

LE WORLD WIDE WEB : L'HYPERMEDIÀ SUR INTERNET

E. Houmel (*)

Abstract

The telecommunication's networks technology linked to the electronic document has changed abroad the information specialists' methods of work. The Internet network did a lot in these big changes and especially after the World Wide Web integration which is a high hypermedia distributed information System.

In Algeria lots of documentation institutions are not informed of these advanced technologies. Consequently, the goal of this paper is the vulgarisation of the World Wide Web, that is the most remarkable Internet tool that tends to be taken for the Internet itself.

L'avènement du réseau Internet (acronyme de Intercommunication Network) a fait que l'humanité presque entière est rentrée de plein pied dans l'ère de l'information. La quasi totalité des réseaux offrent actuellement un accès Internet, accentuant de ce fait son formidable essor à l'échelle planétaire et de nouveaux outils technologiques et ressources informationnelles relatifs à ce réseau viennent quotidiennement enrichir la panoplie de ceux déjà existants.

Internet, c'est le réseau de tous les réseaux. A l'origine, c'était l'idée d'un réseau ne possédant aucun centre vital : tous ses noeuds ont le même statut. L'information divisée en paquets suit sa route au gré de la disponibilité d'un noeud de transit sur son trajet, elle ne dépend que des points de départ et d'arrivée. Cette idée est à la base même du bi-protocole de communication actuel TCP/IP du Réseau.

(*) Ingénieur Principal. CliRIST

Le protocole IP (Internet Protocol) achemine les paquets d'information d'un ordinateur à un autre, le protocole TCP (Transmission Control Protocol) contrôle l'information transmise et en assure la fiabilité et la mise en ordre.(1)

A partir des années 1990, le réseau Internet connaît une croissance de plus en plus accrue, des colonnes vertébrales à très haut débit se multiplient. Le programme NREN (National Research and Education Network) voit le jour en 1991 à l'initiative d'Albert Gore, vice-président des Etats unis et véritable animateur de ce projet. Ce programme prévoit notamment pour 1995 une épine dorsale de 3 milliards de bits/seconde. En février 1995, les sept pays les plus industrialisés au monde (G7) se partagent à Bruxelles les onze principaux projets relatifs aux autoroutes de l'information.

En Algérie, on commence enfin à recevoir des échos de ce phénomène et en même temps à prendre conscience de l'importance stratégique qu'il revêt dans le développement du pays. A cet effet, des rencontres se sont tenues autour de ce thème pour notamment interpeller les

pouvoirs publics sur la nécessité à mettre en oeuvre tous les moyens (ligne spécialisée, routeurs...) permettant à nos institutions d'en tirer profit.

Les premiers qui doivent être sensibilisés dans ce sens sont incontestablement nos professionnels de l'information (documentalistes, bibliothécaires...) car le réseau internet est en train de révolutionner leur métier. Aussi, il est urgent pour eux de connaître, d'utiliser et de maîtriser ses services et outils pour pouvoir suivre son évolution et pourquoi pas y participer comme sont en train de le faire actuellement des milliers de spécialistes de l'information de par le monde. Ainsi, des groupes de discussion (news groups), des pages Web,... sont créés quotidiennement pour alimenter le réseau.

D'après l'Internet Group, le nombre total d'utilisateurs d'Internet était d'environ 25 millions à la fin de l'année 1994. D'après la même source, toutes les personnes connectées au réseau utilisaient le service de messagerie électronique (E-mail). Une moitié d'entre eux, soit près de 12,5 millions, utilisait les services d'informations textuelles (Gopher) et 5 millions seulement utilisaient les services hypermédiés du World Wide Web (WWW). Cependant, c'est ce dernier qui assure actuellement la croissance exponentielle du nombre de services et d'utilisateurs d'Internet (plus de 35 services par jour)(2).

Dans le présent article, nous essayerons de présenter le World Wide Web (traduire la toile d'araignée mondiale), l'outil d'intemet qui connaît depuis pratiquement deux années l'expansion la plus remarquable (voir figures plus loin) à tel point que son impact a entraîné l'apparition de publicité et autres sponsors sur le réseau. La commercialisation

d'Internet est devenu par voie de conséquence une donnée de plus en plus incontournable. Cette réalité ressentie différemment par les uns et les autres, a donné lieu depuis à des débats parfois passionnés sur notamment la démocratisation de l'accès à l'information.

LE WORLD WIDE WEB

Le World Wide Web (WWW) est un système d'information hypermédia reparti qui a été intégré au réseau Internet durant l'année 1989. La notion d'hypermédia est le fait que l'information est prise au sens large (texte, image, vidéo,, son...) avec au sein du document qui la supporte des liens (ou ancres) qui renvoient vers d'autres informations se trouvant sur le même document ou sur des documents différents. La notion de reparti tient au fait que les ressources d'information peuvent être disponibles dans des sites différents et parfois dans des serveurs autres que Web(Ftp,Gopher,...).

Les premiers jalons du Web ont été posés par Tim Berners-Lee qui a proposé d'étendre la technique de l'hypertext aux réseaux pour la communication de l'information pour la communauté du CERN (Laboratoire Européen de la Physique des Particules) de Genève. La définition officielle du World Wide Web est Wide-area hypermédia information retrieval initiative aiming to give universal access to a large universe of documents.

Le principe de base du Web

Le World Wide Web, à l'instar de la plupart des applications qui fonctionnent sur Internet, s'appuie sur le modèle client/serveur. Dans ce modèle, un programme client (ou client) (3) est à la disposition de l'utilisateur. Son travail consiste à acheminer les requêtes de ce

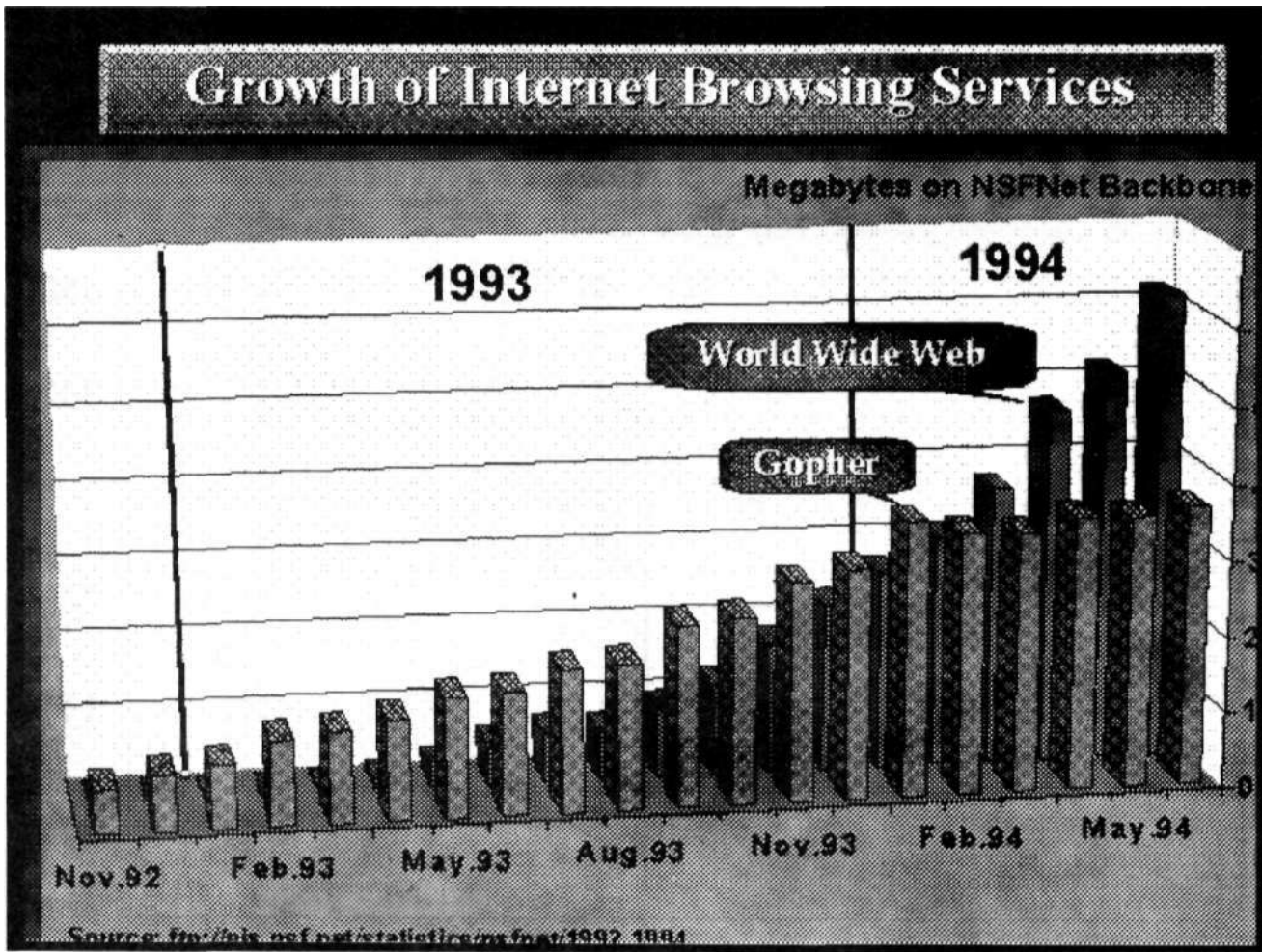
Figure 1: Croissance des serveurs Web [1993 1994]

MttM&tM&iM

Source: Vftidefer, MatlrewGray, Mlne wvrv.itituii]:«r»ijdK's»f>t>lisB.mkpfiwhiwt*) qiiiutii mm



Figure 2: Evolution chronologique nov. 92 - nov. 94
des services Gopher et Web



dernier au programme serveur (ou serveur), de traduire les résultats de ces requêtes pour les rendre compréhensibles par l'utilisateur et de les présenter sur écran, le travail du serveur consiste à répondre à ces requêtes (transfert de fichier, interrogation de bases de données. .) Le client sert en quelque sorte d'interface entre l'utilisateur et le serveur. Il n'existe pas de session permanente entre un client Web et un serveur.

Le dialogue entre un client et un serveur (www.gopher,...) suit le processus simplifié suivant:

- Etablissement de la connexion
- Le client envoie sa requête
- Le serveur fournit ou non l'information demandée
- Fermeture de la connexion

Le langage HTML

. teoq •

"Le protocole de communication entre un client Web et un serveur Web s'appelle hypertext Transfer Protocol (HTTP).

Les documents échangés entre eux sont décrits dans le format HTML (hypertext Markup Language) qui est issue de la norme SGML (Standard Generalized Markup Language). Cette norme devenue officielle en 1986, a été élaborée sous l'égide de l'International Standard Organisation (ISO)(4).

HyperText Language Markup (5) est un langage qui utilise un ensemble de commandes de formatage logique. Ces commandes sont utilisées pour ajouter de l'information aux documents originaux et les transformer ainsi en format HTML pour les rendre accessibles par le Web. Lorsqu'un client reçoit un document HTML d'un serveur Web, il interprète automatiquement toutes les commandes

insérées dans le document. C'est en lisant ces commandes que le client comprend par exemple que telle partie du document représente le titre, telle autre constitue le corps du document, tel mot représente un lien avec un autre document, etc.. Ainsi, le document sera en quelque sorte traduit et présenté sous une forme acceptable à l'utilisateur.

Le langage html connaît des développements très poussés ces derniers temps. La version 2.0 est en cours, la version 3.0 en cours de spécification. Les travaux relatifs aux versions 2.0 (6) et 3.0 (7) sont disponibles sur le réseau.

Les URLS

Comme on l'a souligné plus haut, le Web utilise des liens entre des documents se trouvant dans divers points du réseau Internet. Ces liens contiennent chacun une information indiquant de manière précise la localisation des documents ainsi que le protocole à utiliser pour récupérer ces derniers, c'est ce qu'on appelle communément l'URL (Uniform Resource Locator).

Le format simplifié d'un URL est le suivant:

Methode://nom_de_machine/nom_de_repertoire/nom_de_fichier

Le champ méthode indique le protocole à utiliser. Voici une liste assez exhaustive des protocoles utilisés actuellement:

telnet: pour accéder à une session interactive;

gopher: pour accéder aux ressources du gopherspace;

http: pour accéder aux services du Web;

ftp (file transfer protocol): pour le transfert de fichiers;

wais (wide area information servers): pour accéder aux bases de données utilisant les indexeurs wais;

nttp (net news transfer protocol): pour accéder aux groupes de discussions, maîlto pour les adresses de courrier électronique;

prospero: pour l'accès aux serveurs utilisant les annuaires de ce type;

file: pour l'accès à un fichier local.

Le nom de la machine serveur où est disponible le document, est parfois suivi du caractère ":" et d'un numéro de port lorsque ce dernier est différent du port standard de l'application.

Le chemin d'accès vers le fichier final peut contenir plusieurs répertoires emboîtés au lieu d'un seul.

Le fichier final est éventuellement suivi du caractère "#" et d'une chaîne de caractères. Ceci permet de référencer un lien existant dans le document et permettant d'accéder directement à un endroit précis de ce dernier.

L'utilisation du caractère "?" suivi d'une chaîne de caractères à la fin du nom du fichier permet de passer des paramètres à un programme qui sera exécuté par le serveur Web.

L'URL qu'on vient de définir est parfois appelé URL absolu en comparaison avec l'URL relatif qui a une écriture simplifiée quand il est utilisé dans un document pour référencer un autre document localisé sur le même serveur et accessible par le même protocole.

Dans un URL, le format des caractères (majuscule, minuscule,...) doit être respecté alors que le caractère "espace" est interdit.

Les principaux outils du World wide web

Les programmes clients

Il existe toute une panoplie de clients Web (8) et pour pratiquement toutes les plateformes. Leur niveau de développement et les postes de travail auxquels ils sont destinés sont étroitement liés à leurs fonctionnalités. Les clients les plus utilisés en ce moment sont les suivants :

Mosaïc : Il a été réalisé au NCSA (National Center Supercomputing Applications), un service de l'université de l'Illinois. Ses environnements requis sont Macintosh, Windows et X-Windows. Des passerelles vers les serveurs gopher.news et ftp sont intégrées au sein même du programme client. Le client relatif à l'environnement X-Windows est le plus évolutif.

Lynx : L'intérêt premier de ce client est son adaptation à des postes de travail bas de gamme (terminaux asynchrones, PC et Mac utilisant un émulateur de terminal,...). Il offre des passerelles vers les serveurs gopher, ftp, news,... et se compile sans problèmes sur les machines Unix et VMS.

Netscape : c'est actuellement le logiciel le plus prisé. Son principal attrait réside dans la simplicité de son emploi et sa rapidité. Outre l'offre de passerelles vers des serveurs autres que Web (gopher, ftp,...), le logiciel permet d'envoyer ponctuellement du courrier électronique et de gérer des carnets d'adresses (bookmarks).

Les programmes serveurs

Il existe, à l'instar des programmes clients, plusieurs programmes serveurs. Il

y'a évidemment des serveurs plus performants que d'autres car possédant beaucoup plus de possibilités. Parmi ces derniers, deux logiciels émergent du lot et sont largement utilisés actuellement, il s'agit des serveurs du CERN et du NCSA. Leurs possibilités sont multiples: autorisation d'accès, images cliquables,

documents.il existe actuellement un certain nombre de logiciels d'édition HTML fonctionnant sur divers environnements et disponibles gratuitement sur des sites Ftp. Les plus utilisés en ce moment sont le Simple HTML Editor (SHE) pour Macintosh et le HTML Assistant pour Windows. Il existe

QUELQUES EXEMPLES D'URLS

http://web.urec.fr/docs/WWW/section2_1_1.html est l'URL qui définit le fichier section2_1_1.html lequel est disponible sur le serveur web.urec.fr. Ce fichier est accessible par le protocole http, il est dans le répertoire WWW qui est lui-même dans le répertoire docs.

<http://www.cob.ohio-state.edu/dept/fin/cerninv.htm#Link16> est l'URL qui définit un endroit précis du fichier cerninv.htm lequel est disponible sur le serveur www.cob.ohio-state.edu et accessible par le protocole http. Ce fichier se trouve dans le répertoire fin qui est lui-même dans le répertoire dept.

<ftp://ftp.ibp.fr/pub/www/ncsa/Mosaic/Windows/axp.mosaic.2.0.0a8.exe> est l'URL du fichier source du logiciel client Mosaic lequel est disponible sur le serveur ftp.ibp.fr par le protocole ftp. Ce fichier se trouve dans la série de répertoires emboîtés pub/www/ncsa/Mosaic.

<gopher://alyssa.rsinet.tn> est l'URL du serveur alyssa.rsinet.tn accessible par le protocole gopher.

<wais://quake.think.com:210/directory-of-servers?inria> correspond à l'interrogation de la base WAIS directory of servers sur le serveur quake.think.com en attente sur le port 210, inria étant le paramètre de l'interrogation.

recherche indexée, etc.. Leurs plateformes respectives sont **UNIX** ou **VMS** (pour le CERN) et UNIX (pour le NCSA)

aussi des logiciels d'édition HTML compatibles avec des logiciels de traitement de texte (Microsoft Word 6.0 et Word Perfect 6.0).

Les éditeurs HTML

Il est possible de traduire un document décrit dans un certain format en un document HTML par le simple ajout manuel de commandes de balisage au sein de ce même document. Cependant, cette démarche demeure fastidieuse et engendre parfois des erreurs d'inattention telles que l'omission de certaines commandes par exemple. A cet effet, les éditeurs HTML viennent à point nommé pour remédier à cette situation et offrir ainsi leur assistance pour un traitement rapide et infaillible des

Les convertisseurs

Les convertisseurs (ou, filtres) (9) permettent de traduire un document décrit dans un certain format en un document HTML. Un certain nombre d'outils **est** toutefois nécessaire à leur fonctionnement. Les principaux filtres sont les suivants :

rtfohtml : il permet la conversion de documents RTF et par voie de conséquence de fichiers Word et WordPerfect vers le format HTML.

LaTeX2HTML : c'est un convertisseur de textes écrits en LaTeX vers HTML. Il possède entre autres pour fonctionnalités la création d'un fichier HTML par section et sous section, la navigation par page à l'aide d'icônes et la traduction de figures, footnotes et autres.

Frame2html : ce logiciel convertit un fichier FrameMaker en HTML. Il est écrit en langage C et s'installe sous UNIX. Ses principales fonctionnalités concernent le support de références croisées, la conversion au format GIF des images se trouvant dans les fichiers FrameMaker et la diffusion de documentation électronique à partir d'un seul fichier source FrameMaker.

LE WEB ET LES PROFESSIONNELS DE L'INFORMATION

Le World Wide Web permet une communication rapide de l'information et ce de façon très conviviale. Il offre par rapport aux autres outils d'information d'Internet beaucoup plus de possibilités et donne à travers la notion d'hypermédia une certaine idée de la bibliothèque de demain où l'information au sens large du terme est offerte en libre accès à l'utilisateur, ce dernier navigant selon son besoin ou sa curiosité d'un document à un autre.

La création et la maintenance de sites Web est un marché en pleine expansion. L'installation d'un serveur Web demeure le travail de l'informaticien, cependant, elle peut être faite par un non-informaticien maîtrisant l'outil informatique. Sa maintenance qui comprend la création et la mise à jour de l'information relève du domaine du spécialiste de l'information.

Toute institution peut sans moyens importants mettre en place son propre

serveur. Il suffit pour cela de disposer de tous les outils nécessaires à son installation (logiciel client, logiciel serveur,...) et bien sûr de la matière première qui est l'information.

La création et l'administration de sites WWW ont donné lieu à travers des écrits (10) à de nombreux conseils, notamment ceux relatifs à l'organisation et à la taille des fichiers qui doivent être pensées au départ car une fois qu'un serveur est connu, il sera très vite référencé par d'autres serveurs et sollicité par la suite par des clients.

Le World Wide Web est à travers l'hypermédia une nouvelle conception de l'organisation de l'information. Cependant, cette organisation risque de poser quelques problèmes majeurs parmi

lesquels on pourra citer l'excès d'information qui peut entraîner l'étouffement des informations stratégiques indispensables aux décideurs et le partage des ressources qui peut paradoxalement engendrer une faille dans le système des sites créateurs de services Web devient inopérant pour une raison ou pour une autre (11). Le rôle des spécialistes de l'information et autres bibliothécaires réside justement dans cet apport perpétuel d'idées capable de résoudre même partiellement ce genre de problèmes qui se posent incessamment. Pour ce faire, ils doivent notamment être au diapason de toute technologie touchant de près ou de loin à leur profession.

Les auteurs de ces notes bibliographiques sont : aeb aiotisq 9ibr noissimo'l 9UD usns

NOTES BIBLIOGRAPHIQUES

(1) Bezsonoff (Alexandre). - Internet. un puzzle à reconstituer : le réseau de communication restreint à l'espace

d'information étendu In Bulletin des Bibliothèques de France, 1994,t.39,n° 36, p. 58-63

(2)Ghesquiere (Gilles).- Qui a peur d'Internet. In: Solutions Télématiques,Mai 1995,n°35,p. 16-17

(3)Certains ouvrages utilisent les termes baladeur ou visualiseur,d'autres le terme anglophone browser.

(4)Role (François).- La norme SGML Pour décrire la structure logique des documents. In: Documentaliste-Science de l'information,1991,vol. 28,n° 4-5, p. 187-191

(5)<http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html>

(6)http://WWW.W3.org/hypertext/WWW/MarkUp/html-spec/html-spec_toc.html

(7)<http://ds.internic.net/internet-drafts/draft-ietf-html-specv300.txt><http://WWW.W3.org/hypertext/WWW/MarkUp/html3/html.dtd>

(8)<http://WWW.W3.org/hypertext/WWW/Clients.html>

(9)http://web.urec.fr/docs/WWW/section2_1_1.html

(10)<http://www.w3.org/hypertextA/VWW/Provider/Style/Overview.html>

(11)Le Crosnier (Hervé).- L'hypertexte en réseau : repenser la Bibliothèque. In : Bulletin des Bibliothèques de France,1995,t.40,n°2,p. 23-31