



LE WEB

- *URL*
- *Relation Client – Serveur*
- *Moteurs de recherche*

“

— *Pourquoi je ne trouve pas mon site avec Google ?*

— *Mmh depuis combien de temps votre site est en ligne ?*

— *Deux minutes !*

”

Et oui malheureusement (ou heureusement d'ailleurs), le référencement prend du temps et même pour **apparaître** sur Google après une mise en ligne, cela peut prendre quelques jours.

Sciences Numériques et Technologie
2^{de}

Support de cours :
Jean-Christophe BONNEFOY

Objectifs :

- Savoir décomposer l'URL d'une page
- Savoir reconnaître les pages sécurisées
- Savoir identifier ce qui est exécuté par le client et par le serveur
- Savoir mener une analyse critique des résultats fournis par un moteur de recherche

1. Définition :

Le web est l'abréviation de **World Wide Web**. C'est un système hypertexte (c'est-à-dire qu'il permet d'établir des liaisons directes entre des éléments différents : texte, image, etc.), qui fonctionne sur le réseau internet.

Les pages disponibles sur le web sont liées entre elles par des liens hypertextes, à l'image d'une toile d'araignée (d'où son nom).

Remarque : Il ne faut pas confondre le web avec internet : internet est un réseau alors que le web est un système qui utilise ce réseau.

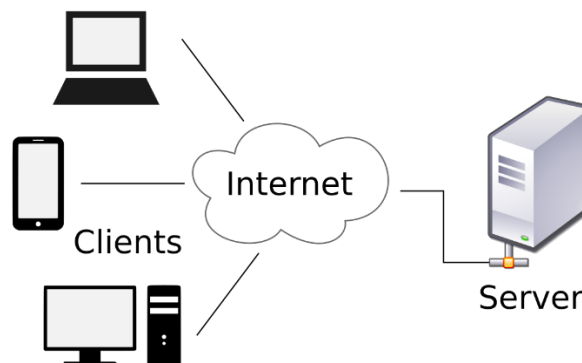
2. Une structure client - serveur :

Un **serveur** est un dispositif informatique qui offre des services à des clients.

Un **client** est un logiciel informatique qui fait des requêtes à un serveur.

Un système client/serveur fonctionne de la manière suivante :

- le client émet une requête vers le serveur par l'intermédiaire de son adresse et de son port (porte d'entrée de communication d'un ordinateur), qui sera différent selon le protocole utilisé (HTTP, HTTPS, FTP, etc.) ;
- le serveur reçoit la demande et répond à la machine client via l'adresse et le port de la machine.



L'accès à l'information, du côté client, se fait via une interface graphique : le navigateur Web.

3. L'URL

Une **URL** (*Uniform Resource Locator*) est une adresse qui est liée à une ressource donnée et qui reste unique sur le web.

Ces ressources peuvent être de plusieurs types : des pages HTML, des images, des vidéos, ... Chacune de ces URL peut être saisie dans la barre d'adresse du navigateur (Chrome, Firefox, Internet Explorer, Edge, Safari, etc.)

Les différentes parties d'une URL sont les suivantes :

- Le **protocole** est un ensemble de règles qui régissent les échanges de données. Il existe différents types de protocoles utilisés sur le web :
 - Le protocole HTTP permet de surfer sur le web de manière non sécurisée : les données ne sont pas cryptées.
 - Le protocole HTTPS permet également de surfer sur le web, mais en cryptant les données : la navigation est sécurisée.
 - Le protocole FTP permet de transférer des fichiers.
- Le **nom de domaine** est une adresse web, c'est-à-dire une adresse « postale » qui nous permet de nous repérer sur la toile.
- Le **chemin d'accès** représente l'accès unique à un fichier système, lequel comporte lui-même le fichier souhaité (**nom du fichier**).

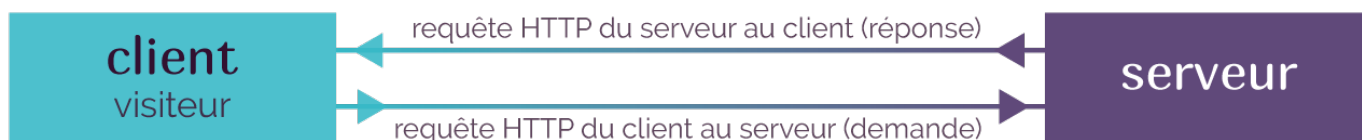
Exemple : <http://www.jcbmathsandco.fr/snt-sciences-numeriques-et-technologie-le-web/>

Protocole	http://
Nom de domaine	www.jcbmathsandco.fr
Chemin d'accès	/snt-sciences-numeriques-et-technologie-le-web/

Remarque : il faut savoir que les noms de domaines ne sont pas gratuits : il faut les acheter auprès d'un opérateur spécialisé et autorisé appelé *registrar*. De plus, ces noms ne peuvent pas être achetés à vie, seulement loués pendant une période donnée (pouvant aller jusqu'à 10 ans).

4. Le protocole HTTP

HTTP est le protocole de base du Web : il est en charge des requêtes d'affichage des pages Web. Une **requête HTTP** est une demande effectuée par le navigateur web (Chrome, Internet Explorer, Firefox, Mozilla, Safari, etc.) au serveur HTTP lorsqu'il souhaite consulter une page web. Les requêtes se font par paires : la demande (du client) et la réponse (du serveur).



Le protocole **HTTPS** est plus avancé et bien plus sécurisé que le protocole HTTP. Le « S » de HTTPS signifie en effet « Secure », ou « sécurisé » en français. Souvent, on aperçoit un petit cadenas dans le navigateur Web.

Ce protocole fonctionne grâce au protocole TLS (Transport Layer Security), qui est une technologie de sécurité standard pour l'établissement d'une **connexion chiffrée entre un serveur web et un navigateur**.

Si le protocole n'est pas sécurisé (HTTP), les données numériques qui sont entrées sur un site (nom d'utilisateur, mot de passe, adresse, téléphone, etc.) sont vulnérables aux interceptions et à l'espionnage par des hackers (pirates informatiques).

5. Les moteurs de recherche

Avec la quantité colossale de sites, et donc de pages web, il s'est rapidement avéré nécessaire d'avoir des outils pour rechercher une donnée sur le Web. Parmi les moteurs de recherche existants, on peut citer le géant *Google* utilisé par 90% des internautes et *Bing*, son concurrent direct développé par Microsoft.

Le concepteur d'un site doit donc indexer correctement son site s'il souhaite que ce site soit ajouté à l'annuaire du moteur de recherche. Le processus **d'indexation** consiste à remplir un formulaire afin de fournir des informations (nom, titre, adresse URL, mots clés, description du contenu du site internet, etc.). Le concepteur de la page web doit veiller à utiliser les bons mots clés puis choisir une ou plusieurs catégories de rattachement, afin d'optimiser la visibilité de son site sur la toile.

Une fois qu'un site a été indexé par son concepteur, il faut que le moteur de recherche le répertorie pour qu'il soit ajouté à l'annuaire du moteur. Afin d'alimenter l'annuaire, le moteur de recherche parcourt la toile et référence les pages internet selon leurs contenus et leurs mots clé, grâce à ce que l'on appelle un **robot d'indexation**. Le rôle de ces robots est d'extraire les liens afin de les visiter ; c'est ainsi qu'en suivant les pages, de liens en liens, il est possible d'inspecter pratiquement tout le web. Chaque moteur de recherche parcourt tous les jours le web.

Dès que le moteur de recherche a terminé de scruter l'ensemble des pages web qui correspondent aux mots clés, un index des pages présentes sur le web est alors confectionné par ces robots.

Il existe des techniques, dites de référencement naturel, ou SEO (Search Engine Optimization) afin d'améliorer le référencement d'un site. Ce domaine est actuellement en plein essor, tant les enjeux sont importants.

6. Impact sur les pratiques humaines

Presque 30 ans d'histoire, c'est assez court ! Et pourtant le Web a pris une place prépondérante dans notre quotidien :

- ❖ Le Web est devenu la première source d'informations. C'est un formidable espace de partage des savoirs.
- ❖ Les outils de communication sont nombreux et variés, les réseaux sociaux ont modifié la vie sociale.
- ❖ Le commerce en ligne « a explosé », ce qui a profondément changer nos habitudes.

Face à cette évolution, nous devons prendre un certain recul et se placer dans son propre contexte :

- ❖ L'information que je consulte est-elle réelle ? Objective ?
- ❖ Les avis que je consulte sont-ils impartiaux ?
- ❖ La sécurité est-elle garantie lors de mon achat en ligne ?
- ❖ Comment sont gérer mes données personnelles ?
- ❖ Quelle publicité dans ce site ?

Dans ce constant mouvement, l'information et la formation de chacun sont nécessaires.