

Les activités et quizz du chapitre sont disponibles sur frederic-junier.org.

1 Web et Internet

En 1989, **Tim Berners-Lee**, qui s'occupe de la gestion des documents scientifiques au **CERN**, propose avec le projet *mesh* (filet) de faire traverser le réseau internet à l'**hypertexte** imaginé par **Ted Nelson** en 1965. C'est la naissance du *World Wide Web* : des liens hypertextes relient des documents stockés sur des machines différentes et l'utilisateur d'une machine cliente peut naviguer de lien en lien à travers ce réseau logique d'informations qui se superpose au réseau physique d'internet.

Avec l'ouverture au public des standards du Web en 1993 et leur standardisation organisée par le **W3C**, qui assure la portabilité des pages Web sur toutes les machines, le Web s'impose à partir du milieu des années 90 comme le service le plus populaire disponible sur Internet.

2 Les fondements du Web

2.1 Architecture client / serveur

Comme la plupart des services d'Internet, le Web repose sur une **architecture client / serveur** : pour obtenir une page Web, le **client** la demande au **serveur** qui la détient.

Le client et le serveur sont des logiciels et peuvent être installés sur n'importe quelle machine.

Le logiciel client s'appelle un **navigateur**, comme par exemple *Firefox* ou *Chrome*.

2.2 URL : l'adressage logique

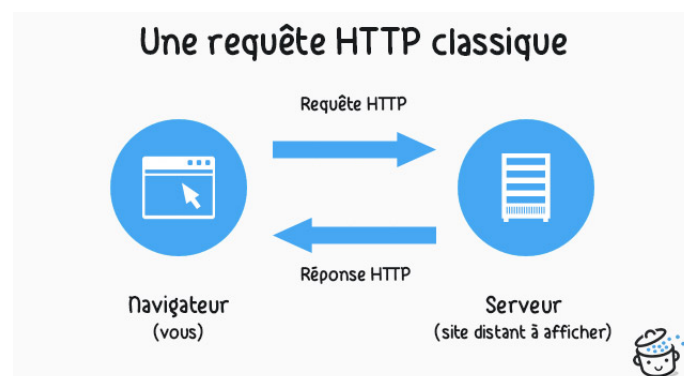
Les documents disponibles sont localisés sur le Web par une adresse logique, **Uniform Resource Locator**. La syntaxe des URL est standardisée, par exemple décomposons :

```
https://www.gnu.org/gnu/linux-and-gnu.fr.html
```

- le protocole est `https` ;
- le nom de domaine sur Internet de la machine est `www.gnu.org`, le préfixe `www` étant le nom usuel du dossier public sur un serveur Web ;
- `gnu/linux-and-gnu.fr.html` est le chemin vers la ressource sur le serveur : le fichier `linux-and-gnu.fr.html` qui se trouve dossier `gnu` lui-même dans le dossier public `www`.

2.3 HTTP : le protocole

Lors d'un échange Web, le **client** et le **serveur** dialoguent selon des règles fixées par un protocole standardisé, **HyperText Transfer Protocol**.



Source : <https://wpmarmite.com/wordpress-https/>

Si la page demandée n'est pas disponible, le serveur retourne le célèbre code d'erreur 404.

2.4 HTML : le langage

Les **premières pages Web** sont des **pages statiques**, c'est-à-dire que la page retournée par le serveur ne dépend pas du client. Le premier langage du Web est **HyperText Markup Language** qui est un langage de description de documents à l'aide de **balises**.

HTML permet de définir la structure d'un document en le découpant en blocs logiques délimités par une **<balise>** ouvrante et une **</balise>** fermante. En particulier il permet de créer des liens hypertextes entre deux blocs d'un même document ou de documents différents.

Code source HTML

```
<!DOCTYPE html>
<html>
<head>
<meta charset="utf-8">
</head>
<body>

<h2>Recommandation</h2>
<p>Un <a href="https://frederic-junier.org/wp">
  Super site</a> à visiter absolument ! </p>

</body>
</html>
```

Affichage dans le navigateur

Recommandation

Un [Super site](https://frederic-junier.org/wp) à visiter absolument !

3 Les langages du Web

3.1 Pages Web statiques

Les **pages statiques** sont écrites en HTML et en **CSS** qui est un langage de feuille de style utilisé pour décrire la présentation d'un document écrit en HTML. La standardisation du CSS a souffert de la guerre des navigateurs à la fin des années 1990 mais CSS s'est imposé ensuite comme un complément indispensable de HTML.

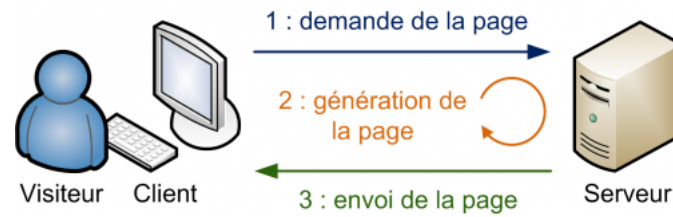
HTML décrit la structure du document et CSS définit son apparence. Le fond et la forme sont ainsi séparés. D'ailleurs, le code CSS, même s'il peut être intégré au code HTML, est plutôt dans une feuille de style séparée au document HTML par une balise `<link rel="stylesheet" href="feuille_style.css">`.

3.2 Pages Web dynamiques

A partir de la fin des années 1990, avec le développement du commerce en ligne, apparaissent des langages permettant de dynamiser les pages Web.

PHP est un langage qui s'exécute côté serveur pour générer des pages Web à la volée, il s'utilise avec une base de données pour générer des pages Web personnalisées et permet l'utilisation de formulaires interactifs.

Javascript est un langage qui s'exécute côté client pour créer des animations et modifier la page en fonction d'événements utilisateurs (souris, clavier ...).



Source : <https://openclassrooms.com>

4 Les moteurs de recherche

Le Web est devenu un gigantesque hyertextet et pour y chercher une information, de puissants logiciels, les **moteurs de recherches** sont nécessaires. Le plus célèbre est *Google search*, qui traite plusieurs milliards de requêtes par jour. Le nom Google a d'ailleurs pour origine *googol* qui désigne le très grand nombre 10^{100} .

Un moteur de recherche doit effectuer plusieurs tâches dont le traitement repose sur des algorithmes souvent secrets et l'utilisation de gigantesques fermes de serveur (pour stocker des copies des pages indexées) :

- des robots (logiciels), les *crawlers* explorent les pages du Web en suivant les liens hypertextes : les pages qui ne peuvent être atteintes constituent le **Web profond** ;
- les pages sont indexées en fonction de mots clefs issus de leur contenu ou de leurs métadonnées (balises meta) et une note de popularité leur est assignée (en fonction des liens qu'elles reçoivent pour le *PageRank de Google*) ;
- répondre aux requêtes très rapidement avec un classement des réponses le plus pertinent possible.

5 Sécurité et confidentialité sur le Web

5.1 Paramétrer son navigateur

Les serveurs Web gardent des traces de leurs visiteurs en stockant des petits fichiers, les **cookies**, sur les machines clientes. Ces **cookies** sont souvent nécessaires au confort de la navigation mais ils sont parfois utilisés de façon abusive par des régies publicitaires pour récolter des données personnelles à des fins de publicité ciblée. On peut néanmoins gérer ces **cookies** à travers Le menu de paramétrage du navigateur et certaines extensions de navigateur comme **Disconnect** permettent de visualiser et bloquer ces traceurs.

Le **mode de navigation privée** ne rend pas anonyme sur Internet (l'adresse IP n'est pas masquée) mais il permet d'effacer l'historique de navigation, les marque pages, les cookies à la fin de la session. Il est recommandé pour une navigation sur une machine multiutilisateurs.

5.2 Le protocole HTTPS

Le protocole **HTTPS** est une combinaison du protocole HTTP avec le protocole TLS qui permet de chiffrer la connexion (avec des algorithmes arithmétiques) et d'authentifier les deux bouts (serveur ou client) avec des **certificats de sécurité** délivrés par des autorités de confiance.

La plupart des serveurs Web ont désormais des certificats et établissent des connexions HTTPS. Lorsque le certificat est vérifié par le navigateur, le petit cadenas devant l'URL dans la barre d'adresses est en vert, sinon il est en jaune (certificat expiré) voir en rouge.



5.3 Notions juridiques et enjeux sociétaux

Le Web offre une gigantesque masse d'informations à disposition de tous. Il peut être un formidable outil de partage de connaissance et son impact sur la société est de plus en plus profond. Même les objets et les programmes sont reliés entre eux sur le Web.

Les valeurs de partage et de liberté aux origines du Web et d'Internet doivent être préservées et de bonnes pratiques sont nécessaires pour gérer intelligemment cette abondance d'informations :

- Les documents disponibles sur le Web (textes, images, sons) sont soumis au droit d'auteur et avant de les rediffuser ou de les modifier, il convient de vérifier leurs droits d'usages. Certaines licences comme les **Creative Commons** assouplissent le droit d'auteur pour favoriser le partage sur le Web.
- Le Web est accessible à tous et tout le monde peut y diffuser de l'information vraie ou fausse. Il convient de manipuler avec prudence ces informations et de vérifier leurs sources.
- La liberté d'expression sur le Web doit être défendue et certains pays appliquent malheureusement des filtres, néanmoins les droits fondamentaux comme le droit à l'image doivent être respectés et la loi s'applique aussi sur le Web.
- La maîtrise de ses données personnelles sur Web est un enjeu majeur, depuis la **loi Informatique et Libertés de 1978** et encore plus avec la **Réglementation Générale sur la Protection des Données** au niveau européen, des dispositifs juridiques existent pour défendre les droits des usagers : droit de consultation, droit à l'oubli, droit à la portabilité. En France, la **Commission Nationale Informatique et Libertés** est chargée d'informer les citoyens, d'anticiper les adaptations du droit et de contrôler les professionnels et les particuliers (HADOPI).

