

Partie I / Le Web

Le **Web** (toile en français) est un service construit sur l'infrastructure du réseau internet pour mettre à disposition des utilisateurs des documents répartis au niveau mondial.

Le Web est géré au niveau mondial par le W3C -*World Wide Web Consortium*. Cet organisme international indépendant définit les standards permettant aux ordinateurs connectés de pouvoir communiquer ensemble.

I/ Le fonctionnement du web

Pour pouvoir aller sur le Web, il faut disposer sur sa machine d'un programme spécifique appelé **navigateur** (Chrome, Mozilla, Opéra, Safari etc.) chargé de communiquer avec les serveurs du Web.

HTTP : protocole de transfert d'hypertexte (en anglais *HyperText Transfer Protocol*) qui définit les règles que doit suivre un navigateur pour obtenir une page d'un serveur.

Ce protocole précise quel message le navigateur doit envoyer par le réseau internet pour demander l'affichage d'une page et sous quel format (texte, image, vidéo etc.) il attend une réponse.

Remarque : https est la version sécurisée du protocole http. En l'utilisant, la communication est chiffrée et ne peut être comprise par un intermédiaire qui intercepterait les messages.

Hypertexte : texte contenant des liens nommés « hyperliens ». Il permet d'associer une partie de texte l'adresse d'une autre page à charger lorsque l'utilisateur clique dessus.

URL : adresse uniforme de localisation d'une ressource (en anglais *Uniform Resource Locator*). Cette adresse unique doit comporter le nom du protocole, le nom de la machine qui héberge la ressource, le nom de la ressource et éventuellement de paramètres optionnels.

Exemple : [https://fr.wikipedia.org/wiki/Hypertext Transfer Protol](https://fr.wikipedia.org/wiki/Hypertext_Transfer_Protol)

Page : dossier :

Serveur : protocole :

II/ Les langages HTML et CSS

Plusieurs langages ont été standardisés par le W3C pour décrire le contenu des pages (HTML), leur forme (CSS) et pour programmer des interactions (Javascript, PHP).

HTML : provenant de l'anglais *HyperText Markup Language*, langage de balisage pour hypertexte, HTML est un langage permettant de créer et représenter le contenu des pages web et de structurer un texte/vidéo/image en lui ajoutant des **balises**.

CSS : langage de style (en anglais *Cascading Stylesheet*) qui permet de préciser la police utilisée, leur taille, couleur etc. Un fichier CSS contribue ainsi à homogénéiser la présentation d'une page Web et à séparer le fond (HTML) de la forme (CSS).

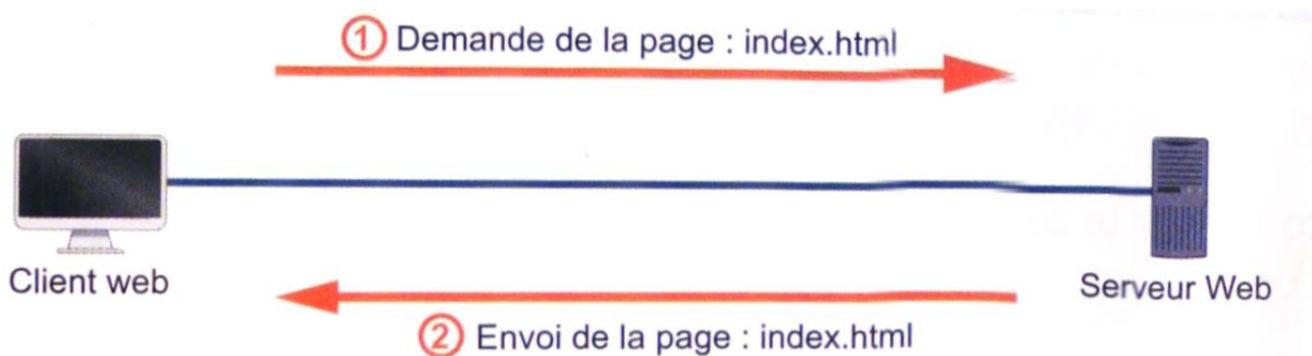
III/ Communication entre serveurs et clients

Il existe deux types de sites :

- ceux qui envoient toujours les mêmes informations à tous les internautes : ce sont des **sites statiques**.
- ceux qui sont capables de personnaliser le contenu en fonction de la demande de l'utilisateur : ce sont des **sites dynamiques**.

Accéder à un site statique

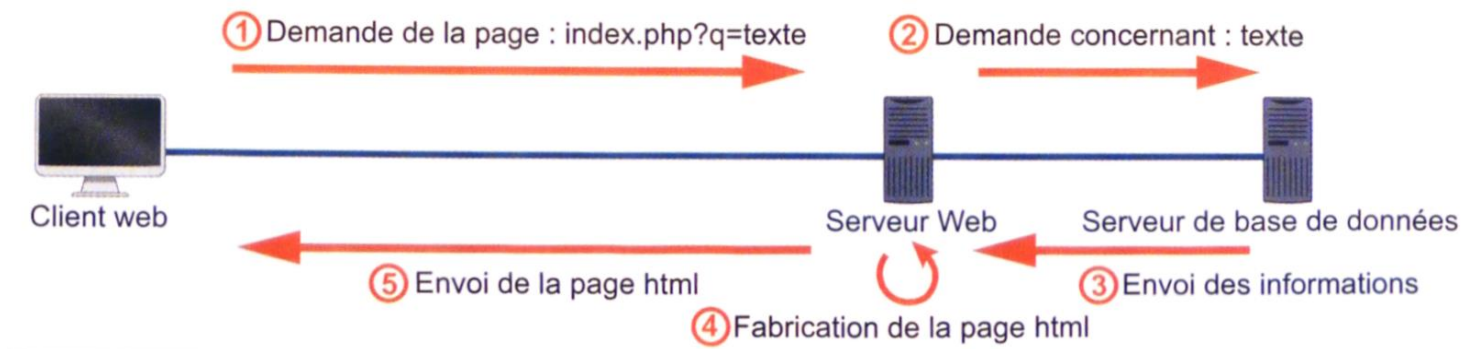
Le navigateur envoie une demande de page en précisant l'URL de la page demandée et l'adresse de retour. Le serveur cherche le fichier correspondant et l'envoie au demandeur.



Accéder à un site dynamique

Tous les sites permettant à leurs utilisateurs de consulter des informations personnalisées (sites bancaires, d'achats en ligne etc.) sont dynamiques.

Pour transmettre des informations à un serveur web dynamique, le navigateur peut utiliser le mécanisme des paramètres : les informations sont alors transmises à la fin de l'URL après un point d'interrogation (envoi d'un formulaire). Le langage PHP est spécialisé pour ces tâches.



IV/ Les moteurs de recherche

Les navigateurs Web proposent à leurs utilisateurs d'accéder à des pages sans connaître leur URL et même sans en connaître l'existence à l'avance. Pour cela, il fait appel à des moteurs de recherche (Google, Qwant, Altavista, Bing etc.). Ils parcourent en permanence le web, répertorient les pages et les classes en fonction de leur popularité.

Indexation : traitement qui consiste à analyser des pages pour détecter des mots clés utilisés couramment dans les demandes des utilisateurs puis à fabriquer un index permettant de retrouver rapidement ces pages à partir des mots clés.

Tout ceci est calculé à l'avance, même en cas de nouvelle demande d'un internaute. En réalité, d'autres personnes l'avaient très probablement déjà effectuée avant !

La principale difficulté des moteurs de recherche consiste à classer, de la manière la plus pertinente qui soit, l'ensemble des pages contenant les mots clés demandés.

Ce sont essentiellement sur des **algorithmes** de classement que les moteurs de recherche se font une concurrence acharnée et cachent leurs secrets de fonctionnement. En 1998, les développeurs informatiques Larry page et Sergey Brin proposent l'algorithme *PageRank* qui a conduit à la création du moteur de recherche Google et ayant permis de proposer une définition à l'idée de **popularité** d'une page.

En somme, plus une page est citée, plus sa popularité est élevée, surtout si elle est citée par des pages ayant elles-mêmes une bonne popularité.

V/ Sécurité et confidentialité du web

Pour surfer sur le Web en toute sécurité et confidentialité, il faut éliminer ses traces en paramétrant son navigateur. Il est possible d'effacer son **historique** à partir des options du navigateur.

S'assurer d'utiliser **des sites sécurisés** (protocole https) lors d'opérations délicates (achats etc.).

Les ressources publiées sur le web ne sont pas toutes libres de droits, leur usage est réglementé par diverses législations en fonction des pays ou par des licences *Creative Commons*.