

## CH 6 - Le Web

### 1. L'architecture du Web

#### 1.1. Web et Internet

Le Web (toile ou réseau) désigne un système donnant accès à un ensemble de données (page, image, son, vidéo) reliées par des **liens hypertextes** et accessibles sur le réseau internet.

C'est en **1965** que **Ted Nelson** a inventé le concept d'hypertexte qui permet d'accéder directement à des documents divers en les interconnectant par des liens.

Le Web naîtra en **1989** grâce à **Tim Berners Lee** qui permettra la navigation de documents en documents grâce à des hyperliens. Pour avoir une portée mondiale il a fallu uniformiser les langages et les protocoles et c'est ainsi que Tim Berners Lee a défini le protocole **HTTP**, le langage **HTML** et l'adresse **URL**.

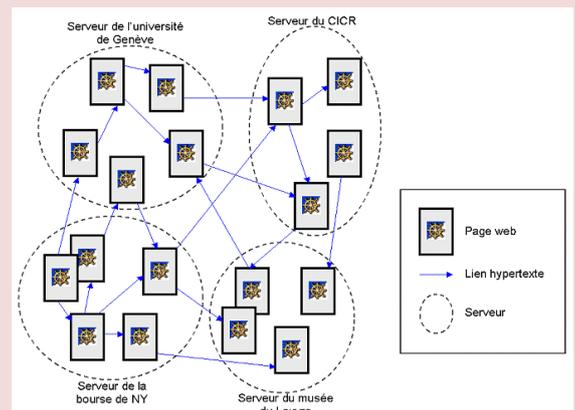
Le premier navigateur (Mosaic) arrive en **1993** et va permettre une diffusion rapide au grand public.

Depuis **2010**, le web se développe sur les applications mobiles et ne cesse de croître.

#### Attention, il ne faut pas confondre le Web et Internet.

Comme nous l'avons vu au chapitre 5, Internet est le réseau des réseaux qui consiste à relier **physiquement** différentes machines entre elles et de leur permettre de communiquer grâce aux protocoles **TCP/IP**.

Le Web quant à lui consiste à connecter **différentes pages** entre elles grâce à des **liens**. C'est en représentant les liens entre les différentes pages que l'on a l'image d'une toile (web en anglais).



<https://tecfa.unige.ch/tecfa/teaching/LME/tech/c1-usage-reseau-internet.html>

#### 1.2. Langages HTML et CSS

Une page Web (souvent appelée « page internet ») est un document contenant du texte, des images, des liens vers d'autres pages ...

Pour ce faire les pages Web sont écrites dans le **langage de balises HTML** (HyperText Markup Language qui en français veut dire «Langage de balisage pour document hypertexte »).

C'est un langage de description, il ne fait pas de programmation comme le langage python, il n'y a donc pas de tests, de boucles, ...

Le langage HTML utilise des balises pour délimiter le début et la fin de chaque instruction lue par le navigateur. Chaque instruction HTML est délimitée par une balise ouvrante **<élément>** et une balise fermante **</élément>**. Ces balises sont invisibles à l'écran mais elles permettent au navigateur d'interpréter ce qu'il doit afficher.

La mise en page ou le style graphique est défini dans le **langage CSS** (Cascading **S**tyle **S**heets ce qui signifie « *Feuille de style en cascade* »). C'est donc lui qui définit l'arrière-plan, le type, la taille, la couleur des caractères.

Le **lien hypertexte** est sans doute la balise la plus emblématique : elle permet un renvoi à un autre endroit d'une page Web, à une autre page du site ou à un autre site. Elle est identifiée par la balise `<a = href>` et se termine par `</a>`.

*Voir activité expérimentale : CH6 - Langage HTML et CSS.*

## 2. Clients - serveurs

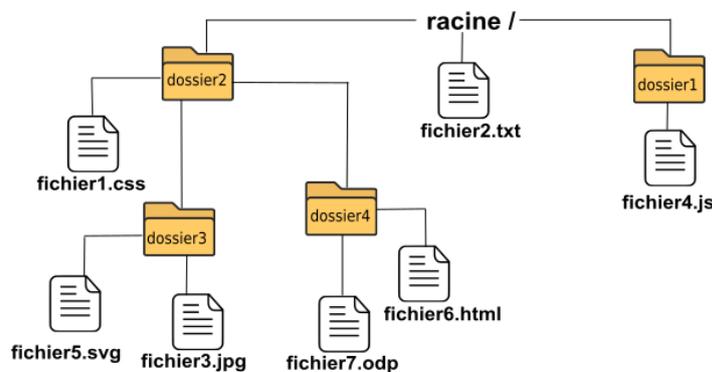
### 2.1. URL

Une **URL** (Uniform **R**esource **L**ocator) est l'adresse d'une page Web. Elle est composée de trois éléments :

- + **http://** ou **https://** (pour la version sécurisée) qui correspond au protocole de communication client-serveur développé pour le Web ;
- + un **nom de domaine**, souvent une marque, une entreprise, une association, etc. ;
- + un **chemin** qui pointe vers une ressource ou une page précise.

<https://www.lycee-saintexupery-fameck.fr/vie-scolaire/>

L'URL indique donc « l'endroit » où se trouve une ressource sur un ordinateur. Un fichier peut se trouver dans un dossier qui peut lui-même se trouver dans un autre dossier... On parle d'une structure en arborescence, car elle ressemble à un arbre à l'envers :



[https://pixees.fr/informatiquelycee/n\\_site/snt\\_web\\_url.html](https://pixees.fr/informatiquelycee/n_site/snt_web_url.html)

Pour accéder au fichier désiré il faut donc donner le chemin. Par exemple le chemin du fichier *fichier3.jpg* sera : `/dossier2/dossier3/fichier3.jpg`.

### 2.2. Requêtes HTTP

Les pages web sont accessibles via internet en utilisant le protocole **HTTP** (Hyper**T**ext **T**ransfer **P**rotocol) ou sa version sécurisée **HTTPS** qui crypte les échanges. Un « dialogue » se met alors en place entre clients (celui qui veut accéder à une ressource) et serveurs (celui qui a la ressource) on dit que le client envoie une **requête http** au serveur.

**Exemple :** Imaginons une page Web dans laquelle existe un lien vers l'image « fichier3.jpg » d'un autre site. En cliquant sur le lien, le navigateur va envoyer la requête suivante :

**GET /HTTP/1.1** indique la version utilisée

**Host : www.monsite.fr/dossier2/dossier3/fichier3.jpg** donne l'URL

Le serveur répond alors avec un code indiquant la réussite de l'accès et donne la ressource. Si le document n'existe pas il renvoie un code d'erreur comme le code bien connu « Erreur 404 : fichier non trouvé ».

De nos jours, les pages Web sont devenues dynamiques, c'est-à-dire que le serveur va également envoyer des requêtes vers le client pour obtenir des informations. Ces dernières seront alors utilisées pour afficher une page Web personnalisée.

### 3. Les usages

#### 3.1. Moteurs de recherche

Les moteurs de recherche permettent de trouver des informations dans des pages dont on ne connaît pas l'adresse, voire dont on ignore l'existence. La méthode de recherche appelée référencement naturel se décompose en trois grandes activités :

- **le parcours automatique du Web** pour collecter les pages visitées par des robots ;
- **l'analyse du contenu des pages** et leur indexation sur les mots qu'elles contiennent (constitution d'un annuaire inversé) ;
- **la requête faite par un internaute**, construit une liste ordonnée des pages comportant les mots clés de la requête.

L'ordre d'apparition des pages proposées suite à la requête dépend notamment de leur popularité (principe des liens), de leur pertinence (aux mots de la requête), et de l'ordre des termes de la requête. Les concepteurs de site Web peuvent améliorer le référencement de leurs pages en choisissant bien les mots et en les plaçant à des endroits stratégiques dans les pages.

Par ailleurs, des moteurs de recherche permettent à certains sites d'acquérir de la visibilité sur la première page des résultats de recherche en **achetant** de la publicité qui apparaîtra parmi les liens promotionnels.

#### 3.2. Notions juridiques

Dans l'histoire de la communication, le *Web* est une révolution : il a ouvert à tous la **possibilité et le droit de publier** ; il permet une coopération d'une nature nouvelle entre individus et entre organisations : commerce en ligne, création et distribution de logiciels libres multi-auteurs, création d'encyclopédies mises à jour en permanence, etc. ; il devient universel pour communiquer avec les objets connectés.

Le *Web* permet aussi de diffuser toutes sortes d'informations dont **ni la qualité, ni la pertinence, ni la véracité ne sont garanties** et dont la vérification des sources n'est pas toujours facile.

Il conserve des informations, parfois **personnelles**, accessibles partout sur de longues durées sans qu'il soit facile de les effacer, ce qui pose la question du droit à l'oubli. Il permet une exploitation de ses données, dont les conséquences sociétales sont encore difficiles à estimer.

#### 3.3. Sécurité

En formulant des requêtes sur des sites *Web* dynamiques et en laissant des programmes s'exécuter sur sa machine, **l'utilisateur prend des risques** : il peut communiquer des informations personnelles à son insu à des serveurs qui en gardent une trace, à distance ou localement par des **cookies**, ou encore charger des pages contenant des programmes malveillants, par exemple permettant d'espionner en continu les actions de l'utilisateur.

Par ailleurs, un navigateur peut garder un historique de toutes les interactions, et le laisser accessible aux sites connectés. L'utilisateur peut utiliser des services qui s'engagent à ne pas garder de traces de ses interactions, par exemple certains moteurs de recherche. Il peut aussi **paramétrer son navigateur** de façon à ce que celui-ci n'enregistre pas d'historique des interactions.

Enfin, de fausses pages peuvent encore être utilisées pour l'hameçonnage (*phishing*) des utilisateurs. Un nom de lien pouvant cacher une adresse *Web* malveillante, il faut examiner cette adresse avant de l'activer par un clic.