

## Le World Wide Web

### 1.Présentation :

Le World Wide Web, littéralement la « toile (d'araignée) mondiale », communément appelé le Web, parfois la Toile ou le WWW, il est un système hypertexte public fonctionnant sur Internet qui permet de consulter, avec un navigateur, des pages accessibles sur des sites. L'image de la toile d'araignée vient des hyperliens qui lient les pages web entre elles.

Le Web n'est qu'une des applications d'Internet. D'autres applications sont le courrier électronique, la messagerie instantanée et Usenet, etc. Le Web a été inventé plusieurs années après Internet, mais c'est lui qui a rendu les médias grand public attentifs à Internet.

### 2. Les bases du Web

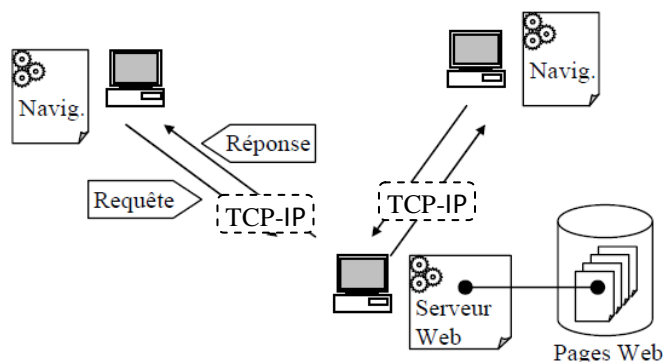
Le Web fait appel à un certain nombre de concepts de base. En ce qui suit, nous verrons les trois les plus importants :

#### 2.1 Architecture Client-Serveur

Le World Wide Web s'appuie sur la notion d'architecture client serveur.

Un serveur est une machine en général assez puissante qui fournit un ou plusieurs services (accès à des sources de données, applications...). Pour fournir ces services, elle fait tourner en permanence des programmes que l'on appelle aussi des serveurs, les plus utilisés sont des serveurs Web ou serveurs HTTP.

De l'autre coté les utilisateurs font tourner sur leur machine (machine cliente) un programme client qui, comme son nom l'indique va être demandeur de services, généralement ce client est un navigateur Web qui va demander des pages Web à un serveur Web. Le dialogue entre le client et le serveur se compose donc de requêtes émises par le client et de réponses données par le serveur.



Le client est couramment appelé un navigateur (browser). Les navigateurs les plus connus étant Mozilla Firefox, Internet Explorer, Google Chrome, Safari , Opera...etc. Les plus courants acceptent des extensions (Plug-In) permettant d'étendre leurs capacités (lire des vidéos, recevoir du son ou des films en flot continu, ...). Ils connaissent aussi fréquemment des langages évolués (JAVA, Javascript, ...) permettent d'élargir le champ des possibilités de l'utilisation des pages Web.

L'utilisateur peut visualiser une page disponible sur sa machine auquel cas le navigateur obtient le fichier par simple lecture directe sur un disque de la machine sur laquelle il s'exécute. Aussi, l'utilisateur peut accéder à une page disponible sur une machine distante auquel cas le navigateur doit se connecter au serveur publiant cette page à travers le réseau.

Chaque ordinateur connecté à Internet est adressé par un numéro IP unique (statique (fixe pour toutes les connections de ce micro à Internet (c'est le cas des micros serveurs) ou dynamique (peut changer à chaque

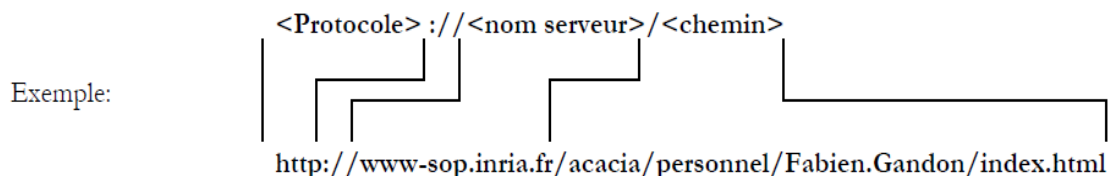
connexion (c'est le cas des micros clients généralement)). Une adresse IP est représentée en quatre nombres compris entre 0 et 255.

**Exemple :** 69.147.76.15, 186.34.223.0

## 2.2 URL (Universal/Uniform Resource Locator):

Pour accéder à une page web, il faut d'abord pouvoir décrire où elle se trouve. Pour repérer un document, un fichier, une source de données... on a développé la notation URL. Un URL peut désigner un serveur FTP, un fichier sur votre disque, une image, une adresse courrier, un serveur de News, un serveur TelNet et bien sûr une page Web publiée par un serveur HTTP, c'est-à-dire un serveur de Web.

En particulier, dans ce dernier cas l'URL contient le nom du protocole d'accès au fichier (HTTP, SHTTP), le nom du serveur (adresse IP ou nom symbolique), le chemin d'accès au fichier et bien sûr le nom du fichier :



Un des problèmes posés par ce système est que si un URL vient à changer, il faut remettre à jour tous les liens qui l'utilisent. Il faut donc maintenir ses pages et régulièrement vérifier que les liens sont toujours corrects ou sinon. Les surfeurs risquent de se retrouver dans une impasse avec une ressource (page web, image...) introuvable.

Les noms d'URL utilisent les lettres de l'alphabet en général en minuscule, les chiffres sont autorisés, certains caractères / . : # ? & ont une signification particulière et sont donc réservés, enfin certains caractères sont dit non sûrs dans la mesure où ils sont interprétés ou interprétables différemment : les blancs, les étoiles, etc.

**Exemple:** si vous lancez une requête sur le mot clef 'Algérie' sous Google!

l'URL est : [https://www.google.dz/?gws\\_rd=cr&ei=0yneVLG6Iobdat3AgdgN#q=algérie](https://www.google.dz/?gws_rd=cr&ei=0yneVLG6Iobdat3AgdgN#q=algérie)

Cela signifie que votre requête appelle un programme de recherche du moteur google avec le paramètre 'Algérie'.

La forme la plus complexe d'un URL étant :

**<service>:[//][nom utilisateur]:[mot de passe][@]<serveur>:<port>/<chemin>**

On trouve d'autres types d'URL, chacun représentant un service donné, un certain nombre d'exemples vous sont donnés dans le tableau ci-dessous

Service	Masque	Exemple
Web	http://serveur:port/repertoire/fichier.html	http://www.chez.com/toto/fichier.html
FTP	ftp://serveur/repertoire/fichier	ftp://inria.ftp.fr
Fichier local sur PC	file:///disque /repertoire/fichier	file:///c /tmp/fichier.txt
Mail	mailto:nom@organisation.domaine	mailto:Fabien.Gandon@sophia.inria.fr
Telnet	telnet://Nom:Password@serveur:port	telnet://gandonf:abcde@gopa.insa.fr:23
Gopher	gopher://serveur:port/repertoire/fichier#marqueur	
Serveur de News	newsr://serveur:port/repertoire/nom.de.la.news	
WAIS	wais://<host>:<port>/<database>	

## 2.3 Hypertexte et Topologie du Web

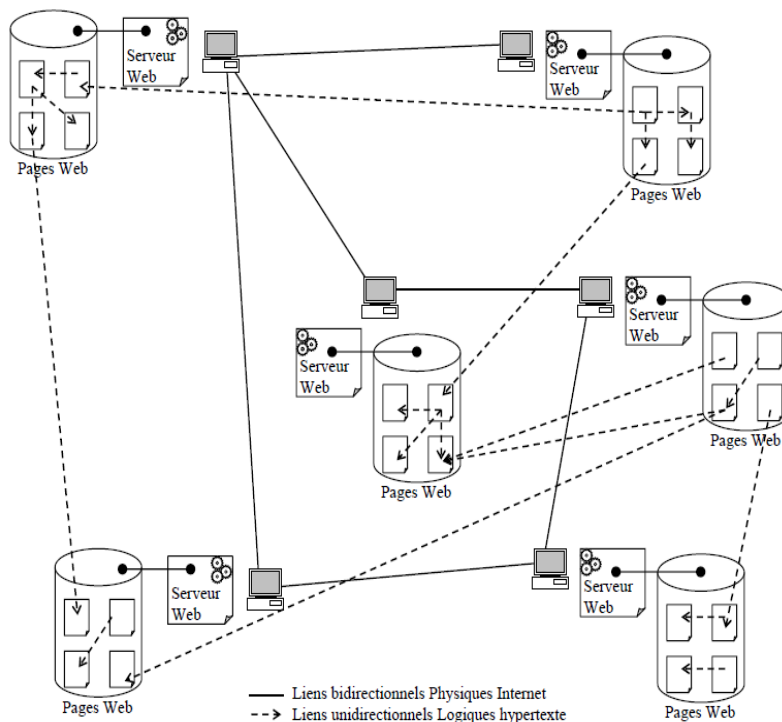
Le World Wide Web est un vaste ensemble de sources d'informations accessibles à travers le réseau Internet. L'information est présentée essentiellement sous forme de texte et d'images, animations, son et vidéo... le multimédia. Etant en pleine explosion, on commence à employer de plus en plus souvent le terme hypermédia au lieu d'hypertexte. On qualifie d'Hypertexte un document essentiellement textuel, dynamique, capable de changer et de réagir en fonction de certains événements comme par exemple un clic à la souris.

Le langage permettant de décrire les pages Web est le HTML (*Hyper Text Markup Language*). Ce langage à balise permet de doter certains mots, ou images d'une propriété d'hyperlien ou plus simplement de lien hypertexte qui est constitué d'une adresse URL que vous atteindrez en cliquant dessus.

Une Page Web contient donc du texte, des images... et des liens vers d'autres pages Web ou d'autres fichiers.

Les liens permettent de naviguer de pages en pages d'un simple clic. L'utilisateur peut passer en un clic d'une page placée sur un serveur à Alger à une autre sur un serveur à Tokyo.

En résumé, le World Wide Web est une toile d'araignée de serveurs d'informations reliés les uns aux autres par des liens physiques (le réseau matériel) et des liens logiques (les liens hypertextes) entre les pages qu'ils publient.



## 3. LA TECHNOLOGIE DU WEB

Le World Wide Web est un ensemble de protocoles (ex : HTTP...) d'outils (ex : navigateurs...) et de normes (ex : URL...) permettant de créer, formater, rechercher et partager de manière interactive des informations hétérogènes à travers Internet sur le principe du Client/Serveur. Le formatage des informations est principalement basé sur la technique des documents hypertextes balisés grâce au langage HTML et diffusés grâce au protocole HTTP. Le Web est aussi capable d'utiliser d'autres protocoles tels que : FTP (File Transfer protocol), Telnet, NNTP (Network News Transfer Protocol), WAIS (Wide Area Information System/Server), Gopher (de 'go fer'), ...

### 3.1 Le HTTP

Le protocole de base du World Wide Web est HTTP (*HyperText Transfer Protocol*) qui peut être utilisé pour n'importe quelle application client-serveur impliquant de l'hypertexte. Ce protocole est capable d'assurer le transfert de texte, hypertexte, fichiers audio, images ou tout autre type d'information pouvant se mettre sous la forme d'un fichier.

Le scénario de dialogue classique entre un navigateur et un serveur Web est le suivant :

- Le navigateur Web client établit une connexion TCP avec le serveur Web qui contient la page qui l'intéresse.
- Une fois la connexion établie, le client émet une requête HTTP contenant une commande, une URL, et parfois d'autres informations.
- Lorsque le serveur Web reçoit la requête il essaie d'exécuter la commande qu'elle contient. Il retourne ensuite comme réponse le résultat obtenu qui peut être des données, un message d'erreur, et d'autres informations.
- Une fois que le client a reçu sa réponse, la connexion est fermée et détruite.

Voici par exemple une requête émise par un navigateur dont l'utilisateur est "Mohamed Islem" et demandant à ce que le document html "cv.html" lui soit envoyé s'il a été modifié depuis le 31 juillet 2000 à 12:30:00.

GET cv.html	→	demande le document "cv.html"
Content-Type: text/html	→	précise que le document est du texte pur ou html
If-Modified-Since: 31 Jul 2000 12:30:00	→	n'envoyer le doc que s'il a été modifié depuis cette date
From: Mohamed Islem < <a href="mailto: Mohamed.Islem@yahoo.fr">Mohamed.Islem@yahoo.fr</a> >	→	la requête est émise par Mohamed Islem

### 3.2 Le HTML

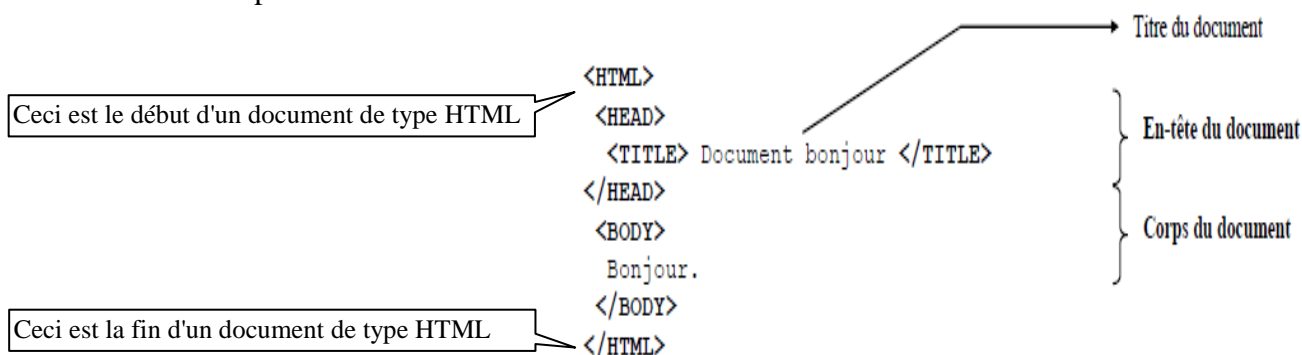
Le **HTML** (*HyperText Markup Language*) est un format de présentation de données permettant de créer des pages web (un fichier **HTML**) pouvant être lues par des navigateurs Web. C'est un **langage de description de données** et **non pas un langage de programmation**. Ce langage se compose d'un ensemble d'annotations, appelées étiquettes ou balises, qui permettent de créer et formater un document hypertexte.

Une page Web peut être créée directement avec un simple éditeur de texte (Microsoft Word, Bloc Notes ...) en tapant des commandes HTML et en enregistrant le fichier en extension htm ou html, ou plus simplement en utilisant un éditeur de page Web (Microsoft Expression Web qui remplace Microsoft Frontpage, Adobe Dream Weaver ..) qui très souvent vous permettra de créer votre document de façon très conviviale et générera pour vous le code HTML correspondant sans que vous ayez à connaître ce langage.

Un fichier HTML est un fichier texte ce qui a l'avantage de le rendre facilement lisible sur n'importe quelle plate-forme/ordinateur. Les balises du HTML sont insérées dans le texte du document et guident son affichage. Le navigateur interprète les commandes HTML contenues dans le document et en déduit le format d'affichage du document.

Une balise HTML est un élément que l'on va ajouter au texte de départ pour dire au navigateur de quelle manière l'afficher. Elle n'est pas affichée telle quelle dans le navigateur, elle est interprétée par celui-ci. Elle est toujours délimitée par les signes < et >.

Un fichier HTML simple ressemble à ceci :



Une balise commence toujours par un signe "<" et se finit toujours par un signe ">". La plupart des balises doivent être ouvertes et fermées pour délimiter leur zone d'influence. La balise fermante contient la même commande que la balise ouvrante, mais précédée d'un caractère /.