

Chapitre 1 : Internet, Web, C/S et Web 2.0

1. Aux origines de L'internet

L'Internet est un système global de réseaux d'ordinateurs interconnectés qui utilise la norme TCP/IP pour relier des milliards de terminaux au niveau mondial.

C'est un réseau de réseaux qui est constitué de millions de réseaux privés et publics, académiques, d'entreprises, gouvernementaux, d'étendue locale/régionale et globale. Ils sont liés par un large ensemble de technologies électriques filaires, sans fil et optiques (laser).

L'Internet est le support d'un large éventail de ressources et de services tels que les documents hypertextes du World Wide Web (WWW), l'infrastructure pour la messagerie électronique, les réseaux pair à pair pour le partage de fichiers et la téléphonie.

A L'origine, L'internet était un réseau commandé par le ministère américain de la défense à travers son agence pour la recherche dans les projets avancés (DARPA). Le projet a été lancé en réaction au premier satellite artificiel lancé par l'URSS, Sputnik (1957). Celui-ci permettait aux unités combattantes de communiquer quelque que soit leur position sur terre et même si les américains attaquaient et détruisaient les infrastructures de télécommunications en URSS. Les Etats-unis voulaient disposer d'un réseau qui s'adapte pour assurer l'acheminement des paquets d'un hôte source vers un hôte destination d'où les algorithmes de routage de l'Internet. Le premier réseau s'appelait ainsi ARPANet (Le réseau de L'ARPA) aux états-unis dans les années 60.

Dans les années 80, Tim Berners-Lee inventa le concept de documents liés grâce à des liens. L'ouverture de l'Internet aux entreprises en 1992 marqua le début de l'Internet tel qu'on le connaît aujourd'hui. Depuis lors, le réseau a connu, en effet,

Internet, Web, C/S et Web 2.0

une croissance exponentielle soutenue, au fur et à mesure que des générations d'ordinateurs institutionnels, personnels et mobiles y furent connectés.

2. Services Internet

L'Internet rend effectifs des « services réseaux », dont les plus éminents sont le World Wide Web, la messagerie électronique (mail), la téléphonie Internet, et le partage de fichiers (FTP) et réseaux Peer 2 Peer.

Le *World Wide Web* est désigné par de nombreux noms et abréviations synonymes : *WorldWideWeb*, *World Wide Web*, *World-wide Web*, *Web*, *WWW*, *W3*, *Toile d'araignée mondiale*, *Toile mondiale*, *Toile*. Le nom désormais préconisé par le *World Wide Web Consortium (W3C)* sépare les trois mots sans trait d'union. Bien que « mondial » s'écrive *world-wide* ou *worldwide* en anglais, l'orthographe *World Wide Web* et l'abréviation *Web* sont maintenant bien établies.

Le sigle *WWW* a été largement utilisé pour abrégé *World Wide Web* avant que l'abréviation *Web* prenne le pas. À l'écrit, les lettres *www* restent très utilisées dans les adresses Web et quelques autres conventions techniques. *WWW* est parfois abrégé en *W3*, abréviation qu'on retrouve dans le sigle *W3C* du *World Wide Web Consortium*.

3. Termes rattachés au web

L'expression « en ligne » signifie « connecté à un réseau », en l'occurrence le réseau informatique Internet. Cette expression n'est donc pas propre au web, mais à Internet dans sa globalité, on la retrouve également à propos des réseaux téléphoniques.

Un **hôte** est un ordinateur connecté au réseau Internet.

Une ressource du World Wide Web est une entité informatique (texte, image, forum Usenet, boîte aux lettres électronique, etc.) accessible indépendamment d'autres

Internet, Web, C/S et Web 2.0

ressources. Une ressource en accès public est librement accessible depuis tout Internet. Lorsqu'une ressource est présente sur l'hôte de l'utilisateur, elle est dite locale ; par opposition elle est distante si elle est présente sur un hôte différent. Lorsque celle-ci est connectée et disponible au réseau, elle est en ligne. On ne peut accéder à une ressource distante qu'en respectant un protocole de communication. Les fonctionnalités de chaque protocole varient : réception, envoi, voire échange continu d'informations.

HTTP (pour *HyperText Transfer Protocol*) est le protocole de communication communément utilisé pour transférer les ressources du Web. HTTPS est la variante avec authentification et chiffrement de ce protocole.

Une **URL** (pour *Uniform Resource Locator*, signifiant littéralement « localisateur de ressource uniforme ») est une chaîne de caractères décrivant l'emplacement d'une ressource. Elle contient généralement dans l'ordre : le nom du protocole, un deux-points (:), deux barres obliques (/), un nom d'hôte, une barre oblique (/), et un chemin composé de texte séparés par des barres obliques. Exemple : <http://example.com/un/chemin/page.html>.

Un **hyperlien** (ou dans le langage courant, un lien) est un élément dans une ressource qui est associé à une URL. Les hyperliens du web sont orientés : ils permettent d'aller d'une source à une destination. Seule la ressource à la source contient les données définissant l'hyperlien. Ainsi, il est possible d'établir un hyperlien sans la moindre intervention, ni coopération, de la ressource cible de l'hyperlien. Ce modèle unilatéral adopté par le Web ne se retrouve pas dans tous les systèmes hypertextes. Il a l'avantage prépondérant, dans un réseau public mondial, de permettre la création d'innombrables hyperliens entre partenaires décentralisés et non coordonnés.

HTML (pour *HyperText Markup Language*) et **XHTML** (*Extensible HyperText Markup Language*) sont les langages informatiques permettant de décrire le contenu d'un document (titres, paragraphes, disposition des images, etc.) et d'y

Internet, Web, C/S et Web 2.0

inclure des hyperliens. Un document HTML est un document décrit avec le langage HTML.

Documents

Une page web (ou page) est un document destiné à être consulté avec un navigateur web. Une page web est toujours constituée d'une ressource centrale (généralement un document HTML ou XHTML) et d'éventuelles ressources liées, automatiquement accessibles (par exemple, des images, des feuilles de style, des scripts javascript ou des fontes de caractères). Il n'est pas exceptionnel qu'un seul document HTML fasse appel à entre quelques dizaines à une centaine de ressources annexes.

Un éditeur HTML (ou éditeur web) est un logiciel conçu pour faciliter l'écriture de documents HTML et de pages web en général.

Un site web (ou site) est un ensemble de pages web et d'éventuelles autres ressources, liées dans une structure cohérente, publiées par un propriétaire (une entreprise, une administration, une association, un particulier, etc.) et hébergées sur un ou plusieurs serveurs web.

4. Le modèle Client/Serveur

Traditionnellement, toute application peut être décomposée en trois parties : Les données, les traitements qui modifient les données et les rapportes à l'utilisateur et l'interface utilisateur. L'interface utilisateur est la partie d'application avec laquelle l'utilisateur interagit : il clique sur des objets graphiques tels que des boutons, des menus, des zones de saisies de texte, etc...L'application réagit aux actions de l'utilisateur sur l'UI en déclenchant les traitements associés à chaque action de l'utilisateur.

Toute application de cette forme peut être décomposée en deux ou plus de parties qu'on appelle *tiers*. Lorsque le nombre de tiers est 2, l'application est dite à deux tiers ou de type client/serveur. La partie client est celle qui du côté de l'utilisateur. L'autre partie est dite la partie serveur. La partie client et la partie serveur sont exécutées sur deux machines différentes, la machine où est exécutée le client est

Internet, Web, C/S et Web 2.0

appelé machine client, et la machine où est exécutée la partie serveur est dite machine serveur ou un serveur. Le dialogue entre la partie client et la partie serveur se fait grâce à des messages envoyés par le client auxquels le serveur répond par des messages aussi. Le dialogue C/S commence toujours à l'initiative du client. On appelle le message du client : requête (query) la réponse elle (answer). A chaque requête correspond une réponse. L'ensemble des messages possibles et leur ordre constitue ce qu'on appelle un protocole de communication ou juste un protocole. Pratiquement tous les protocoles de l'Internet sont de type C/S y compris donc le HTTP.

Dans le mode de communication client-serveur, un serveur est un hôte sur lequel fonctionne un **logiciel serveur** auquel peuvent se connecter des **logiciels clients** fonctionnant sur des *hôtes clients*, un seul hôte peut contenir les deux.

Serveur HTTP, Client HTTP

Un **serveur HTTP** (ou service web) est un programme (Apache ou IIS, etc) utilisé pour servir des ressources à travers le protocole HTTP. Il accepte des connexions HTTP pour servir des requêtes HTTP en envoyant des messages de réponse HTTP.

Un **serveur web** est un hôte sur lequel fonctionne un serveur HTTP (ou service web). Un serveur web peut héberger les ressources qu'il dessert, peut les récupérer sur des hôtes distants, délivrer ces ressources telles quels, on parlera alors de ressources statiques), ou bien les modifier en fonction de différents critères, on parlera alors de ressources dynamiques.

Client HTTP

Un client HTTP est un programme permettant d'utiliser les ressources fournies par un serveur HTTP. Il est chargé en relation avec différents services, de faire traduire le nom d'hôte en adresse IP par un DNS, de rentrer en contact avec le serveur HTTP, et de lui fournir différentes informations telles que le nom du site, et la ressource désirée, ainsi que ses capacités (gestion de différents types de ressource, capacité

Internet, Web, C/S et Web 2.0

d'affichage, possibilité de compression de données, etc.) ou sa volonté (choix des langues, protection de la vie privée) dans l'utilisation des données fournies. Il envoie au serveur HTTP un ou plusieurs messages de type requête.

Un **navigateur web** est un type de logiciel client HTTP à interface homme-machine et conçu pour accéder aux ressources du web. Sa fonction de base est de permettre la consultation des documents HTML disponibles sur les serveurs HTTP. Le support d'autres types de ressources et d'autres protocoles de communication dépend du type de navigateur.

Application lourde vs application légère

Un *application lourde*, est une application réalisée pour s'exécuter sur un système graphique particulier (Windows : WinForms ; Linux : Qt, Mac osx : Cocoa....etc).

Une *application légère* est une application client/serveur dans laquelle la partie interface utilisateur est visible dans un navigateur Web, alors que la partie serveur (traitements + données) est hébergée sur un serveur (web + de base de données).

L'avantage de l'application légère est qu'elle ne nécessite aucune installation sur les machines des utilisateurs, leur mise à jour se fait en un seul endroit (le serveur), et que l'utilisateur peut y accéder à partir de système d'exploitation/terminaux de différents sans effort, ce qui n'est pas le cas des applications lourdes.

5. Le protocole HTTP

Le HTTP est le protocole de communication communément utilisé pour transférer les ressources du web. HTTPS est la variante avec authentification et chiffrement.

Les requêtes et les réponses HTTP suivent la norme RFC 822. Le format de message générique contient les 04 éléments suivants :

+Une ligne de début

Internet, Web, C/S et Web 2.0

+Zéro ou plusieurs champs d'en-tête ,chacun suivi par CRLF

+Une ligne vide (seulement CRLF) indique la fin des champs d'en-tête.

+Un corps de message optionnel

La ligne de début est soit une ligne de requête (lorsque le message provient du client) soit une une ligne d'état (lorsque il vient d'un serveur)

Par exemple : GET /hello.htm HTTP/1.1 (requête)

HTTP/1.1 200 OK. (ligne d'état)

Champs d'en-tête

Ils fournissent des informations requises sur la requête ou la réponse ou sur un objet envoyé dans le corps du message

Exemples :

User-Agent : Mozilla/5.0 (Macintosh, Intel Mac OS X 10_10_2[...]Safari/600.3.18

Indique la version du navigateur, du système d'exploitation, la nature de l'hôte...

Accept-Language : en, fr : liste des langues supportées par le navigateur par ordre de préférence de l'utilisateur

Content-Length : nombre en octets : Longueur du corps du message

Content-Type : Type MIME de contenu (text/plain, text/html, application/x-pdf, etc

Last-modified : Date + heure de dernière modification de la ressource

Cookie : valeur alphanumérique pour supporter un contexte utilisateur

Type de requêtes HTTP

Une requête HTTP a le format suivant :

Méthode espace URI espace Version HTTP CRLF.

La méthode indique l'action à effectuer sur la ressource identifiée par la ressource fournie par l'URI. La version HTTP est notée HTTP/1.0 ou HTTP/1.1.

Les types de méthodes possibles sont : GET, HEAD, POST, PUT, DELETE (HTTP/1.x) et CONNECT, OPTIONS, TRACE (HTTP/1.1)

GET : permet de retourner la ressource à partir du serveur. Cette méthode ne modifie pas les données sur le serveur. Si un en-tête Range (Intervalle) est utilisé, le GET est dit partiel

HEAD : Même chose que GET mais la réponse du serveur contient uniquement la ligne d'état et les en-têtes, mais pas le corps du message

POST : Utilisée pour envoyer des données au serveur. Par exemple, des informations concernant un client d'une entreprise, un fichier uploadé en utilisant les formulaires HTML (balise form)

PUT : Remplacer le contenu de la ressource URI par le corps de ce message

DELETE : Supprimer la ressource cible

Exemple de message GET :

```
GET /hello.htm HTTP/1.1
User-Agent: Mozilla/4.0 (compatible; MSIE5.01; Windows NT)
Host: www.tutorialspoint.com
Accept-Language: en-us
Accept-Encoding: gzip, deflate
Connection: Keep-Alive
```

Exemple de message POST :

```
POST /cgi-bin/process.cgi HTTP/1.1
User-Agent: Mozilla/4.0 (compatible; MSIE5.01; Windows NT)
Host: www.tutorialspoint.com
Content-Type: application/x-www-form-urlencoded
Content-Length: longueur
Accept-Language: en-us
Accept-Encoding: gzip, deflate
Connection: Keep-Alive
```

```
licenseID=string&content=string&/paramsXML=string
```

Exemple de message de réponse (une page HTML est envoyée dans le corps, après les en-têtes, la ligne vide entre le dernier en-tête et le HTML fait partie du message, ce n'est pas une erreur):

```
HTTP/1.1 200 OK
Date: Mon, 27 Jul 2009 12:28:53 GMT
Server: Apache/2.2.14 (Win32)
Last-Modified: Wed, 22 Jul 2009 19:15:56 GMT
Content-Length: 88
Content-Type: text/html
Connection: Closed
```

```
<html>
<body>
<h1>Hello, World!</h1>
</body>
</html>
```

Un exemple d'une ligne d'état indiquant un code d'erreur

```
HTTP/1.1 404 Not Found
Date: Sun, 18 Oct 2012 10:36:20 GMT
Server: Apache/2.2.14 (Win32)
Content-Length: 230
Connection: Closed
Content-Type: text/html; charset=iso-8859-1
```

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<html>
<head>
  <title>404 Not Found</title>
</head>
<body>
  <h1>Not Found</h1>
```

```
<p>The requested URL /t.html was not found on this server.</p>  
</body>  
</html>
```

6. Web 1.0 et Web 2.0

On appelle site web, l'ensemble des ressources HTTP associées à un nom de domaine et un ou plusieurs noms d'hôtes. Une adresse web est une URL de page web, généralement écrite sous une forme simplifiée limitée à un nom d'hôte. Une adresse de site web est en fait l'adresse d'une page du site prévue pour accueillir les visiteurs (souvent elle s'appelle index.html, index.php ou index.extension selon le langage qui la génère).

La première génération de sites web créés permettait de présenter des informations qui changent très rarement. Par exemple, une entreprise créait un site de type vitrine, qui expose la liste de ses produits, leur caractéristiques, une ou plusieurs figures, etc...

Dans un second temps, des sites où le nombre de données est important (des dizaines de produits) ont été présentés aux utilisateurs. Ces sites utilisaient une base de données contenant la liste des produits. En se connectant au site, la page contenant la liste des produits est générée par un programme déclenché par l'utilisateur. Ce type de page est appelée « page générée dynamiquement ». Ainsi est née une deuxième génération de site web dit à contenu dynamique ou « site dynamique ». Avec les sites à contenu statique, l'utilisateur est passif; il consulte les pages du site sans pouvoir y contribuer de quelque façon que ce soit. Avec les sites dynamiques, il peut introduire certaines informations à travers un formulaire, le serveur web déclenche un programme qui génère la page web résultat, puis la retourne au client HTTP (observez par exemple le moteur de recherche google, etc...) Mais le contenu de la base de données est fixé souvent par les entreprises qui possèdent le site web. Tout ces sites sont dits de type Web 1.0.

Avec les années 2000 les notions de blog, de wiki (en 2001, lancement de Wikipédia en anglais) et de réseau social (Myspace en 2003, Facebook en 2004) deviennent populaires. Le contenu généré par les utilisateurs se répand (Wikipédia, YouTube en 2005, Twitter en 2006). L'expression Web 2.0, largement popularisée au milieu

Internet, Web, C/S et Web 2.0

des années 2000, désigne cette transition dans le flux de l'information et la manière d'utiliser le web.

Les utilisateurs créent les contenus et « interagissent » au lieu de seulement accéder à des contenus créés par des entreprises ou des organisations.

Ainsi, l'expression « **Web 2.0** » désigne l'ensemble des techniques, des fonctionnalités et des usages du *World Wide Web* qui ont suivi la forme originelle du web, caractérisée par plus de simplicité et d'interactivité. Elle concerne en particulier les interfaces permettant aux internautes ayant peu de connaissances techniques de s'approprier de nouvelles fonctionnalités du web. Les internautes peuvent d'une part contribuer à l'échange d'informations et interagir (partager, échanger, etc.) de façon simple, à la fois au niveau du contenu et de la structure des pages, et d'autre part entre eux, créant notamment le Web social. L'internaute devient, grâce aux outils mis à sa disposition, une personne active sur la toile.

Le **Web 2.0** est donc l'évolution du Web vers l'interactivité à travers une complexification interne de la technologie mais permettant plus de simplicité d'utilisation, les connaissances techniques et informatiques n'étant pas indispensables pour les utilisateurs.

HTML5

Depuis sa création, le langage HTML a fait l'objet d'une concurrence rude entre les différents éditeurs navigateurs web (Netscape et Microsoft dans un premier temps, puis avec Firefox, safari , etc). En effet, chaque éditeur ajoutait ses propres balises et ses propres attributs ou leur donnait des noms différents de ceux utilisés par d'autres navigateurs. Il en a résulté une grande difficulté pour les créateurs de sites web pour produire des pages compatibles avec tous les navigateurs. Ces problèmes ont commencé à avoir une solution avec la constitution du W3C qui regroupe tous les éditeurs de navigateurs (Microsoft, Apple, Mozilla, etc...). Le W3C produit des standards relatifs à HTML, CSS, et XML en particulier. La dernière version de HTML s'appelle HTML5.

Dès l'année 2012, la majorité des moteurs de rendu HTML permettent d'exploiter ce standard et en 2013, la majorité des Système de gestion de contenu utilisent une

Internet, Web, C/S et Web 2.0

base HTML5. HTML5 apporte la possibilité d'utiliser des vidéos et de l'audio dans les pages sans dépendre d'un format propriétaire, avec les normes MPEG-4, WebM et Ogg Vorbis, d'utiliser davantage de graphiques vectoriels avec l'utilisation de la norme SVG, de faire de la visioconférence avec webRTC26. Le développement de WebGL permet désormais d'utiliser des objets en trois dimensions dans les applications web avec une grande inter-opérabilité. Le format JSON permet d'échanger des données, en relations avec des bases ou des systèmes de fichiers, augmentant au passage les possibilités de référencement, et d'automatisation des tâches entre les applications d'Internet.

7. Conclusion

L'internet connaît une utilisation de plus en plus généralisée pour le divertissement, la communication interpersonnelle ou professionnelle, l'échange d'information et la relation avec les administrations tels que les impôts, les services d'états civil, de justice (services appelés e-gouvernance). Tous les échanges sont réalisés grâce au WWW qui est le principal et le plus populaire des services de l'Internet à côté de la messagerie et du partage de fichiers. Le WWW est basé sur le concept de client/serveur et utilise des langages pour décrire l'interface utilisateurs (HTML ; CSS ; Javascript) et d'autres langages pour coder les traitements sur le serveur web (langages cotés serveur : php, java, perl, python, etc....).