium leaves it up to manufacturers to follow the Recommendations. Many of its standards define levels of conformance, which are required for the developers to follow. Like any standards of other organizations, W3C recommendations are sometimes implemented partially. The Recommendations are under a royalty-free patent license, allowing anyone to implement them.

Unlike the ISOC and other international standards bodies, the W3C does not have a

certification program. A certification program is a process which has benefits and drawbacks. The W3C has decided for now that it is not suitable to start such a program without the risk of creating more drawbacks for the community than benefits.

9.3 Administration

The Consortium is jointly administered by the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) in the USA, the European Research Consortium for Informatics and Mathematics (ERCIM) (in Sophia Antipolis, France), and Keio University (in Japan).

W3C also has World Offices in fifteen regions around the world. The W3C Offices work with their regional Web communities to promote W3C technologies in local languages, broaden W3C's geographical base, and encourage international participation in W3C Activities.

10 Internet Hosting Services

An *Internet hosting service* is a service that runs Internet servers, allowing organizations and individuals to serve content on the Internet. There are various levels of service and various kinds of services offered.

The most common kind of hosting is *web hosting*. Most hosting services offer a combined variety of services; most web hosting services also offer *e-mail hosting service*, for example. *DNS hosting service* is usually bundled with domain name registration.

The most generic and powerful kind of Internet hosting provides a server where the client can do anything he/she wants, and an Internet connection with good upstream bandwidth. The client can run anything he wants, including web servers and other servers. **Full-featured hosting services** include:

- *Dedicated hosting service*, where the hosting service owns and manages the machine, leasing full control to the client
- *Virtual private server*, which appears as a dedicated server but only uses a single physical server
- *Colocation facilities*, which provide just the Internet connection and climate control, but let the client do his own system administration; it is the most expensive

Limited or application-specific hosting services include:

- Web hosting service
- E-mail hosting service
- DNS hosting service
- Game servers

10.1Dedicated hosting service

A **dedicated hosting service**, **dedicated server**, or **managed hosting service** is a type of Internet hosting where the client leases an entire server not shared with anyone. This is more

flexible than shared hosting, as organizations have full control over the server(s), including choice of operating system, hardware, etc.

Server administration can usually be provided by the hosting company as an add-on service. In some cases a dedicated server can offer less overhead and a larger return on investment.

Dedicated servers are most often housed in data centers, similar to colocation facilities, providing redundant power sources and HVAC systems. In contrast to colocation, a dedicated hosting service provides system administration and owns the server itself.

10.2Virtual private servers

Virtual private servers or **Virtual dedicated servers** are a form of virtualization that splits a *single physical server* into *multiple virtual servers*.

The practice of partitioning a single server so that it appeared as multiple servers has long been common practice in mainframe computers, but has seen a resurgence lately with the development of software such as VMware, User-mode Linux, FreeVPS, Virtuozzo and Xen.

10.3Colocation centre

A **colocation centre** ("colo") or **carrier hotel** is a type of data centre where multiple telecommunications network or service providers, such as telcos or internet service providers, site their connections to one another's networks (points of presence).

Increasingly organizations are recognizing the benefits of colocating their mission-critical equipment within a data centre. Colocation is becoming popular because of the time and cost savings a company can realize as result of using shared data centre infrastructure. With IT and communications facilities in safe, secure hands, telecommunications, internet, ASP and content providers, as well as enterprises, enjoy less latency and the freedom to focus on their core business.

Additionally, customers reduce their traffic back-haul costs and free up their internal networks for other uses. Moreover, by outsourcing network traffic to a colocation service provider with greater bandwidth capacity, web site access speeds should improve considerably.

Major types of colocation customers are:

- *Web commerce companies*, who use the facilities for a safe environment and cost-effective, redundant connections to the Internet
- *Major enterprises*, who use the facility for disaster avoidance
- *Telecommunication companies*, who use the facilities to interexchange traffic with other telecommunications companies and access to potential clients

10.4Web Hosting

A **web hosting** service is a type of Internet hosting service that provides individuals, organizations and users with online systems for storing information, images, video, or any content accessible via the Web. Web hosts are companies that provide space on a server they own for use by their clients as well as providing Internet connectivity, typically in a data

center. Web hosts can also provide data center space and connectivity to the Internet for servers they do not own to be located in their data center.

10.4.1Free web hosting

A **free web hosting** service is a web hosting service that is free, usually advertisement-supported and of limited functionality.

Free web hosts will either provide a *subdomain* (**yourname**.hostservicename.com) or a *directory* (www. hostservicename.com/~**yourname**). In contrast, paid web hosts will usually provide a Second-level domain along with the hosting (www.**yourname**.com). Some free hosts do allow use of separately-purchased domains. Rarely, a free host may also operate as a domain name registrar, but their registry services are usually more expensive than others.

A few free web hosts have a basic package for free, and enhanced packages (with more features) for a cost. This allows users to try the service for an initial trial (see how it performs compared to other hosts), and then upgrade when (and if) needed.

10.4.2Shared web hosting

A **shared web hosting** service or virtual hosting service is a form of web hosting service where more than one web site is hosted *on the same server*. It is generally the most economical option for hosting as many people share the overall cost of server maintenance.

The hosting service must include system administration since it is shared by many users; this is a benefit for users that do not want to deal with it, but a hindrance to power users that want more control.

In shared hosting, the provider is generally responsible for management of servers, installation of server software, security updates and other aspects of the service. Servers are often based on the Linux operating systems because more Control Panel products (see above) are made for Linux. However, some providers offer Microsoft Windows based solutions. The Plesk control panel, for instance, has two versions - for Linux and Windows, both with very similar interfaces and functionality, with the exception of OS-specific differences (for example, supporting ASP.NET, or Microsoft SQL Server on Windows).

Shared web hosting can also be done privately by sharing the cost of running a server in a colocation centre; this is called *cooperative hosting*.

10.4.3Clustered hosting

Clustered hosting technology is designed to eliminate the problems inherent with typical shared hosting infrastructures. This technology provides customers with a “clustered” handling of security, load balancing, and necessary website resources.

A clustered hosting platform is data-driven, which means that no human interaction is needed to provision a new account to the platform.

Clustered hosting "virtualizes" the resources beyond the limits of one physical server, and as a result, a website is not limited to one server. They share the processing power of many servers and their applications are distributed in real-time. This means that they can purchase as much computing power as they want from a virtually inexhaustible source, since even the largest customer never consumes more than a fraction of a percent of the total server pool. Customer

account changes (to add new resources or change settings) are propagated immediately to every server in the cluster. This is different from typical shared hosting architectures that usually require changes to a configuration file that becomes live after the server is rebooted during off hours, or are pushed on a cyclic basis every few hours.

10.4.4Other hosting services

Reseller hosting is a form of web hosting wherein the account owner may split up their alloted hard drive space and bandwidth and resell web hosting.

An **image hosting service** allows individuals to upload images to an Internet website. The image host will then store the image onto its server, and show the individual different types of code to allow others to view that image.

A **file hosting service** or **online file storage service** is an Internet hosting service specifically designed to host static content, typically large files that are not web pages. Typically they allow web and FTP access. They can be optimized for serving many users (as is implied by the term "hosting") or be optimized for single-user storage (as is implied by the term "storage"). A related service is remote backup.

A **remote backup service** or **online backup service** is a service that provides users with an online system for backing up and storing computer files. Remote Backup Service Providers are companies that the software and server space for storing files. Remote backup is similar to file hosting service; files are usually guaranteed not to be lost for a high premium. The client runs a scheduled software program (typically once a day). This program compresses, encrypts, and transfers to the remote backup service provider's servers.

A **game server** is a server used by game clients. Any video game played over the internet generally requires a connection to a game server.

A **DNS hosting service** is a service that runs Domain Name System servers. Most, but not all, domain name registrars include DNS hosting service with registration. Free DNS hosting services also exist. Almost all DNS hosting services are "shared"; except for the most popular Internet sites, there is no need to dedicate a server to hosting DNS for a single website. Many third-party DNS hosting services provide Dynamic DNS.

An **e-mail hosting service** is an Internet hosting service that runs e-mail servers. E-mail hosting services differ from typical end-user e-mail providers such as webmail sites. They cater to large businesses, allow custom configuration, large number of accounts, use of own domain name.

11 References

[1] Wikipedia: http://en.wikipedia.org

[2] Wikipedia W3C: http://en.wikipedia.org/wiki/World_Wide_Web_Consortium

[3] W3C homepage: http://www.w3.org/

[4] About the World Wide Web Consortium: http://www.w3.org/Consortium/