



1. Un petit bout d'histoire

• 1965 : Naissance du concept de l' hypertexte

Ted Nelson, publie le concept de l'hypertexte numérique (lien cliquable vers d'autres textes ou d'autres pages....)

Première connexion informatique à longue distance entre le Massachusetts et la Californie. Cette expérience montra que des ordinateurs pouvaient travailler ensemble à distance.



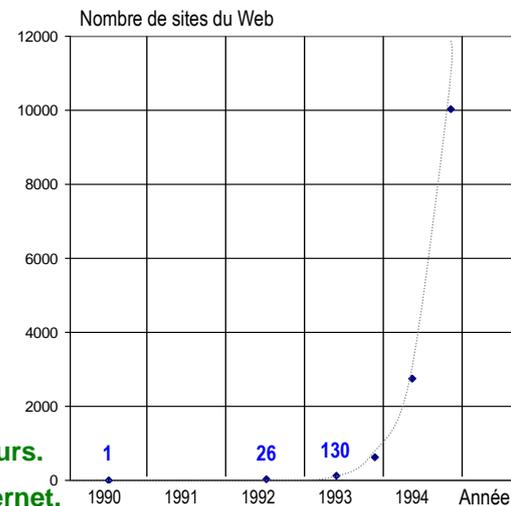
• 1990 : Tim Berners-Lee Invention du Web

Il propose d'utiliser le principe des hypertextes sur internet et donne ainsi naissance au World Wide Web. Le WWW est une collection de documents électroniques accessible sur internet et reliés entre eux par des liens hypertextes.

La première page Web voit le jour au CERN ainsi que le premier navigateur baptisé Nexus.

• 1993 : Naissance du WEB

Internet est ouvert au domaine public. 130 sites web en juin de la même année. Apparition des premiers « Navigateurs », applications développées pour lire les pages Web



- ▶ **Internet** correspond au réseau de communication établi entre les ordinateurs.
- ▶ **Le Web** correspond aux différents sites mis à disposition sur le réseau Internet.

2. Les Protocoles du web

En 1989, *Tim Berners-Lee*, un informaticien britannique, propose de développer le système hypertexte pour améliorer la diffusion des informations au sein du réseau interne au CERN.

Il met au point les **URL (adresses web)**, l'**HyperText Transfer Protocol (HTTP)** et l'**HyperText Markup Language (HTML)**.

Hypertexte : texte contenant des liens appelés hyperliens permettant de passer à une autre page d'information sur internet.

Pour que ce système de liens puisse fonctionner, il faut standardiser la manière d'écrire les liens vers d'autres pages. On met donc en place les **URL (Uniform Resource Locator)**.

L'**URL** d'une page d'information correspond à son adresse.



L'**HyperText Transfer Protocol (HTTP)**, littéralement « protocole de transfert hypertexte » est un protocole de communication client-serveur développé pour le World Wide Web.

La version HTTPS (avec S pour secured) est la variante du HTTP sécurisée par l'usage de communications cryptées.

En 1994, *Tim Berners-Lee* quitte le CERN et fonde le W3C (*World Wide Web Consortium*). Cet organisme international, indépendant et à but non lucratif, définit les standards (protocoles, langages pour coder, ...) permettant aux machines connectées sur Internet de pouvoir communiquer entre elles.



3. Les pages web, principe Client Serveur

Les pages web de base sont faites avec le code HTML pour gérer entre autre leur affichage et les hyperliens qu'elles contiennent. Ce langage utilise des balises pour spécifier les fonctions à utiliser par le navigateur client.

Exemple : La page de base en HTML5 a la forme suivante :

```
<!doctype html> ①
<html lang="fr"> ②

<head> ③
  <meta charset="utf-8">
  <title>Titre de la page</title>
  <link rel="stylesheet" href="style.css">
  <script src="script.js"></script>
</head> ③

<body> ④
  contenu de la page (textes, liens,
  images, tableaux, ...)
</body> ④

</html> ②
```

- ① Cette ligne précise le type de document qui est codé sur la page pour permettre au navigateur (*le client*) de savoir quel langage il est sensé interpréter.
- ② Les balises `<html>` `</html>` indique le début du code html. L'attribut `lang="fr"` précise la langue utilisée dans la page.
- ③ Entre les balises `<head>` et `</head>` se trouvent les méta-informations sur la page, comme des données pour son indexation (titre) et des paramètres pour son affichage (*UTF-8, feuille de Style*)
- ④ Entre les balises `<body>` et `</body>` se trouve le contenu de la page web à afficher.

• Les Echanges Client-Serveur

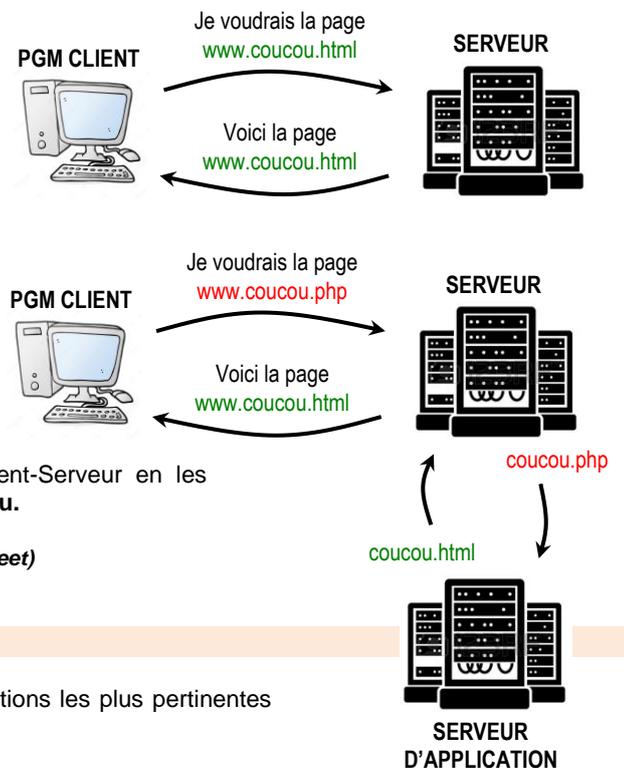
Les pages codées en html (avec ou sans JavaScript ou CSS) sont envoyées par le serveur sur la machine ayant émis la requête, et sont exécutées par le navigateur client. Ces pages, dites **statiques**, sont identiques pour tous les clients qui cherchent à y accéder.

Le Langage « html » (HyperText Marking Langage) est constitué de balises permettant l'affichage d'une page Web par un client.

exemple `<p>` `</p>` pour un paragraphe, `
` pour un retour à la ligne
`` Texte cliquable `` pour insérer un lien hypertext
`` pour insérer une image

Le Protocole « **https** » permet de sécuriser les communications entre Client-Serveur en les cryptant et en identifiant les protagonistes. Il est identifiable par un **petit verrou**.

Le Style de la page Web est inscrit dans le fichier CSS (*Cascading Style Sheet*)



4. Les moteurs de recherche

Un moteur de recherche est un logiciel permettant de rechercher les informations les plus pertinentes sur le web suite à une requête.

Indexation :

Pour être efficace, un robot d'indexation (ou *web spider* comme *BingBot*, *GoogleBot* ou *Slurp*) des serveurs d'un moteur de recherche indexe par mots clés toutes les pages du web accessibles, et ceci de manière continue.

Les méthodes d'indexations actuelles utilisent des algorithmes propre à chaque moteur.

Certaines ressources du Web restent inexplorés par les *web spider* et sont appelé le web profond (deep web).

Popularité :

Pour être populaire et faire partie des premières ressources disponibles sur un moteur de recherche, une page web doit être la cible d'un maximum de liens venant d'autres sites, et si possible de sites eux-mêmes populaires.

PageRank est, par exemple, un indicateur parmi d'autre de l'algorithme qui permet à *Google* de classer les pages de web dans son moteur.

Moteurs de recherche solidaires ou écologiques :

Ce sont des moteurs qui reversent une partie de leurs revenus à des causes écologiques, sociales ou humanitaires (*Ecosia*, *Lilo*, ...).

